

Handling confounding variables in statistical shape analysis - application to cardiac remodelling

Gabriel Bernardino^{a,*}, Oualid Benkarim^b, María Sanz de la Garza^c, Susanna Prat-González^{c,d}, Alvaro Sepulveda-Martinez^{e,f}, Fátima Crispi^{d,e,g}, Marta Sitges^{c,d,h}, Constantine Butakoffⁱ, Mathieu De Craene^j, Bart Bijmens^{a,d,k}, Miguel A. González Ballester^{a,k}

^aBCN Medtech, Dept. of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

^bMcGill University, Montreal, Canada

^cCardiovascular Institute, Hospital Clínic, Barcelona, Spain

^dIDIBAPS, Barcelona, Spain

^eBCNatal, Hospitals Clinic and Sant Joan de Déu, Universitat de Barcelona, Barcelona, Spain

^fFetal Medicine Unit, Department of Obstetrics and Gynecology, Hospital Clínico de la Universidad de Chile. Santiago de Chile, Chile

^gCIBER-ER, Barcelona, Spain

^hCIBERCV, Barcelona, Spain

ⁱBarcelona Supercomputing Center, Barcelona, Spain

^jPhilips Medisys, Paris, France

^kICREA, Barcelona, Spain

Abstract

Statistical shape analysis is a powerful tool to assess organ morphologies and find shape changes associated to a particular disease. However, imbalance in confounding factors, such as demographics might invalidate the analysis if not taken into consideration. Despite the methodological advances in the field, providing new methods that are able to capture complex and regional shape differences, the relationship between non-imaging information and shape variability has been overlooked.

We present a linear statistical shape analysis framework that finds shape differences unassociated to a controlled set of confounding variables. It includes two confounding correction methods: confounding deflation and adjustment. We applied our framework to a cardiac magnetic resonance imaging dataset, consisting of the cardiac ventricles of 89 triathletes and 77 controls, to identify cardiac remodelling due to the practice of endurance exercise. To test robustness to confounders, subsets of this dataset were generated by randomly removing controls with low body mass index, thus introducing imbalance.

The analysis of the whole dataset indicates an increase of ventricular volumes and myocardial mass in athletes, which is consistent with the clinical literature. However, when confounders are not taken into consideration no increase of myocardial mass is found. Using the downsampled datasets, we find that confounder adjustment methods are needed to find the real remodelling patterns in imbalanced datasets.

Keywords: Confounder correction, statistical shape analysis, computational anatomy, cardiac remodelling

*Corresponding author: Tel.: +34-000-000-0000; fax: +34-000-000-0000; Bernardino)

Email address: gabriel.bernardino@upf.edu (Gabriel)

1. Introduction

Analysing the shapes of parts of biological organs and organisms has been the object of extensive study for over a century (Thompson, 1942). This interest in shape is also present in medicine: several studies have focused in the relationship between organ morphology and illness. For instance, cardiac shape remodels to improve cardiac pressure/volume output under abnormal working conditions, and it is used to assess the presence/evolution of illness (Arts et al., 1994; Grossman et al., 1975). In a nutshell, pressure overload produces concentric remodelling (thickening of the myocardium without dilation of the ventricle) to maintain wall stresses low, and volume overload dilates the ventricle without a myocardial mass thickening. This is an oversimplification, as a volume overload will also increase pressure, and the exact remodelling mechanisms and triggers at a cellular level are still under research and discussion. The classical way of analysing shape in the clinical community consists in manually extracting hand-crafted features, and analysing these shape descriptors. These measurements are usually standardised and defined in guidelines (Lang et al., 2015), and usually refer to global characteristics of the shape that carry little regional information.

Nowadays, it is possible to acquire 3D images in clinical routine. Furthermore, advances in computing permit to automatically segment the images and to generate personalised 3D models of the organs (González Ballester et al., 2000; Mitchell et al., 2002; Ecabert et al., 2006; Bernard et al., 2018). Statistical shape analysis (SSA) is a set of techniques to represent both shapes and images and do the analysis directly with these objects, not being limited to only analyse previously defined measurements. SSA is used in the medical imaging field, in order to identify and represent shape variability of those organs (Cerrolaza et al., 2015; Blanc et al., 2012; Rajamani et al., 2007; Sierra et al., 2006). This allows expressing and quantifying regional shape patterns in a robust and objective manner, instead of working with a small set of predefined measurements on the shapes (like volumes and diameters). SSA can be used together with statistical learning techniques to construct mod-

els that find regional differences in anatomy that are associated to pathologies (Zhang et al., 2014a; Singh et al., 2014; Varano et al., 2017; Sarvari et al., 2017), based on a control and a pathological population. Roughly, the typical framework consists of first using SSA to construct an atlas of all shapes in the population, then use principal components analysis (PCA) or another dimensionality reduction (DR) technique to find a low-rank representation of the shape space, and finally use a classification algorithm on that space to train a model that predicts the control/pathologic status.

Beyond pathological variability, the shape of an organ also exhibits variability due to other factors, like lifestyle, gender, ethnicity, or size. The framework described above uses the implicit hypothesis that differences in shape are only due to the pathology. In some cases the remodelling associated to the pathology is prominent and easily identifiable. However, in others, like subclinical or early-stage studies, differences can often be very subtle and less pronounced than demographic-related variability. Given the difficulty to acquire medical data, many imaging studies are cross-sectional and observational, and participants are recruited prospectively. As imaging datasets have typically low sample sizes, it is not always possible to obtain balanced sub-datasets, and researchers are forced to analyse imbalanced datasets. If left uncorrected this imbalance in the population characteristics may result in wrong conclusions/models (Cohen et al., 2018). Even in cases where the pathological remodelling is significant, and the populations are similar in terms of demographics, demographic-related variability will add noise to the analysis. A number of authors have explored the usage of non-imaging information. For instance Singh et al. proposed a procedure similar to partial correlation between shape and several clinical variables while correcting for confounding variables (Singh et al., 2014). Zhang et al. adjusted by demographics in their studies of cardiac remodelling in myocardial infarction (Zhang et al., 2014b), and Zhang et al. and Mauger et al. explored the relationship between shape and classical clinical measurements (Zhang et al., 2017, 2014b; Mauger et al., 2019). However, not all authors include corrections

for confounders, and their effect in shape analysis studies has not been yet quantitatively tested.

In this paper, we present a SSA framework to find differences between control and pathological populations that outputs the most discriminating shape pattern, that can be visualised for interpretability. We quantitatively and qualitatively show the effect an imbalance in the confounding variables has in the analysis, and propose techniques to reduce that effect. The proposed model consists of the following steps: (1) the construction of an atlas of the personalised 3D meshes automatically generated from images; (2) identification and removal of shape variability due to confounding variables; (3) dimensionality reduction and classification.

To illustrate the framework, we use a dataset of cardiac magnetic resonance imaging (MRI) involving sedentary controls and triathlon athletes. This dataset was collected to study the remodelling due to the extended practice of endurance sport, which produces a volume overload to the heart. This volume overload triggers compensatory mechanisms to improve cardiac output and withstand the increased pressure during exercise. The whole of this remodelling is called the Athlete’s Heart and involves substantial changes in function and geometry at both rest and during exercise (D’Andrea et al., 2015; Schiros et al., 2013). Although the remodelling is not yet completely understood, researchers have established a strong relationship between cardiopulmonary performance during exercise and cardiac geometry at rest (La Gerche et al., 2012; Scharhag et al., 2002).

2. Methodology

The full process to compute the confounder invariant most discriminating shape pattern between two populations is summarised in Figure 1. The presented framework consists of 3 main steps:

1. Compute the mean shape of the population, and register all shapes to this template.
2. Identify and remove shape variability attributable to confounding variables.
3. Train a classification model to obtain the most discriminating shape pattern between both pop-

ulations and generate a visual representation of the most discriminating shape pattern

2.1. Atlas construction

From the short axis (SA) MRI sequence, we use a model-based automatic segmentation method to obtain personalised meshes with point-to-point correspondence (Ecabert et al., 2006). The method deforms a full heart (4 chambers) template mesh using a polyaffine deformation to match the myocardial boundaries. The method includes slice correction to remove misalignment between consecutive slices.

