Sparsity based morphological characterisation of heartbeats

Laura Rebollo-Neira
email: l.rebollo-neira@aston.ac.uk
Mathematics Department
Aston University
B4 7ET Birmingham, UK

Khalil Battikh and Amadou Sidi Watt
ENSIIE - École nationale supérieure d'informatique pour l'industrie et l'entreprise
1 Rue de la Résistance, 91000 Évry-Courcouronnes, France

Abstract

Background: The electrocardiogram (ECG) is one of the most common primary tests to evaluate the health of the heart. Reliable automatic interpretation of ECG records is crucial to the goal of improving public health. It can enable a safe inexpensive monitoring. This work presents a new methodology for morphological characterisation of heartbeats, which is placed outside the usual machine learning framework.

Method: The proposal considers the sparsity of the representation of a heartbeat as a parameter for its morphological characterisation. The approach involves greedy pursuit strategies for selecting elements from redundant dictionaries, which should be previously learnt from examples of the classes to be characterised. Using different metrics of sparsity, the dictionary rendering the smallest sparsity value, for the equivalent approximation quality of a new heartbeat, characterises the morphology of that beat. This study focuses on a procedure for learning the dictionaries and compares several metrics of sparsity for morphological characterisation on the basis of those metrics.

Results: The suitability of the method is illustrated by binary differentiation of Normal and Ventricular heartbeats in the MIT-BIH Arrhythmia data set. In general classification the sensitivity and predictivity scores for Ventricular beats are 97.6% and 99.7%, respectively. In interpatient assessment these scores drop to 92.4% and 97.8%. Due to the huge difference in the number of Normal and Ventricular beats in the whole data set even a small proportion of misidentified Normal beats amounts to a much larger proportion of false Ventricular beats. This produces a comparatively lower value of the F_1 metric for Ventricular beats.

Conclusions: The results are encouraging because the proposed binary identification is realised outside the usual machine leaning framework. Thus, extensions of the approach to allow for combination with other features and other machine leaning techniques are readily foreseen. It is from this perspective that the approach should be appreciated, because a morphological feature in isolation does not provide complete information for classification of different types of heartbeats. The numerical tests, designed to emphasise the

interpretability and reliability of the approach, demonstrate the potential of the method to contribute towards the development of a well grounded expert system for classification of heartbeats in ECG records.

Keywords. *Automation of heartbeat identification. Sparse representations. Greedy pursuit strategies. Computerised ECG interpretation.*

1 Introduction

Cardiovascular diseases (CDVs) are a group of disorders of the heart and blood vessels, including coronary heart disease, rheumatic heart disease and other conditions. These diseases are the leading cause of global death. The World Health Organisation (WHO) estimates that 32% of all deaths worldwide are caused by CDVs. In addition to preventing CDVs by maintaining a healthy lifestyle, early detection can curb mortality rates for heart disease patients and also for people with increased risk of this condition.

An electrocardiogram (ECG) is a safe and non-invasive procedure that detects cardiac abnormalities by measuring the electrical activity generated by the heart as it contracts. Sensors attached to the skin are used to detect the electrical signals produced by the heart each time it beats. These signals are analysed by a specialist to assess if they are unusual. Computerised ECG analysis plays a critical role in clinical ECG workflow. It is predicted that in the future full automated analysis of ECG may become more reliable than human analysis.

Automation of ECG interpretation has been successfully addressed by Artificial Intelligence (AI) methods which fall within the end-to-end deep learning framework [1–4]. However, due to the over-parameterised black-box nature of this framework [5] it is difficult to understand how deep models make decisions. Besides, interpretation using Deep Neural Networks (DNN) has been shown to be susceptible to adversarial attacks [6,7]. Automatic classification of heartbeats has also been addressed by different techniques of machine learning which rely on the ability to extract distinctive features of the beats in each class. Reviews and summaries of these techniques can be found in [8–11].

In this work we present an alternative viewpoint for the specific task of morphological identification of heartbeats. We base the proposed method on the ability to model the morphology of a heartbeat as superposition of elementary components. We work under the assumption that there exits a set of elementary component which is better suited for representing heartbeats of certain morphology than others. Inspired by a previous work on face recognition [12], we consider that the sets of elementary components are redundant. However, we differ from [12] in the methodology and also in that here the redundant sets are to be learnt from data. The common ground between the approaches is that both base the identification criterion on the concept of sparsity.

In the signal processing field, a signal approximation is said to be sparse if it belongs to a subspace of much smaller dimensionality than that of the space where the signal comes from. The search for the approximation subspace is carried out using a redundant set, called a 'dictionary'. Only a few elements of the dictionary, called 'atoms', are finally used for constructing the signal approximation, called 'atomic decomposition'. Sparse approximation of heartbeats is applied in [13] for classification using Gabor dictionaries. The particular Gabor's atoms involved in the representation of the beats, as well as the coefficients in the atomic decomposition, are taken as features to feed machine learning classifiers. Contrarily, our proposal relies on the possibility of learning dictionaries, which play themselves a central role in the decision making process. To the best of our knowledge this setup has not been considered before. We focus on the layout of the approach and test it for binary classification.

The paper is organised as follows. Sec. 2 discuses the greedy pursuit strategies to be considered, as well as the method for learning the dictionaries. The sparsity metrics supporting the discrimination criteria are also introduced in this section. The numerical tests illustrating the approach are placed in Sec. 3. The conclusions are drawn in Sec. 4.

2 Signal representation using dictionaries

We start this section by introducing some basic notation. Throughout the paper $\mathbb{R}$ represents the set of real numbers, and $\mathbb{N}$ and $\mathbb{Z}$ the sets of natural and integer numbers respectively. Boldface lower case letters indicate Euclidean vectors and boldface capital letters indicate matrices. The corresponding components are indicated using standard mathematical fonts e.g., $\mathbf{d} \in \mathbb{R}^N$ is a vector of components $d(i) \in \mathbb{R}$, $i = 1, \ldots, N$ and $\mathbf{G} \in \mathbb{R}^{N \times M}$ is a matrix of components $G(i,j)$, $i = 1, \ldots, N$, $j = 1, \ldots, M$. The Euclidean inner product $\langle \cdot, \cdot \rangle$ between vectors in $\mathbb{R}^N$ is defined as

$$\langle \mathbf{d}, \mathbf{g} \rangle = \sum_{i=1}^{N} d(i) g(i).$$

This definition induces the 2-norm $\|\mathbf{g}\| = \sqrt{\langle \mathbf{g}, \mathbf{g} \rangle}$. The 1-norm of a vector $\mathbf{g} \in \mathbb{R}^N$, is indicated as $\|\mathbf{g}\|_1$ and calculated as $\|\mathbf{g}\|_1 = \sum_{i=1}^{N} |g(i)|$.

The inner product between matrices in $\mathbb{R}^{N \times M}$ is the Frobenius inner product, which is defined as

$$\langle \mathbf{F}, \mathbf{G} \rangle_F = \sum_{i=1}^{N} \sum_{j=1}^{M} F(i,j) G(i,j).$$

This definition induces the Frobenius norm $\|\mathbf{G}\|_F = \sqrt{\langle \mathbf{G}, \mathbf{G} \rangle_F}$. The transpose of a matrix $\mathbf{G}$ is indicated as $\mathbf{G}^\top$.

