Estimating 3D ventricular shape from 2D echocardiography: feasibility and effect of noise

Gabriel Bernardino[1,*], Constantine Butakoff[1], Marta Nuez-Garcia[1], Sebastian Imre Sarvari[2], Merida Rodriguez-Lopez[3,4], Fatima Crispi[3,4], Miguel A. González Ballester[1,5], Mathieu De Craene[6], and Bart Bijnens[1,5]

[1] DTIC, UPF, Barcelona, Spain
gabriel.bernardino@upf.edu
[2] Department of Cardiology, Oslo University Hospital, Rikshospitalet, Oslo, Norway
[3] Fetal i+D Fetal Medicine Research Center, IDIBAPS, Barcelona, Spain
[4] BCNatal, Hospital Clínic and Hospital Sant Joan de Déu, Barcelona, Spain
[5] ICREA, Barcelona, Spain
[6] Philips Research, Paris, France

Abstract. Many cardiac diseases are associated with changes in ventricular shape. However, in daily practice, the heart is mostly assessed by 2D echocardiography only. While 3D techniques are available, they are rarely used. In this paper we analyze to which extent it is possible to obtain the 3D shape of a left ventricle (LV) using measurements from 2D echocardiography. First, we investigate this using synthetic datasets, and afterwards, we illustrate it in clinical 2D echocardiography measurements with corresponding 3D meshes obtained using 3D echocardiography. We demonstrate that standard measurements taken in 2D allow quantifying only the ellipsoidal shape of the ventricle, and that capturing other shape features require either additional geometrical measurements or clinical information related to shape remodelling. We show that noise in the measurements is the primary cause for poor association between the measurements and the LV shape features and that an estimated 10% level of noise on the 2D measurements limits the recoverability of shape. Finally we show that clinical variables relating to the clinical history can substitute the lack of geometric measurements, thus providing alternatives for shape assessment in daily practice.

Keywords: LV shape, LV measurements, Echocardiography, Shape prediction, LV measurement accuracy

1 Introduction

Many cardiac pathologies are associated with changes in the cardiovascular system and in particular the shape of cardiac ventricles[1]. Some pathologies are associated to very particular changes in the shape, such as infarctions producing a concavity/aneurysm in one of the walls of the left ventricle (LV) [2]. Additionally, some conditions determined in fetal life (e.g. tetralogy of Fallot[3] or intra-uterine growth restriction [4]) shown to result in (sometimes subtle) shape

changes that are related to the severity of remodeling and thus might have clinical predictive value above more commonly used parameters like volumes and ejection fraction.

In daily practice, cardiac imaging is primarily carried out by 2D echocardiography. 2D echo has a good temporal resolution and a reasonable spatial resolution. It is also cheap and can be operated at bedside. Its problem is that it only allows to image a slice of the heart in each acquisition. This makes it difficult to assess the shape, as the whole heart is not observable from a single acquisition, but several views must be combined.

Even if 3D imaging techniques are available that give a complete view of the heart and might thus be preferable, they have important drawbacks in daily practice. 3D echo is nowadays not used in daily clinical practice due to its limited temporal and spatial resolution and need for extra transducers and high-end scanners. It is also more prone to having poor acoustic window and poor image quality. Other 3D technologies, MRI and CT also have drawbacks. They have better volume reproducibility and less imaging artifacts as compared to 2D echo, but their temporal resolution is much worse. Moreover, they are not portable, expensive and, in case of CT, use ionizing radiation.

In this paper, we investigate whether it is possible to infer the 3D shape of the human left ventricle from the 2D echo measurements with a linear regression approach. We will use not only the 2D measurements related to the geometry, but also functional and clinical parameters. This approach reveals which parts of the shape can be recovered in 2D and what confidence we can expect on those predictions. We aim to know in which situations 3D imaging is required, and in which 2D gives enough information.