Since only the ventricles were visible in our images, we discarded the atria and big vessels from each segmented shape. Each resulting mesh has 4446 vertices (the left ventricle (LV) has 3052 vertices and the right ventricle (RV) has 1776 vertices, the right-most part of the septum belongs to both ventricles) in point-to-point correspondence, and 9004 triangles. Only the end diastolic (ED) frame is selected for the analysis. Since meshes are in point-to-point correspondence and share the same connectivity, they can be analysed using the point distribution model (PDM) (Cootes et al., 1995). In PDM, the shape of each patient j is associated to a vector with the concatenated position (x, y, z) of its nodes, giving a shape vector:

$$X_j = (x_0, y_0, z_0, x_1, y_1, z_1 \dots x_N, y_N, z_N) \quad (1)$$

A problem with this parametrisation is that it does not avoid degenerate self-intersecting shapes. More complex techniques have been developed, in which meshes are represented by the action of diffeomorphism to a given reference (Ashburner and Friston, 2011; Miller et al., 2006; Durrleman et al., 2014). These representations avoid self-intersections and provide a more accurate shape representation, but have a bigger computational cost. We chose PDM over these representations since we have not observed any of the non-diffeomorphic artefacts. In supplementary material S1 we have compared the metric induced by these methods with the PDM and found only minor differences. We applied generalised Partial Procrustes Analysis (Dryden and Mardia, 1998) to align the meshes thus removing the positioning and orientation variability. We maintained size during

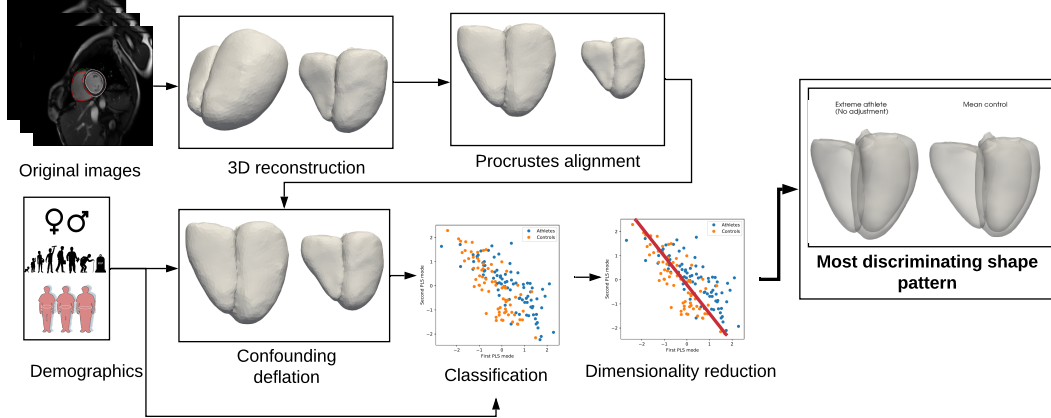


Figure 1: Schema of the framework and its different components. The input are the short axis MRI, and the demographics of the population. From the image, we generate a personalised 3D mesh of the ventricles, and align them using Procrustes analysis. Afterwards, we remove confounding-related shape variability using confounding deflation. This is followed by a dimensionality reduction and a classification steps. The final step is to compute the most discriminative shape pattern from the classification model coefficients.

this step, but differences in size due to anthropometric variables will be identified and removed in other steps of the framework. This algorithm is an iterative method, at each iteration computes an estimation of the mean shape \bar{X}^j and then rigidly registers each shape to that estimated mean. This is repeated until convergence, in order to obtain an unbiased mean. Here we show a full iteration j of the algorithm:

$$\bar{X}^j = \frac{1}{n} \sum_{i=1}^n X_i^j \quad (2)$$

$$\forall i : R_i^j, t_i^j = \arg \min_{R, t} \|R(X_i - t) - \bar{X}^j\|^2 \quad (3)$$

$$X_i^{j+1} = R_i^j(X_i - t_i) \quad (4)$$

2.2. Confounding deflation

To identify and remove the shape variability related to the confounding variables (M), and not to the studied condition, we use a procedure similar to partial correlation, as done in (Singh et al., 2014) for a regression problem. We assume that shape X can be decomposed in the sum of the population mean (μ_X) and some deformation from that mean that is

composed of noise (ϵ), variability caused by the confounding variables (X_M) and variability from other sources (X_i). The last component includes remodelling due to pathology :

$$X = \mu_X + X_M + X_i + \epsilon \quad (5)$$

The first step consists in estimating X_M by building a linear model (whose coefficients are w_M) that predicts the expected shape from the confounding variables M . The training data of this model is only one population (the controls), to avoid introducing possible inter-population differences in the model. This prediction will be our X_M . Then, for the shape of each individual, we subtract the predicted shape from the actual shape obtaining the prediction residuals. Residuals represent the part of the shape variability that cannot be explained by the confounding variables. To maintain the residual vectors in the same range as the original shape vectors, we summate the original population shape mean. The final formula reads:

$$X_{res} = X - X_M = (X - w_c \cdot M) + \mu_X \quad (6)$$

The regression model coefficients w_c are computed

using partial least squares (PLS) with the shapes X and confounding variables M . PLS is a regression method that projects the input and output data to two low-dimension subspaces (called embeddings) that have maximal covariance (Wegelin, 2000). The confounding variables are standardised to have 1 standard deviation and 0 mean, but shapes are only standardised to have 0 mean. The full process is described below, and consists in an iterative process where at each iteration a new regression dimension of the low-dimensional spaces is computed. The predicted part is then removed from the input and output spaces, and the process is repeated until the desired number of iterations is reached. Several versions of PLS exist, and we used Wold’s version. All PLS versions agree in the first iteration, but give different result for embedding spaces of more than one dimension. Here we show the algorithm for the r -th iteration of a PLS associating an input space X with some response Y , both being matrices. In our setting, and only for this part X will be the confounding variables (M in the rest of the paper) and Y the shape vectors (X in the rest of the paper).

First, we compute the new dimensions of the embedding spaces by solving an eigenvector problem:

$$\hat{u}^r, \hat{v}^r = \arg \max_{\|u^*\|=1, \|v^*\|=1} u^{*t} X^t Y v^* \quad (7)$$

We compute the rank one approximations (hat variables), and the regression coefficient w_r ; for OLS refers to the classical ordinary least squares.

$$\begin{aligned} \hat{X}^r &= OLS_{prediction}(X^r \hat{u}^r, Y^r \hat{v}^r) \\ w^r &= \hat{v}^r (\hat{u}^r)^t \frac{(X^r \hat{u}^r)^t Y^r \hat{v}^r}{X^r \hat{u}^r}, \\ \hat{Y}^r &= OLS_{prediction}(Y^r \hat{u}^r, X^r \hat{v}^r) \end{aligned} \quad (8)$$

These rank one optimisations are used to update the input X and its output Y by removing the part that is already predicted.

$$\begin{aligned} X^{r+1} &= X^r - \hat{X}^r \\ Y^{r+1} &= Y^r - \hat{Y}^r \end{aligned} \quad (9)$$

The final result of the algorithm is the regression coefficients w from X to Y ; and it is obtained by sum-

ming all the one-rank approximation regression coefficients.

$$w = \sum_r w^r \quad (10)$$

The coefficients w_i of the prediction model, associated to confounder variable M^i , of this model can also be visualised and interpreted. We can consider the partial determination coefficient, and visualise the shape pattern most associated to a certain confounder, as if it were a discriminating shape (see subsection 2.3).

2.2.1. Dimensionality reduction

Given the high dimensionality of the shape vectors and low number of samples, we use a DR method to find a subspace that contains the most relevant shape patterns. We choose linear methods over non-linear alternatives due to their higher interpretability, as the projection function is computed exactly. Additionally, an initial exploration of a non-linear approach (see supplementary material S2) did not lead to different results. The general linear DR model reads:

$$X_{orig} = \mu_X + K X_{red} + \epsilon \quad (11)$$

Each shape X_{orig} is expressed as the population mean (μ_x) plus some shape-specific associated low-dimensionality vector X_{red} and a noise term ϵ . The embedding matrix K is constructed depending on the dimensionality reduction method, and contains the most interesting shape directions according to a certain metric.

We use three different methods of linear DR: PCA, PLS and a combination of both. PCA and PLS have been reported to be used in conjunction to classification methods, and in particular logistic regression (Bastien, 2005). We have described the regression modality of PLS in subsection 2.2, but PLS also computes a DR space. The difference is that we discard the regression coefficients w and only use embedding vectors u^r in equation 7 of the input space. These vectors define a vector subspace, but are not guaranteed to be orthogonal, so we use QR decomposition (Lord et al., 1999) to obtain an orthonormal base. The combined method is based on prefiltering the

shape using PCA, keeping a high number of components ($> 90\%$ of the variance), to then use as input to a PLS. This decreases the computation time, and adds stability by denoising the data. Contrary to the typical procedure in machine learning, we chose not to standardise the PCA modes by variance before applying PLS, as the variance of each PCA mode carries important information of the signal-to-noise-ratio.

