AUTO-ENCODING OF DISCRIMINATING MORPHOMETRY FROM CARDIAC MRI

Dong Hye Ye[1], Benoit Desjardins[2], Victor Ferrari[2], Dimitris Metaxas[3], Kilian M. Pohl[4]

Department of Electrical and Computer Engineering, Purdue University[1]
Department of Radiology, University of Pennsylvania[2]
Department of Computer Science, Rutgers University[3]
SRI International & Department of Psychiatry and Behavioral Sciences, Stanford University[4]

ABSTRACT

We propose a fully-automatic morphometric encoding targeted towards differentiating diseased from healthy cardiac MRI. Existing encodings rely on accurate segmentations of each scan. Segmentation generally includes labour-intensive editing and increases the risk associated with intra- and inter-rater variability. Our morphometric framework only requires the segmentation of a template scan. This template is non-rigidly registered to the other scans. We then confine the resulting deformation maps to the regions outlined by the segmentations. We learn a manifold for each region and identify the most informative coordinates with respect to distinguishing diseased from healthy scans. Compared with volumetric measurements and a deformation-based score, this encoding is much more accurate in capturing morphometric patterns distinguishing healthy subjects from those with Tetralogy of Fallot, diastolic dysfunction, and hypertrophic cardiomyopathy.

Index Terms— Morphometry, Manifold learning, Cardiac MR, Disease Classification

1. INTRODUCTION

To reduce the number of cardiac-related deaths, researchers are searching for new phenotypes from high-resolution cardiac MRI. Most morphometric studies rely on volumetric scores, such as [1], due to the difficulties associated with extracting more refined measurements. These scores are often insensitive to the heterogeneous and subtle manifestation of cardiac diseases. For example, the left ventricular ejection fraction is normal in 50% of heart failure cases. In this paper, we describe an automatic encoding for aiding the identification of disease-specific morphometric patterns from MRI.

Deformation maps frequently serve as descriptors of cardiac anatomy and function. For example, [2, 3] register all cases to a template and then regress the corresponding deformation maps with respect to clinical diagnosis. Alternatively, [4] align the series of images of a tagged MRI to identify disease specific motion patterns. All of these methods rely on accurate segmentations of the images to be analysed. Label maps enable explicit tracking of the cardiac boundaries and simplify the registration task as non cardiac structures are ignored. To ensure accuracy, medical experts manually refine these segmentations. This task increases the risk of intra- and inter-subject variability. It is also time-consuming, which is a challenge for smaller studies and prohibitive for larger ones. To the best of our knowledge, our proposed morphometric encoding is the first approach that only requires a segmentation of a single cardiac MR scan, i.e. the template.

Similar to existing morphometric frameworks of brain image analysis [5, 6], we encode the anatomy of an individual scan by non-rigidly registering the template to that scan. Until recently, registering cardiac MRIs was considered difficult as scans can greatly differ with respect to the anatomy captured in their field of view and the shape of the anatomy. Our algorithm in [7] accurately registers those scans through online learning of the smoothing parameters and stitching together maps from different image regions. Assuming the registration between the template and a target scan is accurate, applying the corresponding deformation map to the segmentation of the template results in an accurate label map of the target scan. This observation motivates our morphometric encoding. We encode anatomy in the target scan by the deformation associated with that region, which is defined by the segmentation of the template scan. Thus, our encoding only requires a single segmentation unlike existing ones, such as [5, 6].

Learning disease-characteristic patterns from the resulting high-dimensional morphometric encoding is difficult due to the relatively small sample size of cardiac MR studies. We therefore derive a low-dimensional descriptor by applying manifold learning [8] to the encodings. We then further prune the resulting coordinates via minimum Redundancy Maximum Relevance (mRMR) [9], which only keeps the most informative ones with respect to distinguishing the populations under study. The outcome is a morphometric encoding that applicable to a variety of data sets as its important parameters are determined automatically.

