HAL
open science

Characterization of myocardial motion patterns by unsupervised multiple kernel learning

Sergio Sanchez-Martinez, Nicolas Duchateau, Tamas Erdei, Alan Fraser, Bart Bijnens, Gemma Piella

To cite this version:

Sergio Sanchez-Martinez, Nicolas Duchateau, Tamas Erdei, Alan Fraser, Bart Bijnens, et al.. Characterization of myocardial motion patterns by unsupervised multiple kernel learning. Medical Image Analysis, 2017, 35, pp.70-82. 10.1016/j.media.2016.06.007 . hal-01331550

HAL Id: hal-01331550

https://inria.hal.science/hal-01331550

Submitted on 14 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

Characterization of Myocardial Motion Patterns by Unsupervised Multiple Kernel Learning

Sergio Sanchez-Martinez[a,*], Nicolas Duchateau[c], Tamas Erdei[d], Alan G. Fraser[d], Bart H. Bijnens[a,b], Gemma Piella[a]

[a]*Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain*
[b]*ICREA, Barcelona, Spain*
[c]*Inria Asclepios research project, Sophia Antipolis, France*
[d]*Wales Heart Research Institute, Cardiff University, United Kingdom*

Abstract

We propose an independent objective method to characterize different patterns of functional responses to stress in the heart failure with preserved ejection fraction (HFPEF) syndrome by combining multiple temporally-aligned myocardial velocity traces at rest and during exercise, together with temporal information on the occurrence of cardiac events (valves openings/closures and atrial activation). The method builds upon multiple kernel learning, a machine learning technique that allows the combination of data of different nature and the reduction of their dimensionality towards a meaningful representation (output space). The learning process is kept unsupervised, to study the variability of the input traces without being conditioned by data labels. To enhance the physiological interpretation of the output space, the variability that it encodes is analyzed in the space of input signals after reconstructing the velocity traces via multiscale kernel regression. The methodology was applied to 2D sequences from a stress echocardiography protocol from 55 subjects (22 healthy, 19 HFPEF and 14 breathless subjects). The results confirm that characterization of the myocardial functional response to stress in the HFPEF syndrome may be improved by the joint analysis of multiple relevant features.

Keywords: Myocardial motion, echocardiography, multiple kernel learning, pattern analysis

1. Introduction

Multiple feature analysis has proved to be helpful to understand the information embedded in complex data. We demonstrate the value of this kind of analysis to unravel patterns of cardiac mechanical dysfunction, focusing on the heart failure with preserved ejection fraction (HFPEF) syndrome, which is pathophysiologically heterogeneous and difficult to diagnose (Shah et al., 2014; Komajda and Lam, 2014). We illustrate the relevance of our approach to improve the understanding of this syndrome.

1.1. Multiple features analysis

Clinical guidelines provide consensus indications to guide diagnosis in daily clinical practice. However, the measurements that they suggest are kept to simple parameters that have been shown to correlate with disease status (Borlaug and Paulus, 2011; Penicka et al., 2014). The number of studies that include more advanced statistical tools is rather limited. Multiple stepwise logistic regression was previously used within the context of our application (Mädler et al., 2003), and multiparametric approaches were proposed to predict response to cardiac resynchronization therapy (CRT) (Santaularia-Tomas and Abraham, 2009; Lafitte et al., 2009). Nonetheless, all these approaches have notable limits for the advanced diagnosis of complex diseases. Indeed, they all seek for simple key markers of the disease, such as peak measurements or specific timings, while induced changes tend to be more complex, both spatially and temporally. A more complete analysis should therefore target the study of cardiac function at the pattern level, as was previously recommended to improve the prediction of response to CRT (Fornwalt et al., 2009) and to reach a deeper understanding of myocardial mechanics and physiology (Bijnens et al., 2012).

*Corresponding author at: DTIC, Universitat Pompeu Fabra (office 55.123), Roc Boronat 138, E08018 Barcelona, Spain. Tel.: +34 93 542 1346
Email address: sergio.sanchezm@upf.edu
(Sergio Sanchez-Martinez)

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

Such an analysis should also allow jointly analyzing multiple variables, eventually at different hierarchical levels.

In the computer vision community, the fusion of high-dimensional heterogeneous descriptors has become a wide field of research, successfully applied to a large variety of pattern recognition applications (object, face and handwriting recognition). However, its application to medical imaging is rather new. Tiwari et al. (2013) quantitatively combined structural and metabolic imaging data for prostate cancer characterization. Castro et al. (2014) combined the phase and magnitude of magnetic resonance acquisitions to improve the characterization of schizophrenia. Wolz et al. (2012) fused image-derived descriptors and relevant clinical parameters to improve the classification of Alzheimer's disease and mild cognitive impairment.

Different trends have recently emerged in machine learning to address the fusion of different descriptors. The most common approach consists in directly analyzing a concatenated set of the input features (Beckmann and Smith, 2005), which can be improved by taking into account the structure of the manifold to which these features belong (Lombaert and Peyrat, 2013). However, this simple approach does not handle possible differences in the nature or distribution of the analyzed features. Besides, the possible redundancy in the different input descriptors may bias the final result. Multi-set canonical correlation analysis (Correa et al., 2010) deals with redundancy by finding cross-descriptor associations and performing source separation. Nonetheless, all these linear approaches are not suitable for the type of data used in our application. Indeed, it has been pointed out that non-linear operations may be more adequate to prevent artifacts when processing medical images (Gerber et al., 2010) and more precisely cardiac motion patterns (Duchateau et al., 2012).

Kernel methods are appropriate to deal with these disadvantages, since these non-linear approaches do not make any assumption on the nature of the data. They all use a kernel-based affinity matrix that codifies pair-wise similarities between samples. The eigendecomposition of this matrix leads to a space of reduced dimension, which encodes the most relevant characteristics of the data (Yan et al., 2007).

This framework can be extended to fuse high-dimensional heterogeneous descriptors. Feature space fusion and operator fusion are two approaches to tackle this problem (Cloninger et al., 2013). The first strategy consists of finding a common eigenbasis to the multiple eigenspaces associated to each input feature. This can be done by joint diagonalization of Laplacians (Eynard et al., 2015; Dodero et al., 2014) or changing basis (Coifman and Hirn, 2014). However, none of these methods allows determining the relative optimal contribution of each input feature to the final result.

The approach presented in this paper belongs to the second category and consists in fusing different features at the operator level. It builds upon a recently proposed framework, known as multiple kernel learning (MKL) (Lin et al., 2011). By a combination of feature-based kernels, it allows optimally fusing heterogeneous information and weighting the contribution of each input to the final result. Several supervised examples that applied this technique to clinical applications are found in the literature (Tiwari et al., 2013; Castro et al., 2014). It has to be noted that the formulation of Lin et al. (2011) is a broad point-of-view to the MKL problem, originally introduced for the supervised problem of a support vector machine (SVM) (Lanckriet et al., 2004; Rakotomamonjy et al., 2008). Despite the high discriminative power of these kinds of approaches, unsupervised methods are more suitable for our application, for two reasons. First, they are more suitable to extract the hidden structure of the data (data spread), which has proven to correlate with disease in several applications (Wolz et al., 2012; Ye et al., 2014). Second, their result is not influenced by possibly incorrect labels from ambiguous diagnosis.

In addition, our approach allows determining the optimal weight to attribute to each feature. Other methods also pursue a similar objective. Automatic Relevance Determination (Neal, 1996) is a Bayesian approach that aims at selecting features of interest among a larger set of features. Nonetheless, its decision is categorical, while MKL proposes a relative weighting of the different features. Relevant Component Analysis (Shental et al., 2002) adopts a slightly different philosophy, being a semi-supervised approach where irrelevant data variability is reduced while the relevant one is amplified. However, it requires a priori knowledge about the content of the input, which makes this approach inappropriate for unsupervised analysis.

1.2. Heart failure with preserved ejection fraction

We apply our methodology to characterize myocardial velocities in the context of HFPEF. This syndrome presents signs of heart failure but still maintains the ejection fraction (EF) within a normal range. Left ventricular diastolic dysfunction has been identified as one of the leading mechanisms causing it (Paulus et al., 2007; Erdei et al., 2014). Nonetheless, recent studies

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

suggest that HFPEF is a rather heterogeneous condition consisting of several pathophysiological subtypes (Lekavich et al., 2015; Shah et al., 2014). The current diagnosis relies on simple measurements (Paulus et al., 2007), such as EF and E/e' (ratio of the early transmitral flow velocity by pulsed Doppler and the early mitral annular velocity by myocardial velocity imaging). This leads to a suboptimal characterization, since these parameters do not entirely capture the complexity of the observable mechanical abnormalities (Erdei et al., 2014; Santaularia-Tomas and Abraham, 2009; Tan et al., 2009). Thus, there is still a need for an improved understanding of the syndrome. This challenge is not limited to the HFPEF syndrome, as the limitations of peak or time-to-event measurements have been largely discussed in other cardiac applications (Fornwalt et al., 2009; Fornwalt, 2011). In our work, the analyzed data are velocity traces measured from a stress protocol, where subjects are examined while performing exercise on a semi-supine bicycle. This modality has been recommended in many clinical studies to assess HFPEF (Erdei et al., 2014; Donal et al., 2015). The idea behind it is that signs and symptoms may only be revealed during exercise. In this regard, we implement a technique that enables to jointly analyze these data at different stress stages. This results in a tool to support clinical diagnosis and to uncover mechanical interrelations, not necessarily noticed before.

1.3. Proposed approach

In this paper, we illustrate the usefulness of fusing different high-dimensional descriptors to improve the characterization of HFPEF. Our work is based upon the unsupervised MKL method introduced in (Lin et al., 2011). We explicitly detail its formulation and adapt its optimization towards the data spread rather than towards classification (e.g., healthy or diseased). This allows us to study the variability of velocity patterns—which is our primary objective. Therefore, our method differs from other supervised attempts to characterize the HFPEF syndrome (Shuai et al., 2011; Zordoky et al., 2015) since we do not compromise the learning process by involving possibly unreliable labels. In our case, the agreement with clinical labels only serves to determine the configuration and parameters of the algorithm and to check the coherence of the data spread. Previous works have derived an unsupervised formulation of the MKL problem (Tzortzis and Likas, 2012; Zhuang et al., 2011). Tzortzis et al. used it to learn a robust ensemble kernel that ultimately leads to a clustered space in which to perform classification. On the contrary, based on a Laplacian formulation, we seek to preserve the distribution of the data as a continuum, which still provides distance tools to perform inter-subject comparisons and variability analysis. Zhuang et al. used unsupervised MKL in combination with kernel PCA for dimensionality reduction. However, both the MKL and the dimensionality reduction steps were executed separately, whereas our formulation jointly optimizes them. In terms of application, Zhuang et al. also focused on classification, whereas we prioritise the variability analysis. Our method makes a direct contribution to the understanding of the HFPEF syndrome, by: (i) giving insights into which are the important features to look at within the characterization, (ii) uncovering mechanical inter-relations that impair cardiac function and (iii) supporting clinical diagnosis.

