HAL
open science

Hierarchical Multi-modality Prediction Model to Assess Obesity-Related Remodelling

Gabriel Bernardino, Patrick Clarysse, Álvaro Sepúlveda-Martínez, Mérida Rodríguez-López, Susanna Prat-Gonzàlez, Marta Sitges, Eduard Gratacós, Fàtima Crispi, Nicolas Duchateau

To cite this version:

Gabriel Bernardino, Patrick Clarysse, Álvaro Sepúlveda-Martínez, Mérida Rodríguez-López, Susanna Prat-Gonzàlez, et al.. Hierarchical Multi-modality Prediction Model to Assess Obesity-Related Remodelling. Statistical Atlases and Computational Modelling of the Heart, 2021, Strasbourg, France. pp.103-112, 10.1007/978-3-030-93722-5_12 . hal-03536022

HAL Id: hal-03536022

https://hal.science/hal-03536022v1

Submitted on 19 Jan 2022

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

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

Hierarchical multi-modality prediction model to assess obesity-related remodelling

Gabriel Bernardino[1], Patrick Clarysse[1], Álvaro Sepúlveda-Martínez[2,3], Mérida Rodríguez-López[2,4], Susanna Prat-Gonzàlez[5], Marta Sitges[6,7], Eduard Gratacós[3,6], Fàtima Crispi[3,6], Nicolas Duchateau[1]

[1] Univ Lyon, Université Claude Bernard Lyon 1, INSA-Lyon, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Lyon, France
[2] Fetal Medicine Research Center, BCNatal - Barcelona Center for Maternal-Fetal and Neonatal Medicine (Hospital Clínic and Hospital Sant Joan de Déu), Institut Clínic de Ginecologia Obstetricia i Neonatologia, Universitat de Barcelona, Centre for Biomedical Research on Rare Diseases (CIBER-ER), Barcelona, Spain
[3] Fetal Medicine Unit, Department of Obstetrics and Gynecology, Hospital Clínico de la Universidad de Chile, Santiago de Chile, Chile
[4] Pontificia Universidad Javeriana seccional Cali, Cali, Colombia
[5] Centre de Diagnòstic per la Imatge, Hospital Clínic, Universitat de Barcelona, Barcelona, Spain
[6] Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain.
[7] Institut Clínic Cardiovascular, Hospital Clínic, Universitat de Barcelona, Centre for Biomedical Research on CardioVascular Diseases (CIBERCV), Barcelona, Spain

Abstract. The diagnosis of cardiovascular illnesses uses multiple modalities in order to obtain a complete and as robust as possible assessment of the heart. However, when addressing distinct pathologies, not all information might be needed in order to achieve a confident-enough diagnosis. We propose a probabilistic machine learning method to identify the patients for which the acquisition of more complex data would be useful. We hypothesise that there exists a hierarchical relationship between modalities: echocardiography is more accessible and has a lower economical cost than other modalities (like magnetic resonance imaging (MRI)). The framework consists of two classifier models, each predicting the illness from the echocardiographic and MRI views, and a sample-weighting model that combines both predictions. This weighting model is used to decide which individuals will not need an MRI acquisition additional to the echocardiographic examination.

We illustrated this on a dataset of asymptomatic individuals with an echocardiographic study (N = 480), a subset of those also includes a MRI (N = 159). We analyse the effect of being overweight on cardiac geometry. We identified that the type of remodelling depended on blood pressure: overweight combined with high blood pressure resulted in an increase of ventricular mass, while only size changes were preserved for low-pressure individuals. With our method, we established that boundary cases of the former group could be correctly classified after incorporating MRI, while it was not the case for the latter.