2.3. Classification

In this section, we present the method used to train the classifier model and use the model coefficients to obtain the most discriminating shape. The shape features obtained from the DR are combined with the confounding variables in a logistic regression model. We choose logistic regression because we expect not to have a complete separation between both populations, and we want the model to be simple and interpretable. The logistic model gives a probability that an individual j with shape X_j and confounding variables M_j belongs to the pathological or control populations.

$$Pr(y_j = \text{control} | X_j, M_j) = \text{logit}(\langle X_j, w_X \rangle + \langle M_j, w_M \rangle + b) \quad (12)$$

Logit refers to the logistic function $x \mapsto 1/(1 + \exp(-x))$. w_X and w_M are the logistic regression coefficients for the shape and confounding variables respectively, and are chosen to minimize the log-loss of the probability of the training data X . The log-loss is the logarithm of the probability that the model is inconsistent with the observed data:

$$\frac{1}{n} \sum \log(|y_j - Pr(y_j = 1 | X_j, M_j)|) \quad (13)$$

The logistic regression coefficients associated to the shape w_X , can be mapped from the reduced shape space to the full space by using the pseudoinverse of the dimensionality reduction matrix. Let $w_{X,red}$ be the coefficients associated to the reduced shape models, and K_{PCA} and K_{PLS} the projection matrices of PCA and PLS respectively. Then the coefficients associated with the full shape are:

$$w_X = K_{PCA}^T K_{PLS}^T w_{X,red} \quad (14)$$

Then, we can visualise and interpret the shape pattern. Analogously to multivariate regression, where the coefficients are indexed by the standard deviation to allow comparison among them, we need to adjust for differences in variance of the different coordinates. Since node coordinates carry no meaning on their own, we treat shape as an object itself and do PCA whitening of the shape, as is typically done with other multidimensional signals (Kessy et al., 2018). Since we are only interested in the remodelling direction, we normalise the vector to be unitary in the L_2 norm. With these corrections, all the shape features are correctly scaled by their importance in prediction. The full process to find the standardised shape pattern \hat{w} reads:

$$\hat{w} = \frac{\Sigma^{1/2} w_X}{\|\Sigma^{1/2} w_X\|} \quad (15)$$

Where Σ is the covariance matrix estimated using PCA. For visualisation, we can generate shapes that are representative of that shape pattern by adding the remodelling shape pattern, scaled with a parameter λ , to the mean shape. To keep the shapes within the original range, we impose that λ has to be within 3 standard deviations of the variance associated to the shape pattern.

$$X_{repr}(\lambda) = \mu_X + \lambda \hat{w} \quad (16)$$

We can quantify the presence of remodelling in each shape, obtaining a scalar score for each individual, by computing the dot product of the shape vectors with the raw logistic regression coefficients associated to the shape only. The previous PCA whitening is only done for visualisation and comparison of modes. If the shape pattern needs to be quantified in a population, the original one without standardization needs to be used.

$$\text{score}_i = \langle w_X, X_i - \mu_X \rangle \quad (17)$$

The shape patterns can be compared using the L_2 dot product between standardised shape vectors, which coincides with the correlation of the scores associated to each pattern.

3. Experimental setup

3.1. Clinical dataset

The study comprises 77 controls and 89 athletes that underwent a MRI, to study the cardiac remodelling triggered by the practice of endurance sport. The study was approved by a local ethical board, and all participants gave written informed consent for the handling of their data. Recruited athletes had been training an endurance sport, triathlon, over 10h a week during the last 5 years. All the study participants were Caucasian and none had previous cardiovascular disease, nor were any detected during the study. More details on the recruitment protocol are found in Bernardino et al. (2020). Table 1 shows the demographics of both populations.

The controls and athletes come from different studies, and the demographics of both populations did not match exactly in age, but roughly represent the same general population in age and gender. The study protocol and radiologist were the same for both cases. Age is statistically different between athletes and controls, but the difference is very small (2 years). We do not expect big differences due to this imbalance, since both athletes and controls are middle-aged. There are also statistically significant differences in both weight and body surface area (BSA), but these correspond to physiological remodelling since endurance athletes are obviously fitter than the general population. We used as possible confounders age, BSA and gender, the typical adjustment variables in cardiology studies.

The MRI acquisition was ECG-gated from the R-peak during breath-hold. The MRI machines were Siemens Aereo and Siemens Magnetom, with an in-plane spatial resolution ranging from 0.5mm to 1mm. The spacing between slices range between 8mm and 10.4mm, and the slice thickness was 8mm. MRI sequences were acquired with 25 frames per cardiac cycle. The ventricular contours (epicardium and endocardium in the case of the LV, and only epicardium of the RV) were automatically segmented from the MRI SA using the automatic procedure described in the methodology. ED images were selected as the ones with maximal LV volume, and the rest of the cardiac cycle was discarded.

The quality of the automatic segmentations was assessed by one of our experts, but no manual refinement was performed in order to preserve point-wise correspondences. Cases where errors could not be considered to be small were discarded: two individuals (both of them athletes) were discarded because the segmentation was inconsistent with the image. The segmentation and registration errors are handled as noise in our study. The meshes were very uniform and we found no self-intersecting artefacts. As a consequence of the thick slices, the apex was not correctly imaged and presented much more noise than the basal part of the ventricles. No extra processing was done to correct for potential apex artefacts. We decided to leave the apex in the study to avoid boundary effects.

3.2. Automatic measurements

We computed automatic measurements of the 3D shapes that are analogous to the classical clinical measurements, using the point-to-point correspondence and labelling coming from the model-based registration and segmentation. This allowed a better understanding of the discriminative shape patterns by assessing how these measurements vary in response to the remodelling score λ on the synthetically generated meshes according to equation 16. We computed the ED volumes of both ventricles, as well as the myocardial mass of the LV.

3.3. BMI-based downsampling

Obesity (defined as body mass index (BMI) > 30) and overweight (defined as BMI > 25) have been reported as risk factors to cardiovascular illnesses in the literature and have a clear influence in cardiac shape and function (Alpert et al., 2018). Surprisingly, overweight and athletic remodelling share similarities. Even if one might expect them to be opposite, as endurance athletes and overweight body fat are in the opposite sides of the spectrum, both remodellings are triggered by an increase of the heart's loading. In the case of the athletes, it is the increase of cardiac output (CO) during exercise that produces a volume overload, and in the obese it is a mix of increased CO needs at rest to account for the bigger

Table 1: Population demographics of the study participants. Athletes have a lower heart rate and weight than controls. The age is significantly different, but both cohorts are middle-aged and we do not expect major age-related differences. The p-values are obtained using a Mann-Whitney test.

	Athletes	Controls	p-value
Age [y]	35.4(6.1)	33.4(3.8)	0.013
BSA [m ²]	1.78(0.19)	1.86(0.20)	0.005
Weight [kg]	66.8(11.3)	73.5(15.1)	0.001
Height [m]	1.71(0.09)	1.73(0.08)	0.151
Women [%]	0.48	0.44	0.938
Resting HR [bpm]	57.2(8.4)	65.8(10.6)	<0.001

body size (Lavie et al., 2007) and a ventricular pressure overload due to an increase of arterial pressure (Messerli et al., 1982).

To study the effect of an imbalance in a shape-affecting variable between the control and case populations, we biased our population to increase the BMI of the controls, as overweight has a well known effect on the heart. To generate the imbalance, the control class was downsampled to maintain only 25% of its controls, favouring keeping the ones with higher BMI. The individuals to remove were selected randomly among the controls, with a probability of being kept proportional to the rank of its BMI in the control population. Athletes were not downsampled. This procedure was repeated with 100 different seeds to obtain different imbalanced datasets and add robustness to the results.

We analysed how this imbalance affected the remodelling pattern found, and to which extent could confounding adjustment and confounding deflation correct this effect. To study the stability of the shape pattern, we computed the L_2 product between the discriminative shape patterns obtained with the downsampled datasets and the results obtained with the full dataset, that serves as groundtruth. This was performed for the different DR methods. We also tested the L_2 product of the discriminative patterns with the BMI-shape pattern, obtained with an adaptation of our framework to the regression problem. Additionally, we compared qualitatively the automatic measurement response to the remodelling score of both the downsampled and original dataset. We tested both the covariate adjustment, and the con-

founding deflation. For the confounding deflation, we also evaluated how different choices during the training of the shape prediction model affected the obtained most discriminative shape pattern.

4. Results

4.1. Dimensionality reduction

To choose the best configuration of parameters for the DR method, we used 10-fold cross validation (CV) and computed the mean log-loss of all the validation set, defined in equation 13, over a wide combination of parameters. In table 2 we find the log-loss of the best parameter choice for the 3 different DR methods (PCA, PLS and PCA + PLS). They correspond to a PCA with 5 modes, a PLS with 3 and the PCA + PLS with 20 PCA modes and 3 PLS modes. In the supplementary material S3, Table 3 shows the results of all the parameter combinations tested. This experiment also provided an overview of how the use of demographics affected the classification metrics: there is a considerable improvement in terms of log-loss when the confounding variables are used in the model, and a minor improvement when using PLS instead of PCA. When using PCA + PLS, the metric is very similar to PLS. Confounding deflation gave a worse result than the raw (non-deflated) shapes when adjustment is used. However, confounding deflation improved the resulting classification metric with respect to the raw when confounders are not added to the logistic regression model.

Table 2: 10-fold CV log-loss scores of the best choice for each DR method.

