Dimensionality reduced cortical features and their use in predicting longitudinal changes in Alzheimer's disease

Hyunjin Park[a], Jin-ju Yang[b], Jongbum Seo[c], Jong-min Lee[b,*], the ADNI[1]

[a] School of Electronic Electrical Engineering, Sungkyunkwan University, Republic of Korea
[b] Department of Biomedical Engineering, Hanyang University, Seoul, Republic of Korea
[c] Department of Biomedical Engineering, Yonsei University, Republic of Korea

HIGHLIGHTS

- Cortical features provide information to distinguish Alzheimer's from normal.
- Cortical thickness and sulcal depth were used.
- Manifold learning using PCA reduced the dimensionality of the cortical feature.
- The dimensionality reduced features were used to predict conversion to AD.
- SVM classifiers were used with the reduced cortical features.

ARTICLE INFO

ABSTRACT

Neuroimaging features derived from the cortical surface provide important information in detecting changes related to the progression of Alzheimer's disease (AD). Recent widespread adoption of neuroimaging has allowed researchers to study longitudinal data in AD. We adopted cortical thickness and sulcal depth, parameterized by three-dimensional meshes, from magnetic resonance imaging as the surface features. The cortical feature is high-dimensional, and it is difficult to use directly with a classifier because of the "small sample size" problem. We applied manifold learning to reduce the dimensionality of the feature and then tested the usage of the dimensionality reduced feature with a support vector machine classifier. Principal component analysis (PCA) was chosen as the method of manifold learning. PCA was applied to a region of interest within the cortical surface. We used 30 normal, 30 mild cognitive impairment (MCI) and 12 conversion cases taken from the ADNI database. The classifier was trained using the cortical features extracted from normal and MCI patients. The classifier was tested for the 12 conversion patients only using the imaging data before the actual conversion. The conversion was predicted early with an accuracy of 83%.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Recent advances in neuroimaging techniques such as high-resolution magnetic resonance imaging (MRI) have revealed detailed information about the cortex. Study of shape, known as morphometry, has been successfully applied to distinguish AD from normal controls [11]. Recent widespread adoption of neuroimaging has allowed researchers to study longitudinal data in AD. Researchers can assess shape information from different time points to compute the temporal changes in shape information for a given patient in longitudinal data. Many researchers have adopted a method known as voxel based morphometry (VBM) [1]. VBM is typically applied to 3D volume data. Surface features derived from brain cortex have also been adopted to distinguish demented patients from normal controls. Cortical thickness, sulcal depth, surface area, and mean curvature computed from the cortical surface have been applied to distinguish AD from normal controls [10,19]. The cortex is a complexly folded two 2D sheet embedded in a 3D space, typically modeled using a set of 3D polygonal meshes. Tens of thousands of vertices are used to form a cortical surface in many implementations. All the features defined on the cortical surface can be pooled and treated as one high-dimensional feature vector, which is commonly done in machine learning. N feature vectors, whose dimensionality is in the tens of thousands, would be generated to study N patients. Neuroimaging studies tend to have less than few hundred patients in a study. Thus, the number of

* Corresponding author. Tel.: +82 2 2220 0697; fax: +82 2 2296 5943.
E-mail addresses: ljm@hanyang.ac.kr, jmlee@bme.hanyang.ac.kr (J.-m. Lee).