To that end, we first create a synthetic dataset for studying the influence of noise on the measurements and, afterwards, we illustrate the use of our regression approach on a population of pre-adolescents born after intrauterine growth restriction (IUGR) where both 3D shapes and 2D measurements are available. Specifically, we want to study how the following shape patterns reported in the literature for these patients [5] can be best predicted from 2D measurements: the overall size of the Left Ventricle (LV), its sphericity, and a lateral shift of the apex with respect to the LV base. We also study in this paper the influence of increasing noise levels on the 2D measurements.

Although the problem of inferring the complete 3D shape from 2D images or an incomplete part of the 3D shape has already been addressed by the community (e.g. for the liver[6], or the femur[7]), this is, to our knowledge, the first attempt to infer the LV shape from 2D clinical measurements obtained through echocardiography, including the analysis of the impact of measurement noise on the prediction.

Table 1: Classification of the 2D measurements included in the study

Class	#measurements	examples
Geometry of the LV	18	ES volume, ED volume, internal dimension, long axis length, basal diameter
Function of the LV	14	cardiac output, cardiac index, mitral A speed
Geometry of the other parts	17	LA area, RV volume
Function of the other parts	18	TAPSE, Tricuspid E speed, AR duration
Clinical history	13	age, sex, IUGR label, height, weight

Table 2: Summary of the IUGR population.

Variable	mean	std	min	max
Age (years)	10.51	1.70	7.24	13.23
Weight (kg)	38.32	9.96	19.10	64.00
Height (m)	1.42	0.12	1.16	1.74
LV Volume (ml)	80.16	18.17	44.00	126.00
LV axial diameter (mm)	66.90	6.68	48.30	87.00

2 Data description

2.1 IUGR patient data

We used the dataset of [5] of 152, 7-13 years old, individuals (58 with Intrauterine Growth Restriction (IUGR) and 94 controls). For each individual we have a 3D mesh of the end-diastolic LV acquired using 3D echo (Echopac version 108.1.6) and an independent complete 2D echo study with the standard clinical measurements. After removing cases with incomplete measurements, 116 patients remained. The summary of the population can be found in Table 2. The meshes in this dataset have point correspondence, containing 720 nearly regular faces with an average face area of $13.9mm^2$. Each patient has 90 measurements related to the function and geometry of the LV, other chambers, as well as clinical history. Some examples of measurements included in the previous classification are showed in Table 1. A more complete overview of the variables included can be found in [8].

2.2 Generation of the synthetic database

In order to study the effect of noise on the measurements, we have generated a synthetic database matching the main sources of variability found in the IUGR population: size, sphericity and septal bulging (an asymmetrical *half moon* deformation). We first created a *template LV* as a sphere cropped at 3/4 of its height. We estimated the distribution of the scaling coefficients that the IUGR meshes follow by affinely registering the *template LV* to each LV mesh, and storing the scaling coefficients of the affine transform. With those we estimated a Multivariate Normal distribution C. We randomly assigned every individual to either the pathological (P) or control (H) group. Individuals belonging to the pathological group were given a stronger *half moon* shape pattern. The *template LV* was first positioned so that the cropped section is perpendicular to the y

axis, the *half moon* deformation was further introduced along the x axis using the following deformation for very point $\mathbf{p}(x, y, z)$ of the mesh:

$$\phi(x, y, z) = (|x - x_{mean}|(y - y_{min})(y - y_{max}), 0, 0) \tag{1}$$

where x_{mean} refers to the mean value of the x coordinates, y_{min} and y_{max} refer to the minimal and maximal y coordinate over the points of the template. The deformation is then normalized to have unit L_2 norm. The generation of a synthetic mesh was performed as follows:

1. Sample scaling coefficients c from C, scale the *template LV* using c and obtain the cropped ellipsoid e;
2. Generate the label H/P, with 50% probability for each class;
3. Sample μ from $D_H = \mathcal{N}(0cm, 0.3cm^2)$ if the label is H, or from $D_P = \mathcal{N}(1.5cm, 0.4cm^2)$ if the label is P;
4. The synthetic mesh is $m := e + \mu\phi$, where ϕ is the *half moon* deformation vector. We modeled the noise in the 3D mesh by a random displacement at every vertex with mean norm of 1mm.