Method	No deflation		No deflation	
	No adj.	Confounders adj.	No adj.	Confounders adj.
PCA ₅	0.58	0.46	0.48	0.46 ^T
PLS ₃	0.59	0.44	0.52	0.48
PCA ₂₀ + PLS ₃	0.59	0.43	0.50	0.46

4.2. Athletic model

We applied our framework to identify the athletic remodelling in our dataset, and compared the effect of using the different DR method and confounding-bias correction methods (confounding deflation and adjustment). We used the DR parameters found in the previous section via CV. Figure 2 shows the L_2 product between all combinations of DR methods and the adjustment or not by confounding variables. For the original (no confounding deflated) shapes (left), we found big differences due the inclusion or not of the confounding variables. The DR choice gave only minor differences in this dataset: the resulting most discriminating modes of the DR were very similar. When the confounding deflation was applied (Figure 2b), the use of the confounding adjustment became redundant and did not influence the resulting shape.

Figure 3 depicts the most discriminative shape patterns, expressed as mean + 2 STD, obtained by the PCA model with and without adjustment. Figure 4 shows their associated measurement response. Both the adjusted and unadjusted models found an increase of the ventricular volumes that is the same for both LV and RV. Both models found shape changes in the RV: the outflow region dilated more than the inlet base and apex. Regarding myocardial mass, however, the models gave different results. After confounder adjustment, we observed a large increase of LV myocardial mass similar to ventricular dilation, while the unadjusted model found no difference in the mass. The increase of mass of the adjusted model is concentrated in the base and lower in the apex: the base of the LV is flatter and is exposed to bigger mechanical stress, so it is more prone to compensatory myocardial thickening. Finally, the increase of ventricular volume in the adjusted model is more pro-

nounced than in the unadjusted model.

The adjustment or not by confounders resulted in big differences between the remodelling patterns found. On the other hand, differences due to DR were smaller, yet noteworthy. Figure 5 shows the remodelling patterns found for the different DR techniques, with a colormap showing the local amount of remodelling: red indicates substantial changes and blue no remodelling. The figure shows a view of the RV free wall, which is the region that experienced more shape changes. All 3 shape patterns were similar and followed the same global trends, but they presented regional differences. We observed that both PLS and PCA + PLS found a remodelling that was localised in the RV outlet and , at a lower degree, in the apex, while PCA showed a more spatially distributed remodelling that also affected the base and had a smoother transition between the affected and unaffected areas.

4.3. BMI model

To have a better understanding of the effect of an elevated BMI in the ventricles, we constructed a regression model that predicts BMI from the cardiac shape. As athletes and controls had different ranges of BMI, and to avoid finding any interference with the remodelling due to endurance sport, we only use the controls to build the BMI model. Also, since BMI and body size are related, we did not use the confounding variables in this model. The model was built using a PLS in its original regression mode, using 3 dimensions. To evaluate the model prediction capability, we computed the determination coefficient R^2 using 5-fold CV, which was 0.44. Similar to classification, we obtained synthetic representative shapes (\hat{x}_b) associated with certain BMI value b . The synthetic mesh associated with a certain BMI was the

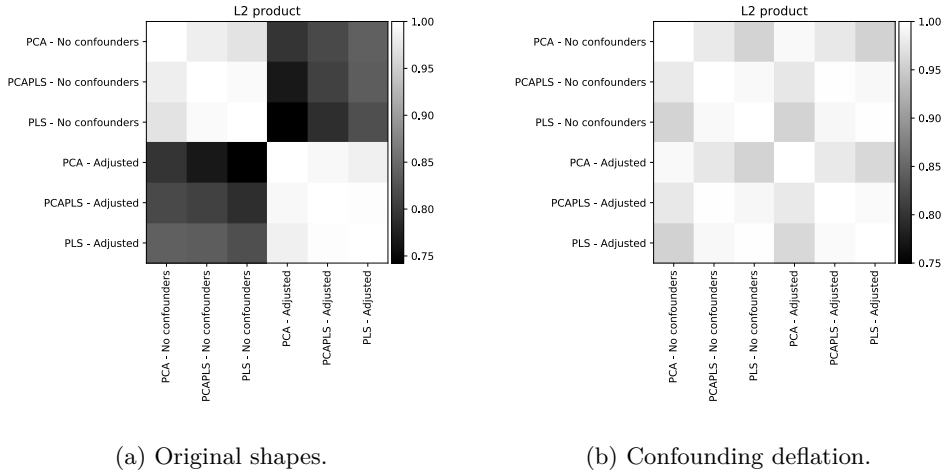


Figure 2: L_2 product between the most discriminative shape pattern (the L_2 product between the most discriminative shape coincides with the correlation between scores) obtained using different combinations of DR techniques and using or not the confounding adjustment in the logistic model. In subfigure 2a, we can see that differences due to the DR method were secondary to the adjustment when the shapes are not deflated. In subfigure 2b, we can see that the differences due to adjustment disappear after the shapes are corrected using confounding deflation.

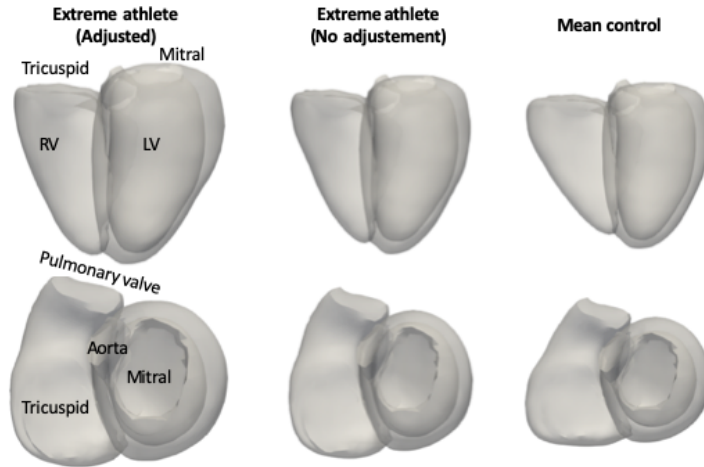


Figure 3: The picture depicts the mean shape of the population (right), and that mean shape after applying 2STD of the athletic remodelling obtained after adjusting by confounding (left) and the same remodelling without the adjustment (center). Both remodelling patterns show a dilation of the ventricles, with a bigger dilation of the RV outlet, but the one adjusted by confounders has a more pronounced dilation, and also a clear increase of the basal LV wall thickness. The different rows correspond to two different views: the top row depicts a longitudinal view of the anterior wall of both ventricles and the bottom corresponds to a short axis view of the base, with the observer located in the atria.

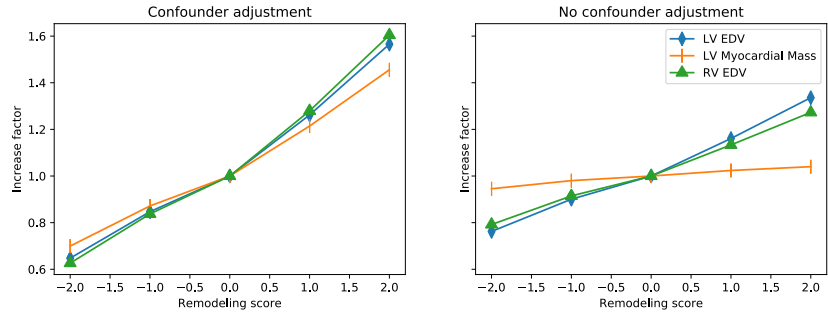


Figure 4: The plot shows the measurement response to the remodelling: for each of the synthetic meshes generated by adding the shape remodelling pattern to the mean shape with different magnitude, we compute classical measurements and show the variation ratio with respect to the mean shape.

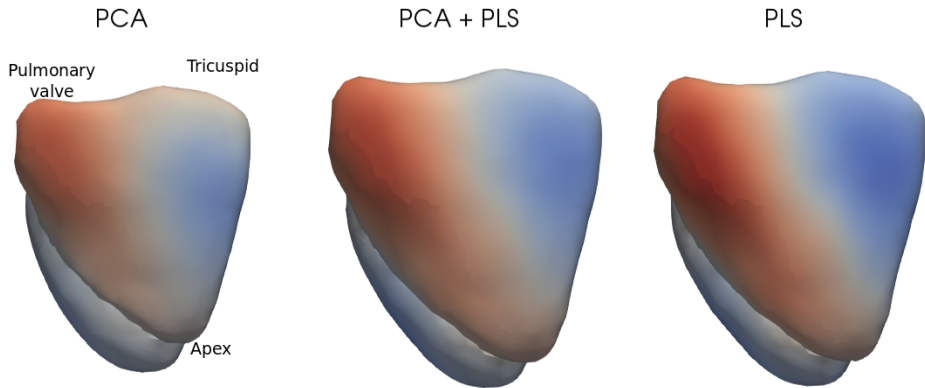


Figure 5: Close-up on the right ventricular free wall of the different models predicting the athletic remodelling shape pattern. The red-blue color map encodes the regional amount of remodelling: red means big differences compared to the control population, and blue small/ no remodelling. We can see that PCA remodelling is smoothly distributed through the whole ventricle, while the results obtained with PLS and PCA + PLS present a more localized remodelling with a sharper red-blue transition: the remodelling is concentrated in the outflow.

one having minimal-distance to the mean shape (using the Mahalanobis distance), constrained to having the required predicted BMI.