observations is far less than the dimensionality of the features, which would lead to unreliable classifier performance [6,7,15]. This problem of unsatisfactory classifier performance is known as the "small sample size" problem. The dimensionality of the feature vector may be reduced by using manifold learning methods. As a result, handling of shape information becomes more tractable in a low dimensional space. PCA was applied to the high dimensional cortical features in this study. We chose cortical thickness and sulcal depth as our features since these two features have been effective for reflecting shape changes related to AD and MCI [10]. PCA was applied to the features defined on the whole cortex and the features defined on a specific ROI. ROI was automatically defined by choosing a set of vertices with high discriminative power to distinguish demented scans from normal controls. The features whose dimensions were reduced were fed into a support vector machine (SVM) classifier. SVM is one of the state-of-the-art classifiers currently available with a proven track record [2]. Many excellent papers applied PCA or methods similar to PCA for feature extraction and then used the extracted features with a modern classifier to distinguish AD from normal controls [4,8,14,15,17]. López et al. adopted kernel PCA and linear discriminant analysis (LDA) for feature extraction of SPECT voxel values and used the extracted features with the SVM classifier [14]. Ramírez et al. adopted partial least square (PLS) regression for feature extraction of SPECT voxel values and used random forest method for prediction [17]. Graña et al. applied Pearson's correlation to diffusion tensor imaging and selected voxels with high discriminative power [8]. Then the selected voxels were used as input to the SVM classifier. They adopted fractional anisotropy and mean diffusivity as features derived from the diffusion tensor imaging. Chaves et al. applied voxel-wise *T*-tests to find voxels with good normalized mean squared error features and used the selected voxels as input to the SVM classifier in SPECT [4]. López et al. applied PCA and LDA for feature extraction of SPECT or PET voxel values and used the extracted features with SVM or neural network classifier [15]. The SPECT and PET scans were obtained from the ADNI database. The main difference between our approach and the existing work is that we used cortical features while the others adopted volumetric features, which are either functional (i.e., SPECT/PET) voxel values or structural (i.e., MRI) voxel values. Most existing research for classifying AD/MCI cases from normal controls did not consider longitudinal data [10]. Each patient was scanned once to retrospectively determine what class the patient's imaging data belonged to. Our study is a longitudinal one whose imaging data comes from the ADNI brain database [3]. The goal of this study is to apply dimensionality reduced features to detect conversion to AD early by using SVM. We want to predict the conversion to AD before the actual conversion to AD was confirmed. A secondary goal of this study is to classify MCI and normal by using dimensionality reduced features. For early detection of AD, the SVM classifier is trained with normal and demented scans and tested for conversion patients, who converted from the normal to demented state, by only using the imaging data before the actual conversion. Existing research that had analyzed longitudinal data considered cortical features as well but did not consider dimensionality reduction with the cortical features [19]. This study builds on a previous article where dimensionality reduced cortical features were adopted for distinguishing non-longitudinal AD/MCI scans from normal controls [16]. Here, we adopt the same approach proposed but extend its application to longitudinal data to address the issue of early AD detection. The very combination of (1) using two cortical features, (2) using PCA for manifold learning, and (3) predicting early conversion to AD could not be found in existing research to the best of our knowledge. We achieved prediction accuracy of early AD conversion comparable to that in a recent study [19].

2. Materials and methods

2.1. Subjects and MRI images

Image data were obtained from the ADNI database [3]. Each subject was scanned on two or more visits, separated by at least 6 months between visits. We selected 30 normal cases, 30 MCI cases, and 12 conversion cases from the database if they were clinically classified as (1) normal controls – individuals who were cognitively normal (CDR = 0) with Mini-Mental State Examination (MMSE) scores between 28 and 30, (2) MCI – individuals with a memory complaint who experienced a very mild cognitive decline with a CDR of 0.5 and MMSE scores between 24 and 27, and (3) conversion – individuals who converted from normal to MCI. We were limited to 12 cases of conversion as there were 16 conversion cases in the database and 4 cases were deemed unsuitable for our pre-processing procedures. We randomly chose 30 normal and 30 MCI cases from the database to have adequate statistical power for each group. The MCI patients were 77.12 ± 6.83 (mean $\pm$ STD) years old and of gender ratio of 19/11 (M/F). The normal control patients were 75.13 ± 5.17 years old and of gender ratio of 18/12 (M/F). The conversion patients were 75.92 ± 5.30 years old and of gender ratio of 8/4 (M/F). All MRIs were sagittal T1-weighted scans and had typical dimensions of 256×256 and resolutions of $0.94\,mm \times 0.94\,mm \times 1.2\,mm$.