Given a signal $\mathbf{f}$ as a vector in $\mathbb{R}^N$, the K-term atomic decomposition for its approximation is of the form

$$\mathbf{f}^K = \sum_{n=1}^{K} c(n) \mathbf{d}_{\ell_n}, \tag{1}$$

where the elements $\mathbf{d}_{\ell_n}$, called atoms, are chosen from a redundant set $\mathcal{D} = \{\mathbf{d}_n, \in \mathbb{R}^N, \|\mathbf{d}_n\| = 1, n = 1, \ldots, M\}$, called dictionary. The set of all linear combination of elements in $\mathcal{D}$ is denoted as span$\{\mathcal{D}\}$. The problem of how to select from $\mathcal{D}$ the K elements $\mathbf{d}_{\ell_n}$, $n = 1 \ldots, K$ such that $\|\mathbf{f} - \mathbf{f}^K\|$ is minimal is intractable (there are $\frac{N!}{(N-K)!K!}$ possibilities). In practical applications the problem is addressed by 'tractable' methods. For the most part these methods are realised by

(a) Expressing $\mathbf{f}^K = \sum_{n=1}^{M} c(n) \mathbf{d}_n$ using only K-nonzero coefficients minimising the 1-norm $\|\mathbf{c}\|_1$ [14].

(b) Using a greedy pursuit strategy for stepwise selection of the K elements $\mathbf{d}_{\ell_n}$, $n = 1, \ldots, K$, for producing the approximation (1).

We restrict consideration to greedy pursuit algorithms, because for the application we are considering these types of methods are effective and faster than those based on minimisation of the 1-norm.

2.1 Pursuit Strategies

In the context of signal processing the simplest pursuit strategy is known under the name of Matching Pursuit (MP) [15]. Depending on their implementation and context of application variations of the MP approach can also be found under different names. We discuss here a refinement of MP known as Orthogonal Matching Pursuit (OMP) [16] as well as the stepped wise optimised version termed Optimised Orthogonal Matching Pursuit (OOMP) [17].

2.1.1 From MP to OMP

The MP algorithm evolves by successive approximations as follows: Setting $k = 0$ and starting with an initial approximation $\mathbf{f}^0 = 0$ and residual $\mathbf{r}^0 = \mathbf{f}$, the algorithm progresses by sub-decomposing the k-th order residual in the form

$$\mathbf{r}^k = \left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle \mathbf{d}_{\ell_{k+1}} + \mathbf{r}^{k+1}, \tag{2}$$

where $\mathbf{d}_{\ell_{k+1}}$ is the atom corresponding to the index selected as

$$\ell_{k+1} = \underset{n=1,\ldots,M}{\arg\max} \left| \left\langle \mathbf{d}_n, \mathbf{r}^k \right\rangle \right|. \tag{3}$$

This atom is used to update the approximation $\mathbf{f}^k$ as

$$\mathbf{f}^{k+1} = \mathbf{f}^k + \left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle \mathbf{d}_{\ell_{k+1}}. \tag{4}$$

From (2) it follows that $\|\mathbf{r}^{k+1}\| \leq \|\mathbf{r}^k\|$, since

$$\|\mathbf{r}^k\|^2 = \left| \left\langle \mathbf{d}_{\ell_{k+1}}, \mathbf{r}^k \right\rangle \right|^2 + \|\mathbf{r}^{k+1}\|^2. \tag{5}$$

It is easy to prove that in the limit $k \to \infty$, the sequence $\mathbf{f}^k$ given in (4) converges to $\mathbf{f}$, if $\mathbf{f} \in V_M = \text{span}\{\mathcal{D}\}$, or to $\hat{P}_{V_M}\mathbf{f}$ the orthogonal projection of $\mathbf{f}$ onto V_M, if $\mathbf{f} \notin V_M$ (a pedagogical proof can be found on [18]). However, the method is not stepwise optimal because it does not yield an orthogonal projection at each step. Accordingly, the algorithm may select linearly dependent atoms, which is the main drawback of MP when applied with highly coherent dictionaries. A refinement to MP, which does yield an orthogonal projection approximation at each step is called OMP [16]. The method selects the atoms as in (3) but at each iteration produces a decomposition of the signal as given by:

$$\mathbf{f} = \mathbf{f}^k + \tilde{\mathbf{r}}^k = \sum_{n=1}^{k} c^k(n)\mathbf{d}_{\ell_n} + \tilde{\mathbf{r}}^k, \tag{6}$$

where the coefficients $c^k(n)$ are computed in such a way that it is true that

$$\sum_{n=1}^{k} c^k(n)\mathbf{d}_{\ell_n} = \hat{P}_{\mathtt{V}_k}\mathbf{f}, \quad \text{with} \quad V_k = \text{span}\{\mathbf{d}_{\ell_n}\}_{n=1}^{k}.$$

The superscript of $c^k(n)$ in (6) indicates the dependence of these quantities on the iteration step k. Thus, in addition to selecting linearly independent atoms, OMP yields the unique element $\mathbf{f}^k \in V_k$ minimising $\|\mathbf{f} - \mathbf{f}^k\|$. As discussed next, a further refinement to OMP, called OOMP [17], selects also the atoms in order to minimise in a stepwise manner the norm of the residual error.

2.1.2 The OOMP method

The OOMP approach iterates as follows. The algorithm is initialised by setting: $\mathbf{r}^0 = \mathbf{f}$, $\mathbf{f}^0 = 0$, $\Gamma = \emptyset$ and $k = 0$. The first atom is selected as the one corresponding to the index ℓ_1 such that

$$\ell_1 = \underset{n=1,\ldots,M}{\arg\max} \left| \left\langle \mathbf{d}_n, \mathbf{r}^0 \right\rangle \right|^2. \tag{7}$$

This first atom is used to assign $\mathbf{w}_1 = \mathbf{d}_{\ell_1} = \boldsymbol{\beta}_1$, calculate $\mathbf{r}^1 = \mathbf{f} - \mathbf{d}_{\ell_1} \langle \mathbf{d}_{\ell_1}, \mathbf{f} \rangle$, and iterate as prescribed below.

1) Upgrade the set $\Gamma \leftarrow \Gamma \cup \ell_{k+1}$, increase $k \leftarrow k+1$, and select the index of a new atom for the approximation as

$$\ell_{k+1} = \underset{\substack{n=1,\ldots,M \\ n \notin \Gamma}}{\arg\max} \frac{|\left\langle \mathbf{d}_n, \mathbf{r}^k \right\rangle|^2}{1 - \sum_{i=1}^{k} |\left\langle \mathbf{d}_n, \tilde{\mathbf{w}}_i \right\rangle|^2}, \quad \text{with} \quad \tilde{\mathbf{w}}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|_2}. \tag{8}$$

2) Compute the corresponding new vector $\mathbf{w}_{k+1}$ as

$$\mathbf{w}_{k+1} = \mathbf{d}_{\ell_{k+1}} - \sum_{n=1}^{k} \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|_2^2} \left\langle \mathbf{w}_n, \mathbf{d}_{\ell_{k+1}} \right\rangle. \tag{9}$$

including, for numerical accuracy, the re-orthogonalisation step:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_{k+1} - \sum_{n=1}^{k} \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|_2^2} \left\langle \mathbf{w}_n, \mathbf{w}_{k+1} \right\rangle. \tag{10}$$

3) Upgrade vectors $\boldsymbol{\beta}_n^k$ as

$$\boldsymbol{\beta}_{k+1}^{k+1} = \frac{\mathbf{w}_{k+1}}{\|\mathbf{w}_{k+1}\|^2}, \quad \boldsymbol{\beta}_n^{k+1} = \boldsymbol{\beta}_n^k - \boldsymbol{\beta}_{k+1}^{k+1} \left\langle \mathbf{d}_{\ell_{k+1}}, \boldsymbol{\beta}_n^k \right\rangle, \quad n = 1, \ldots, k. \tag{11}$$

4) Upgrade vector $\mathbf{r}^k$ as

$$\mathbf{r}^{k+1} \;=\; \mathbf{r}^k - \langle \mathbf{w}_{k+1}, \mathbf{f} \rangle \frac{\mathbf{w}_{k+1}}{\|\mathbf{w}_{k+1}\|^2}. \tag{12}$$

5) If for a given ρ value the condition $\|\mathbf{r}^{k+1}\| < \rho$ has been met stop the selection process. Otherwise repeat steps 1) - 5).