Preliminary results using this method have already been presented (Sanchez-Martinez et al., 2015). The present paper extends this previous work in several aspects. The main one is the reconstruction of signals from the output space of coordinates through multiscale kernel regression (Bermanis et al., 2012). This allows analyzing the data variability in the space of input signals. Such an analysis enables the detection of class-related differences in the velocity traces, which is of interest to improve the diagnosis. Another improvement is the extension of the database to 55 subjects, including some that reported breathlessness. Such subjects—who may lie in-between the healthy and diseased populations—do not formally fulfill heart failure symptoms. However, some of them might be already showing the syndrome but remain undetected given current criteria, which reinforces the need for a quantitative analysis of their motion patterns. An additional improvement is the incorporation of new descriptors. One of them is the iso-volumic contraction phase, whose duration was a contributor to predict future major cardiovascular events (Biering-Sørensen et al., 2015). The other is a descriptor that captures the temporal transformation performed during temporal alignment.

The relevance of the method is investigated in two ways. First, by a detailed quantitative analysis of the agreement between the output space coordinates and the diagnosis based on the most recent clinical guidelines. We use this agreement to choose the configuration of the algorithm that results in the most precise characterization according to clinical diagnosis. Second, by an analysis of the joint variability of the different descriptors, which is our primary objective. This variability is further interpreted in light of the physiological characteristics of the HFPEF syndrome, which may uncover mechanical interrelations probably overlooked to date.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

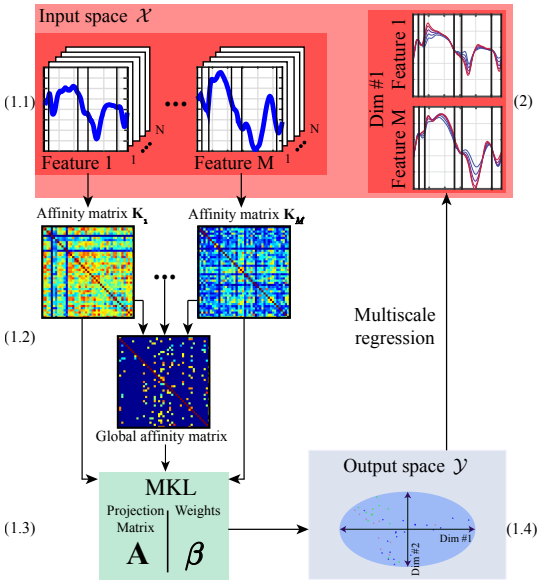

Figure 1: Pipeline of the proposed method: (1.1) input space with M input features for each subject, (1.2) feature-specific and global affinity matrices used as input to the MKL optimization, (1.3) MKL optimization, (1.4) output space of reduced dimensionality and (2) reconstructed input signals through multiscale kernel regression.

Thus, we explore the power of jointly analyzing different velocity traces acquired at rest and exercise to improve the characterization of the HFPEF syndrome.

2. Methodology

The method we propose consists of two steps: (1) the combination of multiple features by MKL, which finds the optimal combination of input features and maps every subject to a space of reduced dimensions; (2) a multiscale kernel regression (Bermanis et al., 2012), which reconstructs input signals from the output space of coordinates. A basic schema describing the pipeline of the algorithm is shown in Fig.1.

2.1. Dimensionality reduction by MKL

The input data $\{\mathbf{x}_{m,i}\}_{(m,i)\in[1,M]\times[1,N]} \in \mathcal{X}_m$, consist of N samples with M features each. For each feature an affinity matrix $\mathbf{K}_m$ is computed, using a Gaussian kernel function:

$$\mathbf{K}_m(\mathbf{x}_{m,i}, \mathbf{x}_{m,j}) = \exp\left(-\frac{\|\mathbf{x}_{m,i} - \mathbf{x}_{m,j}\|^2}{2{\sigma_m}^2}\right), \qquad (1)$$

which relates each pair of samples (i, j) according to their similarity with respect to feature m. Here, σ_m stands for the kernel bandwidth, whose choice is discussed in Sec.3.2. A global affinity matrix is defined from these feature-based affinity matrices. It serves to quantify the closeness of individuals regarding all the input data and is defined as:

$$\mathbf{W} = \sum_{m=1}^{M} \widehat{\mathbf{K}}_m, \qquad (2)$$

where $\widehat{\mathbf{K}}_m = (\mathbf{K}_m)^{1/\alpha_m}$ and α_m results from dividing the variance of kernel $\mathbf{K}_m$ by the variance of the smallest variance kernel among the M kernels ($\alpha_m \geq 1, \forall m \in [1, M]$). This amounts to normalizing the features by their variance, thus balancing their contribution to the neighborhood information encoded in the global affinity matrix ($\mathbf{W}$). In other words, this prevents the highest variability features from dominating the rest of the features in the construction of $\mathbf{W}$. Then, this matrix is made sparse by retaining the entries within a fixed neighborhood ($\widehat{\mathbf{W}}$), whose size is discussed in Sec.3.2.

$\widehat{\mathbf{W}}$ and $\mathbf{K}_m$ are the inputs to the MKL block, which is the core of the algorithm. Contrary to most applications of MKL (Tiwari et al., 2013; Castro et al., 2014; Lanckriet et al., 2004; Rakotomamonjy et al., 2008; Gönen, 2013), we keep an unsupervised formulation for the problem, as our end objective is not classification but studying variability. Thus, the algorithm relies on a Laplacian formulation, which aims at preserving the topology of the data without being conditioned by a given class. In contrast, supervised approaches mainly use SVM formulations, more intended for classification. In the case of a single feature formulation (Yan et al., 2007), the optimal embedding is obtained through the minimization of:

$$\begin{aligned} &\min_{\mathbf{v}} \sum_{i,j=1}^{N} \|\mathbf{v}^\top \mathbf{x}_i - \mathbf{v}^\top \mathbf{x}_j\|^2 \widehat{\mathbf{W}}_{i,j}, \\ &\text{s.t.} \sum_{i=1}^{N} \|\mathbf{v}^\top \mathbf{x}_i\|^2 \mathbf{D}_{i,i} = 1, \end{aligned} \qquad (3)$$

where $\mathbf{D}$ is a diagonal weight matrix, whose entries are the result of a row-wise summation of $\widehat{\mathbf{W}}$ ($\mathbf{D}_{i,i} = \sum_{j=1}^{N} \widehat{\mathbf{W}}_{i,j}$), N is the number of samples, $\mathbf{x}_i$ is the value of the only descriptor associated to sample i in the input space and $\mathbf{v}$ is the matrix that projects it to the output space. Similar samples in the input space (i.e., with large $\widehat{\mathbf{W}}_{i,j}$ values) are mapped to nearby points on the output space.

In the case of multiple features, these are combined by a kernelization step. Working with the associated

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

kernels of the different input features and not with their raw content allows combining them, even if they are of different nature. The MKL problem is solved by the minimization of:

$$\begin{aligned} \min_{\substack{\boldsymbol{\beta}\in\mathbb{R}^{M\times 1}\\ \mathbf{A}\in\mathbb{R}^{N\times N-1}}} \quad & \sum_{i,j=1}^{N} \|\mathbf{A}^\top\mathbb{K}^{(i)}\boldsymbol{\beta} - \mathbf{A}^\top\mathbb{K}^{(j)}\boldsymbol{\beta}\|^2\widehat{\mathbf{W}}_{i,j} \\ \text{s.t.} \quad & \sum_{i=1}^{N} \|\mathbf{A}^\top\mathbb{K}^{(i)}\boldsymbol{\beta}\|^2\mathbf{D}_{i,i} = 1, \\ & \beta_m \geq 0, \sum_{m=1}^{M}\beta_m = 1, \end{aligned} \tag{4}$$

where the unknowns are: $\mathbf{A}$, the rotation matrix that maps the input to the output space, which is $N-1$ dimensional since the smallest eigenvalue is 0; and $\boldsymbol{\beta} = [\beta_1 \dots \beta_M]^\top$, the weights given to the different features. Matrix $\mathbb{K}^{(i)}$ is defined for the i-th sample as:

$$\mathbb{K}^{(i)} = \begin{bmatrix} \mathbf{K}_1(1,i) & \dots & \mathbf{K}_M(1,i) \\ \vdots & \ddots & \vdots \\ \mathbf{K}_1(N,i) & \dots & \mathbf{K}_M(N,i) \end{bmatrix} \in \mathbb{R}^{N\times M}. \tag{5}$$

The values of $\mathbf{A}$ and $\boldsymbol{\beta}$ are calculated by means of an iterative two-step optimization strategy. The first step aims at optimizing $\mathbf{A}$, while $\boldsymbol{\beta}$ is fixed. It is initialized by defining $\beta_m = 1/M$, $\forall m \in [1, M]$ and is solved by a generalized eigenvalue decomposition, which has an explicit solution. The second step aims at optimizing $\boldsymbol{\beta}$, while fixing the previously calculated $\mathbf{A}$. This problem can be solved by quadratically constrained quadratic programming (QCQP), which is computationally expensive. Nevertheless, it can be relaxed to a semidefinite programming problem, which is solved more efficiently. In practice, this is addressed by the use of CVX, a convex optimization package (Grant and Boyd, 2013). Further details about the optimization can be found in the work that presents the supervised formulation of MKL (Lin et al., 2011).

