Hyperbolic Space Sparse Coding with Its Application on Prediction of Alzheimer's Disease in Mild Cognitive Impairment

Jie Zhang[1], **Jie Shi**[1], **Cynthia Stonnington**[2], **Qingyang Li**[1], **Boris A. Gutman**[4], **Kewei Chen**[3], **Eric M. Reiman**[3], **Richard J. Caselli**[6], **Paul M. Thompson**[4], **Jieping Ye**[5], and **Yalin Wang**[1]

[1]School of Computing, Informatics, and Decision Systems Engineering, Arizona State Univ., Tempe, AZ

[2]Dept. of Psychiatry and Psychology, Mayo Clinic Arizona, Scottsdale, AZ

[3]Banner Alzheimer's Institute and Banner Good Samaritan PET Center, Phoenix, AZ

[4]Imaging Genetics Center, Institute for Neuroimaging and Informatics, Univ. of Southern California, Marina del Rey, CA

[5]Dept. of Computational Medicine and Bioinformatics, Univ. of Michigan, Ann Arbor, MI

[6]Dept. of Neurology, Mayo Clinic Arizona, Scottsdale, AZ

Abstract

Mild Cognitive Impairment (MCI) is a transitional stage between normal age-related cognitive decline and Alzheimer's disease (AD). Here we introduce a hyperbolic space sparse coding method to predict impending decline of MCI patients to dementia using surface measures of ventricular enlargement. First, we compute diffeomorphic mappings between ventricular surfaces using a canonical hyperbolic parameter space with consistent boundary conditions and surface tensor-based morphometry is computed to measure local surface deformations. Second, ring-shaped patches of TBM features are selected according to the geometric structure of the hyperbolic parameter space to initialize a dictionary. Sparse coding is then applied on the patch features to learn sparse codes and update the dictionary. Finally, we adopt max-pooling to reduce the feature dimensions and apply Adaboost to predict AD in MCI patients ($N = 133$) from the Alzheimer's Disease Neuroimaging Initiative baseline dataset. Our work achieved an accuracy rate of 96.7% and outperformed some other morphometry measures. The hyperbolic space sparse coding method may offer a more sensitive tool to study AD and its early symptom.

Keywords

Mild Cognitive Impairment; Hyperbolic Parameter Space; Ring-shaped Patches; Sparse Coding and Dictionary Learning

1 Introduction

Mild Cognitive Impairment (MCI) is a transitional stage between normal aging and Alzheimer's disease (AD). Many neuroimaging studies aim to identify abnormal anatomical or functional patterns, their association with cognitive decline, and evaluate the therapeutic efficacy of interventions in MCI. Structural magnetic resonance imaging (MRI) measures

have been a mainstay of AD imaging research, including whole-brain [12], entorhinal cortex [2], hippocampus [15] and ventricular enlargement [14].

Ventricular enlargement is a highly reproducible measure of AD progression, owing to the high contrast between the CSF and surrounding brain tissue on T1-weighted images. However, its concave shape, complex branching topology and the extreme narrowness of the inferior and posterior horns have made ventricular enlargement notoriously difficult for analysis. Recent research has demonstrated that subregional surface-based ventricular morphometry analysis may offer improved statistical power. For example, a variety of surface-based analysis techniques such as SPHARM [13] and radial distance [14] have been proposed to analyze ventricular morphometry abnormalities. To model a topologically complicated ventricular surface, Shi et al. [11] proposed to use the hyperbolic conformal geometry to build the canonical hyperbolic parameter space of ventricular surfaces. After introducing cuts on the ends of three horns, ventricular surfaces become genus-zero surfaces with multiple open boundaries, which may be equipped with Riemannian metrics that induce negative Gaussian curvature. Hyperbolic Ricci flow method was adopted to compute their hyperbolic conformal parameterizations and the resulting parameterizations have no singularities. After registration, tensor-based morphometry (TBM) [11] was computed on the entire ventricular surfaces and used for group difference study. Thus far, no attempt has been made to use the hyperbolic space based surface morphometry features for the prognosis of AD.