We measure the accuracy of our encoding by training a classifier on distinguishing normal from diseased scans on 3 data sets. Each data set contains a different diseased popu-

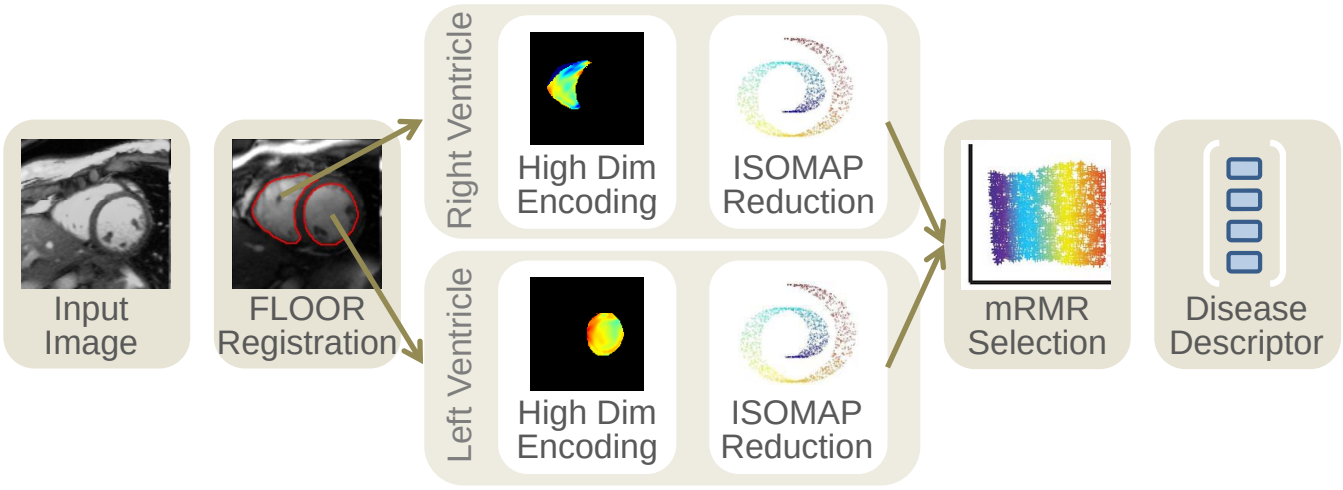

Fig. 1. Steps for automatically creating disease sensitive encoding of images: (1) register template to image via FLOOR [7], (2) embed the resulting deformation map into the low-dimensional manifold embedding via ISOMAP [8], (3) select the embeddings distinguishing health from diseased scans via mRMR [9].

lation, which are Tetrology of Fallot, diastolic dysfunction, and hypertrophic cardiomyopathy. Performing leave-one-out cross validation, the accuracy of our encoding is higher than that of volumetric scores and an existing morphometric encoding, called RAVENS [6], which encodes shape by the deformation map from the scan to the template.

2. ENCODING MORPHOMETRY

We now describe our morphometric encoding, for which we first select a template before following the steps outlined in Figure 1, i.e. registering the template to the image, creating the initial high-dimensional morphometric descriptor, transforming the descriptor to a low-dimensional encoding via manifold learning, and only selecting those encoding coordinates that discriminate between patient populations.

(1) Template Selection: The template can be empirically set beforehand or determined automatically based on the data-driven approach described below. Selecting the template is a critical component in our framework, as a bad choice could negatively impact the accuracy of the encoding. Outliers, for example, are generally difficult to match to other scans and thus are not well suited as templates. We minimize this risk via GRAM [10], which first affinely registers all images to a common coordinate system and then selects the image closest to the *geodesic mean* across the scans of healthy subjects. This image is most similar and thus easiest to register to all the other scans of the healthy population. We complete the template by manually outlining the endocardium of each ventricle in the scan (see Fig. 1) resulting in a binary map B_{RV} of the Right Ventricle cavity (RV) and a binary map B_{LV} of the Left Ventricle cavity (LV). Note, that clinicians frequently measure the volume of the ventricles based on their cavities.

(2) Template Registration: We non-rigidly register the template to all the other MRIs of the study via FLOOR [7]. We choose FLOOR as it learns the important parameters online and can register cardiac MRI that differ in their field of view and the shape of the ventricles. It determines the mapping at each image location by choosing the optimal solution from a set of *candidates*. Each candidate minimizes the energy function of the approach according to a specific smoothing parameter and image region. The result is a smooth deformation map, D, whose stiffness is regulated throughout the image domain by the shape difference of the underlying anatomy between the scans. In addition, missing correspondences are contained locally instead of corrupting the mapping throughout the entire image domain. It thus produces accurate results even if non-cardiac structures are not removed.