Keywords: Data fusion · Probabilistic model · Machine learning · Cardiac imaging · Cardiac remodelling

1 Introduction

Medical imaging is a central component in the diagnosis of cardiovascular illnesses. Several modalities can be used to assess cardiac function and geometry, each offering a different view on the properties of the heart, and each being more or less appropriate for detecting different pathologies. In order to have the most complete assessment of the heart, it could be desirable to use as many modalities as possible, however multiple acquisitions result in higher economic cost, time, discomfort/danger to the patient. Moreover, the heart is a dynamic organ, and changes from beat to beat, difficulting the integration of data acquired in different beats.

Echocardiography is the most common modality used in cardiology, due to its low cost and accessibility, being able to be operated at bedside or in the clinician's office. Other modalities, such as magnetic resonance imaging (MRI) are also available, but are often used after echocardiographic findings are not conclusive enough to produce a diagnosis. The study of what kind of information can be determined from each modality, and when it is more appropriate to use one or the other, has not been quantitatively discussed beyond some consensus established by clinical experts [5].

In this work, we explore the feasibility of identifying the individuals who would benefit from more complex (and therefore costly) acquisitions. Our rationale is that there are individuals for whom echocardiography is not able to provide an accurate diagnosis, yet acquiring MRI will not improve diagnosis. We aim to identify those patients to avoid unnecessary MRI. This problem shares similarities with active learning [15, 6], which is a subfield of machine learning that allows to identify in a semi-annotated dataset the unannotated samples that would be more helpful for training a model, and to Bayesian optimisation [4], a black-box optimisation method for functions that are expensive to evaluate.

Our approach is similar to data fusion, which is still a challenging research topic in machine learning [9]. Grossly, there are two strategies: obtaining a fused representation of the different views and then perform predictions in that space, or perform predictions independently and then combine them. Some methods, specially the ones based on a probabilistic framework such as [13] allow missing data (it is mostly assumed that the data is missing at random) and assume no hierarchical relationship between views. This is the main difference with our method, since our assumption is that we can identify the individuals for which the more complex acquisition is unnecessary, and therefore "missing" data will not be at random.

We illustrate the framework on a dataset including echocardiography and MRI from healthy asymptomaptic individuals, for which we study the cardiac changes related to body mass index (BMI). Overweight and obesity are well known cardiovascular risk factors [14]. The most reported risk is an increased

probability of myocardial infarct due to the presence of cholesterol in the vessels, but also obesity is related to pressure (due to an increase of systolic blood pressure (SBP)) and volume loading of the left ventricle (LV) (since it needs to produce higher cardiac output (CO) to sustain the increased body mass). Its effect is most noticeable on the left ventricle: myocardial hypertrophy and also a dilation of the heart [7]. The changes are more easily seen in the cardiac geometry, but also affect the function, specially in the diastolic function [12].

2 Methodology

2.1 Echocardiographic and MRI dataset

More details on the dataset are found in [2]. Its original purpose was to find the effect on the cardiac system of adults (30-40 years old) who were born low-weight. Its effect was found to be subtle, so the individuals with low weight at birth were kept in the current work in order to have a higher population size. We reused this dataset to study the effect of elevated body mass index, using the WHO classification for overweight (BMI > 25) [1]. We used overweight (BMI > 25) over strictly obese (BMI > 30) because our MRI population contained only 20 obese individuals. Overweight definition also includes individuals who are in the transition zone between obesity and appropriate weight, and they are therefore expected to have minor remodelling. A short description of the demographics of the adequate weight and overweight population can be found in Table 1. We did not consider the interaction between weight at birth and overweight, and its effect is treated as noise in this analysis.

	BMI < 25		BMI ≥ 25	
	Echo	MRI	Echo	MRI
N	320	83	178	64
Weight [kg]	60.6 ± 10.	61.5 ± 9.8	84.4 ± 13.0	85.2 ± 11.1
Height [cm]	167 ± 9	168 ± 9	169 ± 10	171 ± 9
BSA [m^2]	1.67 ± 0.18	1.69 ± 0.17	1.99 ± 0.19	2.01 ± 0.16
Male gender [%]	39	36	63	67
Age [y]	30.0 ± 4.9	33.2 ± 3.7	31.7 ± 4.9	33.6 ± 4.1
Heart rate [bpm]	68.4 ± 11.2	66.9 ± 11.3	69.7 ± 12.3	71.2 ± 12.9
Systolic BP [mmHg]	114.6 ± 11.8	115.1 ± 13.4	121.9 ± 12.2	122.8 ± 13.3
Ejection Fraction [%]	65.4 ± 6.8	66.7 ± 7.9	64.5 ± 6.5	64.3 ± 6.7

Table 1: Demographic description of the population. Variables are described by their mean and STD.