In this paper, we propose a new hyperbolic space sparse coding and dictionary learning framework, in which a Farthest point sampling with Breadth-first Search (FBS) algorithm is proposed to construct ring-shaped feature patches from hyperbolic space and patch based hyperbolic sparse coding algorithm is developed to reduce feature dimensions. Max-pooling [1] and Adaboost [10] are used for finalizing features and binary classification. We further validate our algorithms with AD prediction in MCI using ventricular surface TBM features. The major contributions of this paper are as follows. First, to the best of our knowledge, it is the first sparse coding framework which is designed on hyperbolic space. Second, the hyperbolic space sparse coding empowers the AD prediction accuracy through ventricular morphometry analysis. In our experiments with the ADNI data ($N = 133$), our ventricular morphometry system achieves 96.7% accuracy, 93.3% sensitivity, 100.0% specificity and outperforms other ventricular morphometric measures in predicting AD conversion for MCI patients.

2 Hyperbolic Space Sparse Coding

The major computational steps of the proposed system are illustrated in Fig. 1. The new method can be divided into two stages. In the first stage, we perform MRI scan segmentation, ventricular surface reconstruction, hyperbolic Ricci flow based surface registration and surface TBM statistic computation. In the second stage, we build ring-shaped patches on the hyperbolic parameter space by FBS to initialize original dictionary, SCC based sparse coding and dictionary learning and max-pooling are performed for dimension reduction. Following that, Adaboost is adopted to predict future AD conversion, i.e. classification on MCI-converter group versus MCI-stable group.

2.1 Hyperbolic Space and Surface Tenser-based Morphometry

We applied hyperbolic Ricci flow method [11] on ventricular surfaces and mapped them to the Poincaré disk with conformal mapping. On the Poincaré disk, we computed a set of consistent geodesics and projected them back to the original ventricular surface, termed as *geodesic curve lifting*. Further, we converted the Poincaré model to the Klein model where the ventricular surfaces are registered by the constrained harmonic map. The computation of canonical hyperbolic spaces for a left ventricular surface is shown in Fig. 2.

In Fig. 2, geodesic curve lifting used to construct a canonical hyperbolic space for ventricular surface registration. γ_1, γ_2, γ_3 are some consistent anchor curves automatically located on the end points of each horn. On the parameter domain, τ_1 is an arc on the circle which passes one endpoint of $\frac{\gamma_1}{2}$ and one endpoint of γ_2 and is orthogonal to $|z| = 1$. The initial paths τ_1 and τ_2 can be inconsistent, but they have to connect consistent endpoints of γ_1, γ_2 and γ_3, as to guarantee the consistency of the geodesic curve computation. After slicing the universal covering space along the geodesics, we get the canonical fundamental domain in the Poincaré disk, as shown in Fig. 2(b). All the boundary curves become geodesics. As the geodesics are unique, they are also consistent when we map them back to the surface in $\mathbb{R}^3$. Furthermore, we convert the Poincaré model to the Klein model with the complex function [11]: $z = 2z/1 + \overline{z}z$. It converts the canonical fundamental domains of the ventricular surfaces to a Euclidean octagon, as shown in Fig. 2 (c). Then we use the Klein disk as the canonical parameter space for the ventricular surface analysis.

After that, we computed the TBM features [11] and smooth them with the heat kernel method [3]. Suppose $\phi = S_1 \rightarrow S_2$ is a map from surface S_1 to surface S_2. The derivative map of ϕ is the linear map between the tangent spaces $d\phi : TM(p) \rightarrow TM(\phi(p))$, induced by the map ϕ, which also defines the Jacobian matrix of ϕ. The derivative map $d\phi$ is approximated by the linear map from one face $[\upsilon_1, \upsilon_2, \upsilon_3]$ to another one $[w_1, w_2, w_3]$. First, we isometrically embed the triangles $[\upsilon_1, \upsilon_2, \upsilon_3]$ and $[w_1, w_2, w_3]$ onto the Klein disk, the planar coordinates of the vertices, denotes by $\upsilon_i, w_i, i = 1,2,3$, which represent the 3D position of points $\upsilon_i, w_i, i = 1,2,3$. Then, the Jacobian matrix for the derivative map $d\phi$ can be computed as $J = d\phi = [w_3 - w_1, w_2 - w_1][\upsilon_3 - \upsilon_1, \upsilon_2 - \upsilon_1]^{-1}$.

Based on the derivative map J, the deformation tensors $S = \sqrt{J^T J}$ was defined as TBM, which measures the amount of local area changes in a surface. As pointed out in [3], each step in the processing pipeline including MRI acquisition, surface registration, etc., are expected to introduce noise in the deformation measurement. To account for the noise effects, we apply the heat kernel smoothing algorithm proposed in [3] to increase the SNR in the TBM statistical features and boost the sensitivity of statistical analysis.