The representative shapes associated to BMI values of 17.5 and 30, which are very extreme values, can be seen in Figure 6, and Figure 7 shows the measurement response of the remodelling. The BMI-associated remodelling consists of a moderate increase of ventricular size and a bigger increment of the myocardial mass.

4.4. Confounding adjustment

Figure 8a depicts the L_2 products between the athletic remodelling shape pattern (normalised to be a unitary vector) derived from the full population, and the shape pattern derived from the BMI-imbalanced populations. We found that without adjustment the imbalance confused the method, and it mixed the differences due to BMI with the ones related to endurance training. The adjustment by confounding variables resulted in a better agreement between the

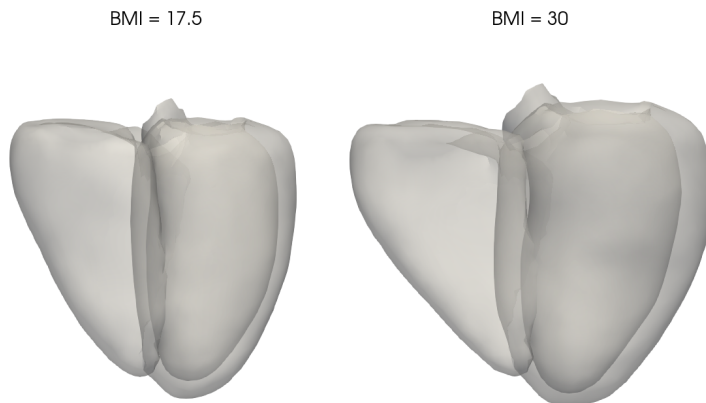


Figure 6: Synthetic representative shapes of a patient with BMI of 17.5 and another with BMI 30 according to the BMI predicting model. The remodelling consisted of an increase of volume (specially in the axial directions) and LV myocardial mass.

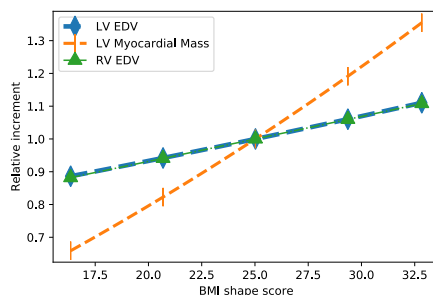


Figure 7: Measurement response of the BMI-related shape changes. As seen in the visual representation of the shapes, its main component is an increase of the myocardial mass, complemented with a smaller increase of volumes.

downsampled-derived remodelling patterns and the full-population pattern than the unadjusted model, as can be seen in Figure 9a. Figure 8b shows the dot product between the most discriminating shape pattern obtained with the downsampled (boxplot) and full (thick solid lines) datasets, and also the shape pattern associated to BMI, computed as described in section 4.3. There, we can see that in the downsampled datasets, the BMI-associated remodelling and the athletic one had lower correlation. This drop of correlation was larger for the models unadjusted by confounder variables.

Figure 9a shows the measurement response to the athletic remodelling of several models trained with a downsampled population. It shows different DR methods with confounding adjustment (upper row) and without confounding adjustment (lower row). We can see that when no confounding correction is used, the remodelling found is associated with lower myocardial mass, contrary to what is observed with the full dataset experiment in which myocardial mass was maintained. When the shape is adjusted by confounders, athletic remodelling was again associated with an increase of myocardial mass, consistent with

the full population model. Figure 10 shows the discriminating shape patterns for a randomly selected downsampled dataset. There, we can observe that the model associates athletic remodelling with a decrease of the wall thickness in the septum and apical regions of the LV when confounders are not considered.

Resulting shape patterns were affected by the DR methods, PLS was more unstable than PCA and PCA + PLS, who had a better correlation to their full-population discriminating shape. In Figure 9a, we can see that PLS was not able to recover the increase of myocardial mass, and its mode presented no variation in the myocardial mass while both PCA and PCA + PLS could.

4.5. Confounding deflation

In this subsection we analyze the effect of the confounding deflation. As stated above, the shape prediction model of the confounding deflation was trained using only the controls. We tested two possible scenarios: the case in which the athletes population was downsampled, and therefore the building of the shape residual model was not affected; and when the controls were downsampled (as in the previous subsection). The former was the most appropriate situation to apply confounding deflation, since there would not be extra unstability/bias introduced during the confounding deflation step, while the latter could introduce the bias in the dataset during the confounding deflation process. Finally, we show the potential danger of training the residual model using both populations and how the confounding deflation could even increment bias.

When controls were not downsampled, the population used to train the shape prediction model was relatively unbiased. Instead of downsampling the controls, we downsampled the athletes, removing athletes with high BMI analogously to the procedure used to downsample the controls. Figure 11a and 11b depict the same experiments as in the previous section: the dot product between the most discriminating shape obtained with the downsampled data and the remodelling obtained considering the whole population, and also the dot product with the BMI mode. Results showed a considerable decrease in variability

and a higher correlation with the full-dataset-derived remodelling than the confounder adjustment experiments (Figure 8a). Adding confounder adjustment on top of confounding deflation did not produce an increase of accuracy.

Figure 13a shows the results when the population in which the residual is trained is downsampled. We downsampled the controls based on their BMI. Results were much worse than when athletes were downsampled, and even worse than a simple confounder adjustment. Adding confounding adjustment on top of confounding deflation had a beneficial effect. Figure 13b shows the effect of using both athletes and controls in the training of the shape-prediction model for confounding deflation: there was a drop in stability compared to the use of a single population when PLS and PCA + PLS were used (Figure 8a). Strangely there was an improvement compared to the baseline (where confounders are completely ignored).

Therefore we observed that using both populations in the training of the shape prediction model during the confounding deflation step resulted in worse results. We explored the reason why using both populations can create a confounding effect, associating an imbalance in a variable to the inter-class shape differences. This can happen even when the variable is not associated to any shape remodelling. To illustrate this effect, we created a dummy synthetic variable which is just the athlete label plus Gaussian noise. Obviously, by generation we knew that this variable did not have any direct relationship to shape. To evaluate this effect, we computed the L_2 dot product with the most discriminating shape between athletes and controls and the shape pattern associated to the dummy variable in the shape-prediction model. We repeated this full variable generation and L_2 products computation process 100 times to remove randomness of the analysis. Figure 14 (left) shows the distribution of this dummy variable for a certain seed. We constructed two residual models one only with the controls, and the other with both populations. Figure 14 (right) shows the distribution of the dot product of the shape prediction model coefficients associated to the dummy variable when both populations are used, and when only the controls are used. The shape associated to the dummy variable is independent to

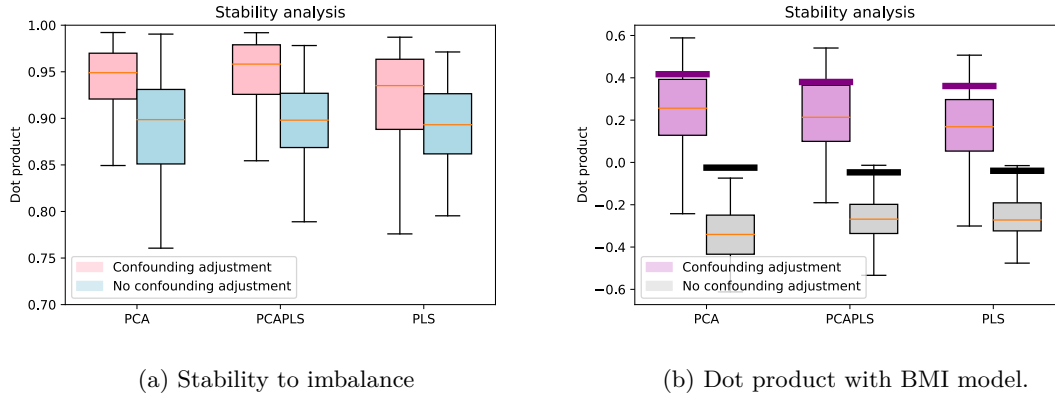


Figure 8: Effect of using confounding adjustment in the stability of the models derived from the imbalanced datasets for the 3 different DR methods. a) agreement of the model trained on downsampled data with the one on full data, measured via its dot product. b) The thick solid lines correspond to the dot product between the BMI shape remodelling pattern, and the athletic shape remodelling derived from the full population. The boxplots show the dot product between the downsampled derived shape patterns and the BMI shape remodelling pattern. When adjustment is used, the athletic and BMI mode have a positive relation (ie, they have remodelling partially in the same direction), due to both partially responding to an increase of pressure; however, that relation can become negative after downsampling, since the controls become more overweight.