Calculate the coefficients in (1) as

$$c(n) = \left\langle \boldsymbol{\beta}_n^k, \mathbf{f} \right\rangle, \quad n = 1, \ldots, k.$$

Calculate the final approximation $\mathbf{f}^k$ as

$$\mathbf{f}^k = \mathbf{f} - \mathbf{r}^k.$$

Remark 1. *The difference between our implementations of OMP and OOMP is only that for OMP the denominator in* (8) *is eliminated. The selection is realised as in MP (c.f.* (3)*) which is not a stepwise optimal selection. Instead,* (8) *selects the atom* $\mathbf{d}_{\ell_{k+1}}$ *which, fixing the previously selected atoms* $\mathbf{d}_{\ell_i}$, $i = 1, \ldots, k$, *minimises the distance* $\|\mathbf{f} - \mathbf{f}^{k+1}\|$. *The proof can be found in [17].*

Remark 2. *Restricting consideration to sparse solutions,* $K \ll M$ *with* K *and* M *as in Sec. 2, the computational complexity of the Pursuit Strategies is dominated by the computation of the inner product in the selection process, i.e. is* $O(NM)$. *This complexity is high, in comparison with constructing the* K*-term approximation with a fast transform. The complexity of selecting the* K*-coefficients of larger magnitude of a discrete wavelet transform, for instance, is dominated for the sort operation, which has average computational complexity* $O(N \log N)$. *Unfortunately the wavelet decomposition would not be applicable in the proposed framework.*

2.2 Sparsity criterion for morphological feature extraction of heartbeats

A digital ECG signal is a sequence of heartbeats, each of which is characterised by a combination of three graphical deflections, known as QRS complex, and the so called P and T waves. An idealised illustration of these deflections are given in Fig. 1.

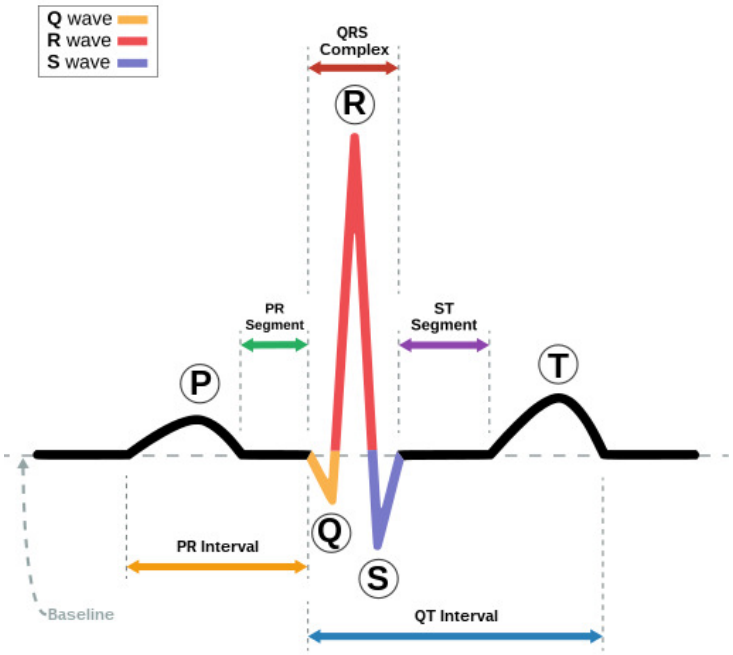

Figure 1: Public domain graph of the QRS complex and P and T waves representing a normal heartbeat.

However, as illustrated in Fig. 2, in real ECG records the shape of the beats in the same class may vary.

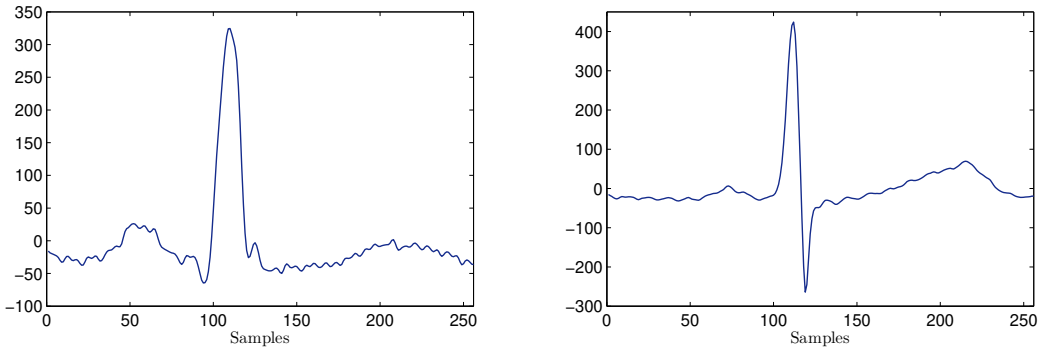

Figure 2: Two normal heartbeats from different ECG records

Approximation for morphological feature extraction of a heartbeat requires the segmentation of the QRS complex. In this work the segmentation is realised by taking a fixed number of samples to left and right of the location of the R-peak, to have beats of equal support, say N_s.

Assuming that one wishes to morphologically differentiate heartbeats of, say class A, from heartbeats of class B, and that a dictionary $\mathcal{D}_{\rm a}$, specially constructed to approximate a heartbeat of class A, and a dictionary $\mathcal{D}_{\rm b}$, specially constructed to approximate a heartbeat of class B, are given, we discuss next several possibilities of using sparsity as a distinguishability criterion.

In order to decide whether a heartbeat $\mathbf{f}_q \in \mathbb{R}^{N_s}$ belongs to class A or class B this beat is approximated, up to the same precision, using both dictionaries, i.e.

$$\mathbf{f}^a = \sum_{n=1}^{K_a} c^a(n)\mathbf{d}_{\ell_n^a}, \tag{13}$$

where the atoms $\mathbf{d}_{\ell_n^a}$ are chosen from $\mathcal{D}_{\rm a}$ and

$$\mathbf{f}^b = \sum_{n=1}^{K_b} c^b(n)\mathbf{d}_{\ell_n^b}, \tag{14}$$

where the atoms $\mathbf{d}_{\ell_n^b}$ are chosen from $\mathcal{D}_{\rm b}$.

Criterion I (a) and (b)

If $K_a < K_b$ the beat $\mathbf{f}_q$ is assigned to class A.
If $K_b < K_a$ the beat $\mathbf{f}_q$ is assigned to class B.
If $K_a = K_b$ the criterion does not make a decision.

In the event that $K_a = K_b$ a decision could still be made by recourse to a different metric.

