Dimensionality reduced cortical features and their use in the classification of Alzheimer's disease and mild cognitive impairment

Hyunjin Park, Jin-ju Yang, Jongbum Seo, Jong-min Lee

School of Electronic Electrical Engineering, Sungkyunkwan University, Republic of Korea
Department of Biomedical Engineering, Hanyang University, Republic of Korea
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.
- The dimensionality reduced features obtained by PCA were applied to a SVM classifier.
- The classifier performance was on par or better compared to recent studies.

ABSTRACT

Features defined on the cortical surface derived from magnetic resonance imaging provide important information to distinguish normal controls from Alzheimer's disease (AD) and mild cognitive impairment (MCI). We adopted cortical thickness and sulcal depth, parameterized by three dimensional meshes, as our feature. The cortical feature is high dimensional and direct use of it is problematic in a modern classifier due to 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. A leave-one-out cross-validation was adopted for quantifying classifier performance. We chose principal component analysis (PCA) as the manifold learning method. We applied PCA to a region of interest within the cortical surface. Our classification performance was at least on par for the AD/normal and MCI/normal groups and significantly better for the AD/MCI groups compared to recent studies. Our approach was tested using 25 AD, 25 MCI, and 50 normal control patients from the OASIS database.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

A detailed scan of the brain is possible with recent advances in neuroimaging technology, especially with high fidelity magnetic resonance imaging (MRI). Alzheimer's disease (AD) is an important neurodegenerative disease with global implications. Study of shape, known as morphometry, has been successfully applied to distinguish AD MRI images from normal control MRI images [14]. A method known as voxel based morphometry (VBM) has been widely accepted [2]. VBM is typically applied to three dimensional (3D) volume data. Others have considered features defined on the brain cortex to distinguish AD patients from normal controls.

Cortical thickness, sulcal depth, surface area, and mean curvature computed on the cortical surface have been applied to distinguish AD from normal controls [12,9,14]. The cortex is a highly folded two dimensional (2D) sheet embedded in a 3D space, which may be represented using a set of 3D polygonal meshes. The number of vertices used to form the surface is in the tens of thousands. One may link the feature values, be it grayscale or not, defined on the cortical surface into a long vector and use it as one feature vector, which is a common framework adopted in machine learning/classification literature. To study N patients, one would have N feature vectors whose dimensionality is in the tens of thousands. Typically, one would have less than one hundred patients in a study. Thus, researchers are left with a situation where the number of observations (i.e., less than hundred) is far less than the dimensionality of the features, which leads to unreliable classifier performance [4,15]. This issue regarding degraded classifier performance is known as the “small sample size” problem. One way to overcome this problem is to reduce the dimensionality of
the feature vector using manifold learning methods. Analysis of shape in the reduced dimensionality space is preferable because shape analysis in a high dimensional space is noisy and unreliable. Principal component analysis (PCA), a well-known manifold learning method, was applied to high dimensional observations of cortical thickness to distinguish schizophrenic T1-weighted MRI from normal controls [17]. Others applied PCA to high dimensional observations of grayscale values for the SPECT scan of AD patients [15]. In this study, we considered two features defined on the cortex, cortical thickness and sulcal depth, and applied PCA to reduce dimensionality. Existing research using cortical thickness mainly centered on identifying regions of atrophy related to the pathology of AD using group-wise differences [14]. Few studies focused on using cortical thickness for classification tasks. Existing research using sulcal depth mainly focused on quantifying age related morphology and measuring abnormalities in cortical folding [6]. Group-wise differences of sulcal depth were assessed, but it was not for a classification task [9]. We chose cortical thickness and sulcal depth as our features since these two features have good track records for reflecting shape changes related to AD. We applied PCA to the features defined on the whole cortex and features defined on specific region of interest (ROI). ROI was automatically defined by choosing a set of vertices with high discriminative power to distinguish AD/MCI from normal controls. The reduced dimensionality features were used as an input to a support vector machine (SVM) classifier to distinguish AD and mild cognitive impairment (MCI) MRI images from normal controls. SVM is one of the state-of-the-art classifiers currently available with a proven track record. Existing research for classifying AD/MCI from normal controls adopted anatomical features like volume or shape of the hippocampus with SVM [1]. The very combination of (1) using two cortical features, (2) using PCA for manifold learning, and (3) distinguishing AD/MCI from normal control, could not be found in existing research to the best of our knowledge. We were able to show an improvement of classification performance for the following scenarios; (1) going from using all of the high dimensional features to using features with reduced dimensions, (2) going from using a single feature to using two features, and (3) going from analyzing the whole cortex to analyzing the vertices of ROI. We were able to achieve better or comparable accuracy when classifying AD/normal, MCI/normal, and AD/MCI groups.