Once the unknowns are calculated, the input samples are mapped to the output space by:

$$\mathbf{Y} = \mathbf{A}^\top \sum_{m=1}^{M} \mathbf{K}_m\beta_m, \tag{6}$$

where $\mathbf{Y} \in \mathbb{R}^{N-1\times N}$ contains on each column the coordinates in the output space of each input sample $\mathbf{x}_i$. For the sake of simplicity, the matrix $\mathbf{Y}$ is cut down to a $d\times N$ version, therefore reducing the dimensionality of the output space by just considering the eigenvectors associated to the d smallest eigenvalues (see Sec.3.5 for further clarification).

2.2. *Performance analysis*

Statistics over the data samples are first computed by examining the distribution of the samples in the output space (Sec.2.2.1). Further analysis complements it by studying the variability that this space encodes (Sec.2.2.2).

2.2.1. *In the output space - discriminative analysis*

In the output space ($\mathcal{Y}$), the data are sorted by their most relevant characteristics, which often tends to a clustered-like organization of the samples, although this is not explicitly imposed in our algorithm. An approach to quantitatively assess the quality of this output space is to measure the agreement of the data embedded in it with provided clinical labels. To do so, the distance from each sample to the normal and diseased populations is computed using the Mahalanobis distance in a leave-one-out configuration. This means that the evaluated sample is left out from the subgroups when computing its Mahalanobis distance to them, expressed as:

$$d(\mathbf{y}_i, S_c) = \sqrt{(\mathbf{y}_i - \mu_c)\boldsymbol{\Sigma}_c^{-1}(\mathbf{y}_i - \mu_c)^\top} \quad \forall\, i \in [1, N], \tag{7}$$

where S_c are the class-related subgroups and $c \in$ {'healthy','diseased'}. Here, $\mu_c \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}_c \in \mathbb{R}^{d\times d}$ are the mean vector and covariance matrix of S_c. Based on this distance, each sample is labeled to its closest subgroup. The agreement between the algorithm and the clinical labels is quantitatively expressed using the Cohen's kappa, and the sensitivity and specificity coefficients.

2.2.2. *In the input space - variability analysis*

The output space can be further studied by analyzing the variability that it encodes in the input space $\mathcal{X}_m$ of each feature. To this end, the features corresponding to any given point lying in the output space $\mathbf{y} \in \mathcal{Y}$ can be estimated all at once via a multiscale adaptation of kernel regression (Bermanis et al., 2012) (see Fig.1 for better interpretability). This technique is designed to properly fit fine details even when the local density of the samples is not uniform. In a single-scale formulation (following an inexact matching description), this corresponds to looking for the functions f_m solutions of:

$$\operatorname*{argmin}_{f_m\in\mathcal{F}_m} \quad \Big(\frac{1}{2}\|f_m\|^2_{\mathcal{F}_m} + \frac{\gamma_m}{2}\sum_{i=1}^{N}\|f_m(\mathbf{y}_i) - \mathbf{x}_{m,i}\|^2\Big), \tag{8}$$

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

where $m \in [1, M]$, γ_m is the trade-off between the regularization term (first term) and the adherence to the data (second term), and $\|.\|_{\mathcal{F}_m}$ is the norm equipping the reproducible kernel Hilbert space $\mathcal{F}_m$ of functions $\mathcal{Y} \to \mathcal{X}_m$. The analytical solution to this problem is:

$$f_m(\mathbf{y}) = \sum_{i=1}^{N} k(\mathbf{y}, \mathbf{y}_i) \cdot \mathbf{b}_{m,i}, \tag{9}$$

where $\mathbf{y} \in \mathcal{Y}$ and $\mathbf{b}_{m,i}$ corresponds to the i-th column of the matrix:

$$\mathbf{B}_m = \left(\mathbf{K} + \frac{1}{\gamma_m}\mathbf{I}\right)^{-1}\mathbf{X}_m, \tag{10}$$

where $\mathbf{X}_m = (\mathbf{x}_{m,1}, \ldots, \mathbf{x}_{m,N})^\top$, $\mathbf{I}$ is the identity matrix, and $\mathbf{K} = (k(\mathbf{y}_i, \mathbf{y}_j))$ is the affinity matrix defining the similarity of the input samples in the output space. This affinity matrix is defined from a Gaussian kernel function of bandwidth σ, namely: $k(\mathbf{y}_i, \mathbf{y}_j) = \exp(-\|\mathbf{y}_i - \mathbf{y}_j\|^2/\sigma^2)$.

The multiscale extension of the previous problem consists in iterating the process across scales, where the bandwidth σ is reduced by a factor of two at each iteration, until reaching a scale with resolution equal or less than the density (average neighborhood size) of the samples in the output space. The algorithm described in this section is further detailed in (Duchateau et al., 2013).

3. Experiments and results

3.1. Echocardiographic data

The method was applied to the data of 55 subjects who underwent a stress echocardiography protocol using a semi-supine bicycle (Erdei et al., 2014) (average age 69 ± 6 years). They were categorized using standard clinical criteria (Paulus et al., 2007) as: 19 patients presenting HFPEF, 22 age-matched healthy controls and 14 subjects who reported breathlessness but did not fulfill the criteria for HFPEF. 2D myocardial velocity images were acquired at rest and submaximal exercise (63 ± 9 bpm and 97 ± 10 bpm, respectively), using a transthoracic probe (M4S, GE Healthcare, Milwaukee, WI) and a Vivid E9 echocardiographic system (GE Healthcare) with a sampling rate of 200 ± 15 Hz. Velocity traces were extracted from four chamber view acquisitions—the easiest to acquire at a high quality—using commercial software (EchoPAC, v.113, GE Healthcare), by defining a region of interest (ROI, with size 1 × 10 mm, located approximately 10 mm above the mitral annulus) at the basal septum and basal lateral wall of the left ventricle (ellipses in Fig.2 (a)). The analysis was limited

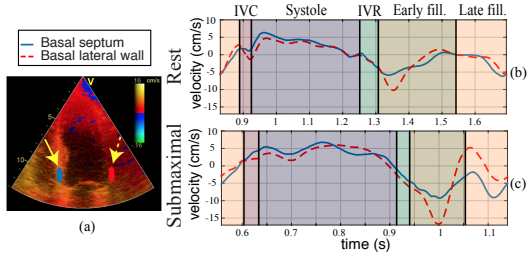

Figure 2: Snapshot of a color-coded myocardial velocity imaging acquisition from a healthy subject (a)—where red and blue, respectively, indicate positive and negative velocities in the probe direction (see the electronic version of the figure)—and its corresponding velocity patterns at (b) rest and (c) submaximal stages. The patterns correspond to the basal septum and lateral wall regions, indicated by arrows in (a). The cardiac phases are indicated by different background colors: iso-volumic contraction (IVC), systole, iso-volumic relaxation (IVR), early and late filling.

to these basal regions, which are the most reproducible and sensitive to detect motion abnormalities in the longitudinal direction and should be sufficient to capture the global longitudinal changes possibly present in the ventricles of the HFPEF subjects (Mor-Avi et al., 2011; Teske et al., 2007). The ROI was kept static throughout the heart cycle to maintain the analysis and the interaction as simple and reproducible as possible. Non-significant dissimilarities in the velocity traces were observed against tracking the tissue either manually or by speckle-tracking, both methods that the used software proposes. Manual tracking requires the user to define a ROI on the myocardium at end-systole and end-diastole. This ROI is interpolated in space to mimic the myocardial movement. Speckle-tracking uses block matching techniques to follow speckles from B-mode images and estimate the myocardial velocities. For better clarification on this point, we have included as *supplementary material* a demo video showing how the three extraction methods work. Myocardial velocity imaging was preferred over (speckle-)tracking on B-mode images due to its high temporal resolution, reproducibility and robustness to noise. With this acquisition protocol we chose the simplest approach according to our philosophy, which is to exploit the echocardiography-derived velocities while minimizing manual intervention. An example of the data extracted for a given subject is illustrated in Fig.2.

The total number of velocity traces was 4 (septal/lateral at rest/submaximal). In the remainder of this paper, these velocities are referred to as cardiac features. In our case, these features are 1D vectors indexing discrete velocity samples. However, the analysis also al-

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

lows including features of other types (e.g., scalar features such as the E/e' ratio), which reinforces its interest and applicability. We considered two types of analysis: *global* and *local*. The *global* analysis involves the velocity signals along the whole cycle, leading to 4 features. *Local* analysis extends this by isolating different cardiac phases (iso-volumic contraction, systole, iso-volumic relaxation, early and late diastole) to treat them independently, leading to 20 features. Class labels based on clinical diagnosis (following standard guidelines (Paulus et al., 2007)) were provided and were used to characterize the output of the MKL algorithm. However, they cannot be fully considered as ground truth, since effective criteria for diagnosing HFPEF are still not established (Komajda and Lam, 2014). Experiments examining the distribution of the samples in the output space will therefore be reported in terms of agreement (Cohen's Kappa), to counterbalance the standard accuracy measurements (sensitivity and specificity), which suppose that ground truth can be trusted.

Temporal normalization: Since our aim is to assess the alterations of the velocity profiles during the different phases of the cardiac cycle, the analyzed velocity traces needed to be expressed within a common temporal reference, in order to be quantitatively compared. The temporal non-correspondence of the traces is due to inter-subject differences in heart rate and timing of cardiac phases (e.g., systole, diastole). These differences are even more pronounced at different levels of physical activity (as noted in Fig.2). To deal with this problem, a two-stage temporal normalization was applied. The first stage consisted in a piece-wise linear warping of the timescale, based on physiological events normalized to the heart cycle. These events were: start/end of the heart cycle and atrial activation (onset of QRS and onset of P-wave, from the ECG) and mitral/aortic valves opening/closure (identified from Doppler flows, acquired separately but at similar heart rates). The timescales were redefined towards a common reference. This reference was selected from the healthy controls, as the one with the most central velocity patterns across a range of tests where each subject was considered successively as reference. The most central pattern is the one for which the sum of Euclidean distances to the remaining patterns is minimized. Then, after achieving a common temporal reference, the second stage consisted in resampling the velocity data to the new common temporal reference, through cubic spline interpolation (Duchateau et al., 2011).

After setting the evaluated velocity traces into a common temporal reference, the timing of cardiac phases is lost. As previously discussed in the literature (Weissler et al., 1968; Sogaard et al., 2002), pathology-related information is not only captured by the amplitude of the velocity, but may also be encoded by the timing of different cardiac phases. In the case of the HFPEF syndrome, the interaction between systole and diastole has been previously discussed (van Zalen et al., 2015). Consequently, we kept this information by adding two extra features to the analysis. They capture the difference in timing of each analyzed subject with the reference, at rest and submaximal stages. Thus, two vectors were defined for each subject: the first one contains the shift in the time events from the studied subject to the reference, and the second one measures the ratio between the duration of each cardiac phase with respect to the reference.