In short, the data consist of echocardiographic measurements from 480 individuals, and MRI biventricular shapes from a subset 159 individuals. The

echocardiographic measurements were: LV long axis, LV basal dimension, RV long axis and LV Mmss (from M-mode), all measured using the recommendations of the European Society of Cardiology [8]. The shapes are the result of fitting a deformable model to each individual, producing meshes at end-diastole (ED) in point-to-point correspondence. Both the measurements and the meshes were indexed by body size area (BSA) to remove cardiac size variability due to the body dimensions, as is commonly used in the clinics. The data was preprocessed using unsupervised dimensionality reduction in order to reduce the noise: principal component analysis (PCA) for the MRI shapes (keeping 5 dimensions), and Isomap for the echocardiographic measurements (keeping 2 dimensions).

2.2 Hierarchical framework

We propose a probabilistic model to express the probability of the label (Y) given the simple ($\mathbf{X}_0$) and complex ($\mathbf{X}_1$) data features. The estimated probability is a weighted combination of two single-view classification models, with a sample weight. We note $\mathbf{w}$ as the vector containing the weights for all individuals, and w^i its value for the i-th individual. This weight is close to 1 when the information in X_0^i is enough to make the prediction and close to 0 when considering the next level of the hierarchy X_1^i is compulsory. The novelty in our approach lies in these weights, for which we add a prior on $\mathbf{w}$ forcing it to be a smooth function of $\mathbf{X_0} = [X_0^i]$. These weights serve as an estimator of the need of acquiring $\mathbf{X_1} = [X_1^i]$.

$$P(Y|\mathbf{X_0}, \mathbf{X_1}) = (\prod_i (w^i P(Y^i|X_0^i, \theta_0) + (1 - w^i) P(Y^i|X_1^i, \theta_1)))P(\mathbf{w}|\mathbf{X_0}), \quad (1)$$

where θ_0 and θ_1 are the parameters of the single view prediction models. Any probabilistic classification model for the single view models can be chosen, as long as sample-weights can be incorporated in such model. In our case, we arbitrarily chose a Gaussian process (GP) Classifier, with a Bernoulli likelihood [11] using a publicly available implementation [10]; but the framework allows the use of any classification method that estimates probabilities.

To ensure smoothness of $\mathbf{w}$ from $\mathbf{X_0}$, we use a GP prior:

$$P(\mathbf{w}|\mathbf{X_0}) \sim MVN(\mathbf{0}, \mathbf{K_{X_0}}), \quad (2)$$

where MVN is the density function of the multivariate normal, and $\mathbf{K_{X_0}}$ is the classical kernel matrix with a Gaussian kernel.

2.3 Optimisation

After including the prior on w^i and the parameters of the classification models θ_0 and θ_1, the posterior log-probability model for the full dataset is:

$$\log P = \sum_i \log \left(w^i P(Y^i|X_0^i, \theta_0) + (1 - w^i) P(Y^i|X_1^i, \theta_1) \right) + \log P(\mathbf{w}|\mathbf{X_0}). \quad (3)$$

We optimise a lower bound L of the expression obtained by applying Jensen's inequality on the logarithm:

$$\log P \geq \sum_i w^i \log P(Y^i|X_0^i, \theta_0) + (1 - w^i) \log P(Y^i|X_1^i, \theta_1) + \log P(\mathbf{w}|\mathbf{X_0}) = L. \tag{4}$$