(3) High-Dimensional Encoding: Inspired by [6], we encode the anatomy of the input image I by analyzing the corresponding deformation map, D_I, with respect to the binary maps B_{RV} and B_{LV} of the template. Specifically, we compute the Jacobian Determinant $J(D_I)$ of D_I. We then multiply $J(D_I)$ with the warped binary maps $B_{RV}(D_I)$ and $B_{LV}(D_I)$ to determine the structure specific encoding:

$$M_{I_{RV}} \triangleq J(D_I) \cdot B_{RV}(D_I) \text{ and } M_{I_{LV}} \triangleq J(D_I) \cdot B_{LV}(D_I).$$

In difference to RAVENS [6], an example of a deformation-based encoding, we compute the encoding in the image space instead of the template space. This allows us to infer local expansion or contraction of the shape with respect to the template using only a single segmentation (of the template).

(4) Derive Low-Dimensional Encoding for each Ventricle: For each ventricle, we determine a low dimensional encoding by learning the corresponding morphometric manifold

Disease	Volume			RAVENS			Our Encoding		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
TOF	88.9	66.7	75.0	77.8	80.0	79.2	100.0	80.0	87.5
HCM	88.9	66.7	75.0	88.9	80.0	83.3	100.0	86.67	91.7
DD	88.9	46.3	62.5	77.8	73.3	75.0	88.9	73.3	79.2

Table 1. Outcome of the linear classifier with respect different encodings. For all three populations our sensitivity, specificity, and accuracy scores are equal or higher than the ones based on Volume and RAVENS.

from the encodings $M_{I,*V}$ across all images. We choose ISOMAP [8] for this task it preserves morphometric characteristics, i.e., image regions with similar morphometry are close to each other on the manifold. We now explain in further detail ISOMAP with respect to the RV, noting that the same approach applies to the LV.

ISOMAP first arranges the RV region of the images as a graph, where each node represents the RV of a specific image and the length of an edge is defined by the distance $d(\cdot,\cdot)$ between the corresponding image regions. We define the pairwise distance $d(\cdot,\cdot)$ between two image regions $I_{RV}^{(1)}$ and $I_{RV}^{(2)}$ by the L_2 norm of the corresponding morphometric encodings $M_{I^{(1)},RV}$ and $M_{I^{(2)},RV}$:

$$d(I_{RV}^{(1)}, I_{RV}^{(2)}) \triangleq \sum\nolimits_x (M_{I^{(1)},RV}(x) - M_{I^{(2)},RV}(x))^2$$

with the sum over the voxels x inside the RV of the template. We note that this distance captures morphometric differences between image regions, as for images with similar RVs the corresponding encodings $M_{I^{(*)},RV}$ are similar, so that $d(\cdot,\cdot)$ is small, whereas $d(\cdot,\cdot)$ is large for image regions with large morphometric differences.

Having defined $d(\cdot,\cdot)$, ISOMAP constructs a graph via the k nearest neighbor (kNN) algorithm. We set k, the size of the neighbourhood, to the smallest value so that the graph is still connected. ISOMAP then determines the low-dimensional embedding of this graph by computing the eigensystem of the matrix defined by the geodesic distances inferred from that graph. In doing so, it preserves the neighbourhood relationship of the graph, i.e., image regions with similar morphometry are in close proximity in the low dimensional space. According to [10], we automatically set the dimension of this space so that at least 90% of the original information is conserved, which we approximate by the *normalized compactness*, i.e., the sum of first n_{RV} over the sum of all eigenvalues.

(5) Derive Discriminating Encoding: We further reduce the dimension of the encoding by accounting for the importance of individual coordinates in distinguishing the diseased from the healthy population. We do so by concatenating the coordinates of the LV and RV for each image and then apply mRMR [9] to the corresponding vectors across all images. mRMR keeps those coordinates that have the highest relevance with respect to separating the two populations while simultaneously minimizing the dependency between these coordinates. We automatically set the number of selected coordinates n_S via parameter exploration, i.e., we measure the accuracy of the encoding on a training data set with respect to 20,40,60 and 80% of the combined coordinates ($n_{LV}+n_{RV}$) and then choose the dimension that achieved the highest accuracy score. An example of an accuracy score is that of a classifier identifying diseased scans based on the encoding. Having determined the reduced dimension, we now defined a low-dimensional encoding of the morphometry of the input image, which is optimized towards encoding morphometric patterns distinguishing diseased from healthy populations.