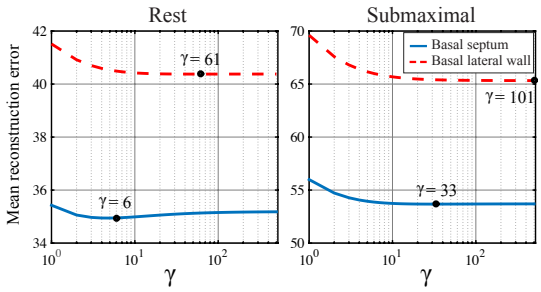

Figure 3: Mean reconstruction error of all the subjects (interpolated following a leave-one-out procedure) as a function of γ for the 4 *global* features. The optimal value of γ is highlighted at the minimum of each curve.

3.2. Parameters setting

The bandwidth σ_m of the kernel $\mathbf{K}_m$ (Eq.1) was calculated feature-wise as the average of the pairwise Euclidean distances between each sample and its k-th nearest neighbour (looking at the corresponding feature). In this case $k = 8$. Then, the number of neighbours used to define the global affinity matrix was fixed to 3. These two values were established heuristically, looking at the maximization of the spread of the samples in the output space. Within the optimization, the number of iterations was conditioned by the convergence rate. Depending on the input features chosen for the analysis, this number of iterations ranged from 5 to 15, which in time translates to less than half a minute for this population. Convergence was considered when the difference between the cost function (Eq.4) from consecutive iterations was smaller than a threshold of value 0.001. The different scales within the multiscale regression iterated from the spread of the samples in the output space until getting

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

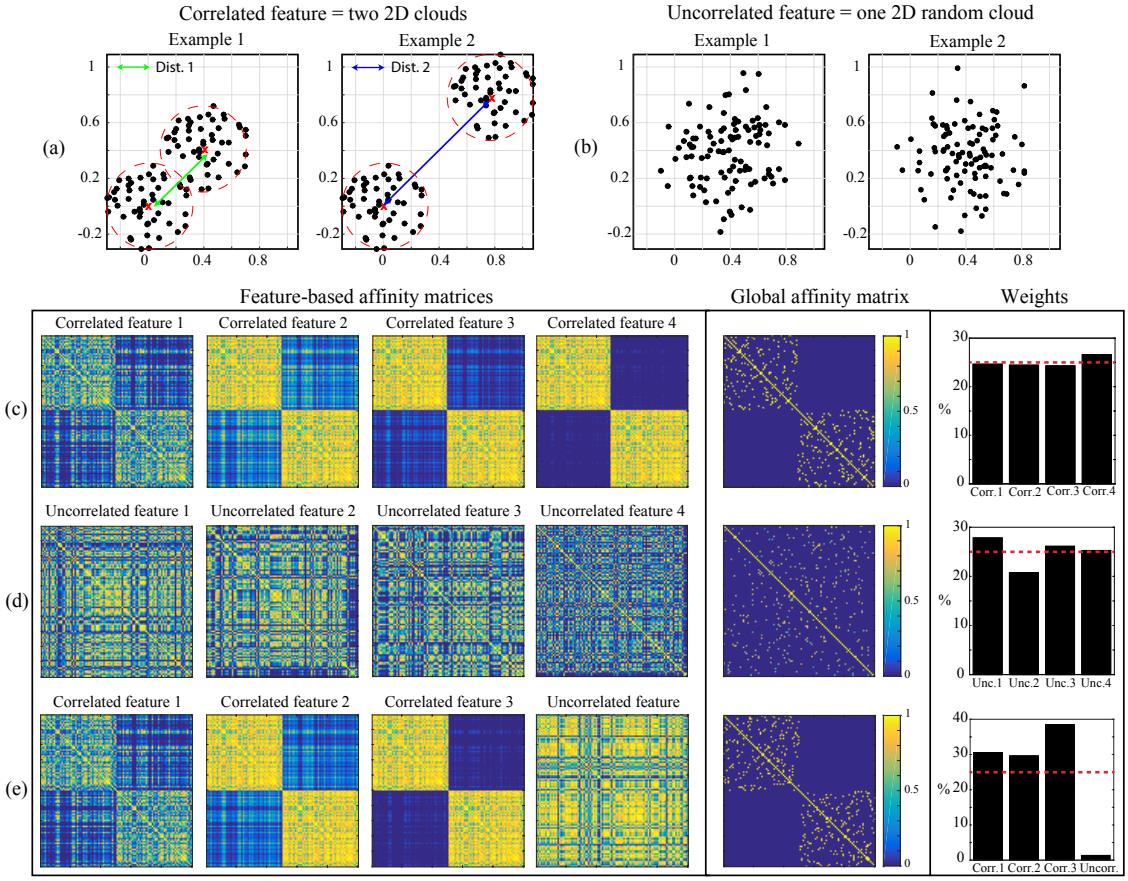

Figure 4: Synthetic experiments to better understand the feature weightings.The top row shows two examples of each type of synthetic features generated: (a) correlated, consisting of 2 identical clouds of points translated at a different distance for each feature; (b) uncorrelated, consisting of a different cloud of random points for each feature. From the second row and from left to right, representation of the feature-based affinity matrices ($\mathbf{K}_m$ in Eq.1), representation of the sparse global affinity matrix ($\widetilde{\mathbf{W}}_{i,j}$ in Eq.4) and weight values obtained after convergence of the MKL algorithm for synthetic experiments using different sets of 2D synthetic features: (c) four correlated features (as in (a)); (d) four uncorrelated features (as in (b)); (e) three correlated features (as in (a)) and a completely uncorrelated feature (as in (b)). The horizontal dashed line in the weights bar graphs corresponds to the theoretical weight value ($1/4 = 25\%$) that would correspond to 4 perfectly correlated features.

lower than their density, and were divided by a factor of two at each iteration. Here, density stands for the average neighborhood size (Duchateau et al., 2013). In practice, we computed it as the average Euclidean distance to the closest neighbor among the studied subjects. The trade-off parameter γ_m for this regression was independently determined for the 4 *global* features considered (septal/lateral velocity patterns at rest/submaximal) following a leave-one-out procedure. Values were chosen as those minimizing the reconstruction error of unseen samples, thus maximizing the generalization ability of the regression model (Davies et al., 2010). The optimal γ_m for each of the 4 *global* features is specified in Fig.3. Note that the values for the submaximal stage are noticeably higher, which means more adherence to the data, allowing to better deal with the higher variability among velocity traces recorded at this stage.

3.3. Features weighting

The second stage of the optimization (Sec.2.1) automatically finds the weights that are solutions of Eq.4. These weights account for the relative importance of the feature-based kernels in the construction of the projection. We hypothesize that their values correlate with the

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

contribution of each feature to the sparse global affinity matrix ($\widehat{\mathbf{W}}_{i,j}$), and that this contribution depends on whether or not the input features are correlated with one another.

We designed different synthetic experiments to test our hypothesis (see Fig.4), and to better understand which weight values should be expected at the output of our algorithm. These experiments intended to mimic the trends observed in the real data, i.e., partial correlation between different velocity profiles and almost no correlation between the different cardiac phases (apart from the temporal causality of these data). To this end, we created correlated and uncorrelated features using 2D clouds of points (feature size $\in [2, N]$, i.e., N samples lying in a two-dimensional space). In the remainder, the term "correlation" refers to the correlation between different features. Specifically, we wanted to evaluate whether our hypothesis—that the correlated features contribute more to the projection—was correct. The correlated features consisted of a fixed cloud of points plus a second identical cloud translated at a different distance for each new feature (see Fig.4 (a) for two examples of correlated features). The distribution within a cloud remained the same for the different correlated features. Each uncorrelated feature consisted of a single cloud of random points (see Fig.4 (b) for two examples of uncorrelated features). The experiments consisted in applying our MKL algorithm to different combinations of four of these features, and examining the obtained weights distributions. These experiments showed that many correlated features lead to almost uniform weight distributions (Fig.4 (c)). The same happened for many uncorrelated (random) features (Fig.4 (d)). Conversely, the algorithm favors several correlated features over a single feature uncorrelated from the rest (Fig.4 (e)).

Then, the weights obtained on real data (Fig.5) were examined in light of the previous synthetic experiments and their conclusions. In this case, the joint analysis of the different velocity traces was done using the *local* configuration of our algorithm detailed in Sec.3.4.2. When analyzing 4 synthetic correlated features (Fig.4(c)), they ended up with weights around $1/4 = 25\%$. In our case we analyzed 22 features. The features corresponding to the same cardiac phase are partially correlated while those from different cardiac phases are uncorrelated. If perfectly correlated, they should end up with weights around $1/22 \approx 4.5\%$ (dashed line in Fig. 5). The obtained weights lie around this value, but correlations are only partial and differences still exist. They can go up to 40% in relative terms between the IVC and the early filling phases. The highest weight values correspond to the early and late filling

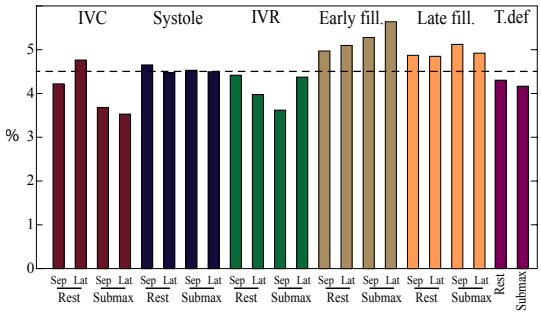

Figure 5: Feature weights proposed by the algorithm for the best setting (#12 in Table 1). T.def stands for "temporal deformation" (Sec. 3.1). The horizontal dashed line corresponds to the theoretical weight value $1/22 \approx 4.5\%$ that would correspond to 22 perfectly correlated features.