(a) *Smaller entropy criterion:* Assigning $p^a(n) = \frac{|c^a(n)|}{\|\mathbf{c}^a\|_1}$ and $p^b(n) = \frac{|c^b(n)|}{\|\mathbf{c}^b\|_1}$ calculate the corresponding Shannon's entropies

$$S^a = -\sum_{n=1}^{K_a} p^a(n)\ln(p^a(n)),$$

$$S^b = -\sum_{n=1}^{K_b} p^b(n)\ln(p^b(n)).$$

The entropy of, say $p^a(n)$, $n = 1, \ldots, K_a$, would be smaller if the components are fewer and the magnitude of some $p^a(n)$ much larger than others. In particular the minimum entropy value occurs if $p^a(n) = \frac{|c^a(n)|}{\|\mathbf{c}^a\|_1} = 1$ for some n and zero otherwise, in which case $S^a = 0$. Accordingly, if $S_a < S_b$ the beat $\mathbf{f}_q$ might be assigned to class A and if $S_b < S_a$ the beat $\mathbf{f}_q$ might be assigned to class B. Let us recall that the criterion of smaller entropy was introduced for basis selection in the context of wavelet packets [23].

(b) *Smaller norm-1 criterion.*
This criterion is in line with the *basis pursuit approach* [14] which adopts the minimisation of the norm-1 as a way of producing a tractable sparse solution from a given dictionary. Consequently, if $\|\mathbf{c}^a\|_1 < \|\mathbf{c}^b\|_1$ the beat $\mathbf{f}_q$ is assigned to class A and if $\|\mathbf{c}^b\|_1 < \|\mathbf{c}^a\|_1$ the beat $\mathbf{f}_q$ if be assigned to class B.

Criterion II

Use always the smaller entropy criterion.

Criterion III

Use always the smaller norm-1 criterion.

2.3 Dictionary learning

So far we have assumed to know the dictionaries $\mathcal{D}_{\rm a}$ and $\mathcal{D}_{\rm b}$ best suited for each class of heartbeat. In this section we discuss how these dictionaries can be learnt from a data set of annotated heartbeats.

The adopted strategy to learn a dictionary as a matrix $\mathbf{D}$, using the heartbeats placed in an array $\mathbf{F} \in \mathbb{R}^{N_s \times Q}$, proceed through a 2 step process. All the superscripts k within the algorithm indicate the corresponding quantity at iteration k. Setting $k = 0$, and given an initial dictionary $\mathbf{D}^k$, the algorithm iterates as below using the selected pursuit method.
Step 1

- Normalise the columns of the matrix $\mathbf{D}^k \in \mathbb{R}^{N_s \times M}$ and apply the greedy pursuit method to approximate each beat $\mathbf{f}_q \in \mathbb{R}^{N_s}$, for $q = 1, \ldots, Q$, as in (13) or (14). Collect the resulting vectors of coefficients $\mathbf{c}_q^k \in \mathbb{R}^{K_q}$.
- Place each vector $\mathbf{c}_q^k$ as a column of a matrix $\mathbf{C}^k \in \mathbb{R}^{M \times Q}$ having nonzero elements $C(\ell_n^q, q)^k = c_q(n)^k$, $n = 1, \ldots, {K_q}^k$, $q = 1, \ldots, Q$ and set all the other elements equal to zero.
- Using matrix $\mathbf{C}^k$ and dictionary $\mathbf{D}^k$ calculate the approximation $\tilde{\mathbf{F}}$ of $\mathbf{F}$ as

$$\tilde{\mathbf{F}}^k = \mathbf{D}^k \mathbf{C}^k$$

Step 2

- Using the approximation $\tilde{\mathbf{F}}^k$ find the updated dictionary $\tilde{\mathbf{D}}$ minimising $\|\mathbf{F} - \tilde{\mathbf{F}}^k\|_F^2$. Thus,

$$\tilde{\mathbf{D}} = \arg\min_{\mathbf{D}} \|\mathbf{F} - \mathbf{D}\mathbf{C}^k\|_F^2 = \mathbf{F}{\mathbf{C}^k}^\top (\mathbf{C}^k {\mathbf{C}^k}^\top)^{-1}. \tag{15}$$

- Given a maximum number of iterations, MaxIt say, and a tolerance tol for the error norm, if $\|\tilde{\mathbf{D}} - \mathbf{D}^k\|_F < \text{tol}$ or $k >$ MaxIt stop. Otherwise set $k \to k + 1$, $\mathbf{D}^k = \tilde{\mathbf{D}}$, and repeat *Steps 1* and *2*.

Remark 3. *The dictionary updating equation* (15) *requires that matrix* $\mathbf{C}^k {\mathbf{C}^k}^\top$ *should have an inverse. This would not be true if some elements in the dictionary were not chosen in the previous step. In that case, the unselected atoms should be removed from the dictionary before implementing equation* (15)*. In practice, as long as* Q *is significantly larger than* M *and the examples are independent, matrix* $\mathbf{C}^k {\mathbf{C}^k}^\top$ *has an inverse.*

Remark 4. *The problems of determining matrix* $\mathbf{C}^k$ *at Step 1 and matrix* $\tilde{\mathbf{D}}$ *at Step 2 are convex problems with unique solution. However, the combined problem of determining matrix* $\mathbf{C}^k$ *and* $\tilde{\mathbf{D}}$ *is not jointly convex. Thus, the solution depends on the initial dictionary. As will be demonstrated by the simulations, this is not crucial within the context of the proposed approach.*

3 Binary morphological differentiation of heartbeats

In this section we test the proposal by differentiating the classes `N` (Normal) and `V` (Ventricular Ectopic) in the MIT-BIH Arrhythmia data set [19, 20]. As per the recommendations given by the Advancement of Medical Instrumentation (AAMI) the MIT-BIH Arrhythmia database is projected into five classes. Within these classes the `N` beats considered here include: The normal beats N, the Left Bundle Branch Block Beats (LBB) and the Right Bundle Branch Block Beats (RBB). The `V` class comprises: the Premature Ventricular Contraction (PVC) and Ventricular Escape Beats (VEB). While the number of `V` beats is much less than the number of `N` beats, the former is still enough to learn the corresponding dictionary. The locations of the R-peaks are retrieved from the annotations provided with the dataset, using the Matlab software [21] available on [19]. The peaks are then segmented in order to capture the QRS complex and P and T waves in a normal heartbeat (c.f. Fig. 1). Since the distance from the R-peak to the end of the T wave is normally slightly larger than the distance to the end of the P wave, the segmentation of a heartbeat is realised by taking 145 samples to the right and 110 samples to the left of the R-peak location (please see Fig. 2). Before learning the dictionary a quick check of the training set is realised, in order to find out if there are some peaks that would not qualify as 'good' examples. For this end we proceed as explained below.

Screening the data training sets

For automatic screening of a set the segmented beats of the same class are approximated, up to the same quality, using a greedy pursuit algorithm and a general wavelet dictionary. We illustrate the process by giving the details for screening the training set for the numerical Test I.

Wavelet dictionaries arise from translations, by a parameter $b = 2^{-l}$, $l \in \mathbb{N}$ of scaling prototypes [24, 25]

$$\phi_{j_0,k,b}(x) = \phi\left(2^{j_0}x - bk\right), \quad k \in \mathbb{Z}, \tag{16}$$

and wavelet prototypes at different scales

$$\psi_{j,k,b}(x) = \psi\left(2^{j}x - bk\right), \quad k \in \mathbb{Z}, \quad j \geq j_0. \tag{17}$$

For screening of the training set we have used the 9/7 Cohen-Daubechies-Feauveau (**cbf97**) wavelet family with $b = 0.25$, $j_0 = 2$ and $j = 3, 4, 5, 6$, which introduces a redundancy factor of 2.67. This dictionary was generated with the software described in [26] available on [27]. The scaling and wavelet prototypes for the **cbf97** wavelet family are shown in Fig. 3.