This completes our description of the encoding of cardiac morphometry. We note that all important parameters are automatically determined by our morphometric framework. Thus, applying our encoding to a new data set is fairly simple assuming the segmentation of the template scan exist.

3. MEASURING THE ENCODING'S ACCURACY

We measure the accuracy of our mophometric encoding with respect to a linear SVM classifier [11] distinguishing normal MRIs from those with cardiac diseases on three different data sets. All data sets contain the same 9 healthy subject scans. The remaining 15 scans in each set are of patients diagnosed with different diseases, which are Tetrology of Fallot (TOF) for the 1^{st} set, Hypertrophic Cardiomyopathy (HCM) for the 2^{nd} set, and Diastolic Dysfunction (DD) for the 3^{rd} set. The medical literature (using volumetric scores) views TOF and HCM as diseases impacting the anatomy, while DD is characterized by Doppler velocity parameters. We therefore expect DD to be more difficult to detect via morphometric encodings. The MRIs of all data sets were acquired as part of standard clinical care at the Hospital of the University of Pennsylvania based on the following protocol: balanced steady-state-free-precession short-axis 1.5 Tesla acquisition, breath hold, 1.25 mm in-plane resolution, and 8 mm slice thickness. The fully automatic pre-processing of each MRI included anonymization, identifying the center of the LV in each slice via Hough transform [12], correcting for slice misalignment by stacking the slices so that the LV center aligned across the slices, correcting for bias field via N4ITK [13], keeping the 12 slices centered at the mid-portion of the ventricles and cropping each of those slices to a $150\mathrm{mm} \times 150\mathrm{mm}$ image centered around the heart. The remainder of this section describes

the experimental setup and results in further detail, showing that our encoding outperforms ventricular volume scores and RAVENS [6], both of whom not only require a segmentation of the template but also of each scan to be analyzed.

To fully parametrize our encoding, GRAM (Step 1) selects a (healthy) template from 100 MRI with the parameter setting defined in [10]. We then measure the accuracy of our encoding by applying the classifier to the three data sets via leave-one-out cross validation, i.e. we repeatedly train on 23 cases and test on one case until the method was tested on the complete data set. During training, we not only fix the size k of the neighborhood of the graph and the number of embedding coordinates (varying between 2 and 12 for each ventricle and between 5 and 22 for each data set) according to the criteria defined in Step 4 but also use parameter exploration to choose the optimal number of selected coordinates n_S of mRMR (see Step 5). We score each setting of n_S based on five-fold cross validation on the training data set of a linear SVM classifier. During five-fold cross validation, we also find the optimal penalty C of the classifier by exploring the search space $\{2^{-3}, 2^{-1}, \ldots, 2^{7}\}$. With the exception of the template segmentation, all important parameters are thus automatically determined during training.

Measuring the accuracy of the other two encodings follows the same mechanism. However, those encodings require the segmentations of the LV and RV for each subject, which were semi-automatically generated by an expert. In addition, the volume scores of LV and RV are directly fed into the SVM (i.e., without manifold learning and mRMR), whose outcome we refer to as 'Volume'. Finally, we replace our high-dimensional encoding of Step 3 with those of RAVENS maps [6] and refer with RAVENS to the outcome of the corresponding implementation.

The results of the experiments are summarized in Table 1. As expected, Volume receives the lowest accuracy scores out of the three encodings. While its score for TOF and HCM is 75%, the accuracy drops down to 62.5% for DD. These are in-line with the clinical findings that TOF and HCM typically impact the size of ventricles while DD affects cardiac function. The accuracy scores of RAVENS are at least 4% higher than those of Volume (see TOF). Our encoding receives the same specificity as RAVENS for TOF and DD, and the same sensitivity as Volume for DD. It clearly outperforms the other encodings with respect to all remaining scores. The highlight is a 91.7% accuracy for HCM. These results indicate that the variability introduced by the segmentations negatively impacts the accuracy of the morphometric encodings such as RAVENS. For DD, both deformation-based encodings achieve the largest gain in accuracy compared with Volume ($< 12\%$). This indicates that DD not only impacts the cardiac function but also the shape. However, the shape changes due to DD are subtle so that the global volume scores are insensitive to those changes.