2. Materials and methods

2.1. Subjects and MRI images

Data were obtained from the Open Access Series of Imaging Studies (OASIS) database [8]. We randomly selected 25 AD, 25 MCI, 50 normal controls from the database. We wanted to have enough samples for each patient group for statistical significance and maintain the overall composition of the OASIS database, which contains more normal cases than AD/MCI. The AD patients were aged 75.72 ± 19.48 (mean ± STD) and had a gender ratio of 18/7 (F/M). The MCI patients were aged 79.16 ± 7.20 years and of gender ratio of 17/8. The normal control patients were aged 65.68 ± 13.98 years and of gender ratio of 34/16. All MRIs were sagittal T1-weighted scans and had typical dimensions of 256 × 256 and resolutions of 1 mm × 1 mm × 1.25 mm. The scans were collected using a Siemens Vision scanner with an MP-RAGE acquisition sequence.

2.2. Cortical surface extraction and surface registration

Input images were processed using the MNI image processing software, which includes cortical surface extraction and a non-rigid registration of cortical surfaces [7]. Using an affine transformation, native MR images were first normalized into a standardized stereotaxic space and then corrected for intensity non-uniformity. The registered and corrected volumes were then classified into white and gray matter, cerebrospinal fluid, and background using a neural-net classifier. Next, the Constrained Laplacian-based Automated Segmentation with Proximities (CLASP) algorithm, which reconstructs the inner cortical surface by deforming a spherical mesh onto the white matter/gray matter boundary, was used for automatic extraction of the hemispheric surfaces of the cortex [11]. A surface model for each brain hemisphere was constructed using 81,924 high resolution polygonal 3D meshes. With the extracted surface, we applied a 2D non-rigid registration algorithm based on a geodesic distance from the gyral crown vertex with an appropriate smoothing term [16]. The registration software established spatial correspondence between a given image and a predefined template image. In the end, all 75 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. The first one was cortical thickness and the other was sulcal depth. Cortical thickness refers to the thickness of the gray matter. The CLASP algorithm results in two cortical surfaces, inner and outer surface, which are deformed from a common sphere parameterized with 3D meshes. Cortical thickness may be computed as Euclidean distance between the linked vertices of inner and outer cortical surface [13]. Sulcal depth refers to the distance between the outer cortical surface and the cerebral hull. We measured the Euclidean distance from each vertex in the cortical surface to the nearest voxel on the volume of the cerebral hull. Details of computing sulcal depth can be found [10]. We refer to the collection of features computed on the cortical surface as a feature map. Each subject’s feature maps were blurred using a 20 mm full width half maximum (FWHM) surface-based diffusion kernel to improve the ability for detecting group-wise differences. The computed feature maps were resampled onto the spatial frame of the common template using the results of the surface registration described before. This process ensured that all the feature maps reside on the same spatial frame for vertex by vertex comparison. In sum, we obtained two features (i.e., cortical thickness and sulcal depth) defined on the common cortical surface for all 81,924 vertices with regards to each subject.