2.2. Cortical surface extraction and surface registration

We applied the same procedures described in the previous work to the longitudinal MRI data [16]. A brief summary follows. We used the MNI image processing software to produce the cortical surface and to register the extracted surfaces [20]. MR images were first registered in an affine fashion and then corrected for intensity non-uniformity. The registered and corrected volumes were then classified into white and gray matter, cerebrospinal fluid, and background. The Constrained Laplacian-based Automated Segmentation with Proximities (CLASP) algorithm was used to extract the cortical surface [12]. A surface model for each brain hemisphere was constructed using 81,924 polygonal 3D meshes. With the extracted surface, we applied a 2D surface based non-rigid registration algorithm based on a geodesic distance from the gyral crown vertex with a smoothing term [18]. The registration software established spatial correspondence between a given image and a predefined template image, and thus all 72 images were registered onto a common template so that they could all be compared on a vertex-by-vertex basis.

2.3. Computation of cortical features

We computed two cortical features, cortical thickness and sulcal depth. The CLASP algorithm provides two cortical surfaces, the inner and the outer surface, which are deformed from a common sphere parameterized with 3D meshes. Cortical thickness was computed as the Euclidean distance between the linked vertices of the inner and the outer cortical surface. We measured the Euclidean distance from each vertex on the cortical surface to the nearest voxel on the volume of the cerebral hull as the sulcal depth [9]. We refer to the collection of features computed on the cortical surface as a feature map. The computed feature maps were resampled onto the spatial frame of the common template using the surface registration described, so that all the feature maps reside on the same spatial frame for vertex-by-vertex comparison. In summary, we obtained two features (i.e., cortical thickness and sulcal depth) defined on the common cortical surface for all 81,924 vertices for each subject. A sample of the adopted cortical features, cortical thickness and sulcal depth, is given in Fig. 1.
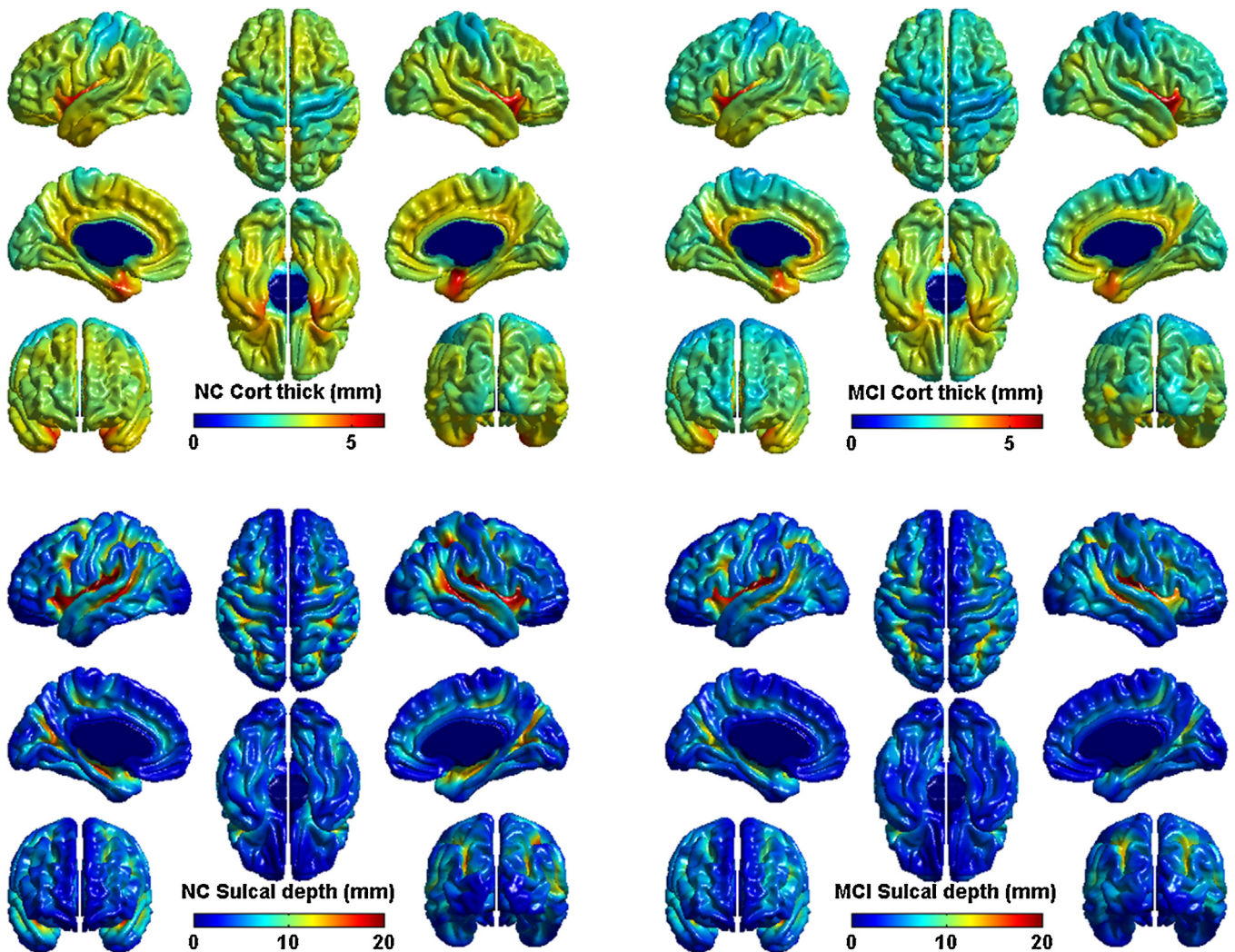
Fig. 1. Plot of sample cortical features. The top left subfigure shows the cortical thickness measurements of a normal patient and the top right subfigure shows the cortical thickness measurements of an MCI patient. The bottom left subfigure shows the sulcal depth measurements of a normal patient and the bottom right subfigure shows the sulcal depth measurements of an MCI patient.