The quality in the approximation of each beat is fixed using the prdn metric as defined by

$$\text{prdn}(q) = \frac{\|\mathbf{f}_q - \mathbf{f}_q^k\|}{\|\mathbf{f}_q - \bar{\mathbf{f}}_q\|} \times 100\%, \quad q = 1, \ldots, Q,$$

where each $\mathbf{f}_q$ is a heartbeat, $\mathbf{f}_q^k$ its corresponding approximation by k atoms, and $\bar{\mathbf{f}}_q$ the mean value of $\mathbf{f}_q$. The approximation of different beats, up to the same prdn, is achieved for different values of k. The left graph of Fig.4 shows the histograms of k values obtained when approximating, up to prdn = 9, the `N` beats in the training set of the numerical Test I (c.f. Sec. 3.1).

Even if not very noticeable, in the histogram there are a few values of k very far from the main support. These values are from 'rare' signals that should be investigated. In this set they

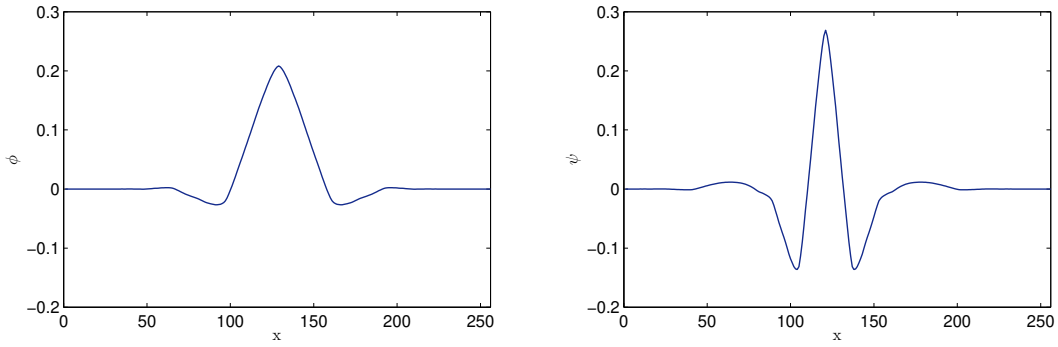

Figure 3: Graph of the scaling prototype $\phi(x)$ (left) and wavelet prototype $\psi(x)$ (right) for the **cbf97** family.

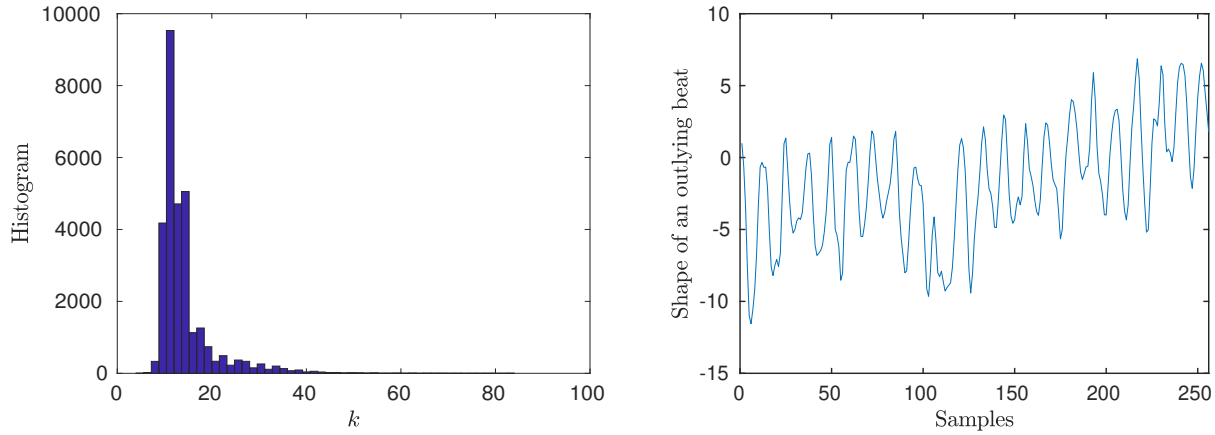

Figure 4: Histogram of k values for approximating (up to prdn = 9) the N-signals in the training set (left graph). Signal corresponding to the outlying value $k = 63$ (right graph).

are just signals looking as pure noise, e.g. the signal in the right graph of Fig 4 corresponds to $k = 61$, very far from the mean value, $\overline{k} = 14.31$ with standard deviation std $= 5.90$. When learning the dictionary for classification of N beats we disregard beats in the training set producing values of k outside the range $(\overline{k} - 3\,\text{std}, \overline{k} + 3\,\text{std})$. This amounts to disregarding 2.58% of the total N beats in the training set.

The left graph of Fig.5 shows the histogram of k values obtained when approximating, up to prdn = 9, the V beats in the training set of the numerical Test I. For this class $\overline{k} = 10.61$ with std $= 3.28$. Because the data set of V beats is smaller than the previous one, when learning the dictionary for classification of V beats we disregard beats in the training set outside the range $(\overline{k} - 2\,\text{std}, \overline{k} + 2\,\text{std})$. This amounts to disregarding 2.65% of the V beats in the training set.

3.1 Numerical Test I

The purpose of this test is to assess the suitability of binary morphological differentiation of heartbeats using dictionaries learnt from examples of N and V shapes in the dataset. For this end we randomly split the $Q = 87277$ N beats into two groups: 30000 N beats are used for training the dictionary, $\mathbf{D}_n$ say. The remaining 57277 N beats are reserved for testing. Since the V beats are much less than the N ones, 50% of the V beats are used for training the dictionary

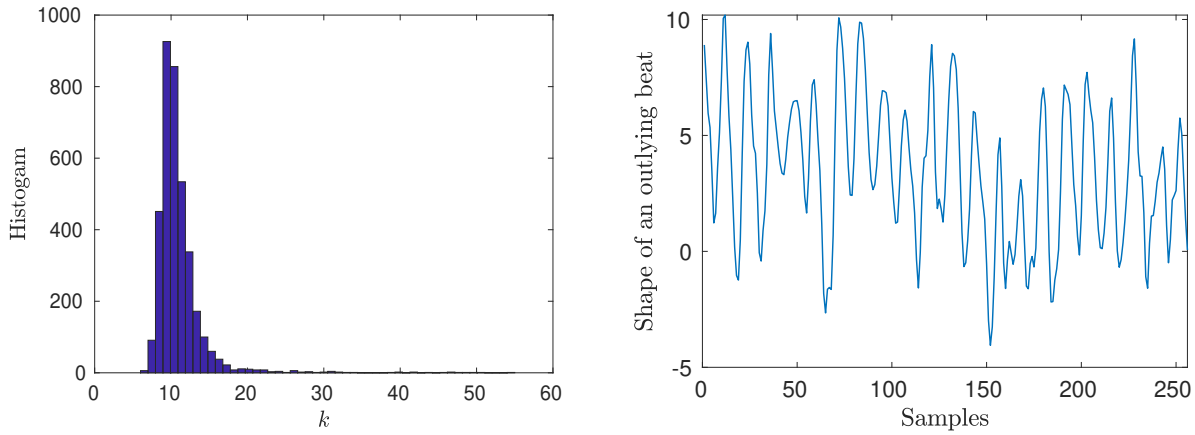

Figure 5: Histogram of k values for approximating (up to prdn = 9) the V-signals in the training set (left graph). Signal corresponding the outlying value $k = 41$ (right graph).

Sets	Training		Testing	
Class	N	V	N	V
Number of beats	30000	3359	57277	3359

Table 1: Number of heartbeats of classes N and V in the training and testing sets for Test I.