To follow up the previous findings, we investigate the impact of a disease on either ventricle by recording in Table 2 the average number of embedding coordinates selected by mRMR with respect to each ventricle and deformation-based encoding. Regardless of the encoding, mRMR clearly finds the coordinates of one ventricle to be more informative for disease identification, i.e. the RV for TOF and the LV for HCM and DD. These findings are also in-line with the medical diagnosis of these diseases. Quite striking is the difference in the combined average number of selected coordinates with respect to each encoding and disease. While for RAVENS, the total varied quite a lot (low: 2.2, high: 4.6), the spread was much smaller for our encoding (low: 3.5, high: 4.1). This spread could indicate the robustness of an encoding (with smaller spreads being better) but confirming this notion would require additional experiments. We end the experiment by qualitatively assessing the outcome of the SVM classifiers based on our encoding. The first row of Figure 2 shows the most extreme cases of each population, i.e., the cases farthest away from the decision boundary defined by SVM. The second row lists the borderline cases closest to the boundary. The borderline cases for TOF and DD look very similar to a healthy one, where the borderline HCM shows the thickening of the myocardium. This explains why our encoding achieved the highest accuracy for HCM. As expected, the most healthy case (first row) looks very similar to the borderline scan of the same population. For the other populations, the extreme cases are quite different from the borderline images. The extreme TOF case clearly shows the enlargement of the RV and the myocardium is very thick in case of the extreme HCM case. In case of DD, the size of the ventricles compares to healthy cases. The low score of Volume is thus not surprising. Unlike for HCM and TOF, we did not find a disease-specific morphometric patterns for DD cases further away from the boundary. This reflects the notion that DD is primarily a disease impacting the function of the heart. We also note that misclassified test cases were always close to the boundary. If these findings persist on larger data sets of HCM and TOF then this distance could indicate the degree of impact by those diseases.

Disease	RAVENS			Our Encoding		
	LV	RV	Total	LV	RV	Total
TOF	0.9	3.7	4.6	0.6	2.9	3.5
HCM	2.2	0	2.2	2.1	1.5	3.6
DD	2.4	0.9	3.3	3.3	0.8	4.1

Table 2. Average number of selected embedding coordinates by mRMR. The numbers reflect disease impact, i.e. TOF mostly impacts the RV, and HCM and DD impact the LV

4. CONCLUSION

We described a fully-automatic morphometric encoding targeted towards differentiating diseased from healthy cardiac

