Regional MRI Measures and Neuropsychological Test for Multi-dimensional Analysis in Alzheimer’s Disease


Abstract—Regional brain atrophy is a typical structural symptom of Alzheimer’s disease (AD). Magnetic resonance imaging (MRI) scans capture brain structure with high resolution and are often processed with automated segmentation and parcellation algorithm (e.g. Freesurfer) to generate regional measures, like cortical volume, cortical thickness and surface area, which are widely used as inputs in classification algorithms. This study aims to find out which combination of MRI measures and neuropsychological test coupled with different normalization techniques can best predict AD using a proposed multivariate feature selection and classification method. A total of 189 subjects with 60 Alzheimer’s disease (AD) and 129 cognitively normal (CN) are included in this study. Freesurfer was used to obtain 34 cortical thickness measures and 35 surface area measures for each hemisphere and 55 regional volumes across the brain. Statistically significant variables selected for each model were used to construct a support vector machine (SVM) classifier. Different normalization approaches were explored to gauge if the classification performance could be improved. Results indicate neuropsychological test score contains the most discriminative information among single measure models, and out of the three MRI measures, cortical volume is a better predictor than the other two. Normalization approaches are seen not to enhance the performance much if any. Hierarchical model of neuropsychological test and cortical volumes without normalization yield the best classification accuracy for this study.

I. INTRODUCTION

Alzheimer’s disease (AD) is a neurodegenerative disease, and is the most common form of dementia. Estimates from the Alzheimer Association as of March 2012 indicate that 5.4 million Americans are diagnosed with AD, and over 95% of this population are 65 years of age or older. Also, nearly half of the population over 85 years of age is affected by AD [1]. AD patients suffer from regional cerebral atrophy, which can be distinguished from normal aging [2, 3]. In AD atrophy is often observed in regions which are closely related to neurodegeneration. It has been shown that atrophy in regions like the hippocampus [4, 5] and amygdala[4] can serve as a predictor of disease progression. Moreover, determination of the key atrophied regions across the entire brain could be used as parameters for the delineation of AD patients from cognitively normal subjects (CN).

Freesurfer is a popular brain imaging software widely used to generate regional volumetric measures from MRI scans. The advantages of Freesurfer over traditional manual segmentations and measures are its high automation and independence from operator subjectivity. Freesurfer is also accurate, precise and has been tested on a large cohort of studies in AD classification research [6, 7].

Important tasks to be considered in AD classification studies include the choice of parameters, the way these parameters ought to be combined, and determining the pre-processing techniques to be employed in order to enhance the prospects of classification. Two essential questions that need to be addressed for AD classification studies are (1) what is the statistical meaningfulness of the measures produced by Freesurfer for predicting AD?; (2) Which normalization approach should be used to potentially minimize the bias incurred by factors like difference in head size and brain structure to enhance classification performance? Westman and his colleagues have investigated this issue using a supervised multivariate data analysis - orthogonal projections to latent structures (OPLS) model [8]. OPLS is similar to Principal Component Analysis (PCA) as they are both linear decomposition techniques and project the original data to the found latent variables.

The current study is based on the method of previous study [9], proposes to construct for each classification model a decisional space using the most statistically significant variables, where the number of dimensions is determined by an incremental error analysis defining those statistically significant variables to be used in an SVM-based classification process. In this study, single measure models and hierarchical models with and without normalization are both examined to find the optimal model. Single measure models include one of the regional MRI measures (cortical volume, cortical thickness and surface area) or the neuropsychological test, the Mini-Mental State Examination (MMSE). MMSE is a neuropsychological test that is most often administered to screen patients for cognitive impairment and dementia. A hierarchical model combines two or more of the single-measure models to examine if the interaction is beneficial towards the classification. The specific aims of this study are thus to determine: (1) the impact of neuropsychological test (MMSE) towards the classification; (2) the combination of regional measures and MMSE that yields the best classification performance; and (3) which normalization scheme should be employed to achieve better classification performance.

II. MATERIALS AND METHODS

2.1 Subjects

A total of 189 subjects are included in this study as shown in Table 1. All participants are from the Wien Center for Alzheimer’s Disease and Memory Disorders, Mount Sinai
Medical Center, Miami Beach, FL, USA. All subjects have taken the Folstein Mini-Mental State Examination[10] with a minimum score of 15. The study was approved by the Mount Sinai Medical Center Institutional Review Board with informed consent provided by the subjects or legal representatives.