As direct joint optimisation of all the parameters is troublesome and leads to instability, we optimise iteratively the classification model and the combination weights until convergence, until the weights $\mathbf{w}$ of two consecutive iterations are not significantly different. The optimisation of the θ paremeters of the classification model is:

$$\begin{aligned} \max_{\theta_0,\theta_1} \ \log(L) = & \max_{\theta_0} \sum_i w^i \log P(Y^i|X_0^i, \theta_0) \\ & + \max_{\theta_1} \sum_i (1 - w^i) \log P(Y^i|X_1^i, \theta_1)) + \sum_i \log(P(w^i)), \end{aligned} \tag{5}$$

which corresponds to two standard sample-weighted classification problems that can be solved using classical classifiers.

The weights $\mathbf{w}$ will be estimated by maximising the joint probability, together with the prior. Putting everything together, we can see that L has terms that do not include the w^i, and therefore do not affect the optimisation. We can pull them out of the optimisation as a term C:

$$\max_{\mathbf{w}} \ \log(L) = \max_{\mathbf{w}} \sum_i w^i (\log P(Y^i|X_0^i, \theta_0) - \log(P(Y^i|X_1^i, \theta_1)) - \mathbf{w}^t \mathbf{K}_{\mathbf{X_0}}^{-1} \mathbf{w}/2 + C, \tag{6}$$

where $C = -d/2 \log(2\pi) - d/2 \log(\det \mathbf{K}_{\mathbf{X_0}}) + \log(P(Y^i|X_1^i, \theta_1))$ is independent of $\mathbf{w}$ and does not need to be considered for optimisation. The previous equation corresponds to a quadratic optimisation problem for which the exact solution can be computed.

2.4 Prediction

The previous formulation needs both $\mathbf{X_0}$ and $\mathbf{X_1}$ for all samples to compute the label probability. However, we would like to obtain some of the predictions without having to use $\mathbf{X_1}$, thus avoiding a costly acquisition. For that, we will fix a threshold k for the weights w^i (which can be computed from $\mathbf{X_0}$ only), and only use $\mathbf{X_1}$ when the predicted value of w^i is below that threshold. A higher value of the threshold will result in lower usage of the second level data. The estimated probability is as follows:

$$P(Y^i|X_0^i, X_1^i) = \begin{cases} P(Y^i|X_0^i) & w^i \geq k, \\ w^i P(Y^i|X_0^i) + (1 - w^i) P(Y^i|X_1^i) & w^i < k. \end{cases} \tag{7}$$

Note that since we have imposed a GP prior over the w^i, these weights can be estimated from X_0^i only.

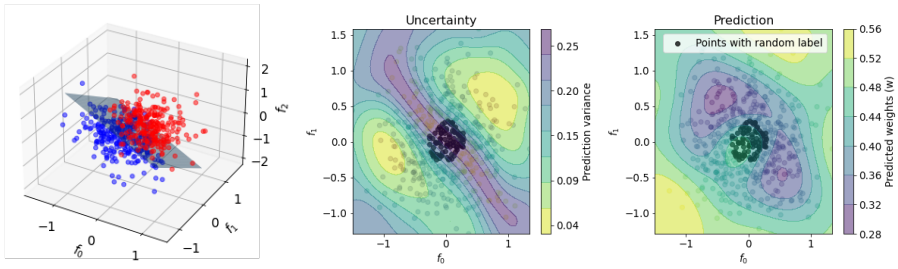


Fig. 1: Scatter plots with the synthetic dataset. We can see that uncertainty (left) is concentrated in the decision boundary $x = -y$, while the weight (right) is also contained near the decision boundary, but excludes the points for which the label is random (black).