2.2 Ring-shaped Patch Selection

The hyperbolic space is different from the original Euclidean space, the structure is more complicated and demands more efforts for selecting patches based on its topological structure. The common rectangle patch construction cannot be directly applied to the hyperbolic space. Therefore, we proposed a Farthest point sampling with Breadth-first Search (FBS) on hyperbolic space to initialize original dictionary for sparse coding. Fig. 3

(right) is the visualization of patch selection on hyperbolic parameter domain. And Fig. 3 (left) projects the selected patches on hyperbolic parameter domain back to the original ventricular surface, which still maintains the same topological structure as the parameter domain.

First, we randomly selected a point center on the hyperbolic space, denotes by p_{x1}, $p_{x1} \in X_r$, where X_r is the set of all discrete vertices on hyperbolic space. Then, we find all points $p_{x1,i}$ ($i = 1, 2, \ldots, n$), where n is the maximum number of connected points connecting with the patch center p_{x1}. The procedure is called breadth-first search (BFS)[8], which is an algorithm for searching graph data structures. It starts at the tree root and explores the neighbor nodes first, before moving to the next level neighbors. Then, we used the same procedure to find all connected points with $p_{x1,i}$, which are $p_{x1,ij}$ ($j = 1, 2, \cdots, m_i$). Here, m_i represents the maximum number of connected points with each specific point $p_{x1,i}$. The points $p_{x1,ij}$ are connected with $p_{x1,i}$ by using same procedure–BFS– between p_{x1} and $p_{x1,i}$. Finally, we get a set P_{x1} as follows, which is a selected patch with patch center p_{x1} and do not contain duplicate points.

$$P_{x_1} = \{p_{x_1}, p_{x_1,1}, p_{x_1,1_1}, \cdots, p_{x_1,1_{m_1}}, \cdots, p_{x_1,n}, p_{x_1,n_1}, \cdots, p_{x_1,n_{m_n}}\}. \quad (1)$$

Algorithm 1

Farthest point sampling with Breadth-first Search (FBS)

Input: Hyperbolic parameter space.

Output: A collect of different amount overlapped patches on topological structure.

1: Start with $X' = \{p_{x_1}\}$, X_r denotes all discrete vertices on the hyperbolic space.

2: **for** t=1 to T **do**

3:   **for** r **do** determine sampling radius

4:    Find all connected components $p_{x_t,i}$ of p_{x_t} by using one step BFS.

5:    Find set P_{x_t} similar with Eq. 1 by using one step BFS.

6:    $r = \max_{p_{x'} \in X_r} dX_t(p_{x'}, p_{x_t})$

7:    **if** r $10e^{-2}$ **then** STOP

8:   **end if**

9:    Find the farthest point from X'

10:    $p_{x_{t+1}} = \arg \max_{p_{x'} \in X_r} dt(p_{x'}, X')$

11:    Add $p_{x_{t+1}}$ to X'

12:   **end for**

13: **end for**

We can find all connected components of the center point p_{x1} which are all in set p_{x1}. After that, we reconstruct the topological patches based on hyperbolic geometry and connected edges between the different points within p_{x1} according to topological structure. We use Φ_1 denotes the first selected patch of the root (patch center) p_{x1}. Since we randomly select patches with different degree overlapped, we use radius $r = \max_{p_{x'} \in X_r} dX_t(p_{x'}, p_{x1}$ to determine next patch's root p_{x2} position.

In this way, we can find the second patch root $p_{x_2} \in X_r$ with the farthest distance r of p_{x_1}. We apply farthest point sampling [7], because the sampling principle is based on the idea of repeatedly placing the next sample point in the middle of the least known area of the sampling domain, which can guarantee the randomness of the patches selection. Here, d is hyperbolic distance in the Klein model. Given two points p and q, draw a straight line between them; the straight line intersects the unit circle at points a and b, so d is defined as follows:

$$d(p,q) = \frac{1}{2}(\log\frac{|aq||bp|}{|ap||bq|}), \quad (2)$$

where $|aq| > |ap|$ and $|bp| > |bq|$.

Then, we can calculate:

$$p_{x_2} = \arg\max_{p_x \in X_r} dX_r(p_x, X), \quad (3)$$

where X' denotes the set of selected patch centers. Then, we add p_{x_2} in X' and iterate the patch selection procedure $T = 2000$ times, because it will cover all vertexes according to the experimental results. The details of FBS are summarized in Algorithm 1.