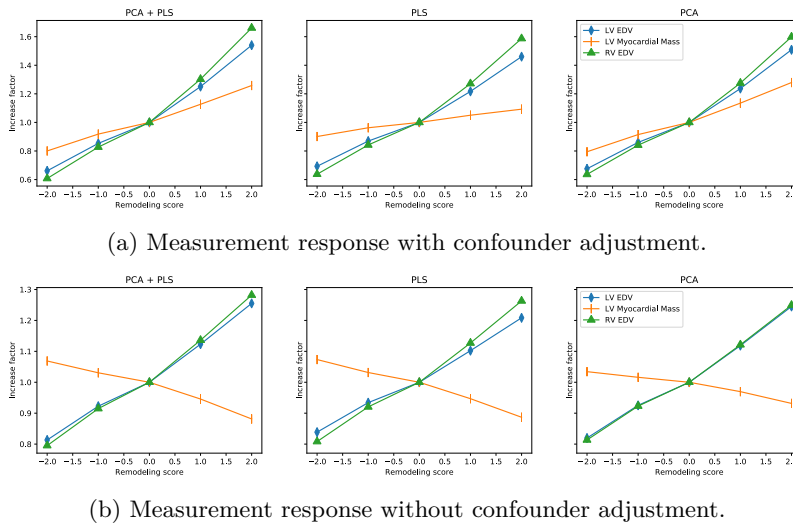


Figure 9: Measurement response of the downsampled population with and without adjustment. We can see that the unadjusted methods find a negative relationship between athletic remodelling and LV mass, but the adjusted methods find a positive relationship. Figure 4 shows the equivalent plots for the models trained with the full population, which we use as groundtruth. The adjusted models are more similar to the groundtruth than the unadjusted.

the athletic remodelling when only one population is used, but becomes very similar to the control-athlete difference when both populations are used in training due to confounding effect.

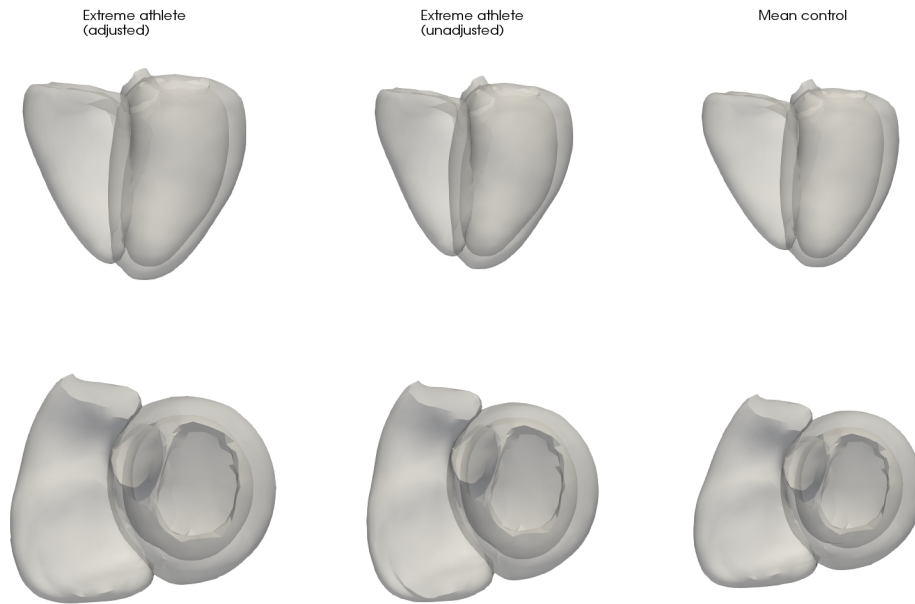
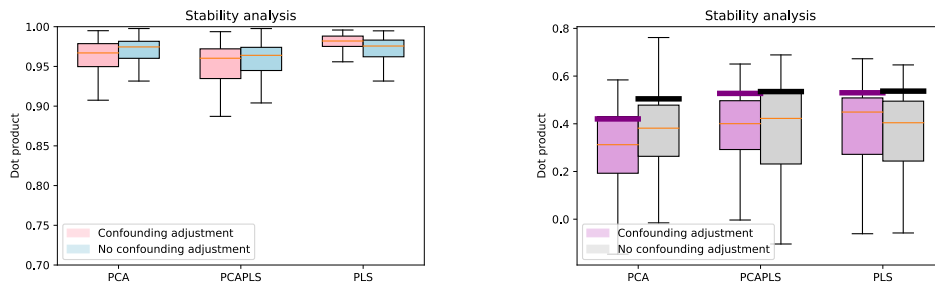


Figure 10: Mean shape (right), and mean shape + 2STD of the athletic shape remodelling pattern(left), derived from a BMI-imbalanced population. The pattern in the center column was obtained without adjustment, while the one on the right was adjusted. We can that the unadjusted model finds a decrease of myocardial mass in the apical and septal walls, while the adjusted finds an increase of mass



(a) Dot product with full data model.

(b) Dot product to BMI model for residual.

Figure 11: Stability analysis of the confounding deflation method when the athletes (that are not used in the construction of the residual model) are downsampled.

5. Discussion

The shape models corresponding to the athletic and overweight remodelling corroborated the current clinical literature. The BMI remodelling consisted in

mostly a concentric remodelling, both the LV mass and LV EDV increased, increasing more the mass than the volume: this coincides with the hypothesis that remodelling is concentric to cope with elevated pressure (increase of myocardial mass) and CO de-

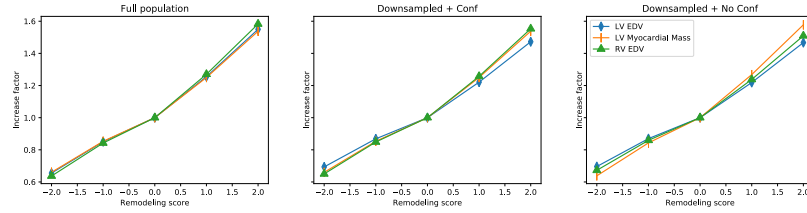
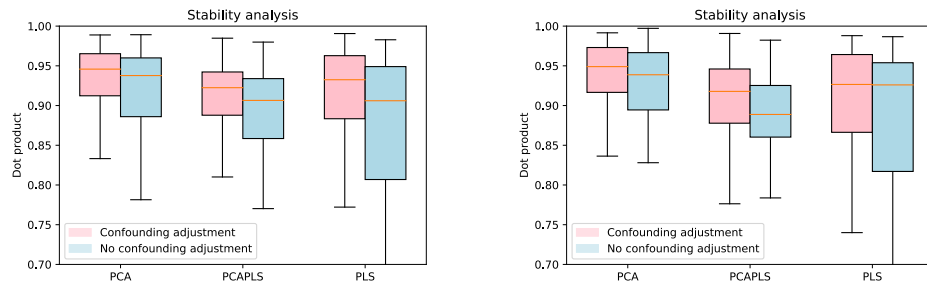


Figure 12: Measurements response of the shape pattern found with the athlete-downsampled data. There were very small differences between the different DR methods, and all methods were able to recover the original shape pattern.



(a) Confounder deflation trained on controls only. (b) Confounder deflation trained on both classes.

Figure 13: Effect of the population used to train the confounder deflation model on the discriminative pattern stability, assessed via its L_2 product with the result obtained using the full population. Subfigure a) shows when the training of the shape prediction model in the confounder deflation step is trained using the downsampled class, and subfigure b) shows when both controls and athletes are used for training.

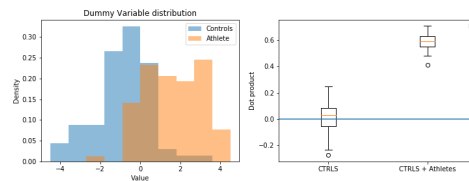


Figure 14: This figure shows the distribution of the dummy variable used as a confounder, and the dot product between the regression coefficients associated to the dummy variable in the residual model and the athletic remodelling shape pattern.

mands (increase of volume). Another hint that the remodelling was mostly pressure driven is that the septum of the mesh representing a high BMI individual presents a bulge below the aorta, that has been described as an early indicator of elevated pressure (Baltaeva et al., 2007; Gaudron et al., 2016). The athletic remodelling consisted of a predominant increase of volume, and if confounder adjustment was

performed, an increase of myocardial mass. There is some controversy if endurance athletes remodelling is more eccentric (they increase their volume more than their mass) or concentric (they increase their mass more than their volume), but there is a consensus that there is both an increase of mass and volume (Scharhag et al., 2002), and therefore we considered the adjusted model more accurate. The increase of

ventricular volume in the confounder adjusted model is considerably larger than the unadjusted. These differences between the adjusted and not adjusted models can be explained because the variability in size and myocardial mass can come also from the patients morphometrics: big persons have big hearts. Without considering this extra non-imaging information, it is not possible to discern when the size of the heart is due to the patient being big, or the heart had a dilation as a remodelling reaction to exercise.

The artificially generated imbalanced sets allowed us to validate our hypothesis: without any correction, the athletic remodelling presented a reduction in LV myocardial mass. This is obviously false, since all studies have found that endurance exercise provokes an increase of myocardial mass. The downsampled control population had a high percentage of overweight, who had a concentric remodelling and causing the previous association with controls and higher myocardial mass in the unadjusted model. The stability analysis, also showed a bigger match between the downsampled-derived remodelling and the full-dataset one when confounding adjustment was used.