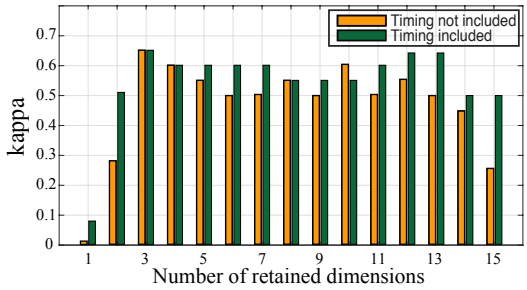

Figure 6: Cohen's kappa coefficient for different number of retained dimensions considering the best experiment (#12 in Table 1). The influence of including or not the features related to the temporal transformation was also tested (average improvement of 8%).

phases, especially at submaximal stress stage. This is even more pronounced when analyzing other sets of features, e.g., the features corresponding to a single region (septal or lateral). The algorithm gives more importance to these features to characterize the data, which agrees with clinical observations (Erdei et al., 2014), namely that at exercise the HFPEF subjects may present filling abnormalities not observed in normal controls.

3.4. Output space

The output of the algorithm consists of a d-dimensional Euclidean space. Linear methods, such as principal component analysis (PCA), find a compact output space, without redundancy. PCA captures the variance in a specific direction and its orthogonal projections, which generally show a rapid decrease. In contrast, a kernel analysis may result in several eigenvectors encoding the same spatial "direction" of a manifold

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

Table 1: Summary of the different runs of the algorithm, varying the region, stress stage and analysis. The number of dimensions of the subspace yielding the best agreement and its corresponding Cohen's kappa, sensitivity and specificity are also indicated.

Configuration	Region	Stage	Analysis	dim	kappa (%)	sens. (%)	spec. (%)
#1	Septal	Rest	*Global*	9	36.5	68.4	68.2
#2	Septal	Rest	*Local*	2	31.3	63.2	68.2
#3	Septal	Submax	*Global*	2	46.2	73.7	72.7
#4	Septal	Submax	*Local*	5	46.2	73.7	72.7
#5	Septal	Rest/Submax	*Global*	5	61.3	**89.5**	72.7
#6	Septal	Rest/Submax	*Local*	3	61.0	84.2	77.3
#7	Sept./Lat.	Rest	*Global*	8	46.2	73.7	72.7
#8	Sept./Lat.	Rest	*Local*	9	50.6	68.4	81.8
#9	Sept./Lat.	Submax	*Global*	3	55.0	63.2	**90.9**
#10	Sept./Lat.	Submax	*Local*	6	55.4	68.4	86.4
#11	Sept./Lat.	Rest/Submax	*Global*	3	65.3	73.7	**90.9**
#12	Sept./Lat.	Rest/Submax	*Local*	3	**65.5**	78.9	86.4

(Nadler et al., 2008), and such a decrease is not necessarily guaranteed. This complicates any process of cutting across the dimensions if assuming that the removed ones mainly correspond to noise in the data. Nevertheless, the top dimensions were found to be the most relevant when characterizing the data, as detailed in Sec.3.5.

This output space allows performing classical measurements over data samples, since it disentangles the complexity of the input towards a Euclidean embedding. Notably, the output space presents two intrinsic properties: (i) it spreads the data attending to their more salient characteristics; (ii) this spread of the data naturally leads to a clustered distribution of the input samples, which facilitates the assessment of its quality with regard to clinical labels.

3.4.1. Quantitative assessment

Given the intrinsic properties of the output space, it permits a quantitative assessment of the way samples are arranged. As previously commented, the unsupervised nature of the algorithm allows learning the distribution of the data samples without being conditioned by any provided (and possibly incorrect) class. However, we still want to evaluate whether the output space estimated in an unsupervised way is coherent with these labels (see Fig.7 for an example). Considering the output space as reference, the algorithm-based labels relied on the Mahalanobis distances between each subject and the normal and diseased distributions, which were obtained using a leave-one-out configuration. Each healthy / HFPEF subject was categorized as {'healthy','diseased'} according to the smallest among these distances (Sec.2.2.1). The breathless subjects were not included in this evaluation: they are an ambiguous population and thus cannot be defined using categorical arguments. Cohen's kappa coefficient, sensitivity and specificity were used to summarize the agreement between clinical and algorithm-based labels. Based on these measures of quality, different configurations of the algorithm were compared to determine which one coincides the most with clinical diagnosis.

We found that including the timing information in the analysis (as defined in Sec.3.1) generally entails a substantial improvement in the agreement. Indeed, after the temporal transformation, which expresses the velocity traces into a common temporal framework, class-related temporal differences may be buried. These results are summarized in Fig.6, where the agreement values for the cases of including and not including the timing information are compared, considering a different number of retained dimensions (the average improvement is 8%). Note however, that for the best case (3 retained dimensions) the inclusion of the timing information does not result in a better agreement.

3.4.2. Benefits of multi-feature analysis

The benefits of a joint analysis of multiple features was tested by changing the algorithm input features in an incremental and structured way. In particular, we set up the input either considering the different regions independently or combining them. The same configuration was followed for the stages of the stress protocol. Another aspect that we considered is whether to analyze the velocity traces as a whole (*global* temporal analysis) or to treat the different cardiac phases of the heart cycle independently (*local* temporal analysis). The agree-

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

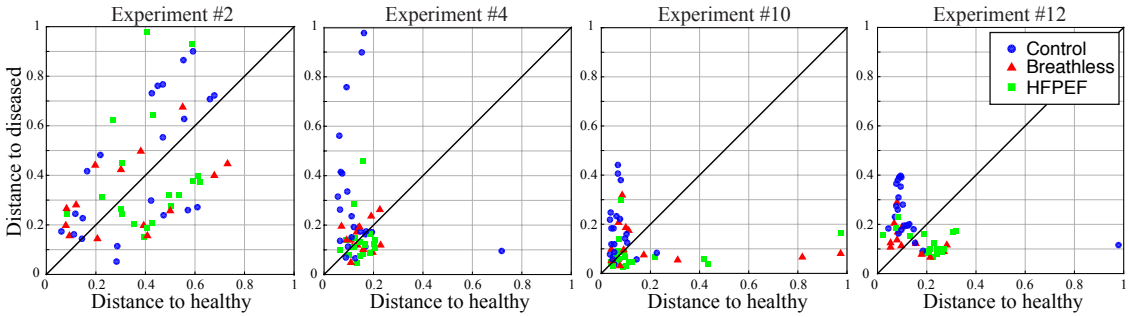

Figure 7: Mahalanobis distances from each subject to the healthy (x axis) and diseased (y axis) distributions. The decision threshold is depicted as a diagonal line. Each sample is colored and shaped based on its clinical label.

ment values for this hierarchical study are summarized in Table 1, in terms of Cohen's kappa coefficient, sensitivity and specificity. It is worth reminding that these values should be carefully handled, as the clinical labels cannot be fully considered as ground truth. They correspond to the best result for each setting, i.e., the one obtained when considering the subset of the output space dimensions that maximized the agreement with the clinical labels. Note that the number of retained dimensions varies from one configuration to another (column "dim" in Table 1). The experiments included the features corresponding to the temporal transformation, which proved to generally improve the agreement value.

Several observations can be made from Table 1. First, single submaximal analysis yields a better result than single rest. Jointly analyzing rest and submaximal returns the best result. Then, septal + lateral analysis performs better than just looking at septal. Finally, a *local* analysis is slightly better than a *global* one in the cases where more information is jointly analyzed (from the experiment #7 on).

To better illustrate the improvement due to a multi-feature analysis, some representative examples are presented in Fig.7. They correspond to experiments #2, 4, 10 and 12 in Table 1. Note that the value of kappa more than doubles across them. The coordinates of each subject are the Mahalanobis distances to the healthy (x axis) and diseased (y axis) distributions and its shape/color stands for the clinical label. The diagonal line is the threshold used for assigning the algorithm-based labels, 'healthy' (above) or 'diseased' (below).

3.4.3. *Posterior re-examination*

We further validated whether or not we could trust the characterization that yielded the best agreement with clinical labels (experiment #12 in Fig.7 and Table 1). We focused on: (i) the discrimination ability between healthy and HFPEF subjects, and (ii) the situation of the breathless subjects within the obtained representation. To this end, a clinically experienced observer blinded to the clinical labels re-examined the echocardiographic records of all the subjects included in the study. This observer confirmed:

(i) A difference on the underlying disease process between the "misclassified" and the well-classified HFPEF subjects. The latter group presented hypertensive-related remodeling abnormalities, while the "misclassified" HFPEF subjects presented abnormalities such as left bundle branch block or right ventricular dysfunction. In a similar way, the "misclassified" healthy controls showed hypertensive remodeling, while their well-classified counterparts showed a perfectly normal function.

(ii) A correlation between the situation of the breathless subjects and the cause of breathlessness (cardiac/non-cardiac). In particular, breathless subjects lying on the "healthy" region had breathlessness suspected to come from non-cardiac reasons. On the other hand, breathless subjects lying on the "diseased" region showed major filling abnormalities and abnormal exercise response, which occur with HFPEF.

3.5. *Joint variability in the input space*

Any kind of interpretation based on the output space can be regarded as an abstract approach to characterize the cardiac mechanical function of a population. Therefore, to facilitate its interpretation, the variability in the input signals as encoded by the output space has been examined. In particular, we considered the output space corresponding to the experiment #12 in Table 1, which yielded the best agreement.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

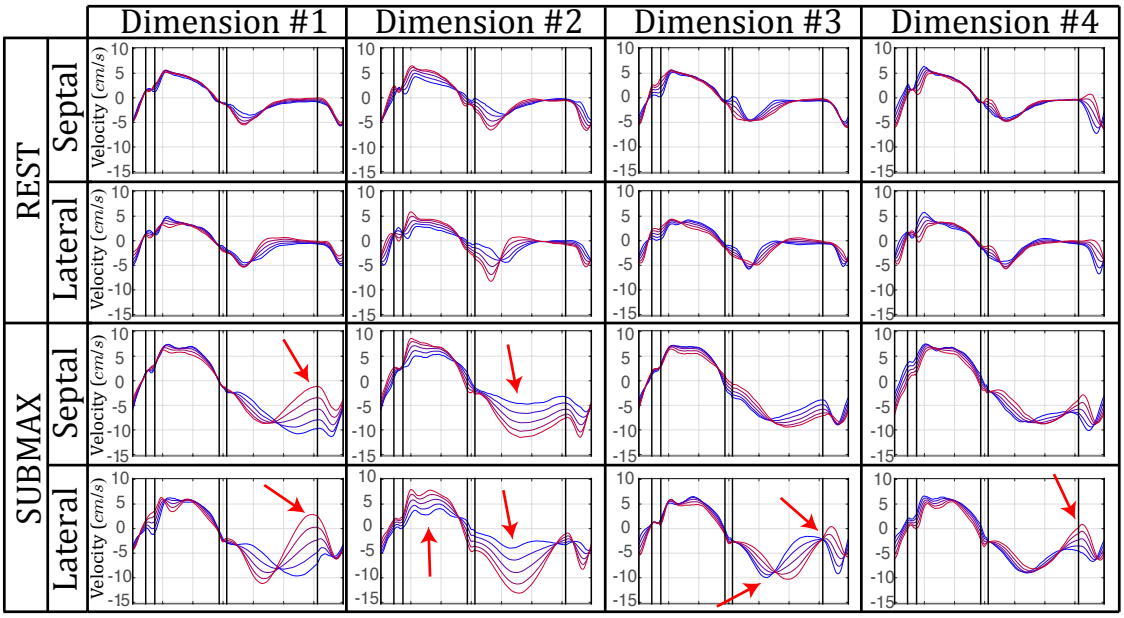

Figure 8: Variability encoded by the 4 top dimensions of the output space. Curves ranging from −2 (blue) to +2 (red) standard deviations along the different dimensions of the output space. Red arrows have been placed to highlight portions that present higher variability and indicate events of physiological interest.