2.4. PCA and dimensionality reduction

A feature map was treated as one high dimensional vector of dimension 81,924. We obtained two (i.e., cortical thickness and sulcal depth) 81,924 long feature vectors for each subject, all resampled onto a common template. Direct use of the high-dimensional cortical feature is troublesome because of the "small sample size" problem, where the number of observations for a given feature map (i.e., number of subjects) is far less than the dimension of the feature map (i.e., number of vertices to form the cortical surface). We applied PCA, a well-known manifold learning method, to reduce the dimension of the cortical features. A major parameter when applying manifold learning is the value of the reduced dimension known as the intrinsic dimension. Intrinsic dimension is difficult to estimate especially in neuroimaging applications [5]. Here, we extensively searched to find the intrinsic dimension that would result in the best classifier performance.

2.5. ROI selection

Not all vertices on the cortical surface are statistically equal for distinguishing between demented cases from normal controls. For example, vertices in the hippocampus region were considered more

informative in distinguishing demented cases from normal controls than considering all vertices of the cortex [15,19]. By considering the features of a ROI, one can remove many non-informative vertex values and shorten the overall feature vector. The observed feature vector is easier to handle since it is shorter and less noisy with many non-informative values removed. Here, we specified the ROI using the statistical discriminative power of vertices. A two-sample T-test was performed on a vertex-by-vertex basis on the feature maps from two subject groups being compared (e.g., MCI and normal) and the vertices whose p-values were lower 10% of the maximum p-value were selected to comprise the ROI. The ROI contained 8912 vertices for using either cortical thickness or sulcal depth alone and contained 16,384 vertices for using the combined feature. Vertices of the ROI were scattered all over the cortical surface but many of them were in the hippocampus region.