We also studied another strategy: confounding deflation. Confounding deflation consists of generating a model that predicts shape from the confounders, and working with the residual of that prediction. Our results showed that this strategy worked well when there was access to a good population to train the shape prediction model: otherwise it can actually increase bias. This was shown when training on the downsampled population, or when both the case and control were used simultaneously.

Finally, we compared different linear methods for DR: PCA and PLS. In the original dataset, we obtained a better classification accuracy using PLS. We also showed that PLS was able to capture more localised remodelling than PCA, that was limited to smooth and global remodelling. However, PCA outperformed PLS in stability tested during the down-sampling analysis. Both models could be combined by using a coarse PCA, that only removes the modes encoding very little variance, with a PLS. With this method we were able to capture both localised remodelling and have high stability to imbalances. In our work, we have focused on linear DR methods,

but the same results are obtained using non-linear methods: an experiment using U-MAP can be found in supplementary material S2 and we observed the same confounding effect that was partially corrected after using confounding adjustment.

A limitation of our work is that only shape features were considered. It remains to be seen if intensity-based features, such as radiomics (Cetin et al., 2018), can provide added value to the classification problem. Another limitation is that only ED was analysed using this technique, which might be enough for studying athletic remodelling. For other conditions that primarily affect the systolic phase, the full cardiac dynamics should be analysed using spatio-temporal atlases. Furthermore, since only SA images were used for constructing the model, the apical region could not be analysed accurately: for assessing apical remodelling, 3D-echocardiography might be better suited.

6. Conclusion

We have presented a SSA framework to find regional shape differences between two populations, taking special care to correct for any potential bias related to demographics parameters. The framework is fully linear: it used a PCA sequentially combined with a PLS as dimensionality reduction of the shapes, followed by a logistic regression model. The linearity allows to easily interpret and visualise the model and build synthetic representative shapes of the model. To correct for confounding effects, it incorporates adjustment, where the confounding variables are added to the logistic model and confounding deflation, which consists of building a regression model that predicts shape from confounding variables, and we used it to remove the shape variability associated to these variables.

We applied our framework to a real dataset consisting of athletes and controls to find the remodelling due to the practice of endurance exercise. Our results confirmed the current literature on endurance-sport remodelling in the LV: ventricular dilation and increment of myocardial mass, specially in the basal area. In the RV, we found that the volume increase was not homogeneous but concentrated in the outflow.

In the controls, we used an adaptation of our classification framework to regression to explore obesity remodelling and found it to be mainly an increase of myocardial mass.

In this population, we analysed the effect of confounders in a semi-synthetic dataset obtained by downsampling the control population non-uniformly, keeping individuals with high BMI. Even if athletic remodelling is very prominent, we were able to bias the model to output that athletes have lower myocardial mass than controls. This was corrected when adjustment was used. However, we found that we could only use confounding deflation when the control population was relatively big and balanced, and if that is not the case using confounding deflation can actually increment bias. Here, we presented an example of a cardiology application, but the proposed method is not unique for the heart. Future studies could identify which demographic variables influence the shape of other organs, based on the same methodology.

In our work we have tested only linear SSA methods, but this confounder-related problems might appear even more with the use of more complex frameworks, able to capture non-linear shape patterns to capture subtler morphology differences in populations that are almost indistinguishable from controls and differences cannot be found through traditional means.

Acknowledgements

We thank Dr Weese and Dr Groth from Philips Research Hamburg for the segmentation tool and Dr Piella for fruitful discussions.

Funding

This study was partially supported by the Spanish Ministry of Economy and Competitiveness (grant DEP2013-44923-P, TIN2014-52923-R; Maria de Maeztu Units of Excellence Programme - MDM-2015-0502), el Fondo Europeo de Desarrollo Regional (FEDER) , the European Union under the Horizon 2020 Programme for Research, Innovation (grant agreement No.

642676 CardioFunXion) and Erasmus+ Programme (Framework Agreement number: 2013-0040), “la Caixa” Foundation (LCF/PR/GN14/10270005, LCF/PR/GN18/10310003), Instituto de Salud Carlos III (PI14/00226, PI17/00675) integrated in the “Plan Nacional I+D+I” and AGAUR 2017 SGR grant nº 1531.

References

- Alpert, M.A., Karthikeyan, K., Abdullah, O., Ghabban, R., 2018. Obesity and Cardiac Remodeling in Adults: Mechanisms and Clinical Implications. *Progress in Cardiovascular Diseases* 61, 114–123. doi:[10.1016/j.pcad.2018.07.012](https://doi.org/10.1016/j.pcad.2018.07.012).
- Arts, T., Prinzen, F.W., Snoeckx, L.H.E.H., Rijcken, J.M., Reneman, R.S., 1994. Adaption of cardiac structure by mechanical feedback in the environment of the cell: A model study. *Biophysical Journal* 66, 953–961. doi:[10.1016/S0006-3495\(94\)80876-8](https://doi.org/10.1016/S0006-3495(94)80876-8).
- Ashburner, J., Friston, K.J., 2011. Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *NeuroImage* doi:[10.1016/j.neuroimage.2010.12.049](https://doi.org/10.1016/j.neuroimage.2010.12.049).
- Baltaeva, A., Marcinak, M., Bijnens, B., Moggridge, J., He, F., Antonios, T., MacGregor, G., Sutherland, G., 2007. Regional left ventricular deformation and geometry analysis provides insights in myocardial remodelling in mild to moderate hypertension. *European Journal of Echocardiography* doi:[10.1016/j.euje.2007.08.004](https://doi.org/10.1016/j.euje.2007.08.004).
- Bastien, P., 2005. PLS generalised linear regression. *Computational Statistics & Data Analysis* 48, 17–46. doi:[10.1016/j.csda.2004.02.005](https://doi.org/10.1016/j.csda.2004.02.005).
- Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M.A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V.A., Krishnamurthi, G., Rohe, M.M., Pennek, X., Sermesant, M., Isensee, F., Jager, P., Maier-Hein, K.H., Full,

- P.M., Wolf, I., Engelhardt, S., Baumgartner, C.F., Koch, L.M., Wolterink, J.M., Isgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.M., 2018. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Transactions on Medical Imaging* 37, 2514–2525. doi:[10.1109/TMI.2018.2837502](https://doi.org/10.1109/TMI.2018.2837502).
- Bernardino, G., Sanz de la Garza, M., Domenech-Ximenes, B., Prat-González, S., Perea, R.J., Blanco, I., Burgos, F., Sepulveda-Martinez, A., Rodriguez-Lopez, M., Crispi, F., Butakoff, C., González Ballester, M.A., De Craene, M., Sitges, M., Bijmens, B., 2020. Three-dimensional regional bi-ventricular shape remodeling is associated with exercise capacity in endurance athletes. *European Journal of Applied Physiology* doi:[10.1007/s00421-020-04335-3](https://doi.org/10.1007/s00421-020-04335-3).
- Blanc, R., Seiler, C., Székely, G., Nolte, L.P., Reyes, M., 2012. Statistical model based shape prediction from a combination of direct observations and various surrogates: Application to orthopaedic research. *Medical Image Analysis* 16, 1156–1166. doi:[10.1016/j.media.2012.04.004](https://doi.org/10.1016/j.media.2012.04.004).
- Cerrolaza, J.J., Reyes, M., Summers, R.M., González-Ballester, M., Linguraru, M.G., 2015. Automatic multi-resolution shape modeling of multi-organ structures. *Medical Image Analysis* 25, 11–21. doi:[10.1016/j.media.2015.04.003](https://doi.org/10.1016/j.media.2015.04.003).
- Cetin, I., Sanroma, G., Petersen, S.E., Napel, S., Camara, O., González Ballester, M.A., Lekadir, K., 2018. A radiomics approach to computer-aided diagnosis with cardiac cine-MRI, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi:[10.1007/978-3-319-75541-0_{9}](https://doi.org/10.1007/978-3-319-75541-0_{9}).
- Cohen, J.P., Luck, M., Honari, S., 2018. Distribution matching losses can hallucinate features in medical image translation, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi:[10.1007/978-3-030-00928-1_{60}](https://doi.org/10.1007/978-3-030-00928-1_{60}).
- Cootes, T., Taylor, C., Cooper, D., Graham, J., 1995. Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding* 61, 38–59. doi:[10.1006/cviu.1995.1004](https://doi.org/10.1006/cviu.1995.1004).
- D’Andrea, A., La Gerche, A., Golia, E., Teske, A.J., Bossone, E., Russo, M.G., Calabrò, R., Baggish, A.L., 2015. Right heart structural and functional remodeling in athletes. *Echocardiography* 32, 11–22. doi:[10.1111/echo.12226](https://doi.org/10.1111/echo.12226).
- Dryden, L., Mardia, K.V., 1998. Statistical Shape Analysis, in: John Wiley & Son.
- Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., Trounev, A., 2014. Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage* 101, 35–49. doi:[10.1016/j.neuroimage.2014.06.043](https://doi.org/10.1016/j.neuroimage.2014.06.043).
- Ecabert, O., Peters, J., Weese, J., Lorenz, C., Von Berg, J., Walker, M.J., Olszewski, M.E., Vembar, M., 2006. Automatic heart segmentation in CT: Current and future applications. *Medical Image Analysis* 10, 12–17.
- Gaudron, P.D., Liu, D., Scholz, F., Hu, K., Florescu, C., Herrmann, S., Bijmens, B., Ertl, G., Störk, S., Weidemann, F., 2016. The septal bulge—an early echocardiographic sign in hypertensive heart disease. *Journal of the American Society of Hypertension* 10, 70–80. doi:[10.1016/j.jash.2015.11.006](https://doi.org/10.1016/j.jash.2015.11.006).
- González Ballester, M.A., Zisserman, A., Brady, M., 2000. Segmentation and measurement of brain structures in MRI including confidence bounds. *Medical image analysis* 4, 189–200.
- Grossman, W., Jones, D., McLaurin, L.P., 1975. Wall stress and patterns of hypertrophy in the human left ventricle. *Journal of Clinical Investigation* 56, 56–64. doi:[10.1172/JCI108079](https://doi.org/10.1172/JCI108079).
- Kessy, A., Lewin, A., Strimmer, K., 2018. Optimal Whitening and Decorrelation. *The American Statistician* 72, 309–314. doi:[10.1080/00031305.2016.1277159](https://doi.org/10.1080/00031305.2016.1277159).