2.4. PCA and dimensionality reduction

A feature map obtained from the previous step may be linked to a long vector whose length is 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. A major goal in this study is to use the extracted cortical features for a modern classifier. Here, we are burdened with a "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 a cortical surface). One way to overcome this problem is to apply manifold learning and reduce the dimensionality of the feature vector. We adopted PCA among many well-known manifold learning methods. Manifold learning assumes that a high dimensional observation actually lies within a low dimensional manifold. A major parameter when applying manifold learning is the value of the reduced dimension, also called the intrinsic dimension. There are many algorithms including the maximum likelihood estimation and packing numbers algorithms to estimate the intrinsic dimension but their application in neuroimaging has been met with limited success [3]. Here, we searched exhaustively to find the
value of the reduced dimension, which resulted in the best classifier performance.

2.5. Automatic ROI selection

A feature map has values defined on all vertices on the cortical surface. Not all vertices on the cortical surface have equal statistical capability to distinguish between AD/MCI images from normal controls [14,15]. Some cortical regions including the parahippocampal region were shown to be more informative in distinguishing AD/MCI images from normal controls [5]. Considering the features from a ROI eliminates many non-informative vertex values and makes the overall feature vector shorter. The observed feature vector is easier to handle since it is shorter and less noisy as many non-informative values are removed. Manual specification of ROI is certainly feasible. A better approach is to specify a ROI using the statistical discriminative power of the individual vertex. A two-sample *T*-test was performed on a vertex by vertex basis on the set of feature maps from two subject groups being compared (e.g., AD and normal) and the vertices were sorted from small *p*-values to large *p*-values. We assigned only vertices with whose *p*-values fell into lower 10% of maximum *p*-value as the ROI. Vertices included in the ROI were scattered around the cortical surface but many of them were in the hippocampus region.

2.6. SVM classifier

The SVM classifier is one of the state-of-the-art classifiers currently available benefiting from recent advances in statistical machine learning. The goal of this study is applying dimensionality reduced features for classifying AD/MCI from normal controls. We refer the readers concerned with the technical details of SVM to consult many textbooks available. The SVM classifier takes an input feature vector and outputs a binary decision. We adopted the SVM classifier with a linear kernel. The input feature vector was the dimensionality reduced feature vector derived from the whole cortical surface or ROI. The SVM classifier belongs to a supervised category of classifiers where the user needs to provide separate training and test data. With limited subjects available, we adopted the leave-one-out cross validation method for separating training and test data. For example, given 50 MCI and 50 normal cases, we assigned one case as the test case and used the remaining 99 cases as the training data for the SVM classifier. The process was repeated 100 times choosing a different test case each time.

2.7. Experimental setup

We formed three groups to apply the procedures of this study. We considered (1) normal control and AD, (2) normal control and MCI, and (3) AD and MCI. We computed the cortical surface and the two related feature maps, cortical thickness and sulcal depth, for all subjects in each group. Each subject’s feature maps were linked to a long feature vector. The collection of feature vectors was separated into training and test data using the leave-one-out cross validation method. Computation of ROI for each group using only the training data was performed as necessary. 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.

3. Results

3.1. Intrinsic dimension of cortical features

An important parameter when applying PCA is the value of reduced dimension, known as the intrinsic dimension. We tried integer values between 2 and 30 and used the dimension that resulted in the best SVM classifier performance. The best classification performance to classify between AD and normal control was achieved with a dimension of 3 and 9 for adopting cortical thickness and sulcal depth, respectively as cortical features in Fig. 1. Note that dimensionality was reduced from 81,924 to a value of less than ten in these cases.