The variability encoded by each dimension is represented by re-mapping to the input space virtual samples spreading along ±2 standard deviations of this specific output dimension (multiscale regression in Fig.1). We have focused on the four main dimensions of the output space, assuming that the remaining ones are less helpful in characterizing the HFPEF syndrome. In fact, from the 4th dimension on, no significant differences between healthy and diseased distributions were observed after performing the Mann Whitney U-test on individual dimensions. Although only the first three dimensions were needed to reach the best agreement with clinical labels in Table 1, this does not necessarily imply that the remaining dimensions are uninformative and can be discarded (as shown by the results of the statistical test). The variability encoded by the four main dimensions is shown in Fig.8.

We hypothesize that analyzing this variability will highlight discriminative characteristics of the HFPEF syndrome. Therefore, we interpret Fig.8 by linking the variability observed in the velocity traces (red arrows) and some indicators of mechanical dysfunction associated to the HFPEF syndrome. For example, the pattern of fusion between early filling and atrial filling observed in dimension #1 could come from an interaction of relaxation and compliance of the myocardial tissue, or the overall change in amplitude observed in dimension #2 may reflect the differences in long-axis functional reserve of the left ventricle.

To further test whether the previously highlighted characteristics are explained by differences between the "characteristic" pattern of healthy and HFPEF subjects, we compared the variability reconstructed from the two principal directions on which these subgroups extend. These directions corresponded to the main modes recovered by a PCA in the output space over each distribution considered independently. The variability associated to these subgroup directions is shown in Fig.9. Several differences—that are consistent with the clinical knowledge about HFPEF mechanical abnormalities—are identified on the HFPEF with respect to the healthy subgroup: lower systolic and diastolic amplitudes, fusion of early and late diastolic curves (indicated by arrows in dimension #2 of Fig.9), and delayed diastolic dynamics (indicated by arrows in dimension #1 of Fig.9) (Flachskampf et al., 2015). We only plot the submaximal stage in Fig.9, which shows the clearest differences between subgroups. This also supports the value of the stress protocol to characterize the HFPEF syndrome.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

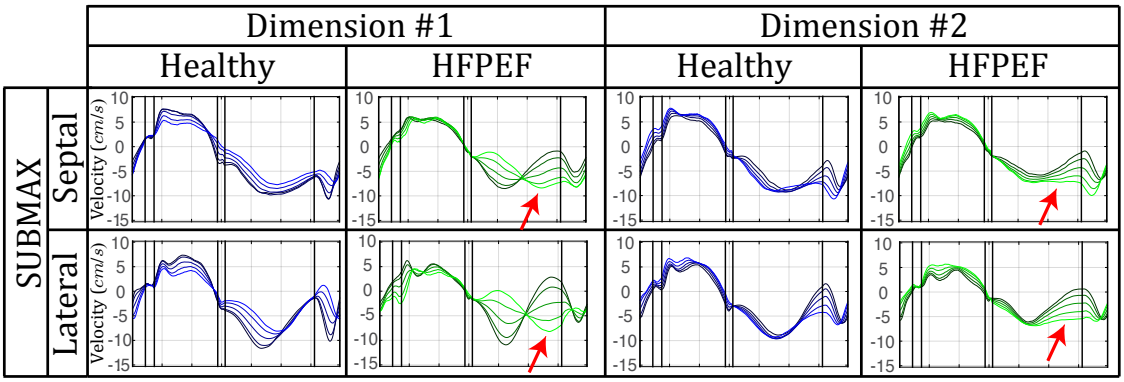

Figure 9: Variability reconstructed from the two principal directions of the healthy and HFPEF subgroups, considered separately. Curves ranging from −2 (darker color) to +2 (brighter color) standard deviations along the mentioned directions. The main differences between the healthy and HFPEF subgroups are highlighted by arrows.

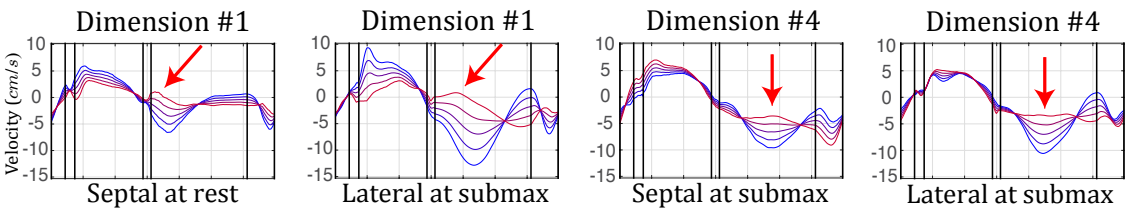

Figure 10: Recovered variability by applying PCA to a concatenated version of the input features. Curves ranging from −2 (blue) to +2 (red) standard deviations along the different dimensions of the output space. The observed physiologically-unrealistic patterns are indicated by red arrows.

The relevance of non-linear vs. linear methods to characterize these data (Sec.1.1) was also tested. A PCA was done using a concatenation of the normalized input features for each configuration in Table 1. The reconstructed variability using this linear method resulted in velocity traces that were physiologically unrealistic (Fig.10). In this figure, the observed irregularities are indicated by red arrows and correspond to almost flat diastolic patterns with absence of E wave and even presenting an unobserved early diastolic contraction, which is highly suspicious, especially during exercise. In terms of agreement compared with PCA, our method resulted in an average improvement in sensitivity across the different experiments in Table 1 of 3.1%. Comparative results between PCA and our method are provided in the Appendix.

4. Discussion

We formulate and evaluate the unsupervised MKL algorithm for dimensionality reduction introduced in (Lin et al., 2011), which we use to jointly analyze multiple features describing the mechanical function of the heart. Other formulations of unsupervised MKL exist, and are optimized for clustering (Tzortzis and Likas, 2012; Lin et al., 2011) or dimensionality reduction (Zhuang et al., 2011). Although the latter might seem convenient to characterize the analyzed population, it tackles the kernel learning and the dimensionality reduction steps separately, whereas we jointly optimize them to account for their interaction in the construction of the projection. Our method was applied to the data of 55 subjects, who underwent a stress protocol. We have highlighted the importance of this protocol to improve the characterization of subjects with HFPEF. By going towards pattern analysis, we improve classical diagnostic approaches often based on simple measurements such as E/e' ratio. Indeed, thresholding the value of this parameter at exercise using the standard E/e'> 8 to detect HFPEF subjects (Paulus et al., 2007), the results of agreement are worse than those obtained with any of the configurations shown in Table 1 (kappa = 26.7%, sens. = 63.2% and spec. = 63.6%). Close in spirit to pattern recognition techniques, we aim at learning the range of ab-

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

normal "patterns" related to the HFPEF syndrome with respect to the healthy pattern, and at estimating their variability. Due to the low statistical power of current diagnostic approaches, some of the provided clinical labels could possibly be incorrect. These labels cannot be further validated, given the absence of more recent guidelines that serve as ground truth. Thus, we prefer to keep the problem unsupervised and to learn the real structure of the data. Despite this, the discriminative capacity of the output space is expressed in terms of agreement with the provided labels, serving as a quantitative indicator of the quality of such space. We have shown the benefits of our multi-feature analysis. Indeed, the results demonstrate the accuracy of our algorithm and support our contributions: (i) the variability analysis, which highlights discriminative markers of the disease; (ii) the weighting of different features, easily transposable to other problems and features; (iii) the detection of misclassified subjects and differences in the underlying disease processes leading to the HFPEF syndrome, which suggests the need for a re-evaluation of the diagnostic guidelines; and (iv) the characterization of the breathless subjects, which is challenging given that they do not fulfill the current diagnostic criteria.

The MKL algorithm converged to an optimal solution using little time, less than half a minute considering the whole set of available input features, and few iterations. Unlike approaches analyzing large databases of very different images (Lin et al., 2011), the relative similarity of the studied features (physiological data in a coherent population) may contribute to this fast convergence. However, the major bulk of work was spent in the extraction and pre-processing of the data, especially in the manual definition of the time events. This could be improved by dedicated software, which is not the purpose of this work.

Absolute differences between the weights in Fig.5 are somehow limited, although relative differences can go up to 40%. From the experiments in Fig.4, we expect that many correlated features would lead to a uniform weight distribution. The same would happen for many uncorrelated features. Similarly, the algorithm would favor a bunch of correlated features against a single uncorrelated one. In our case, the septal/lateral velocity traces at rest/exercise are moderately correlated. Note that this correlation mainly concerns distinct traces and does not occur within the different phases of a given trace (apart from their intrinsic temporal causality, e.g., a high systolic peak will be likely followed by a high diastolic peak). This leads to the subtle differences observed in Fig.5. If the algorithm only considered one among all the correlated features, relevant disease-related characteristics may be lost, thus compromising the final characterization. In the optimized result, the diastolic features at submaximal stage predominate. This agrees with the literature, since at exercise the HFPEF subjects may show filling abnormalities not present in normal subjects (Erdei et al., 2014).