3 Results

3.1 Synthetic experiment

We first verify our framework using a synthetic dataset. This dataset contains two levels of features ($\mathbf{X_0}$ and $\mathbf{X_1}$) and a binary label Y. The label is determined by $\mathbf{X_1}$ and some noise, but to create the situation that not all information can be well predicted from $\mathbf{X_1}$, we set the label of the points of a particular region in the decision boundary of $\mathbf{X_1}$ completely at random. The expected result is that our algorithm gives high weights w^i near the decision boundary, except in the area that the labels are completely random. The data generation is as follows, where $\mathcal{N}$ and Ber denote the normal and Bernoulli distributions respectively:

$$
\begin{aligned}
(f_0, f_1, f_2) &\sim \mathcal{N}(0,1) \times \mathcal{N}(0,1) \times \mathcal{N}(0,1), \\
l &\sim \begin{cases} Ber(0.5) & f_0^2 + f_1^2 < 0.3^2, \\ Ber(sigmoid(f_0 + f_1 + f_2)) & \text{otherwise.} \end{cases}
\end{aligned} \tag{8}
$$

The low and high complexity views are defined as the first two coordinates ($\mathbf{X_0} = (f_0, f_1)$) and the three coordinates ($\mathbf{X_1} = (f_0, f_1, f_2)$) respectively. Figure 1 shows the result of applying the hierarchical model to these synthetic data, comparing the predicted weight $\mathbf{w}$ to the model uncertainty, a classical method in active learning/Bayesian optimisation to decide which samples to annotate/acquire. We can observe that the uncertainty is concentrated in the decision boundary, but is not affected by the central area with no information, while the weight submodel predicting the probability of improvement correctly detects that labels in the central area are random and therefore no further acquisition would be beneficial.

3.2 Real dataset

Using the real dataset, we tested the basic single-view classification models for the echocardiographic measurements and shape embedding only, an aver-

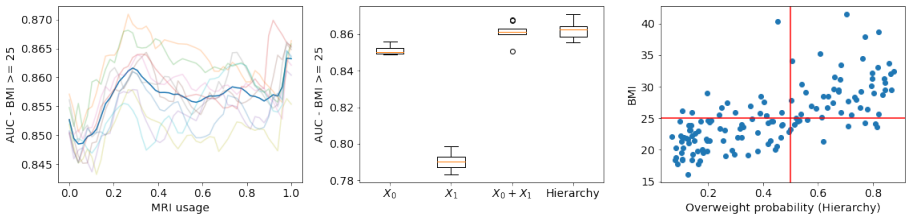
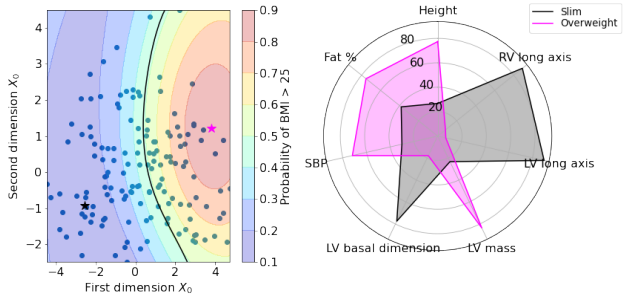


Fig. 2: The *left plot* depicts the AUC for different amounts of MRI data used. Transparent lines are the different bootstrap cross-validations, and the blue line is the mean trace. The maximal AUC is attained at 30%. The *central figure* is a boxplot (orange lines depict the median, the boxes stand for the interquantile range) of the AUC for different models. We can see that there is a small improvement for considering both $\mathbf{X_0}$ and $\mathbf{X_1}$ simultaneously compared to using only $\mathbf{X_0}$ or $\mathbf{X_1}$. This improvement is similar, either when the full population is used, or just a fraction of it. The *right plot* shows the correlation between the probability scores and the raw BMI value.