3.2. Using single feature and combined feature

Classification results using a single feature are reported in Table 1. In general, classification performance to distinguish

<table>
<thead>
<tr>
<th>Groups considered</th>
<th>Features used</th>
<th>Accuracy</th>
<th>Dimension of reduced feature</th>
<th>Accuracy using the long feature vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD/norm</td>
<td>Cortical thickness</td>
<td>0.80</td>
<td>3 (Fig. 1 left)</td>
<td>0.72 (dim = 81,924)</td>
</tr>
<tr>
<td>AD/norm</td>
<td>Sulcal depth</td>
<td>0.80</td>
<td>9 (Fig. 1 right)</td>
<td>0.76</td>
</tr>
<tr>
<td>MCI/norm</td>
<td>Cortical thickness</td>
<td>0.79</td>
<td>4</td>
<td>0.76</td>
</tr>
<tr>
<td>MCI/norm</td>
<td>Sulcal depth</td>
<td>0.75</td>
<td>10</td>
<td>0.69</td>
</tr>
<tr>
<td>AD/MCI</td>
<td>Cortical thickness</td>
<td>0.67</td>
<td>4</td>
<td>0.55</td>
</tr>
<tr>
<td>AD/MCI</td>
<td>Sulcal depth</td>
<td>0.69</td>
<td>3</td>
<td>0.65</td>
</tr>
</tbody>
</table>
between AD/normal was better than that between AD/MCI as expected. Using cortical thickness and sulcal depth did not lead to significant differences in classification performance. The rightmost column of Table 1 shows classification performance of using the non-reduced long feature vector. With dimensionality reduction, we gained significant improvement in accuracy.

Classification results using the combined feature are given in Table 2. 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 rightmost column of Table 2 shows classification performance for using the non-reduced feature vector. With dimensionality reduction, we gained significant improvement in accuracy. The 4th column of Table 2 shows the best accuracy achieved by using a single feature from Table 1 as reference values for comparison. There was a significant gain in performance for distinguishing AD/normal using the combined feature compared to using the single feature. For other groups, differences were not observed.

3.3. Using single feature and combined feature with ROI

Classification results using a single feature with ROI are given in the rows between 2 and 7 within Table 3. Vertices with whose p-values fell into lower 10% of maximum p-value from two-sample T-tests were chosen as the ROI. Classification performance for all three groups (i.e., AD/normal, MCI/normal, and AD/MCI) improved significantly compared to using the whole cortical surface (i.e., not using the ROI). Using cortical thickness fared better than using sulcal depth in general except for distinguishing between AD/MCI. The rightmost column of Table 3, in the rows between 2 and 7, shows the classification performance of using a single feature without ROI from Table 1 as a reference. We achieved an accuracy of 0.94 for the AD/normal group, which is comparable to recent literature [14]. We also achieved accuracy of 0.91 for the AD/MCI group, which is a significant improvement over existing research [5]. Classification results using the combined feature with ROI are given in Table 3, in the rows between 8 and 10. Classification performance for all three groups (i.e., AD/normal, MCI/normal, and AD/MCI) improved significantly compared to using the whole cortical surface. The rightmost column of Table 3, in the rows between 8 and 10, shows the classification performance of using the combined feature without ROI from Table 2 as a reference.

To summarize, the best accuracy achieved for the AD/normal group was 0.94 using cortical thickness with ROI. The best accuracy for the MCI/normal and AD/MCI groups were 0.86 using the combined feature with ROI and 0.91 using sulcal depth with ROI, respectively. Our classification results were at least on par for the AD/normal and the MCI/normal groups and significantly better for the AD/MCI group compared to recent studies [5,14].

4. Discussion

Classification performance using all the cortical features, be it cortical thickness, sulcal depth, or the combined feature, was inferior to that of using dimensionality reduced cortical features. This confirms the notion that all feature values are not equally important and some vertices contain more important features than other vertices. We applied a manifold learning, PCA, to extract such important information from a set of numerous vertices. We observed a similar trend with the ROI based cortical features. Classification performance improved as we move from using all the cortical features within a ROI to using dimensionality reduced cortical features within a ROI. Overall, manifold learning was able to extract more important features with lower dimensions from very high dimensional cortical features, which was verified with classifier performance.