2.3 Sparse Coding and Dictionary Learning

For our problem, the dimension of surface-based features is usually much larger than the number of subjects, e.g., we have approximate 150,000 features from each side of ventricle surfaces on each subject. Therefore, we used the technique of dictionary learning [6] with pooling to reduce the dimension before prediction. The problem statement of dictionary learning is described as below.

Given a finite training set of signals $X = (x_1, x_2, \cdots, x_n)$ in $\mathbf{R}^{n \times m}$ image patches, each image patch $x_i \in \mathbf{R}^m$, $i = 1, 2, \cdots, n$, where m is the dimension of image patch. Then, we can incorporate the idea of patch features into the following optimization problem for each patch x_i:

$$\min_{f_i}(D, z_i) = \frac{1}{2}\|Dz_i - x_i\|_2^2 + \lambda\|z_i\|_1. \quad (4)$$

Specifically, suppose there are t atoms $d_j \in \mathbf{R}^m$, $j = 1, 2, \cdots, t$, where the number of atoms is much smaller than n (the number of image patches) but larger than m (the dimension of the image patches). x_i can be represented by $x_i = \sum_{j=1}^{t} z_{i,j} d_j$. In this way, the m-dimensional vector x_i is represented by an t-dimensional vector $z_i = (z_{i,1}, \cdots, z_{i,t})^T$, which means the learned feature vector z_i is a sparse vector. In Eq. 4, where λ is the regularization parameter,

$\|\cdot\|$ is the standard Euclidean norm and $\|z_i\|_1 = \sum_{j=1}^{t} |z_{i,t}|$ and $D = (d_1, d_2, \cdots, d_l) \in \mathbf{R}^{t \times m}$ is the dictionary, each column representing a basis vector.

To prevent an arbitrary scaling of the sparse codes, the columns d_i are constrained by $C \triangleq \{D \in \mathbf{R}^{t \times m} s.t. \forall j = 1, \cdots, t, d_j^T d_j \leq 1\}$. Thus, the problem of dictionary learning can be rewritten as a matrix factorization problem:

$$\min_{D \in C, \mathbf{Z} \in n \times t} \frac{1}{2} \|X - D\mathbf{Z}\|_F^2 + \lambda \|\mathbf{Z}\|_1. \quad (5)$$

It is a convex problem when either D or $\mathbf{Z}$ is fixed. When the dictionary D is fixed, solving each sparse code z_i is a Lasso problem. Otherwise, when the Z are fixed, it will become a quadratic problem, which is relative time consuming. Thus, we choose the SCC algorithm [6], because it can dramatically reduce the computational cost of the sparse coding while keeping a comparable performance.

3 Dataset of Experiments and Classification Results

We selected 133 subjects from the MCI group in the ADNI baseline dataset [11]. These subjects were chosen on the basis of having at least 36 months of longitudinal data, which consisting of 71 subjects who developed AD within 36 months (MCIc) group and 62 subjects who did not convert to AD (MCIs) group. All subjects underwent thorough clinical and cognitive assessment at the time of aquisition, including Mini-Mental State Examination (MMSE), Alzheimer's disease assessment scale-Cognitive (ADAS-Cog). The statistics with matched gender, education, age, and MMSE are shown in Table 1.

In this work, we employed the Adaboost [10] to do the binary classification and distinguish different individuals in different groups. Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), Positive predictive value (PPV) and Negative predictive value (NPV) were computed to evaluate classification results [4]. We also computed the area-under-the-curve (AUC) of the receiver operating characteristic (ROC) [4]. A five-fold cross-validation was adopted to estimate classification accuracy. For comparison purposes, we computed ventricular volumes and surface areas within the MNI space on each side of the brain hemisphere [9], which are viewed as powerful MRI biomarker that has been widely-used in studies of AD. And we also compared FBS with a ventricular surface shape method in [5] (*Shape*), which built an automatic shape modeling method to generate comparable meshes of all ventricles. The deformation based morphometry model were employed with repeated permutation tests and then used as geometry features. Support vector machine was adopted as the classifier. With our ventricle surface registration results, we followed *Shape* work for selecting biomarkers and classification on the same dataset with our new algorithm. We tested FBS, *Shape*, volume and area on left, right and whole ventricle, respectively. Table 2 shows classification performance in one experiment featuring four methods.