$\mathbf{D}_v$, and the remaining 50% for testing (c.f. Table 1)

The dictionaries are learned to have redundancy two, i.e. each dictionary is a matrix of real numbers of size 256×512. We test the method against:
(i) The initial dictionary.
(ii) The different greedy pursuit algorithms considered in this work: MP, OMP, OOMP.

For this test we randomly take 512 beats from the 30000 N heartbeats in the training set to construct a matrix $\mathbf{D}_n$ of size 256×512, which is used as initial dictionary. The learning curves for the dictionary $\mathbf{D}_n$ with the 3 greedy algorithms are shown in left graph of Fig.6. We repeat the process but taking the 512 dictionary atoms randomly from the 3359 V beats in the training set. The learning curves for this dictionary are shown in the right graph of Fig.6.

The identification of V-beats performance is assessed by means of the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes. These values are used to calculate the following statistics metrics.

Sensitivity (SE): Number of correctly identified heartbeats, among the total number of that class in the set, i.e.

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}.$$

Specificity (SP): Number of correctly rejected heartbeats which are not in that class, i.e.

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}.$$

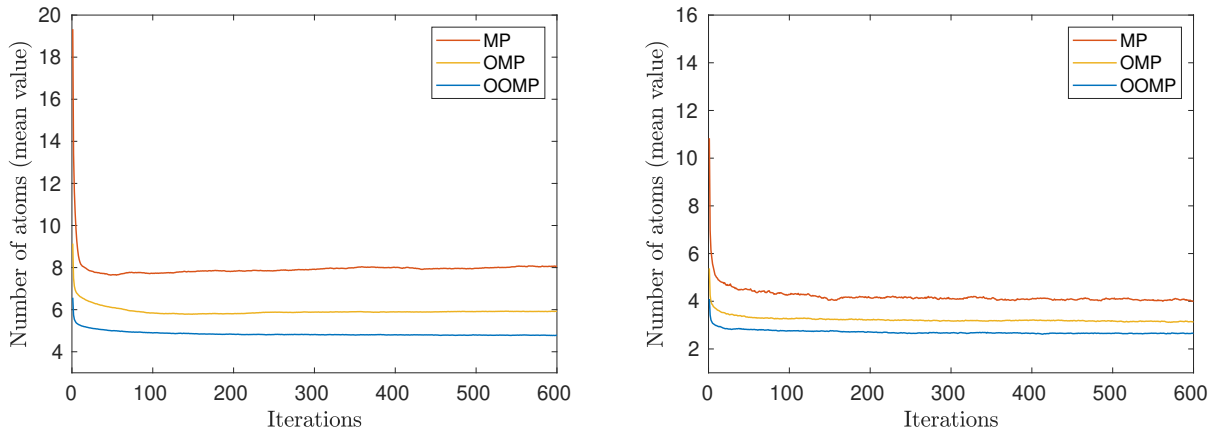

Figure 6: Dictionary learning curves with the three greedy algorithms. The left graph corresponds to dictionary $\mathbf{D}_n$ and the right graph to dictionary $\mathbf{D}_v$.

Positive predictivity (PP): Ratio of correctly identified heartbeats to all the beats classified as `V`.

$$\mathrm{PP} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}.$$

The combination of these metrics produces the F_1 score which is defined as follows:

$$\mathrm{F}_1 = \frac{2\,\mathrm{TP}}{2\,\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} = \frac{2\,\mathrm{SE}\,\mathrm{PP}}{\mathrm{SE} + \mathrm{PP}}.$$

Additionally, the total accuracy (AC) of the classification is calculated as the fraction of correctly classified `N` and `V` heartbeats.

$$\mathrm{AC} = \frac{\text{Correctly classified beats in both classes}}{\text{Total number of beats}}.$$

In Table 2 these scores are given as the mean value of 5 realisations corresponding to 5 random initialisations in the dictionary learning process. The standard deviations (std) are shown in rows 4,6,8,10, and 14 of Table 2. The Matlab codes for implementing the approach with the three greedy algorithms have been made available on [28].

Discussion of results

i) The high SE values for both classes, indicates that: 34.4% of the whole `N` beats in the data set and 50% of the whole `V` beats in the data set provide enough examples in the training set to learn dedicated dictionaries for each class.

ii) From Table 2 we can assert that each of the greedy algorithms performs better with a particular decision criterion.

iii) MP, combined with the decision Criterion III (norm-1), produces the highest statistics scores.

	MP				OMP				OOMP			
Crit.	I (a)	I (b)	II	III	I (a)	I (b)	II	III	I (a)	I (b)	II	III
$SP_{\mathtt{V}}$ (%)	**99.7**	**99.7**	**99.7**	**99.7**	99.1	99.1	99.3	**99.4**	**98.8**	**98.8**	98.5	97.4
std	0.05	0.05	0.04	0.03	0.09	0.09	0.01	0.04	0.14	0.13	0.16	0.45
$SE_{\mathtt{V}}$ (%)	95.7	95.7	**96.9**	**97.6**	96.5	96.5	96.7	**97.0**	**97.6**	**97.6**	97.3	95.3
std	0.65	0.63	0.45	0.34	0.24	0.18	0.27	0.25	0.35	0.21	0.47	1.17
$PP_{\mathtt{N}}$ (%)	99.7	99.7	99.8	**99.9**	**99.8**	**99.8**	**99.8**	**99.8**	**99.9**	**99.9**	99.8	99.7
std	0.04	0.04	0.03	0.02	0.01	0.01	0.02	0.01	0.02	0.01	0.03	0.07
$PP_{\mathtt{V}}$ (%)	**94.9**	**94.9**	94.6	94.1	86.4	86.6	88.6	**90.4**	**82.5**	**82.5**	79.5	68.4
std	0.71	0.71	0.55	0.48	1.27	1.27	0.17	0.55	1.56	1.72	1.72	3.48
$F_{1\mathtt{N}}$ (%)	**99.8**	**99.8**	**99.8**	**99.8**	**99.4**	**99.4**	96.4	97.0	99.3	99.3	99.1	98.5
$F_{1\mathtt{V}}$ (%)	95.3	95.3	95.7	**95.8**	91.2	91.3	92.5	**93.4**	**89.4**	**89.4**	87.5	79.6
AC (%)	**99.5**	**99.5**	**99.5**	**99.5**	99.0	99.0	99.1	**99.3**	**98.7**	**98.7**	98.5	97.3
std	0.03	0.03	0.02	0.03	0.01	0.01	0.02	0.05	0.12	0.13	0.12	0.38

Table 2: Average statistics scores (of 5 different initialisations in the dictionary learning procedure) for classes `N` and `V`, obtained with MP, OMP, and OOMP, and the sparsity criteria discussed in Sec. 2.2. std indicates the corresponding standard deviations.

iv) Being the OOMP approach stepwise optimal, in the sense of selecting at each step the atom which minimises the norm of the residual error, it is not surprising that OOMP secures the least number of atoms for the approximation of beats in th learning process (c.f. Fig.6). It is also not surprising it is the approach that works better with Criterion I. However, there is no reason to expect that OOMP should render all the highest classification scores. This is because, within this framework, what matters for the identification of a heartbeat is the *relative* sparsity value with respect to the 2 dictionaries.

v) For most of the scores, the random initialisation of the learning process does not change the results in any significant manner (low std values).

vi) The comparatively lower values of $PP_{\mathtt{V}}$ reflex the enormous difference of the number of `N` and `V` beats in the testing sets (57277 vs 3359 beats). Thus, even if the percentage of incorrectly classified `N` beats is small, the classification being binary implies that the number of incorrectly identified `N` beats count as false positive `V`. This causes the low $PP_{\mathtt{V}}$ values in comparison to the other statistics metrics and, consequently, also comparatively low values of $F_{1\mathtt{V}}$ scores. The next remark follows up this discussion.