We also demonstrated that the inclusion of temporal deformation features to the analysis generally achieves a significant improvement in the characterization of the studied population (as observed in Fig.6). Indeed, pathology-related information is not only captured by the amplitude and profile of the velocity, but also by the timing of different cardiac phases, as reported in (Weissler et al., 1968). Results on the agreement achieved by *global* and *local* temporal analyses suggest that the latter is slightly better. Chopping off the velocity patterns into physiological temporal windows (systole, iso-volumic relaxation, early and late diastole) enables the algorithm to focus on certain abnormal phases, resulting in a more flexible and thus, more accurate characterization. Certainly, the improvement in sensitivity in the two last experiments could confirm the benefits of the *local* modality to identify diseased subjects.

It has been previously demonstrated (Shah et al., 2014; Erdei et al., 2014) that assessing cardiac function by evaluating just a few parameters does not accomplish a proper characterization of the HFPEF syndrome. This evidence can be further extrapolated to the analysis of patterns. The assumption is that, the more meaningful features examined, the better the quality of the characterization. The benefits of a multi-feature analysis have been demonstrated in the experiments. In particular, it has been shown that jointly analyzing rest and submaximal stages returns a better result than analyzing them independently. This confirms that a stress protocol is helpful when characterizing the HFPEF syndrome (Erdei et al., 2014) and supports the assumption that the joint analysis of longitudinal descriptors (rest + submaximal) improves the characterization (van Zalen et al., 2015). We also demonstrated the added value of jointly analyzing different regions (septal + lateral) within the analysis. Considering both regions, we adequately captured the global longitudinal changes present in the ventricles of the studied HFPEF subjects (Mor-Avi et al., 2011; Teske et al., 2007). As detailed in Sec. 3.4.3, the blind assessment of the characterization confirmed that the algorithm identifies differences in the underlying disease processes leading to the HFPEF syndrome. The hypertensive condition present in most of the HFPEF subjects implies a gradual change from normality to HFPEF, which hinders a clear cut between

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

both states (what clinical diagnosis does). Our unsupervised analysis allows exploring this gradual change and characterizes the data accordingly. These aspects emphasize the need for a re-evaluation of the current diagnostic criteria. The algorithm also situates the breathless subjects according to the cause of their breathlessness condition. This allows characterizing them by looking at their situation vs. other subjects. In this sense, our analysis shares similarities with the concept of *smart atlases*, where new subjects are put into correspondence with the closest known subjects to further refine their diagnosis or prognosis.

Our philosophy is to investigate how much we can take out from the echocardiography-derived velocities to understand the expressions of the HFPEF syndrome. To this end, we study the variability encoded by the most significant dimensions, i.e., those that better discriminate between healthy and diseased populations, which may highlight discriminative markers of the disease. In particular, it shows the (diastolic) dysfunction present in the HFPEF subjects: impaired relaxation, reduced compliance and diminished long-axis functional reserve of the left ventricle (Flachskampf et al., 2015). These characteristics are confirmed when comparing the variability associated to the healthy and HFPEF subgroups independently (Fig.9). Indeed, the HFPEF-related variability differs from the healthy one presenting: lower systolic and diastolic amplitudes, fusion of early and late diastolic curves, and delayed diastolic dynamics. We have limited our variability study to the reconstruction of dimensions treated independently. This is still an open concern, as a combination of dimensions might result in a better discriminability but would not properly address the variability of the patterns. Note that the largest variabilities encoded by the studied dimensions appear within the diastolic phases, especially during exercise. This is in accordance with the examined population, since HFPEF subjects will present diastolic abnormalities under stress (Erdei et al., 2014).

Our database was thoughtfully collected following a well designed protocol, so we did not have to deal with problems arising from missing data. Future extensions of our algorithm should be adapted to address this rather common situation in clinical research and practice.

The heart is a biological engine that dynamically adapts its pumping function to external stimuli and even handles abnormal situations like arterial hypertension or ischemia. However, this adaptation comes at a price. For instance, an elevated after-load in a certain subject may be handled by a change in the contractility during systole, which may affect the early filling of the left ventricle (van Zalen et al., 2015). Current analysis of the cardiac mechanics using absolute values, derived from strain and strain rate curves, still remains suboptimal. With such an approach, inter-relationships between different phases of cardiac mechanical function could easily pass unnoticed, even to the most expert and trained eye. In clinical practice, the detection of abnormalities is usually based on intuitive visual recognition processes. Our approach, going towards pattern analysis, allows a more comprehensive examination of the cardiac function and the variability analysis allows an easy interpretation of it. This analysis may reveal undiscovered interrelations across different phases of the cardiac cycle, increasing our knowledge of the mechanical function and how it is affected by a pathology, in our case the HFPEF syndrome. In future, this gained knowledge could be translated to improved diagnostic tests and classifications that might predict responses to specific therapies.

5. Conclusion

In this paper, we have proposed an unsupervised learning method to jointly analyze the variability of multiple velocity patterns from a stress protocol. It builds upon multiple kernel learning for dimensionality reduction, which weighs the relative importance of the analyzed patterns and reduces their complexity towards a (clinically)-meaningful representation. We illustrated the method on a population of healthy, breathless and HFPEF subjects. The results showed its ability to: identify different disease processes leading to HFPEF; characterize the breathless subjects according to their cause of breathlessness (cardiac/non-cardiac); and reveal the different patterns of functional response to stress between the healthy and HFPEF subgroups. The proposed method has potential to improve our understanding of the pathophysiology associated to the HFPEF syndrome and to pave the way for more quantitative, objective systems to support clinical diagnosis.

Acknowledgment

The authors acknowledge the European Union's Seventh Framework Programme for research, technological development and demonstration (VP2HF FP7-2013-611823 and MEDIA FP7-HEALTH-2010-261409), and the Spanish Ministry of Economy and Competitiveness (TIN2012-35874). The work of S. Sanchez-Martinez was supported by a fellowship from "la Caixa" Banking Foundation. The authors also thank R. Davies for his careful reading of the paper.

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

Appendix. Comparison between PCA and our method

Table 2: Reproduction of Table 1, but comparing the PCA results to those obtained with our method.

				dim	**kappa (%)**	**sens. (%)**	**spec. (%)**
Configuration	**Region**	**Stage**	**Analysis**	PCA — MKL	PCA — MKL	PCA — MKL	PCA — MKL
#1	Septal	Rest	*Global*	2 — 9	20.4 — 36.5	47.4 — 68.4	72.7 — 68.2
#2	Septal	Rest	*Local*	9 — 2	31.3 — 31.3	63.2 — 63.2	68.2 — 68.2
#3	Septal	Submax	*Global*	1 — 2	56.0 — 46.2	78.9 — 73.7	77.3 — 72.7
#4	Septal	Submax	*Local*	3 — 5	45.5 — 46.2	63.2 — 73.7	81.8 — 72.7
#5	Septal	Rest/Submax	*Global*	3 — 5	**65.8** — 61.3	84.2 — **89.5**	81.8 — 72.7
#6	Septal	Rest/Submax	*Local*	4 — 3	56.0 — 61.0	78.9 — 84.2	77.3 — 77.3
#7	Sept./Lat.	Rest	*Global*	2 — 8	45.1 — 46.2	57.9 — 73.7	86.4 — 72.7
#8	Sept./Lat.	Rest	*Local*	6 — 9	16.9 — 50.6	57.9 — 68.4	59.1 — 81.8
#9	Sept./Lat.	Submax	*Global*	2 — 3	61.0 — 55.0	84.2 — 63.2	77.3 — **90.9**
#10	Sept./Lat.	Submax	*Local*	3 — 6	60.5 — 55.4	73.7 — 68.4	86.4 — 86.4
#11	Sept./Lat.	Rest/Submax	*Global*	3 — 3	60.8 — 65.3	78.9 — 73.7	81.8 — **90.9**
#12	Sept./Lat.	Rest/Submax	*Local*	5 — 3	65.3 — 65.5	73.7 — 78.9	**90.9** — 86.4

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

References

Beckmann, C. F., Smith, S. M., 2005. Tensorial extensions of independent component analysis for multisubject fMRI analysis. NeuroImage 25, 294–311.

Bermanis, A., Averbuch, A., Coifman, R. R., 2012. Multiscale data sampling and function extension. Appl Comput Harmon Anal 1, 1–15.

Biering-Sørensen, T., Mogelvandg, R., Jensen, J. S., 2015. Prognostic value of cardiac time intervals measured by tissue Doppler imaging M-mode in the general population. Heart 101 (12), 954–960.

Bijnens, B., Cikes, M., Butakoff, C., Sitges, M., Crispi, F., 2012. Myocardial motion and deformation: What does it tell us and how does it relate to function? Fetal Diagn Ther 32, 5–16.

Borlaug, B., Paulus, W. J., 2011. Heart failure with preserved ejection fraction: pathophysiology, diagnosis, and treatment. Eur Heart J 32, 670–679.

Castro, E., Gómez-Verdejo, V., Martínez-Ramón, M., Kiehl, K. A., Calhoun, V. D., 2014. A multiple kernel learning approach to perform classification of groups from complex-valued fMRI data analysis: application to schizophrenia. NeuroImage 87, 1–17.

Cloninger, A., Czaja, W., Doster, T., 2013. Operator analysis and diffusion based embeddings for heterogeneous data fusion. In: IEEE IGARSS. pp. 1249–1252.

Coifman, R. R., Hirn, M. J., 2014. Diffusion maps for changing data. Appl Comput Harmon Anal 36, 79–107.

Correa, N. M., Eichele, T., Adali, T., Li, Y. O., Calhoun, V. D., 2010. Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI. NeuroImage 50, 1438–1445.

Davies, R. H., Twining, C. J., Cootes, T. F., Taylor, C. J., 2010. Building 3-D Statistical Shape Models by Direct Optimization. IEEE Trans Med Imag 29, 961–981.

Dodero, L., Murino, V., Sona, D., 2014. Joint Laplacian diagonalization for multi-modal brain community detection. PRNI, 1–4.

Donal, E., Lund, L. H., Oger, E., Reynaud, a., Schnell, F., Persson, H., Drouet, E., Linde, C., Daubert, C., 2015. Value of exercise echocardiography in heart failure with preserved ejection fraction: a substudy from the KaRen study. Eur Heart J, 1–8.

Duchateau, N., De Craene, M., Piella, G., Frangi, A. F., 2012. Constrained manifold learning for the characterization of pathological deviations from normality. Med Image Anal 16, 1532–1549.