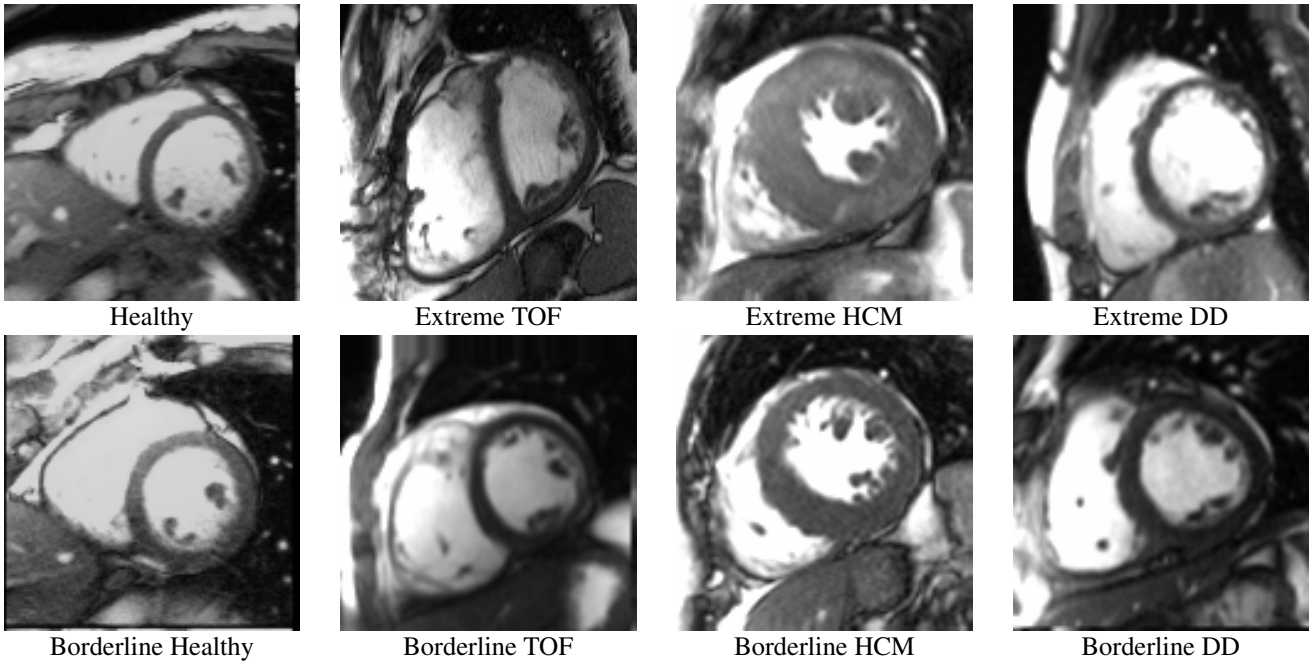

Fig. 2. Plotting scans according to their distances to SVM's decision boundary. The first row shows cases furthest away from the boundary and the second row shows the closest ones. The borderline TOF and DD look similar to the healthy cases.

MR scans. Unlike previous deformation-based scores, we only require the segmentation of a single template MRI. In addition, we reduced the encoding to those dimensions most informative for separating populations. Compared with volumetric scores and RAVENS, our measurements were more accurate in encoding morphometric patterns distinguishing healthy subjects from those with Tetralogy of Fallot, diastolic dysfunction, and hypertrophic cardiomyopathy.

5. ACKNOWLEDGEMENT

The project was supported in part by the NIH grant K05 AA017168.

6. REFERENCES

[1] W.A. Helbing and A. de Roos, "Clinical applications of cardiac magnetic resonance imaging after repair of Tetralogy of Fallot.," *Paediatric Cardiology*, vol. 21, pp. 70–79, 2000.

[2] T. Mansi, I. Voigt, B. Leonardi, X. Pennec, S. Durrleman, M. Sermesant, H. Delingette, A. M. Taylor, Y. Boudjemline, G. Pongiglione, and N. Ayache, "A statistical model for quantification and prediction of cardiac remodelling: Application to Tetralogy of Fallot," *TMI*, vol. 30, no. 9, pp. 1605–1616, 2011.

[3] D.H. Ye, H. Litt, C. Davatzikos, and K.M. Pohl, "Morphological classification: Application to cardiac MRI of Tetralogy of Fallot," in *FIMH*, 2011, vol. 6666 of *LNCS*, pp. 180–187.

[4] J. Xi, P. Lamata, S. Niederer, S. Land, W. Shi, X. Zhuang, S. Ourselin, S.G. Duckett, A.K. Shetty, C.A. Rinaldi, D. Rueckert, R. Razavi, and N.P. Smith, "The estimation of patient-specific cardiac diastolic functions from clinical measurements," *MIA*, vol. 17, no. 2, pp. 133 – 146, 2013.

[5] J. Ashburner and K.J. Friston, "Voxel-based morphometry - the methods," *NI*, vol. 11, pp. 805–821, 2000.

[6] C. Davatzikos, A. Genc, D. Xu, and S.M. Resnick, "Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy," *NI*, vol. 14, pp. 1361–1369, 2001.

[7] D.H. Ye, J. Hamm, B. Desjardins, and K.M. Pohl, "FLOOR: Fusing locally optimal registrations," in *MICCAI*, 2013, vol. 8151 of *LNCS*, pp. 195–202.

[8] J.B. Tenenbaum, V. de Silva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction.," *Science*, vol. 290, pp. 2319–2323, 2000.

[9] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *PAMI*, vol. 27, pp. 1226–1238, 2005.

[10] J. Hamm, D.H. Ye, R. Verma, and C. Davatzikos, "GRAM: A framework for geodesic registration on anatomical manifolds," *MIA*, vol. 14, pp. 633–642, 2010.

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[12] R.O. Duda and P.E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, pp. 11–15, 1972.

[13] Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee, "N4ITK: improved N3 bias correction.," *TMI*, vol. 29, pp. 1310–1320, 2010.