2.3 Measurements

The 2D measurements in the synthetic dataset are taken from the 3D shape, and they contain similar information as real measurements. They are not directly comparable to the real 2D measurements, because the latter are obtained from the echocardiographic image. We used 3 different sets of synthetic measurements:

1. *Basic measurements:* they are equivalent to the volume, long axis length, basal diameter and biggest internal diameter as described in the European Association of Cardiovascular Imaging (EACVI) clinical recommendations ([8]). In Fig. 1a, there is a graphical description of how the linear measurements were taken in the mesh.
2. *Extra measurements:* basic measurements, with the internal diameter divided in two segments from the furthest points to the long axis (Fig. 1b). Those distances are specific to the *half moon* pattern.
3. *Random measurements:* basic measurements and 5 distances between random pairs of points and 5 diameters at different heights along the principal axis.

Optionally the H/P label used in the generation was added to the geometrical measurements.

For the IUGR dataset, 3 types of measurements, obtained from the 2D echo study, were used:

1. Left ventricular end-diastolic volume, obtained using the biplane method, and 3 linear measurements: long axis length, basal diameter and biggest internal diameter.
2. Measurements from the point 1 and the label of the IUGR condition.
3. All the measurements available in the 2D study, including function and geometry of all parts of the heart.
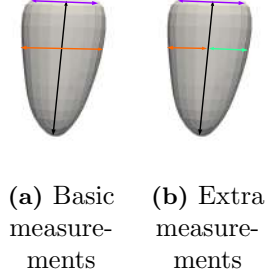
(a) Basic measure- ments   **(b)** Extra measure- ments

Fig. 1: Linear measurements taken for the synthetic meshes. The basic measurements correspond roughly to the long axis (black), internal dimension(orange) and basal diameter(purple). In the extra measurements, the internal dimension is split in the distances from the extremes to the long axis (orange and cyan).

3   3D shape regression from 2D measurements

This section describes how we learn a regression from a set of 2D measurements to the 3D shape. It also defines the metrics that will be used later to quantitatively compare regression results.

3.1   Shape predictor training

In order to carry out 3D shape prediction from the 2D echo measurements, we trained a linear regressor using the standardized 2D measurements as predictor variables. As output, we used the coordinates of the mesh we want to reconstruct after projection on a standard PCA basis. PCA has indeed the advantage of giving both projection and reconstruction linear operators. To make the PCA representation more compact, we kept PCA modes that explained 99% of the variance. As input, the PCA takes the concatenation of all mesh vertex coordinates. We used the Elastic Nets algorithm to learn the matrix of regression coefficients $\hat{\mathbf{W}}^*$ and bias vector $\hat{\mathbf{b}}^*$:

$$\hat{\mathbf{W}}^*, \hat{\mathbf{b}}^* = \arg\min_{\mathbf{W},\mathbf{b}} \left(\frac{1}{n}\|\hat{\mathbf{Y}} - \mathbf{X}\mathbf{W} - \hat{\mathbf{b}}\|_{\mathbf{fro}}^{\mathbf{2}} + \lambda_{\mathbf{1}}\|\mathbf{W}\|_{\mathbf{21}} + \lambda\|\mathbf{W}\|_{\mathbf{fro}}^{\mathbf{2}} \right) \quad (2)$$

where $\|\mathbf{W}\|_{\mathbf{fro}}$ is the Frobinus norm, $\|\mathbf{W}\|_{\mathbf{21}} = \sum_{\mathbf{i}}\|\mathbf{W}_{\mathbf{i}}\|_{\mathbf{2}}$, n is the number of samples used for training, X are the 2D measurements arranged in rows and $\hat{Y}$ are the PCA coordinates after projection. We kept only the modes explaining 99% of the variance. The λ_1 and λ_2 were determined using 3-Fold cross-validation with the training data. From Eqn. 2, we learn a regressor predicting PCA coordinates from the input 2D measurement. The resulting shape is then simply obtained by applying the PCA reconstruction to these regressed coordinates.