Throughout all the experimental results, we can find that the best accuracy (96.7%), the best sensitivity (93.3%), the best specificity (100%), the best positive position value (100%) and

negative position value (88.9%) were achieved when we use TBM features on ventricle hyperbolic space on both sides (whole) for training and testing. The comparison also shows that our new framework selected better features and made better and more meaningful classification results. In Figure 4, we also generated ROC and computed AUC measures in four experiments. The FBS algorithm with whole ventricle TBM features achieved best AUC (0.957). The comparison demonstrated that our proposed algorithm may be useful for AD diagnosis and prognosis research. In the future, we will do more in depth comparisons against other shape analysis modules, such as SPHARM-PDM and radio distance, to further improve our algorithm efficiency and accuracy.

Acknowledgments

References

1. Boureau YL, Ponce J, LeCun Y. A theoretical analysis of feature pooling in visual recognition. Proceedings of the ICML-10. 2010:111–118.

2. Cardenas V, et al. Brain atrophy associated with baseline and longitudinal measures of cognition. Neurobiology of aging. 2011; 32(4):572–580. [PubMed: 19446370]

3. Chung MK, et al. Cortical thickness analysis in autism with heat kernel smoothing. NeuroImage. 2005; 25(4):1256–1265. [PubMed: 15850743]

4. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006; 27(8):861–874.

5. Ferrarini L, et al. Ventricular shape biomarkers for Alzheimer's disease in clinical MR images. Magnetic resonance in medicine. 2008; 59(2):260–267. [PubMed: 18228600]

6. Lin B, et al. Stochastic coordinate coding and its application for drosophila gene expression pattern annotation. arXiv preprint arXiv. 2014; 1407.8147

7. Moenning C, Dodgson NA. Fast marching farthest point sampling. Proc. EUROGRAPHICS. 2003; 2003

8. Patel JR, et al. Comparison between breadth first search and nearest neighbor algorithm for waveguide path planning.

9. Patenaude B, et al. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage. 2011; 56(3):907–922. [PubMed: 21352927]

10. Rojas R. Adaboost and the super bowl of classifiers a tutorial introduction to adaptive boosting. Freie University, Berlin, Tech. Rep. 2009

11. Shi J, et al. Studying ventricular abnormalities in mild cognitive impairment with hyperbolic Ricci flow and tensor-based morphometry. NeuroImage. 2015; 104:1–20. [PubMed: 25285374]

12. Stonnington CM, et al. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. Neuroimage. 2010; 51(4):1405–1413. [PubMed: 20347044]

13. Styner M, et al. Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. Proc. Natl. Acad. Sci. U.S.A. 2005 Mar; 102(13): 4872–4877. [PubMed: 15772166]

14. Thompson PM, et al. Mapping hippocampal and ventricular change in Alzheimer disease. Neuroimage. 2004; 22(4):1754–1766. [PubMed: 15275931]

15. Zhang J, et al. Applying sparse coding to surface multivariate tensor-based morphometry to predict future cognitive decline. IEEE International Symposium on Biomedical Imaging. 2016
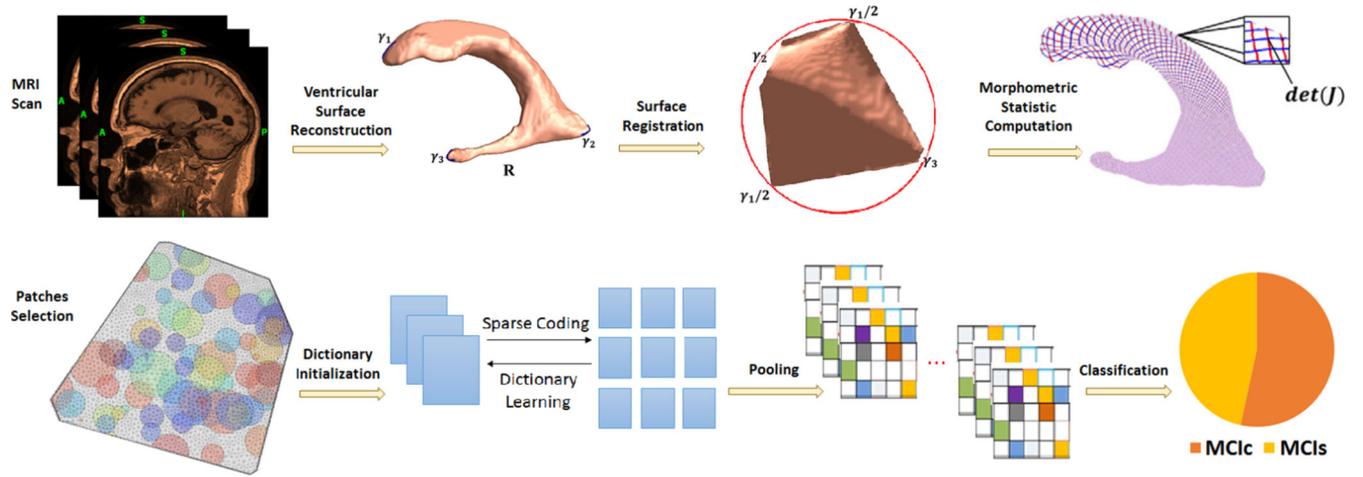
Fig. 1.
The major processing steps in the proposed framework.