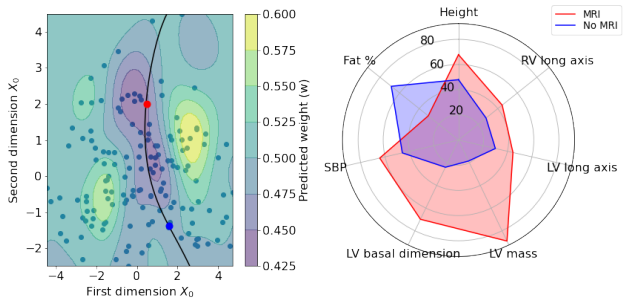
age combination of the previous single-view models, and the combination of the two spaces using our weighting approach. 20-fold cross-validation splits were repeated over 10 different seeds to obtain more statistically-stable results, and we compared the area under curve (AUC) of predicting elevated BMI. We also tested the version of our approach which did not use all the X_1^i, with different values for the threshold k in equation 7. For each value, we report the AUC and the fraction of weights w^i that fell below the threshold, thus the number of times that $\mathbf{X_1}$ data were used. This AUC as a function of $\mathbf{X_1}$ can be found in Figure 2, along with the AUC of the different models.

An interesting result is that MRI shapes ($\mathbf{X_1}$) had lower prediction power than simple echocardiographic measurements ($\mathbf{X_0}$). This can be explained as obesity has a global effect on the heart, then just a few measurements are enough to characterise the changes in size and LV hypertrophy related to obesity, so there is little advantage in using shape models instead of simple measurements. However, the combination of both models, either by simple addition or by sample-weights, showed an improvement of prediction power, indicating benefit on the combination of both modalities. For achieving this improvement, it is not necessary to acquire the MRI for the full population, but it could be achieved with approximately 30% of the cases. In the right side of Figure 2 we can see the correlation between the predicted probability of the hierarchical model, and the BMI. We can see that they share a fair correlation ($\rho = 0.70$), showing that our classification model correctly assigns higher probability to the more obese individuals.

In Figure 3a, we show the probability contours of the BMI model. The BMI models shows a nice separation between the two classes, with overweights being mostly on the right, and slim individuals on the left. The black line represents the decision boundary of the BMI model. We manually selected synthetic representative points of both thin and overweight individuals, that are depicted in



(a) Overweight



(b) Predicted weights

Fig. 3: The subfigures show a scatter plot of the $\mathbf{X}_0$ embedding with contours of the probability of being overweight (top) and of the $\mathbf{w}$ weights indicating the usefulness of the MRI (bottom). The subfigures also contain a spider plot with the characteristics of extreme individuals, selected in the scatter plot as the colored stars.

the same figure as stars. In the right part of the figure, a spider plot compares the two synthetic representative individuals. The characteristics of the synthetic individuals were interpolated using a GP. We examined the characteristics that our model associated with a high BMI: a higher ventricular mass and smaller ventricle (indexed by BSA), specially in the long axis. While this might seem surprising, the smaller ventricle is mostly due not to a shrinking of the heart but to an increase of the BSA.

We also studied what were the characteristics of the individuals for whom the MRI would add more information. In Figure 3b, we can see that far from the decision boundary the weight is high, indicating that $\mathbf{X}_0$ is enough for predicting the output. Near the decision boundary, we can observe a differential behaviour: in the upper part the weight is lower than in the lower part. We manually selected two points representative of a low and high w^i, and we studied their characteristics. We found that high weight (i.e. those for which the MRI is most useful) presented a higher LV mass, specially when compared to the ventricular

size, and SBP, suggesting hypertrophic remodelling due to pressure loading. Low weight was associated with a higher fat percentage, and a smaller ventricle in both dimensions and mass after indexing by BSA, suggesting a remodelling that is preserving the mass-volume relationship: a change that is mostly size-related. Since the former remodelling involves a change in shape, while the later is mostly in size (which can be already be well approximated by echocardiography), it makes sense that adding MRI shape information helps mostly individuals to belong to the first group [3].

4 Conclusion

We have presented a framework to combine predictions from different models, potentially each trained with different data. The main novelty of this model is its ability to identify the cases that are more likely to improve when adding more data, therefore reducing the number of high-cost data used.