- La Gerche, A., Burns, A.T., Mooney, D.J., Inder, W.J., Taylor, A.J., Bogaert, J., MacIsaac, A.I., Heidbüchel, H., Prior, D.L., 2012. Exercise-induced right ventricular dysfunction and structural remodelling in endurance athletes. *European Heart Journal* 33, 998–1006. doi:[10.1093/eurheartj/ehr397](https://doi.org/10.1093/eurheartj/ehr397).
- Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T., Lancellotti, P., Muraru, D., Picard, M.H., Rietzschel, E.R., Rudski, L., Spencer, K.T., Tsang, W., Voigt, J.U., 2015. Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American society of echocardiography and the European association of cardiovascular imaging. *European Heart Journal Cardiovascular Imaging* doi:[10.1093/ehjci/jev014](https://doi.org/10.1093/ehjci/jev014).
- Lavie, C.J., Milani, R.V., Ventura, H.O., Cardeñas, G.A., Mehra, M.R., Messerli, F.H., 2007. Disparate Effects of Left Ventricular Geometry and Obesity on Mortality in Patients With Preserved Left Ventricular Ejection Fraction. *The American Journal of Cardiology* 100, 1460–1464. doi:[10.1016/j.amjcard.2007.06.040](https://doi.org/10.1016/j.amjcard.2007.06.040).
- Lord, N., Golub, G.H., Loan, C.F.V., 1999. Matrix Computations. *The Mathematical Gazette* 83, 556. doi:[10.2307/3621013](https://doi.org/10.2307/3621013).
- Mauger, C., Gilbert, K., Lee, A.M., Sanghvi, M.M., Aung, N., Fung, K., Carapella, V., Piechnik, S.K., Neubauer, S., Petersen, S.E., Suinesiaputra, A., Young, A.A., 2019. Right ventricular shape and function: cardiovascular magnetic resonance reference morphology and biventricular risk factor morphometrics in UK Biobank. *Journal of Cardiovascular Magnetic Resonance* 21, 41. doi:[10.1186/s12968-019-0551-6](https://doi.org/10.1186/s12968-019-0551-6).
- Messerli, F.H., Ventura, H.O., Reisin, E., Dreslinski, G.R., Dunn, F.G., MacPhee, A.A., Frohlich, E.D., 1982. Borderline hypertension and obesity: two prehypertensive states with elevated cardiac output. *Circulation* 66, 55–60. doi:[10.1161/01.CIR.66.1.55](https://doi.org/10.1161/01.CIR.66.1.55).
- Miller, M.I., Trouvé, A., Younes, L., 2006. Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision* doi:[10.1007/s10851-005-3624-0](https://doi.org/10.1007/s10851-005-3624-0).
- Mitchell, S.C., Bosch, J.G., Lelieveldt, B.P.F., Van der Geest, R.J., Reiber, J.H.C., Sonka, M., 2002. 3-D active appearance models: Segmentation of cardiac MR and ultrasound images. *IEEE Transactions on Medical Imaging* 21, 1167–1178. doi:[10.1109/TMI.2002.804425](https://doi.org/10.1109/TMI.2002.804425).
- Rajamani, K.T., Styner, M.A., Talib, H., Zheng, G., Nolte, L.P., González Ballester, M.A., 2007. Statistical deformable bone models for robust 3D surface extrapolation from sparse data. *Medical Image Analysis* 11, 99–109. doi:[10.1016/j.media.2006.05.001](https://doi.org/10.1016/j.media.2006.05.001).
- Sarvari, S.I., Rodriguez-Lopez, M., Nuñez-Garcia, M., Sitges, M., Sepulveda-Martinez, A., Camara, O., Butakoff, C., Gratacos, E., Bijmens, B., Crispi, F., 2017. Persistence of Cardiac Remodeling in Preadolescents With Fetal Growth Restriction. *Circulation: Cardiovascular Imaging* 10. doi:[10.1161/CIRCIMAGING.116.005270](https://doi.org/10.1161/CIRCIMAGING.116.005270).
- Scharhag, J., Schneider, G., Urhausen, A., Rochette, V., Kramann, B., Kindermann, W., 2002. Athlete's heart. *Journal of the American College of Cardiology* 40, 1856–1863. doi:[10.1016/S0735-1097\(02\)02478-6](https://doi.org/10.1016/S0735-1097(02)02478-6).
- Schiros, C.G., Ahmed, M.I., Sanagala, T., Zha, W., McGiffin, D.C., Bamman, M.M., Gupta, H., Lloyd, S.G., Denney, T.S., Dell'Italia, L.J., 2013. Importance of three-dimensional geometric analysis in the assessment of the athlete's heart. *American Journal of Cardiology* 111, 1067–1072. doi:[10.1016/j.amjcard.2012.12.027](https://doi.org/10.1016/j.amjcard.2012.12.027).
- Sierra, R., Zsemlye, G., Székely, G., Bajka, M., 2006. Generation of variable anatomical models for surgical training simulators. *Medical Image Analysis* 10, 275–285. doi:[10.1016/j.media.2005.11.003](https://doi.org/10.1016/j.media.2005.11.003).
- Singh, N., Thomas Fletcher, P., Samuel Preston, J., King, R.D., Marron, J.S., Weiner, M.W., Joshi, S.,

2014. Quantifying anatomical shape variations in neurological disorders. *Medical Image Analysis* 18, 616–633. doi:[10.1016/j.media.2014.01.001](https://doi.org/10.1016/j.media.2014.01.001).
- Thompson, D.W., 1942. On growth and form. doi:[10.2307/2019330](https://doi.org/10.2307/2019330).
- Varano, V., Gabriele, S., Teresi, L., Dryden, I.L., Puddu, P.E., Torromeo, C., Piras, P., 2017. The TPS Direct Transport: A New Method for Transporting Deformations in the Size-and-Shape Space. *International Journal of Computer Vision* 124, 384–408. doi:[10.1007/s11263-017-1031-9](https://doi.org/10.1007/s11263-017-1031-9).
- Wegelin, J., 2000. A survey of Partial Least Squares (PLS) methods, with emphasis on the two-block case. Technical Report 371, Department of Statistics, University of Washington, Seattle, 2000. doi:[10.1.1.30.16](https://doi.org/10.1.1.30.16).
- Zhang, X., Cowan, B.R., Bluemke, D.A., Finn, J.P., Fonseca, C.G., Kadish, A.H., Lee, D.C., Lima, J.A., Suinesiaputra, A., Young, A.A., Medrano-Gracia, P., 2014a. Atlas-based quantification of cardiac remodeling due to myocardial infarction. *PLoS ONE* doi:[10.1371/journal.pone.0110243](https://doi.org/10.1371/journal.pone.0110243).
- Zhang, X., Cowan, B.R., Bluemke, D.A., Finn, J.P., Fonseca, C.G., Kadish, A.H., Lee, D.C., Lima, J.A.C., Suinesiaputra, A., Young, A.A., Medrano-Gracia, P., 2014b. Atlas-Based Quantification of Cardiac Remodeling Due to Myocardial Infarction. *PLoS ONE* 9, e110243. doi:[10.1371/journal.pone.0110243](https://doi.org/10.1371/journal.pone.0110243).
- Zhang, X., Medrano-Gracia, P., Ambale-Venkatesh, B., Bluemke, D.A., Cowan, B.R., Finn, J.P., Kadish, A.H., Lee, D.C., Lima, J.A.C., Young, A.A., Suinesiaputra, A., 2017. Orthogonal decomposition of left ventricular remodeling in myocardial infarction. *GigaScience* 6. doi:[10.1093/gigascience/gix005](https://doi.org/10.1093/gigascience/gix005).