Table 1: Demographic and neuropsychological characteristics of subjects

<table>
<thead>
<tr>
<th>Diagnostic</th>
<th>Subjects</th>
<th>Age</th>
<th>Female/Male</th>
<th>MMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN a</td>
<td>129</td>
<td>72.9 ± 6.4</td>
<td>92 / 37</td>
<td>28.7 ± 1.4</td>
</tr>
<tr>
<td>AD b</td>
<td>60</td>
<td>79.5 ± 6.9</td>
<td>34 / 26</td>
<td>22.6 ± 3.4</td>
</tr>
</tbody>
</table>

Data Presented as mean ± S.D. where applicable

CN: cognitively normal
AD: Alzheimer’s disease
MMSE: Mini Mental State Examination

All subjects had: (1) a neurological and medical evaluation by a physician; (2) a full battery of neuropsychological tests [11], according to the National Alzheimer’s Coordinating Center protocol, and the following additional tests: the Three-Trial Fold Object Memory Evaluation [12] and the Hopkins Verbal Learning Test; as well as (3) a structural volumetrically acquired MRI scan of the brain. The sum of boxes from the Clinical Dementia Rating Scale (CDR-sb) was used as the index of functional ability, and the MMSE was used as the index of cognitive ability. The cognitive diagnosis was made using a combination of the physician’s diagnosis and neuropsychological diagnosis. The etiological diagnosis was made by the examining physician. The diagnosis of cognitively normal controls (CN) required that the physician’s diagnosis was CN and no cognitive test scores were ≥1.5 SD below age- and education-corrected means. A probable AD diagnosis required a dementia syndrome and the National Institute of Neurological Disorders and Stroke (NINDS)/Alzheimer’s Association criteria for AD [13].

2.2 Imaging Protocol

MRI scans were acquired on a 1.5-T machine (Siemens Symphony, Iselin, N.J., USA, General Electric, HDX, Milwaukee, Wisc., USA) using a proprietary 3D-magnetization-prepared rapid-acquisition gradient echo (3D MPRAGE) or 3D spoiled gradient echo sequences (FSPGR). Specifications for 3D MPRAGE include coronal sections with a 1.5-mm gap in thickness; section interval, 0.75 mm; TR, 2190 ms; TE, 4.38 ms; TI, 1100 ms; FA, 15°; NEX, 1; matrix, 256 × 256; FOV, 260 mm; bandwidth, 130 Hz/pixel; acquisition time, 9 minutes; phase-encoding direction, right to left. Specifications for 3D FSPGR were the following: 140 contiguous coronal sections of 1.2-mm thickness; contiguous images with no section interval; TR, 7.8 ms; TE, 3.0 ms; inversion recovery preparation time, 450 ms; flip angle, 12°; NEX, 1; matrix, 256 × 256; FOV, 240 mm; bandwidth, 31.25 Hz/pixel; acquisition time, 6–7 minutes; phase-encoding direction, right to left.

2.3 Regional MRI Measures

Freesurfer pipeline (version 5.1.0) is applied to all the MRI scans producing 55 volumetric variables, which include 45 volumetric measures of white matter parcellation and 10 morphometric statistics. For cortical thickness, 34 variables were determined for each hemisphere, resulting in 68 variables for cortical thickness measures. And there are 35 segmentations for the surface area of each hemisphere, 70 measures for the total brain. The challenge in this study was to come up with a novel feature extraction and selection scheme that will determine that optimal combination of these many variables that will constitute a multidimensional space that is most suitable statistically and spatially for the classification process.

2.4 Feature Extraction and Incremental Error Analysis

A feature extraction method based on statistical testing is proposed to find an optimal set of variables within a model that yields the best classification performance. The different models considered in this study are listed in Tables 2 and 3. All the variables in a model are ranked based on statistical significance between AD and CN. Following this ranking, an incremental error analysis is used whereby the SVM classifier is trained and tested adding a single variable at a time to the classifier to determine the combination of top-ranked variables that yield best classification outcome. This rigorous blind feature selection technique differs from others as it does not rely on prior assumptions of regions of interest (ROI) and assigns equal weights to all the variables. The above process was done for each single measure and hierarchical model to compare the discriminative power of all the models and find good models in terms of classification performance.

It should also be noted that, even though atrophy of regions among AD patients is what is generally sought, the statistical test considers both cases of atrophy and enlargement of these specific brain regions, since volumetric enlargement in ventricles is also shown to be important in differentiating AD and its prodromal stages [14].

2.5 Classification Experiment

Classification in this study was performed using a support vector machine (SVM), which is widely used in classification problems of all types, linear and nonlinear, and is shown to be effective as a classification tool for AD [15, 16]. The kernel function of SVM used in this study is Gaussian Radial Basis Function kernel (rbf) with a scaling factor (sigma) of 3. All the classification results reported are based on a 5-fold cross-validation process. Each classification experiment was run 50 times, the results of which are averaged to evaluate the performance in terms of accuracy, specificity, sensitivity and precision.

III. RESULTS

3.1 Classification Performance of Models

The results of single-measure models using one of the regional MRI measures (cortical volume, cortical thickness and surface area) or neuropsychological test (MMSE) and hierarchical models combining two or more of the single measure models are shown in Table 2. The results for the same models where the data is normalized by an appropriate normalizing factor are shown in Table 3. The results display an average of 50 runs with minimum and maximum values shown in parentheses.