Remark 5. *We would like to highlight with an example the fact that, in binary classification, the difference in the number of beats in each class plays a crucial part in the value of* F_1*. We see from Table 2 that, when the testing set comprises 57277* `N` *beats and 3359* `V` *beats, the highest value of* $F_{1\mathtt{V}}$ *is* 95.8% *and the highest value of* $F_{1\mathtt{N}}$ *is* 99.8%*. Keeping the same* `V` *beats we now reduce, from 30000 to 15000, the number* `N` *beats for training the* `N` *dictionary. In oder to maintain the same proportion of training and testing* `N` *beats as before, we reduce the testing set to have 28638* `N` *beats. The number of testing* `V` *beats remains 3359. Now the best values of* $F_{1\mathtt{V}}$ *and* $F_{1\mathtt{N}}$*, which are achieved with MP and Criterion III, become* 97.6% *and* 99.7%*, respectively,*

while the values of $\mathrm{SE}_{\mathtt{V}}$ *and* $\mathrm{SP}_{\mathtt{V}}$ *remains almost the same (*97.3% *and* 99.8%*, respectively). This example also illustrates the fact that comparisons between statistical quantities produced by different approaches are only relevant when comparing identical classes and identical splitting of a data set.*

In order to place our results in the context of previous works, in Table 3 we reproduce other approaches scores for ventricular contraction detection. However, it should be taken into account that even if all the approaches use the MIT-BIH Arrhythmia data set, [30] and [32] only consider the records involving `V` beats. Whilst [30] considers 27 records [32] considers 22. Moreover, in this test we have included LBB and RBB beats in the `N` category, which increments the number of `N` beats.

	AC (%)	$\mathrm{SE}_{\mathtt{V}}$ (%)	$\mathrm{SP}_{\mathtt{V}}$ (%)	$\mathrm{F}_{1\mathtt{V}}$ (%)
Wavelet statistics [29]	97.8	87.2	99.0	86.0
Six features and SVM (only 27 records) [30]	99.8	99.9	99.4	—
Recurrent Unit Network [31]	97.9	98.0	97.8	—
2D Transfer Learning (only 22 records) [32]	99.9	99.9	99.5	99.8
Our scores with MP (all 44 records)	99.5	97.6	99.7	95.8

Table 3: The first row are the scores arising from statistics on wavelet coefficients, as reported in [29]. The second row are the results, from 27 records, using six features and Support Vector Machine (SVM) classifier [30]. The 3rd row are results produced by a Recurrent Unit Network [31] and the next row by 2D Transfer Learning from 22 records [32]. The last row shows our results from all 44 records with MP.

3.2 Test II

We test now the proposal in a more challenging situation. Because the morphology of `N` and `V` beats depends on the particular ECG signal, a realistic test is carried out by taking the training set and testing set from different records (corresponding to different patients).

Following a classic paper [22], the MIT-BIH Arrhythmia data set is divided for training (DS1) and testing (DS2) as below.

The training set DS1 is taken only from the records:
101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230.

The remaining records provide the testing set DS2. These are the records
100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234.

The same sets for training and testing are used in [33] and [34] for ventricular contraction detection. Since in those papers the LBB and RBB beats are not included, for the sake of comparison in this numerical test we only use the beats originally labelled as `N` beats. For learning the `N` dictionary we take randomly 15,000 `N` beats from the set DS1 and for learning the `V` dictionary the 3578 `V` beats in that set. All the `N` and `V` beats in DS2 are used for testing the approach. The resulting statistics scores are shown in Table 4.

	MP				OMP				OOMP			
Crit.	I (a)	I (b)	II	III	I (a)	I (b)	II	III	I (a)	I (b)	II	III
$SP_{\mathtt{V}}$ (%)	96.6	96.6	96.9	**97.8**	91.2	91.7	92.8	**96.4**	94.0	**94.3**	92.8	92.2
$SE_{\mathtt{V}}$ (%)	86.5	86.6	90.0	**92.4**	91.4	91.0	**91.5**	90.0	93.8	**94.3**	92.8	92.2
$PP_{\mathtt{V}}$ (%)	69.0	69.2	71.7	**75.6**	47.7	49.2	52.9	**69.0**	57.8	**59.4**	45.6	36.7
$F_{1\mathtt{V}}$ (%)	76.7	76.9	79.8	**83.2**	62.7	63.8	67.0	**78.1**	71.5	**72.8**	61.2	52.5
AC (%)	95.7	95.8	96.3	**97.0**	91.2	95.3	95.9	**97.7**	93.9	**94.3**	90.4	86.4

Table 4: Statistics scores for `V` heartbeat characterisation, obtained with MP, OMP, and OOMP, for the numerical Test II, in which training and testing sets correspond to different records.

Discussion of results

i) As expected, the classifications scores are lower than in Test I. This is because not all the shapes of the beats in the testing set bears similarity with shapes present in the training set.

ii) Also in this test each of the greedy algorithms performs better with a particular decision criterion.

iii) The greedy algorithm OOMP is the only one that performs better with Criterion I.

iv) Clearly MP works best with Criterion III and for most scores OMP as well.

v) The low values of $PP_{\mathtt{V}}$ have the same cause as in Test I. Here the values are even lower because the percentage of incorrectly classified `N` beats is higher than in Test I. Consequently, the values of $F_{1\mathtt{V}}$ are also smaller.

With the aim of placing our statistics scores into context, we reproduce in Table 5 the results in [33] and [34]. Table 6 shows the number of `N` and `V` beats reported in those publication. While the splitting of the records is identical, the beats in [34] are less than in [33] and in this work.

	AC (%)	$SE_{\mathtt{V}}$ (%)	$SP_{\mathtt{V}}$ (%)	$F_{1\mathtt{V}}$ (%)
Geometric features and AIS [33]	98.4	91.1	98.7	88.3
Geometric features (average of 8 classifier) [33]	97.2	78.9	98.9	82.6
Deep Metric Learning and KNN [34]	99.7	97.4	99.9	97.8
Our scores with MP	97.0	92.4	97.8	83.2

Table 5: Statistics scores corresponding to the DS2 testing set. The first and second rows are the values arising from geometric features and Artificial Immune System (AIS) classifier and average of 8 classifiers, as reported in [33]. The next row shows the scores produced from automatic feature extraction and K-Nearest Neighbours (KNN) classifier, as reported in [34]. The last row are our scores obtained with MP.

	Number of N beats	Number of V beats
Training set [33]	38,087	3683
Testing set [33]	36,428	3219
Training set [34]	35,640	2851
Testing set [34]	33,868	2548
Training set (this work)	38,086	3578
Testing set (this work)	36,427	3218

Table 6: Number of beats in the DS1 and DS2 sets used in [33], [34], and this work.

4 Conclusions

A set up for binary morphological characterisation of heartbeats on the basis of sparsity metrics and labelled dictionaries has been laid out. The proposal was tested for identification of N and V beats in the MIT-BIH Arrhythmia data set, achieving high scores in true positive identification of the beats in the each class. Due to the huge difference in the number of N and V beats the F_{1V} score is comparatively lower.