3.2 Evaluation

We used 9 different noise levels (between 0 and 0.5) for 2D measurements in the synthetic dataset. For every level, we have added to every measurement a white Gaussian noise with a mean μ_i and a variance $\sigma^2 = (\alpha\mu_i)^2$, where α is the noise level. We generated 10 different synthetic datasets and report in Sect. 4 the mean and 95% CI of theirs results. The quality of the regression in the IUGR dataset was evaluated using Leave-One-Out cross-validation and an independently generated test-set of size 50 for the synthetic dataset. We use the R^2 determination coefficient as a quality metric. It is defined as $R^2 = 1 - var(\mathbf{Y}_{pred} - \mathbf{Y})/var(\mathbf{Y})$, where $\mathbf{Y}$ is the real response and $\mathbf{Y}_{pred} = \mathbf{X}\mathbf{W}^*$. The R^2 coefficient is the percentage of the response space variance that can be predicted via the regression. Given a linear deformation column mode $\mathbf{m}$, the R^2 of the variability associated with that mode is defined similarly: $R^2_m = 1 - var(\mathbf{Y}_{pred}\mathbf{m})/var(\mathbf{Y}\ \mathbf{m})$

4 Results

4.1 PCA modes

As PCA was used as encoding of the 3D shape for learning the regression (Sect. 3.1), we show here PCA modes on both the IUGR and the synthetic data. For IUGR we have previously aligned the meshes using the Procrustes algorithm (preserving the scale). The 4 largest PCA modes and the percentage of variability they represent are shown in the top row of Fig. 2. The two modes with the largest variability are elongation and sphericity. The third mode appears to be associated with a 3D segmentation artifact due to the partial coverage of the outflow tract in many cases. The fourth mode is a non-symmetric deformation where the apex is shifted with respect to the base with a small inclination of the base, creating an impression of a *half moon* ventricular shape due to a bulging septum. The bottom row of Fig. 2 shows the main PCA modes learned on the synthetic data.

4.2 Synthetic data

We first evaluated how the choice of 2D measurements influences the regression quality. Fig. 3 plots the R^2 coefficient (Sect.3.2) for the first 3 PCA modes as a function of noise for all possible combinations of measurements defined in Sect. 2.3. Then, we investigated how training affects the quality of the regression. First we tested whether adding more training data or selecting another regression algorithm would affect the performance. Fig. 4.a shows the effect of changing the training data size for the basic set of measurements and Fig. 4.d for the random set. We also tested how different linear regression algorithms [9] compared in terms of R^2 values: PLS, Ridge, Elastic Nets, classic Linear Regression, Gaussian Kernel Ridge regression for both the basic measurements (Fig. 4.c) and the random ones (Fig. 4.f). Finally we studied the influence of noise on the training
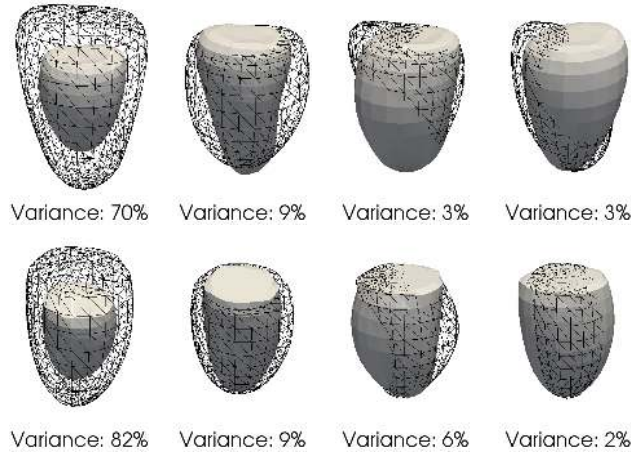
Fig. 2: Top row: Main PCA modes learned on the IUGR dataset. Bottom row: the biggest 4 modes of the PCA of the synthetic dataset. All modes are displayed at ± 3 STD

and testing steps. For the basic set of parameters, we have added to the training set noise with a variance equal to twice the testing set noise and viceversa (Fig. 4.b and Fig. 4.e).