Many studies have explored the usage of cortical thickness. However, sulcal depth has not received such attention. This study explored the usage of sulcal depth defined on the cortical surface as the feature used for the classification of the three groups. Classifier performance using sulcal depth for classifying the AD/MCI group was significantly better than the existing research, which used either cortical thickness or the tissue probability map of cortical structures [5,14]. Existing research reported accuracy values in the 70% for classifying AD and MCI. Sulcal depth perhaps was ignored since it was poor at classifying the AD/normal group, which is the easiest classification task. Sulcal depth may contain important information to distinguish AD and MCI and we believe a more thorough investigation is necessary in the future.

### Table 2

<table>
<thead>
<tr>
<th>Groups considered</th>
<th>Features used</th>
<th>Accuracy (dim used)</th>
<th>Best accuracy using a single feature, Table 1</th>
<th>Accuracy using the long feature vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD/normal</td>
<td>Cortical thickness + sulcal depth</td>
<td>0.85 (dim = 23)</td>
<td>0.80 (dim = 3)</td>
<td>0.77 (dim = 163,848)</td>
</tr>
<tr>
<td>MCI/normal</td>
<td>Cortical thickness + sulcal depth</td>
<td>0.79 (dim = 12)</td>
<td>0.79 (dim = 14)</td>
<td>0.69</td>
</tr>
<tr>
<td>AD/MCI</td>
<td>Cortical thickness + sulcal depth</td>
<td>0.69 (dim = 3)</td>
<td>0.69 (dim = 3)</td>
<td>0.63</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Groups considered</th>
<th>Features used</th>
<th>Accuracy (dim used)</th>
<th>Accuracy achieved without ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD/Normal</td>
<td>Cortical thickness</td>
<td>0.94 (dim = 17)</td>
<td>0.80 (dim = 3)</td>
</tr>
<tr>
<td>AD/Normal</td>
<td>Sulcal depth</td>
<td>0.89 (dim = 6)</td>
<td>0.80 (dim = 9)</td>
</tr>
<tr>
<td>MCI/Normal</td>
<td>Cortical thickness</td>
<td>0.80 (dim = 2)</td>
<td>0.79 (dim = 4)</td>
</tr>
<tr>
<td>MCI/Normal</td>
<td>Sulcal depth</td>
<td>0.85 (dim = 13)</td>
<td>0.75 (dim = 10)</td>
</tr>
<tr>
<td>AD/MCI</td>
<td>Cortical thickness</td>
<td>0.72 (dim = 4)</td>
<td>0.67 (dim = 4)</td>
</tr>
<tr>
<td>AD/MCI</td>
<td>Sulcal depth</td>
<td>0.91 (dim = 22)</td>
<td>0.69 (dim = 3)</td>
</tr>
<tr>
<td>AD/Normal + Sulcal depth</td>
<td>0.90 (dim = 26)</td>
<td>0.85 (dim = 23)</td>
<td></td>
</tr>
<tr>
<td>MCI/Normal + Sulcal depth</td>
<td>0.86 (dim = 12)</td>
<td>0.79 (dim = 12)</td>
<td></td>
</tr>
<tr>
<td>AD/MCI + Sulcal depth</td>
<td>0.90 (dim = 22)</td>
<td>0.69 (dim = 3)</td>
<td></td>
</tr>
</tbody>
</table>
We applied PCA as our choice regarding the manifold learning method. However, there are other promising manifold learning methods such as ISOMAP and Laplacian Eigenmap. Applying other manifold learning methods may result in better classifier performance, which is left for future work. We expect the general trend of improved classifier performance with the use of more relevant information to continue for other manifold learning methods as well. In this study, we considered two features, cortical thickness and sulcal depth. There are other features derived from the cortical surface available such as mean curvature and surface area [12,9]. Applying these new features in addition or instead of the existing two features may improve the final classifier performance.

Acknowledgements

This study was supported by Basic Science Research Program through NRF Korea grants 2012005939, 20100023233, Global Frontier RD Program through NRF Korea grant NRF-M1AXA003-2011-0032035, and KOSEF NLRL Program grant 2011-0028333. Image data collection was supported by NIH grants P50AG05681, P01AG03991, R01AG021910, P50MH071616, U24RR021382, and R01MH56584.

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.2012.09.011.

References