2.6. SVM classifier

The SVM classifier is one of the best performing off-the-shelf classifier available. The goal of this study is not to devise a new classifier. Technical details of the SVM could be found in the reference [2]. The goal of this study is to apply dimensionality reduced features to detect conversion to AD early by using SVM. A secondary
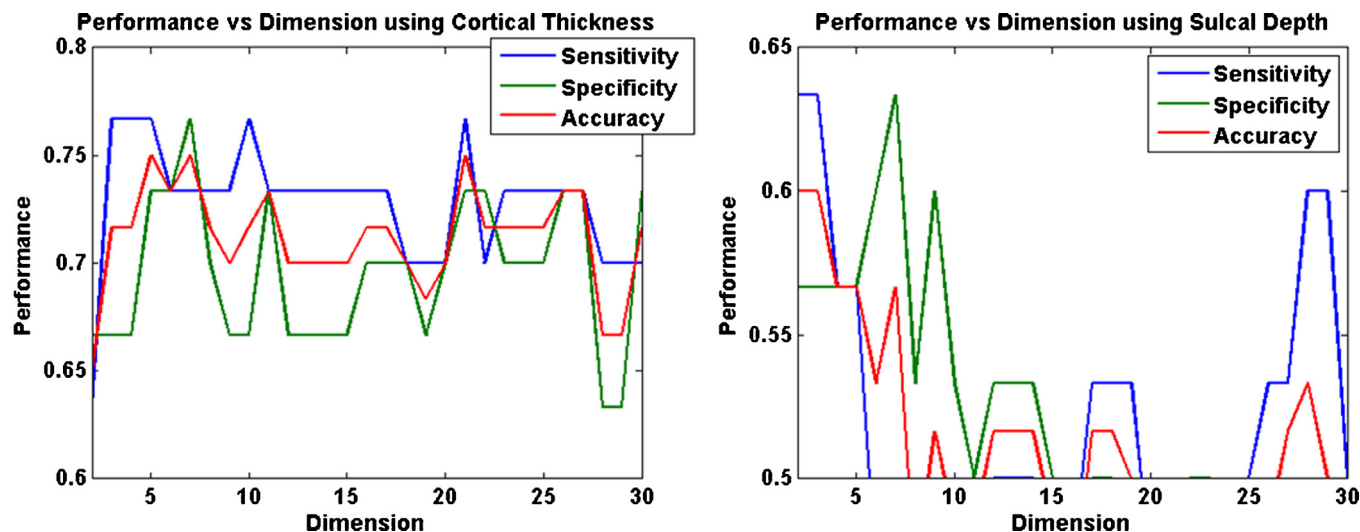
Fig. 2. Plot of classifier performance to distinguish MCI and normal with respect to the value of the reduced dimension. The blue, green, and red lines correspond to sensitivity, specificity, and accuracy values, respectively. The left subplot corresponds to cortical thickness and the right subplot corresponds to sulcal depth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

goal is to classify between MCI and normal by using dimensionality reduced features. We did not consider classifying between AD and normal since it is less related to early detection of AD than classifying between MCI and normal. The SVM classifier was applied with a linear kernel. The input to the SVM was the dimensionality reduced feature vector derived from the whole cortical surface or the ROI. The output is a binary decision, belonging to either the normal or MCI class. As a secondary goal, we wanted to classify between MCI and normal. For this, we adopted the leave-one-out validation method to separate training and test data. Given 30 MCI and 30 normal cases, we assigned one case as the test case and used the remaining 59 cases as the training data for the SVM classifier. The process was repeated 60 times choosing a different test case each time. For early detection of AD conversion, the primary goal of this study, the training data consisted of 30 MCI and 30 normal cases and the test data consisted of 12 conversion cases. For testing conversion cases, we only included imaging data that were obtained before the AD was confirmed. If a test case was classified as normal then it failed at early detection of AD and if a test case was classified as MCI then it was considered as success. Training data and test data never overlapped in all cases.

2.7. Experimental setup

There were three groups in this study: normal, MCI, and conversion groups. We computed the cortical surface and the two related feature maps, cortical thickness and sulcal depth, for all subjects in each group. The feature maps of each subject were linked to a long feature vector. The features were separated into training and

test data by using the leave-one-out validation for distinguishing between normal and MCI. For early detection of AD conversion, training data were taken from normal and MCI groups and test data were from taken from the conversion group. ROI was computed only using the training data. We then applied PCA to reduce the dimension of the feature vector defined on the whole cortex or the ROI using a user-specified dimension as the intrinsic dimension (i.e., value of the reduced dimension). The training and test data were fed into the linear SVM classifier. The accuracy was computed comparing ground truth with the output from the SVM classifier. In detail, we (1) classified between MCI and normal without ROI, (2) classified between MCI and normal with ROI, and (3) performed the early detection of AD with ROI.