4.3 IUGR data

We used 4 measurements (volume, long axis length, basal diameter and internal dimension) in the IUGR dataset that are equivalent to the basic measurements of the synthetic dataset. We have performed regression using those measurements to compare real and synthetic results. In Fig. 5.a we plot the observed R^2 values of the real data for the first two PCA modes along with the synthetic R^2 curves for the same PCA modes. We can see that the noise factor is around 10%, which coincides with the results obtained by D'hooge et al. in [10]. We have also compared the regression results when different types of measurements are used as predictors. Results show that, with 116 data samples, the best combination of parameters is the basic geometric measurements and the clinical label. Using all the measurements available introduces error, as the algorithm is not able to choose which measurements are relevant for the regression.

4.4 Volumes

We computed the volumes of the predicted meshes and compared them with the 3D echo volumes. We evaluated whether they were closer to the volumes acquired from 3D echo than the ones obtained using the biplane method. Results showed that there was no significant precision difference between both volumes estimations. Fig. 6 shows the Bland-Altman plots of the volumes obtained with
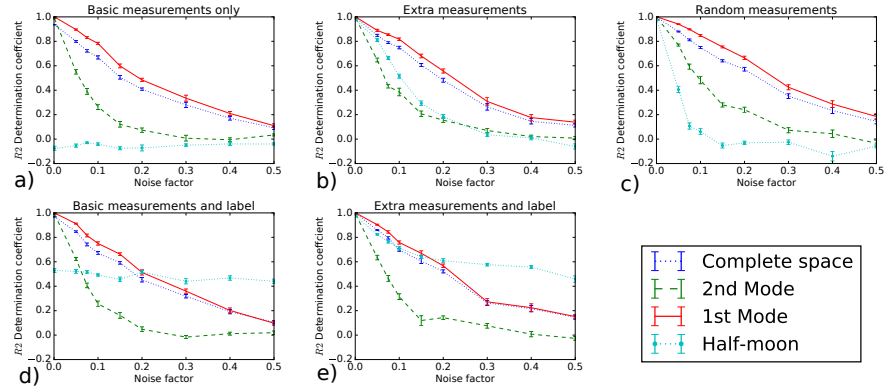
Fig. 3: Influence of the noise and choice of measurements on the regression. For each set of measurements, we plot the R^2 coefficient for specific PCA modes as a function of the noise. All these results were obtained with the synthetic dataset. See Sect. 2.3 for a definition of all types of measurements combined here.

the biplane method and the volumes of the predicted shapes compared with the 3D echo.

5 Discussion

The most common shape representation for the LV is a cropped ellipsoid, as commonly assumed in many measurements done in clinical practice. Our results confirm the validity of this assumption since the PCA analysis of the IUGR dataset (Fig. 2, top row) give modes coinciding with the longitudinal and radial scaling of a mean shape similar to a cropped ellipsoid. Although these modes explain 82% of the total variance, they are not always the most useful parameters for shape assessment to evaluate cardiac pathological remodeling. More complex shape features are encoded by the next modes. As those have much less variability than the main ellipsoidal modes, they are more sensitive to noise and harder to recover. Results in Fig. 3.a show that with the geometrical measurements inspired from EACVI recommendations, only the main ellipsoidal modes can be estimated. A consequence of this might be that the predicted shape does not provide a better volume estimation than the biplane method (Fig. 6). Biplane method is also accurate when the shapes are symmetric, but it is less precise for asymmetric shapes. Findings in the IUGR data were similar (Fig. 3.b).