(a) Topology Optimization of a Pair of Ventricular Surfaces

(b) Parametric Domain of Left Ventricle in (a)

(c) Klein Model of Left Ventricle in (a)
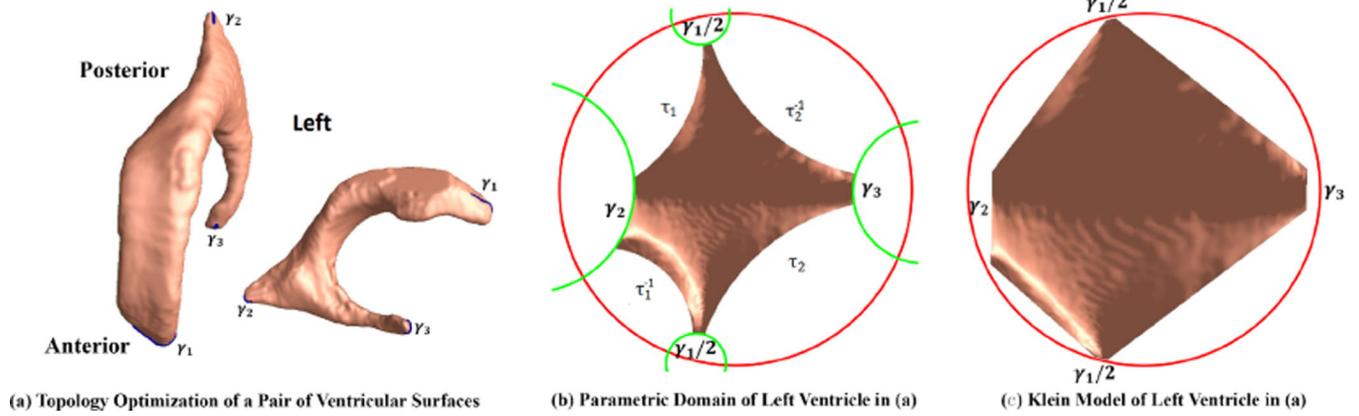
Fig. 2.
Modeling ventricular surface with hyperbolic geometry. (a) shows three identified open boundaries, γ_1, γ_2, γ_3, on the ends of three horns. After that, ventricular surfaces can be conformally mapped to the hyperbolic space. (b)(c) show the hyperbolic parameter space, where (b) is the Poincaré disk model and (c) is the Klein model.
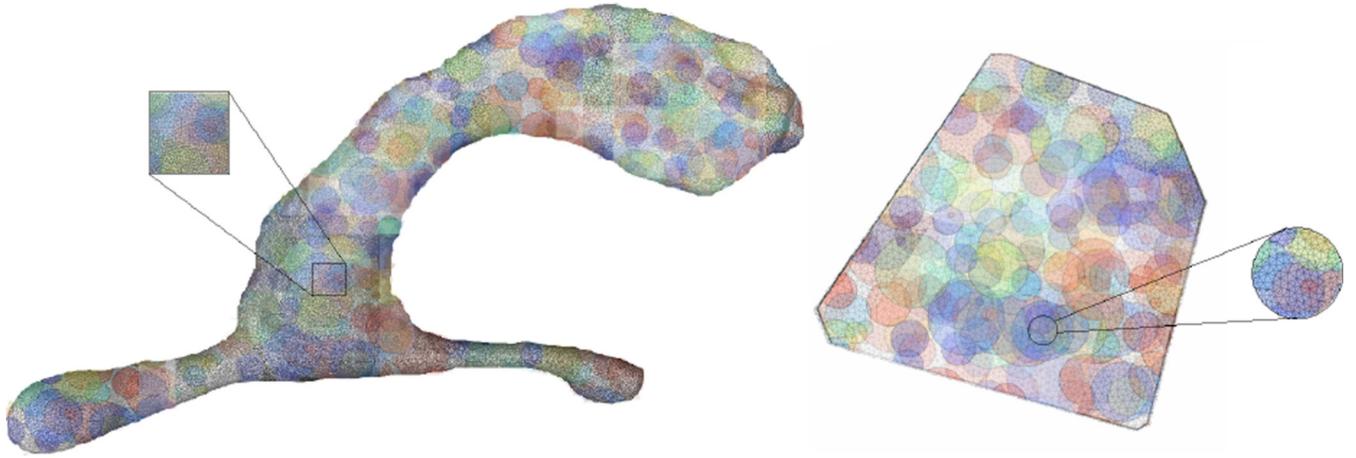
Fig. 3.
Visualization of computed image patches on ventricle surfaces and hyperbolic geometry, respectively. The zoom-in pictures show some overlapping areas between image patches.
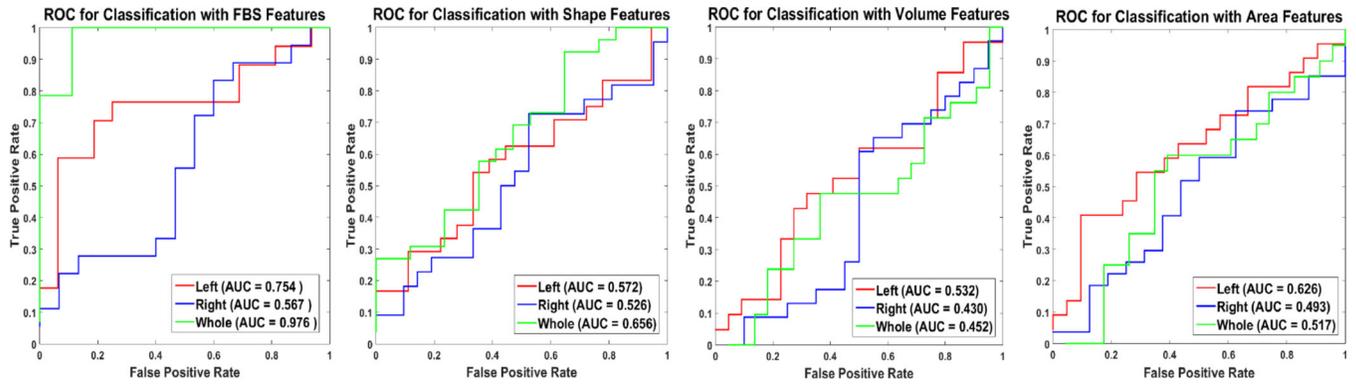
Fig. 4.