Duchateau, N., De Craene, M., Piella, G., Silva, E., Doltra, A., Sitges, M., Bijnens, B., Frangi, A. F., 2011. A spatiotemporal statistical atlas of motion for the quantification of abnormal myocardial tissue velocities. Med Image Anal 15, 316–328.

Duchateau, N., De Craene, M., Sitges, M., Caselles, V., 2013. Adaptation of multiscale function extension to inexact matching. Application to the mapping of individuals to a learnt manifold. In: GSI, LNCS. Vol. 8085. pp. 578–586.

Erdei, T., Smiseth, O. A., Marino, P., Fraser, A. G., 2014. A systematic review of diastolic stress tests in heart failure with preserved ejection fraction, with proposals from the EU-FP7 MEDIA study group. Eur J Heart Fail 16, 1345–1361.

Eynard, D., Kovnatsky, A., Bronstein, M. M., Glashoff, K., Bronstein, A. M., 2015. Multimodal manifold analysis by simultaneous diagonalization of Laplacians. IEEE Trans Pattern Anal Mach Intell 37 (12), 1–14.

Flachskampf, F. A., Biering-Sørensen, T., Solomon, S. D., Duvernoy, O., Bjerner, T., Smiseth, O. A., 2015. Cardiac imaging to evaluate left ventricular diastolic function. JACC: Cardiovascular Imaging 8 (9), 1071–1093.

Fornwalt, B. K., 2011. The dyssynchrony in predicting response to cardiac resynchronization therapy: a call for change. J Am Soc Echocardiogr 24, 180–184.

Fornwalt, B. K., Delfino, J. G., Sprague, W. W., Oshinski, J. N., 2009. It's Time for a Paradigm Shift in the Quantitative Evaluation of Left Ventricular Dyssynchrony. J Am Soc Echocardiogr 22, 672–676.

Gerber, S., Tasdizen, T., Thomas Fletcher, P., Joshi, S., Whitaker, R., 2010. Manifold modeling for brain population analysis. Med Image Anal 14, 643–653.

Gönen, M., 2013. Supervised Multiple Kernel Embedding for Learning Predictive Subspaces. IEEE Trans Knowl Data Eng 25, 2381–2389.

Grant, M., Boyd, S., 2013. CVX: Matlab software for disciplined convex programming, version 2.0 beta. http://cvxr.com/cvx.

Komajda, M., Lam, C. S. P., 2014. Heart failure with preserved ejection fraction: a clinical dilemma. Eur Heart J 35, 1022–1032.

Lafitte, S., Reant, P., Zaroui, A., Donal, E., Mignot, A., Bougted, H., Belghiti, H., Bordachar, P., Deplagne, A., Chabaneix, J., Franceschi, F., Deharo, J. C., Santos, P. D., Clementy, J., Roudaut, R., Leclercq, C., Habib, G., 2009. Validation of an echocardiographic multiparametric strategy to increase responders patients after cardiac resynchronization: A multicentre study. Eur Heart J 30, 2880–2887.

Lanckriet, G. R. G., De Bie, T., Cristianini, N., Jordan, M. I., Noble, S., 2004. A statistical framework for genomic data fusion. Bioinformatics 20, 2626–2635.

Lekavich, C. L., Barksdale, D. J., Neelon, V., 2015. Heart failure preserved ejection fraction (HFpEF): an integrated and strategic review. Heart Fail Rev 20 (6), 643–653.

Lin, Y.-Y., Liu, T.-L., Fuh, C.-S., 2011. Multiple kernel learning for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell 33, 1147–1160.

Lombaert, H., Peyrat, J., 2013. Joint statistics on cardiac shape and fiber architecture. In: MICCAI, LNCS. Vol. 8150. pp. 492–500.

Mädler, C., Payne, N., Wilkenshoff, U., Cohen, A., Derumeaux, G., Piérard, L., Engvall, J., Brodin, L., Sutherland, G., Fraser, A. G., 2003. Non-invasive diagnosis of coronary artery disease by quantitative stress echocardiography: optimal diagnostic models using off-line tissue Doppler in the MYDISE study. Eur Heart J 24, 1584–1594.

Mor-Avi, V., Lang, R. M., Badano, L. P., Belohlavek, M., Cardim, N. M., Derumeaux, G., Galderisi, M., Marwick, T., Nagueh, S. F., Sengupta, P. P., Sicari, R., Smiseth, O. A., Smulevitz, B., Takeuchi, M., Thomas, J. D., Vannan, M., Voigt, J.-U., Zamorano, J. L., 2011. Current and evolving echocardiographic techniques for the quantitative evaluation of cardiac mechanics: ASE/EAE consensus statement on methodology and indications endorsed by the Japanese Society of Echocardiography. Eur J Echo 24, 277–313.

Nadler, B., Lafon, S., Coifman, R., Kevrekidis, I., 2008. Diffusion maps-a probabilistic interpretation for spectral embedding and clustering algorithms. Principal manifolds for data visualization and dimension reduction, 238–260.

Neal, R., 1996. Bayesian Learning for Neural Networks. Lecture notes in statistics.

Paulus, W. J., Tschöpe, C., Sanderson, J. E., Rusconi, C., Flachskampf, F. A., Rademakers, F. E., Marino, P., Smiseth, O. A., De Keulenaer, G., Leite-Moreira, A. F., Borbély, A., Edes, I., Handoko, M. L., Heymans, S., Pezzali, N., Pieske, B., Dickstein, K., Fraser, A. G., Brutsaert, D. L., 2007. How to diagnose diastolic heart failure: a consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the Heart Failure and Echocardiography Associations of the European Society of Cardiology. Eur Heart J 28, 2539–2550.

Penicka, M., Vanderheyden, M., Bartunek, J., 2014. Diagnosis of heart failure with preserved ejection fraction: role of clinical Doppler echocardiography. Heart 100, 68–76.

Rakotomamonjy, A., Bach, F., Canu, S., Grandvalet, Y., 2008. Sim-

Preprint version accepted to appear in Medical Image Analysis.
Final version of this paper available at http://www.sciencedirect.com/science/article/pii/S1361841516300846

pleMKL. J Mach Learn Res 9, 2491–521.

Sanchez-Martinez, S., Duchateau, N., Bijnens, B., Erdei, T., Fraser, A. G., Piella, G., 2015. Characterization of myocardial velocities by multiple kernel learning: application to heart failure with preserved ejection fraction. In: FIMH, LNCS. Vol. 9126. pp. 65–73.

Santaularia-Tomas, M., Abraham, T. P., 2009. Editorial: Criteria predicting response to CRT: Is more better? Eur Heart J 30, 2835–2837.

Shah, S. J., Katz, D. H., Selvaraj, S., Burke, M. A., Yancy, C. W., Gheorghiade, M., Bonow, R. O., Huang, C.-C., Deo, R. C., 2014. Phenomapping for Novel Classification of Heart Failure With Preserved Ejection Fraction. Circulation 131, 269–279.

Shental, N., Hertz, T., Weinshall, D., Pavel, M., 2002. Adjustment learning and relevant component analysis. In: Heidelberg, S. B. (Ed.), ECCV. pp. 776–790.

Shuai, X.-X., Chen, Y.-Y., Lu, Y.-X., Su, G.-H., Wang, Y.-H., Zhao, H.-L., Han, J., 2011. Diagnosis of heart failure with preserved ejection fraction: which parameters and diagnostic strategies are more valuable? Eur J Heart Fail 13 (7), 737–745.

Sogaard, P., Egeblad, H., Pedersen, A. K., Kim, W. Y., Kristensen, B., Hansen, P. S., Mortensen, P. T., 2002. Sequential versus simultaneous biventricular resynchronization for severe heart failure: Evaluation by tissue Doppler imaging. Circulation 106, 2078–2084.

Tan, Y. T., Wenzelburger, F., Lee, E., Heatlie, G., Leyva, F., Patel, K., Frenneaux, M., Sanderson, J. E., 2009. The pathophysiology of heart failure with normal ejection fraction. Exercise echocardiography reveals complex abnormalities of both systolic and diastolic ventricular function involving torsion, untwist, and longitudinal motion. JACC: Cardiovascular Imaging 54, 36–46.

Teske, A. J., De Boeck, B. W. L., Melman, P. G., Sieswerda, G. T., Doevendans, P. A., Cramer, M. J. M., 2007. Echocardiographic quantification of myocardial function using tissue deformation imaging, a guide to image acquisition and analysis using tissue Doppler and speckle tracking. Cardiovasc Ultrasound 5, 27.

Tiwari, P., Kurhanewicz, J., Madabhushi, A., 2013. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. Med Image Anal 17, 219–235.

Tzortzis, G., Likas, A., 2012. Greedy unsupervised multiple kernel learning. In: Artificial Intelligence: Theories and Applications. Vol. 7297. pp. 73–80.

van Zalen, J., Patel, N. R., J Podd, S., Raju, P., McIntosh, R., Brickley, G., Beale, L., Sturridge, L. P., Lloyd, G. W. L., 2015. Prognostic importance of tissue velocity imaging during exercise echocardiography in patients with systolic heart failure. Echo Research and Practice 2, 19–27.

Weissler, A., Harris, W., Schoenfeld, C., 1968. Systolic time intervals in heart failure in man. Circulation 37, 149–159.

Wolz, R., Aljabar, P., Hajnal, J. V., Lötjönen, J., Rueckert, D., 2012. Nonlinear dimensionality reduction combining MR imaging with non-imaging information. Med Image Anal 16, 819–830.

Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., Lin, S., 2007. Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell 29, 40–51.

Ye, D. H., Desjardins, B., Hamm, J., Litt, H., Pohl, K. M., 2014. Regional manifold learning for disease classification. IEEE Trans Med Imag 33 (6), 1236–1247.

Zhuang, J., Wang, J., Hoi, S. C. H., Lan, X., 2011. Unsupervised multiple kernel learning. J Mach Learn Res 20, 129–144.

Zordoky, B. N., Sung, M. M., Ezekowitz, J., Mandal, R., Han, B., Bjorndahl, T. C., Bouatra, S., Anderson, T., Oudit, G. Y., Wishart, D. S., Dyck, J. R. B., 2015. Metabolomic fingerprint of heart failure with preserved ejection fraction. Plos One 10 (5), e0124844.