IV. DISCUSSION

4.1 Classification Performance and Model Selection

Results of the different models are highly consistent as results of the 50 independent repetitions of classification fall within a small range as shown by the minimum and maximum values in Tables 2 and 3. This small range is a clear indication of the replicability and reproducibility of results, both essential attributes in any classification process. Results also
indicate that MMSE is an important factor that should be included in the classification process. Inclusion of MMSE with other measures improves the classification results. For example in case of the best model (Hierarchical Model using MMSE + CV) the inclusion of MMSE with the CV results in an improvement of 9.2% as compared to using CV alone. In general, on an average, an improvement of 13.3% is seen on comparing analogous models with and without MMSE on using raw data and 12.8% on using normalized data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSE</td>
<td>88.3</td>
<td>81.0</td>
<td>91.6</td>
<td>82.6</td>
</tr>
<tr>
<td>Cortical volume (CV)</td>
<td>83.1</td>
<td>77.9</td>
<td>85.6</td>
<td>72.6</td>
</tr>
<tr>
<td>Cortical thickness (CT)</td>
<td>77.7</td>
<td>74.8</td>
<td>79.0</td>
<td>63.0</td>
</tr>
<tr>
<td>Surface area (SA)</td>
<td>71.4</td>
<td>58.7</td>
<td>77.2</td>
<td>55.0</td>
</tr>
<tr>
<td>Average</td>
<td>80.1</td>
<td>73.1</td>
<td>83.4</td>
<td>68.3</td>
</tr>
</tbody>
</table>

**Scaled by the total area of the all the measures**

**Scaled by the mean thickness of the all the thickness measures**

*The results of these models are the same as model of ‘CV’ since the variables extracted for the decisional space are the same as that for ‘CV’ model*

**This model gives identical results as the model of ‘MMSE+CV’ since variables extracted for the decisional space are the same as that for ‘MMSE+CV’**

The results show that using the data prepossessing techniques as proposed in this study, cortical thickness should be normalized by either the mean thickness of all the measured regions or the intracranial volume (ICV) as it increases the accuracies a little, even though it doesn’t have a significant enhancement of performance on most of the models.

Since some models have very close performance in terms of the 4 recorded performance metrics (accuracy, sensitivity, specificity and precision), models that give more than 90% accuracy are considered as good models and are highlighted in gray in Tables 2 and 3. A tradeoff exists between models with some displaying better accuracy at the cost of sensitivity and vice versa. In terms of accuracy, model of ‘MMSE + cortical volume’ (MMSE + CV) is the best; whereas in terms of sensitivity, model of ‘MMSE + cortical thickness’ normalized by mean cortical thickness provides the best classification.

Table 2 and 3 also show that for single models, normalization does enhance classification performance to different degrees. However, for single models of ‘cortical volume’ and ‘surface area’, sensitivity drops by 3.5% and 16.1 % respectively when normalization is performed. On the other hand, for hierarchical models, normalization increases sensitivity by 1.21% on average but has a negative impact on the other three performances.

Table 4 compares the results of the proposed study with other methods in literature, which shows that the proposed technique which uses only MMSE and MRI can yield competitive classification performance as those using two or more imaging modalities or biomarkers.

### 4.2 Representing Classification Data using ‘Best Model’

Model of “MMSE + cortical volume” without normalization gives the highest classification accuracy which utilizes the top 3 variables found within the model (i.e. MMSE, right-hippocampus and left-inferior-lateral-ventricle). One typical distribution of the data points for this classification model is plotted in Fig. 1 to show the clustering
characteristics of the data when MMSE and cortical volumes are employed. Using this optimal decisional space, it can be observed that all the normal subjects are grouped into a very compact cluster, whereas AD subjects are more sparsely distributed in context of these dimensional parameters.

4.3 Model Efficiency and Normalization

Variation in measures can come from many sources, including variation due to AD atrophy ($\sigma^2_{ADa}$), which is of primary interest for this study, as well as other variation noise ($\sigma^2_0$) like individual difference in brain size, structure of brain regions, MRI measurement error, region segmentation error, atrophy due to normal aging and resistance to brain atrophy (e.g., cognitive reserve). Generally, the total variance can be described

$$\sigma^2_{total} = \sigma^2_{ADa} + \sigma^2_0$$  \hspace{1cm} (1)

where $\sigma^2_{total}$ is the total variance of dataset, $\sigma^2_{ADa}$ stands for variance due to AD atrophy and $\sigma^2_0$ is the variance due to what is termed here as an overall source of noise. Also, discriminative power of a model depends on the amount of variance due to AD atrophy captured by it in contrast to the variance due to noise. A relevant term called discriminative power ($Dp$) can be estimated using Eq. 2.

$$Dp = \frac{\sigma^2_{ADa}}{\sigma^2_0}$$  \hspace{1cm} (2)

where $\sigma^2_{ADa}$ is an estimate of the variance due to AD atrophy captured by the model and $\sigma^2_0$ stands for the estimated variance due to noise captured by the model.

Our results show that normalization doesn’t enhance much the classification performance, which could be explained by above formula that normalization does bring down correlated noise ($\sigma^2_0$) incurred by e.g. brain size difference but it also lowers the correlated variance due to AD atrophy ($\sigma^2_{ADa}$). The $Dp$ value could serve as a measure of a model’s performance if relevant sources of the variance are known and are quantifiable, which is not the case in most practical scenarios.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation under grants CNS-0959985, CNS-1042341, HRD-0833093, and IIP-1230661.

REFERENCES