3. Results

3.1. Classification of MCI and normal using cortical features without ROI

An important parameter when applying PCA is the value of intrinsic dimension. We tried integer values between 2 and 40 and used the dimension that resulted in the best SVM classifier performance. The classification performance is reported using sensitivity, specificity, and accuracy where detecting MCI is treated as finding the positive in a binary classification scheme. The best classification performance for distinguishing between MCI and normal without ROI was achieved with dimensions of 5 and 2 with adoption of cortical thickness and sulcal depth, respectively, as cortical features in Fig. 2. We chose the lowest dimension if many intrinsic

Table 1
Classifier performance to distinguish MCI and normal using various features without ROI.

Features used	Accuracy (sensitivity/specificity) [dimension used]	Accuracy (sensitivity/specificity) [dimension used] using the long feature vector
Cortical thickness	0.75 (0.77/0.73) [dim = 5]	0.68 (0.67/0.70) [dim = 81,924]
Sulcal depth	0.60 (0.63/0.57) [dim = 2]	0.52 (0.50/0.53) [dim = 81,924]
Cortical thickness + sulcal depth	0.73 (0.73/0.73) [dim = 29]	0.67 (0.67/0.67) [dim = 163,848]
MRI voxel values	0.60 (0.67/0.53) [dim = 6]	0.57 (0.57/0.57) [dim = 271,633]

dimension values led to the same classifier performance. Classification results using a single feature, be it cortical thickness or sulcal depth, and the combined feature are given in Table 1. The combined feature vector is 163,848 long with cortical thickness and sulcal depth linked into a long vector. PCA was applied to the whole vector, not separately to each feature vector. The right most column of Table 1 shows the classification performance of using the non-reduced long feature vector. With dimensionality reduction, accuracy was improved between 7% and 8%. We applied the same approach to raw MRI voxel values, similar to methods found in recent work, instead of the cortical features to provide a baseline for comparison [14,15]. We used raw MRI voxel values from the registered and skull-stripped volumes, which were by-products of the image processing software for computing cortical features. The performance of using raw MRI voxels as features was inferior to that of using cortical features as reported in the last row of Table 1.

3.2. Classification of MCI and normal using cortical features with ROI

Classification results not using the ROI were in the range of 70%. ROI was selected using discriminative power of individual vertices to improve the classification performance. Vertices whose p-values were lower than 10% of the maximum p-value from two-sample T-tests between normal and MCI groups were chosen as the ROI. The classification performance using the combined cortical feature improved to 0.90 using 3 dimensions, a significant improvement over that not using the ROI, which resulted in accuracy of 0.75. The accuracy of 0.90 to classify between MCI and normal is comparable to values in the recent literature [20].

3.3. Early detection of AD

The primary goal of this study is to test the effectiveness of using dimensionality reduced cortical features for early detection of AD. Using the combined feature with the ROI resulted in the best classification performance between MCI and normal. Thus, we applied the same methodology, i.e., PCA applied to the combined feature with ROI, for early detection of AD. The ROI was defined using the same approach described in the previous subsection. We trained the SVM classifier using dimensionality reduced features of the ROI from 30 MCI and 30 normal cases. The SVM classifier was applied to conversion cases, where only imaging data before the actual conversion were used. We achieved 83% accuracy in the early detection of AD conversion using 2 dimensions, which is comparable to the 82% accuracy reported recent literature [19].

4. Discussion

Classification performance improved as we moved from using the long non-reduced features to dimensionality reduced features for both the single feature (i.e., cortical thickness and sulcal depth) and the combined feature. This shows that most of the useful information can be well represented using few dimensions and that applying manifold learning is an effective method for extracting such information. A similar trend was observed with the ROI based cortical features. Classification performance improved as we moved from using all the cortical features within a ROI to using dimensionality reduced cortical features within a ROI. Thus far, dimensionality reduced features were used to quantify groupwise differences. As a natural progression, we then successfully extended the application of the dimensionality reduced cortical features to longitudinal data with a clinical goal to detect AD conversion early. Our accuracy of early AD detection was comparable

with that reported in the recent literature but our approach has the benefit of shape analysis in the relatively low dimension space because shape analysis in a high dimensional space is noisy and unreliable.