When using only the basic geometrical measurements, neither increasing the training dataset size (Fig. 4.a), nor removing the training noise (Fig. 4.e), nor changing the algorithm (Fig. 4.c) altered the regression quality. This suggests that most of the regression error is induced by the error on 2D input measurements. Adding new measurements allows to improve shape accuracy. An example of this is the *half moon* pattern, that can not be estimated from basic
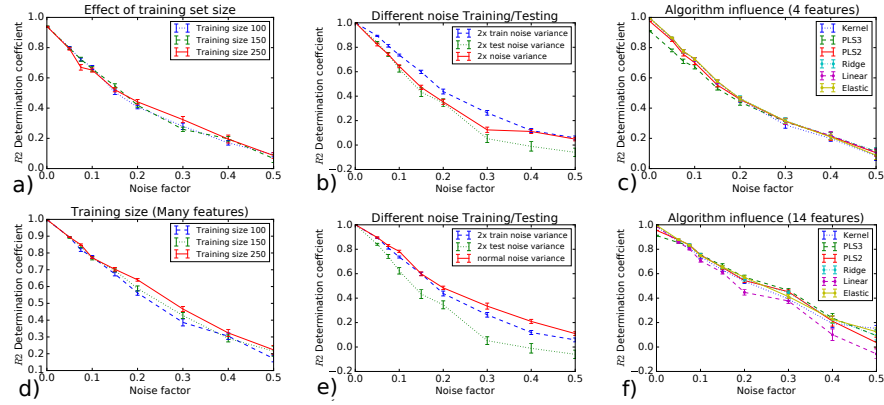
Fig. 4: Effect of: error in training set (a, d), train data size (b, e) and regression algorithm (c, f). The upper row corresponds to the basic measurements and the lower to the random measurements. All these results were obtained with the synthetic dataset.
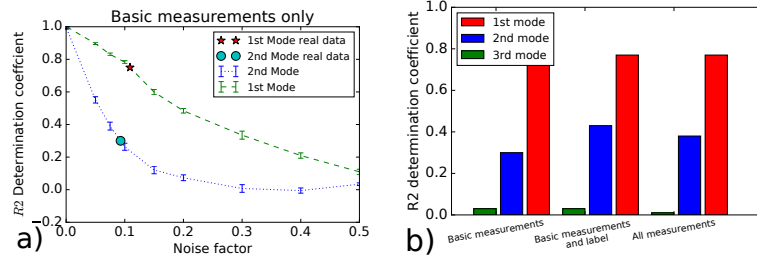

Fig. 5: *a*) R^2 coefficients for the 1st and 2nd PCA mode obtained in the real data compared with the synthetic results for their equivalent PCA modes. *b*) R^2 for the 3 first PCA modes for different sets of measurements.

measurements (dotted cyan graph in Fig. 3.a), but can be retrieved when specific measurements sensitive to this pattern or many random measurements are added. Another positive effect of adding measurements is that we can use redundant information to make the regression more resilient to noise. In Fig. 3 we can see that for the 1st mode, R^2 decreases slower when random measurements are used (c) than when only the basic ones are available (a). We do not have to restrict to geometrical measurements as geometry predictors, we can use measurements related to the clinical history: Fig. 3.d and Fig. 3.e show how adding a label related to the clinical diagnosis improves the regression of the *half moon* feature. A similar behavior was observed for the IUGR dataset in Fig. 4.b.
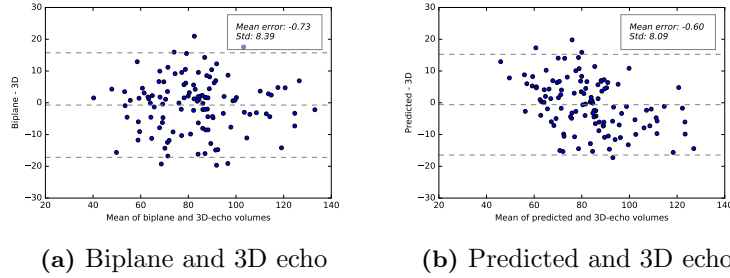
(a) Biplane and 3D echo   **(b)** Predicted and 3D echo