We applied the model to an obesity dataset, and we were able to identify different remodelling patterns in the decision boundary, corresponding to individuals whose diagnosis accuracy would improve after acquiring the MRI data. Our framework reached an accuracy similar to the combination of both models, while only using 30% of the data.

Acknowledgements

This project has been partially funded with support the French ANR (LABEX PRIMES of Univ. Lyon [ANR-11-LABX-0063] and the JCJC project "MIC-MAC" [ANR-19-CE45-0005]), the Erasmus + Programme of the European Union (Framework Agreement number: 2013-0040), "la Caixa" Foundation (LCF/PR/GN14/10270005, LCF/PR/GN18/10310003), the Instituto de Salud Carlos III (PI14/00226, PI17/00675) integrated in "Plan Nacional de I+D+I" and cofinanciated by ISCIII-Subdirección General de Evaluación and Fondo Europeo de Desarrollo Regional (FEDER) "Una manera de hacer Europa", and AGAUR 2017 SGR grant #1531.

References

1. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. Tech. rep. (2000)
2. Crispi, F., Rodríguez-López, M., Bernardino, G., Sepúlveda-Martínez, A.l., Prat-González, S., Pajuelo, C., Perea, R.J., Caralt, M.T., Casu, G., Vellvé, K., Crovetto, F., Burgos, F., De Craene, M., Butakoff, C., González Ballester, M.A., Blanco, I., Sitges, M., Bijnens, B., Gratacós, E.: Exercise Capacity in Young Adults Born Small for Gestational Age. JAMA Cardiol. (2021)
3. Devlin, A., Moore, N., Ostman-Smith, I.: A comparison of MRI and echocardiography in hypertrophic cardiomyopathy. Br J Radiol **72**, 258–264 (1999)
4. Frazier, P.I.: A Tutorial on Bayesian Optimization (2018)

5. Garbi, M., Edvardsen, T., Bax, J., Petersen, S.E., McDonagh, T., Filippatos, G., Lancellotti, P.: EACVI appropriateness criteria for the use of cardiovascular imaging in heart failure derived from European National Imaging Societies voting. European Heart Journal - Cardiovascular Imaging **17**, 711–721 (2016)
6. Gorriz, M., Carlier, A., Faure, E., Giró i Nieto, X.: Cost-effective active learning for melanoma segmentation. CoRR **abs/1711.09168** (2017)
7. Grossman, W., Jones, D., McLaurin, L.: Wall stress and patterns of hypertrophy in the human left ventricle. J Clin Investig **56**, 56–64 (1975)
8. Lang, R., Badano, L., Mor-Avi, V., et al.: Recommendations for Cardiac Chamber Quantification by echocardiography in adults. Eur Heart J Cardiovasc Imaging **28**, 1–39.e14 (2015)
9. Li, Y., Yang, M., Zhang, Z.: A survey of multi-view representation learning. IEEE Trans Knowl Data Eng **31**, 1863–83 (2019)
10. Matthews, A., van der Wilk, M., Nickson, T., et al.: GPflow: A Gaussian process library using TensorFlow. J Mach Learn Res **18**, 1–6 (2017)
11. Nickisch, H., Rasmussen, C.: Approximations for binary Gaussian process classification. J Mach Learn Res **9**, 2035–78 (2008)
12. Russo, C., Jin, Z., Homma, S., et al.: Effect of obesity and overweight on left ventricular diastolic function. J Am Coll Cardiol **57**, 1368–74 (2011)
13. Sabuncu, M., Yeo, B., Van Leemput, K., et al.: A generative model for image segmentation based on label fusion. IEEE Trans Med Imaging **29**, 1714–29 (2010)
14. Wilson, P., D'Agostino, R., Sullivan, L., et al.: Overweight and obesity as determinants of cardiovascular risk. Arch Intern Med **162**, 1867 (2002)
15. Yang, Y., Loog, M.: Active learning using uncertainty information. In: 2016 23rd International Conference on Pattern Recognition (ICPR). pp. 2646–2651. IEEE (2016)