We considered only two features, the cortical thickness and the sulcal depth. Other features can be derived from cortical surface, such as mean curvature and surface area [10]. Applying these new features in addition to or instead of the existing two features may improve the final classifier performance. Other promising manifold learning methods such as ISOMAP and Laplacian Eigenmap may be applicable for early detection of AD [13]. Applying other manifold learning methods may result in better classifier performance, which may be investigated in future work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neulet.2013.06.042.

References

[1] J. Ashburner, K.J. Friston, Voxel-based morphometry – the methods, Neuroimage 11 (2000) 805–821.
[2] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min Knowl Discov 2 (1998) 121–167.
[3] M.C. Carrillo, L.J. Bain, G.B. Frisoni, M.W. Weiner, Worldwide Alzheimer's disease neuroimaging initiative, Alzheimers Dement 8 (2012) 337–342.
[4] R. Chaves, J. Ramírez, J.M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting, Neurosci. Lett. 461 (2009) 293–297.
[5] D. Cordes, R.R. Nandy, Estimation of the intrinsic dimensionality of fMRI data, Neuroimage 29 (2006) 145–154.
[6] R.P.W. Duin, Classifiers in almost empty spaces, in: International Conference on Pattern Recognition, 2000, pp. 1–7.
[7] R. Golland, W.E.L. Grimson, M.E. Shenton, R. Kikinis, Small sample size learning for shape analysis of anatomical structures, LNCS 1935 (2000) 72–82.
[8] M. Graña, M. Termenon, A. Savio, A. Gonzalez-Pinto, J. Echeveste, J.M. Pérez, A. Besga, Computer aided diagnosis system for Alzheimer disease using brain diffusion tensor imaging features selected by Pearson's correlation, Neurosci. Lett. 502 (2011) 225–229.
[9] K. Im, J.M. Lee, O. Lyttelton, S.H. Kim, A.C. Evans, S.I. Kim, Brain size and cortical structure in the adult human brain, Cerbral cortex 18 (2008) 2181–2191.
[10] K. Im, J.M. Lee, S.W. Seo, H.S. Kim, S.I. Kim, D.L. Na, Sulcal morphology changes and their relationship with cortical thickness and gyral white matter volume in mild cognitive impairment and Alzheimer's disease, Neuroimage 43 (2008) 103–113.
[11] S. Kakeda S, Y. Korogi, The efficacy of a voxel-based morphometry on the analysis of imaging in schizophrenia, temporal lobe epilepsy, and Alzheimer's disease/mild cognitive impairment: a review, Neuroradiology 52 (2010) 711–721.
[12] J.S. Kim, V. Singh, J.K. Lee, J. Lerch, Y. Ad-Dabbagh, D. MacDonald, J.M. Lee, S.I. Kim, A.C. Evans, Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification, Neuroimage 27 (2005) 210–221.
[13] J.A. Lee, M. Verleysen, Nonlinear Dimensionality Reduction, Springer-Verlag, 2007.
[14] M.M. López, J. Ramírez, J.M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA, Neurosci. Lett. 464 (2009) 233–238.
[15] M. López, J. Ramírez, J.M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, M. Gómez-Río, Principal component analysis-based

techniques and supervised classification schemes for the early detection of Alzheimer's disease, Neurocomputing 74 (2011) 1260–1271.

[16] H. Park, J.-J. Yang, J. Seo, J.-M. Lee, Dimensionality reduced cortical features and their use in the classification of Alzheimer's disease and mild cognitive impairment, Neurosci. Lett. 529 (2012) 123–127.

[17] J. Ramírez, J.M. Górriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. López, I. Álvarez, P. Padilla, Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification, Neurosci. Lett. 472 (2010) 99–103.

[18] S.M. Robbins, A.C. Evans, D.L. Collins, S. Whitesides, Tuning and comparing spatial normalization methods, Med. Image Anal. 8 (2004) 311–323.

[19] Y. Li, Y. Wang, G. Wu, F. Shi, L. Zhou, W. Lin, D. Shen, Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features, Neurobiol. Aging 33 (2012) 427.e15–427.e30.

[20] A.P. Zijdenbos, R. Forghani, A.C. Evans, Automatic 'pipeline' analysis of 3D MRI data for clinical trials: application to multiple sclerosis, IEEE Trans. Med. Imaging 21 (2002) 1280–1291.