Fig. 6: Bland-Altman plot of the biplane and predicted shape volumes compared with the 3D echo. The dotted lines represent the 95% CI. The volumes of the predicted shapes have a RMSE of 8.11 ml, while the biplanes volume measurements have 8.32 ml.

6 Conclusion

We have shown that it is possible to predict 3D shape from 2D measurements, and how the quality of the prediction is intrinsically limited by the amount of noise in the input measurements. The uncertainty present in the measurements used in daily clinical practice is of such magnitude that it hampers recovering subtle cardiac remodeling features. This can be improved if more measurements specific to the deformation we aim to recover are taken. Alternatively, additional geometric measurements, possibly random, or some clinical history parameters can be used to improve the accuracy of the 3D reconstructed shape. In any case, a sensitivity analysis as done in this paper is important to determine the minimal size of the training population and the optimal choice of features for a given population.

References

1. Konstam, M.A., Kramer, D.G., Patel, A.R., Maron, M.S., Udelson, J.E.: Left Ventricular Remodeling in Heart Failure: Current Concepts in Clinical Significance and Assessment. JACC: Cardiovascular Imaging **4**(1) (2011) 98–108
2. Di Donato, M., Castelvecchio, S., Kukulski, T., Bussadori, C., Giacomazzi, F., Frigiola, A., Menicanti, L.: Surgical Ventricular Restoration: Left Ventricular Shape Influence on Cardiac Function, Clinical Status, and Survival. Annals of Thoracic Surgery **87** (2009) 455–461

3. Ye, D.H., Desjardins, B., Hamm, J., Litt, H., Pohl, K.M.: Regional manifold learning for disease classification. IEEE Transactions on Medical Imaging **33**(6) (2014) 1236–1247

4. Crispi, F., Bijnens, B., Figueras, F., Bartrons, J., Eixarch, E., Le Noble, F., Ahmed, A., Gratacós, E.: Fetal growth restriction results in remodeled and less efficient hearts in children. Circulation **121**(22) (2010) 2427–2436

5. Sarvari, S.I., Rodriguez-Lopez, M., Nuñez-Garcia, M., Sitges, M., Sepulveda-Martinez, A., Camara, O., Butakoff, C., Gratacos, E., Bijnens, B., Crispi, F.: Persistence of Cardiac Remodeling in Preadolescents With Fetal Growth Restriction. Circulation: Cardiovascular Imaging **10**(1) (2017)

6. Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P.: The virtual skeleton database: an open access repository for biomedical research and collaboration. Journal of medical Internet research **15**(11) (1 2013) e245

7. Blanc, R., Seiler, C., Székely, G., Nolte, L.P., Reyes, M.: Statistical model based shape prediction from a combination of direct observations and various surrogates: Application to orthopaedic research. Medical Image Analysis **16**(6) (2012) 1156–1166

8. Lancellotti, P., Cosyns, B.: The EACVI Echo Handbook. The European Society of Cardiology Textbooks Series. Oxford University Press (2015)

9. Murphy, K.: Machine learning: a probabilistic perspective. (2012)

10. D'hooge, J., Barbosa, D., Gao, H., Claus, P., Prater, D., Hamilton, J., Lysyansky, P., Abe, Y., Ito, Y., Houle, H., Pedri, S., Baumann, R., Thomas, J., Badano, L.P.: Two-dimensional speckle tracking echocardiography: Standardization efforts based on synthetic ultrasound data. European Heart Journal Cardiovascular Imaging **17** (2016) 693–701