Classification performance comparison with ROC curves and AUC measures.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Demographic statistic information of our experiment's dataset.

	Number	Gender (F/M)	Education	Age	MMSE
MCIc	71	26/45	15.99±2.73	74.77±6.81	26.83±1.60
MCIs	62	18/44	15.87±2.76	75.42±7.83	27.66±1.57

Table 2

Classification results comparing with other systems.

Name	Region	ACC	SEN	SPE	PPV	NPV	AUC
FBS	Left	0.727	0.786	0.684	0.647	0.813	0.754
	Right	0.652	0.652	0.000	1.000	0.000	0.567
	Whole	**0.967**	**0.933**	**1.000**	**1.000**	**0.889**	**0.976**
Shape	Left	0.535	0.615	0.412	0.615	0.412	0.572
	Right	0.512	0.515	0.500	0.773	0.238	0.526
	Whole	0.605	0.656	0.500	0.731	0.412	0.656
Volume	Left	0.558	0.571	0.552	0.381	0.727	0.532
	Right	0.517	0.536	0.467	0.652	0.350	0.430
	Whole	0.535	0.607	0.400	0.654	0.353	0.452
Area	Left	0.558	0.552	0.571	0.727	0.381	0.626
	Right	0.465	0.625	0.370	0.370	0.625	0.493
	Whole	0.512	0.482	0.563	0.650	0.391	0.517