The results are encouraging because the proposed binary identification is realised outside the usual machine leaning framework, using sparsity as a *single* parameter for characterising the morphology of a beat. This parameter can be used in combination with other features to feed typical machine learning classifiers. It is in this context in which the approach should be appreciated, because a morphological feature in isolation does not provide complete information for classification of different types of heartbeats. The main limitation of the approach is the computational complexity of the approximation process. However, since the approximation is carried out beat by beat, there is scope for straightforward parallelisation of the step.

The possibility of implementing the technique depends on the availability of 'enough' examples to learn the corresponding dictionaries. In the numerical tests realised here, for instance, 3359 examples for beats V were enough to learn the dictionary $\mathbf{D}_v \in \mathbb{R}^{256\times 512}$. The susceptibility of the score F_{1V} to false negative N outcomes indicates that, in binary classification, the proportion between the two classes should be considered in order to assess the suitability of the approach.

Acknowledgements

The authors would like to express their gratitude to an anonymous Reviewer for kindly producing a long list of supportive and useful corrections.

References

[1] U. R. Acharya, Shu Lih Oh, Y. Hagiwara, Jen Hong Tan, M. Adam, A. Gertych, and Ru San Tan, "A deep convolutional neural network model to classify heartbeats", *Computers in Biology and Medicine*, **89** 389–396 (2017).

[2] A. Mincholé and B. Rodriguez, "Artificial intelligence for the electrocardiogram" *Nature Medicine*, **25**, 22–23 (2019) https://doi.org/10.1038/s41591-018-0306-1

[3] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network", *Nature Medicine*, **25**, 65–69 (2019).

[4] A. H. Ribeiro, M. H. Ribeiro, G. M. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. S. Ferreira, C. R. Andersson, P. W. Macfarlane, W Meira Jr., T. B. Schön, and A. L. P. Ribeiro, "Automatic diagnosis of the 12-lead ECG using a deep neural network", *Nature Communications*, **11**, 1760 (2020) https://doi.org/10.1038/s41591-018-0268-3.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning". *Nature* **521**, 436–444 (2015).

[6] X. Han, Y. Hu, L. Foschini, L. Chinitz, L. Jankelson, and R. Ranganath, "Deep learning models for electrocardiograms are susceptible to adversarial attack", *Nature Medicine,* **26**, 360–363 (2020).

[7] L. Ma and L. Liang, "A regularization method to improve adversarial robustness of neural networks for ECG signal classification", *Computers in Biology and Medicine* **144**, (2020) https://doi.org/10.1016/j.compbiomed.2022.105345.

[8] E. J. Luz, W. R. Schwartz, G. Cámara-Chávez, and, D. Mennotti, "ECG-based heartbeat classification for arrhythmia detection: A survey", *Computer Methods and Programs in Biomedicine* **127**, 144–164 (2016) doi: https://doi.org/10.1016/j.cmpb.2015.12.008.

[9] U.R. Acharya, H. Fujita, M. Adam, S.L. Oh, K.V. Sudarshan, J.H. Tan, J.E.W. Koh, Y. Hagiwara, C.K. Chua, C.K. Poo, and R.S. Tan, "Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: A comparative study", *Information Sciences*, **377** (2017) 17–29. doi:https://doi.org/10.1016/j.ins.2016.10.013.

[10] A. Lyon, A. Mincholé, J. P. Martínez, P. Laguna, and B. Rodriguez, "Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances", *Journal of the Royal Society of Interface*, **15** (2018), article No. 20170821 https://doi.org/10.1098/rsif.2017.0821.

[11] E. Merdjanovska and A. Rashkovska, "Comprehensive survey of computational ECG analysis: Databases, methods and applications", *Expert Systems With Applications* **203**, 117206 (2022).

[12] J. Wright, A. Y. Yang, A. Ganesh, S. S, Sastry, and Yi Ma, "Robust Face Recognition via Sparse Representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* **31**, 210–227 (2008).

[13] S. Raj and K. Ch. Ray, "Sparse representation of ECG signals for automated recognition of cardiac arrhythmias", *Expert Systems with Applications*, **105**, 49–65 (2018).

[14] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, **20**, 33–61 (1998).

[15] S. G. Mallat and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries", *IEEE Trans. on Signal Processing*, **41**, 3397–3415 (1993).

[16] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Proceedings of the 27th Annual Asilomar Conference in Signals, System and Computers*, **1**, 40–44, (1993).

[17] L. Rebollo-Neira and D. Lowe, "Optimised orthogonal matching pursuit approach", *IEEE Signal Process. Letters*, **9**, 137–140 (2002).

[18] L. Rebollo-Neira, M. Rozložník, and P. Sasmal, "Analysis of the Self Projected Matching Pursuit Algorithm", *Journal of The Franklin Institute,* **357**, 8980–8994 (2020).

[19] `https://physionet.org/physiobank/database/mitdb/` (Last access Jan 2023).

[20] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database", *IEEE Eng in Med and Biol,* **20**, 45–50 (2001)..

[21] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J.E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals." *Circulation* [Online]. **101** e215—e220 (2000).

[22] P. de Chazal, M. O'Dwyer and R. B. Reilly, "Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features", *IEEE Trans. on Biomedical Engineering*, **51**, 1196–1206 (2004)

[23] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection", *IEEE Trans. on Information Theory* **38**, 713 –718 (1992).

[24] M. Andrle and L. Rebollo-Neira, From cardinal spline wavelet bases to highly coherent dictionaries, *Journal of Physics A* **41** (2008), article No. 172001. doi:10.1088/1751-8113/41/17/172001.

[25] L. Rebollo-Neira and D. Černá, "Wavelet based dictionaries for dimensionality reduction of ECG signals", *Biomedical Signal Processing and Control*, **54** (2019), article No. 101593. doi:https://doi.org/10.1016/j.bspc.2019.101593.

[26] D. Černá and L. Rebollo-Neira, "Construction of wavelet dictionaries for ECG modeling" *MethodsX,* **8**, 101314 (2021).

[27] http://www.nonlinear-approx.info/examples/node013.html

[28] http://www.nonlinear-approx.info/examples/node016.html

[29] Y. Jung, H. Kim, "Detection of PVC by using a wavelet-based statistical ECG monitoring procedure", *Biomedical Signal Processing and Control*, **36**, 176–182 (2017).

[30] H. M. Mazidi, M. Eshghi, M. R. Raoufy, "Detection of premature ventricular contraction (PVC) using linear and nonlinear techniques: an experimental study", *Cluster Computing*, **23**, 759–774 (2020).

[31] J. Wang, "Automated detection of premature ventricular contraction based on the improved gated recurrent unit network", *Computer Methods and Programs in Biomedicine,* **208,** 106284a (2021).

[32] H. Ullah, M. B. B Heyat, F. Akhta, A. Y. Muaad, C. C. Ukwuoma, M. Bilal, M. H. Miraz, M. A. S Bhuiyan, K. Wu, R. Damaševičius R., et al., "An Automatic Premature Ventricular Contraction Recognition System Based on Imbalanced Dataset and Pre-Trained Residual Network Using Transfer Learning on ECG Signal", *Diagnostics* **13,** 87 (2023), https://doi.org/10.3390/diagnostics13010087

[33] B. Rodrigues de Oliveira, C. C. Enside de Abreu, M. A. Queiroz Duarte, J. Vieira Filho, "Geometrical features for premature ventricular contraction recognition with analytic hierarchy process based machine learning algorithms selection", *Computer Methods and Programs in Biomedicine,* **169** 59–69 (2019).

[34] J. Yu, X. Wang, X. Chen, J. Guo, "Automatic Premature Ventricular Contraction Detection Using Deep Metric Learning and KNN", *Biosensors*, **11**, 69, (2021) https://doi.org/10.3390/bios11030069.