Improved multimodal biomarkers for Alzheimer’s disease and Mild Cognitive Impairment diagnosis - data from ADNI

Antonio Martinez-Torteya *a, Víctor Treviño-Alvaradoa, José Tamez-Peñaa, for the Alzheimer’s Disease Neuroimaging Initiative+

aCátedra de Bioinformática, Tecnológico de Monterrey, Monterrey, Nuevo León, México

ABSTRACT

The accurate diagnosis of Alzheimer’s disease (AD) and mild cognitive impairment (MCI) confers many clinical research and patient care benefits. Studies have shown that multimodal biomarkers provide better diagnosis accuracy of AD and MCI than unimodal biomarkers, but their construction has been based on traditional statistical approaches. The objective of this work was the creation of accurate AD and MCI diagnostic multimodal biomarkers using advanced bioinformatics tools. The biomarkers were created by exploring multimodal combinations of features using machine learning techniques. Data was obtained from the ADNI database. The baseline information (e.g. MRI analyses, PET analyses and laboratory essays) from AD, MCI and healthy control (HC) subjects with available diagnosis up to June 2012 was mined for case/controls candidates. The data mining yielded 47 HC, 83 MCI and 43 AD subjects for biomarker creation. Each subject was characterized by at least 980 ADNI features. A genetic algorithm feature selection strategy was used to obtain compact and accurate cross-validated nearest centroid biomarkers. The biomarkers achieved training classification accuracies of 0.983, 0.871 and 0.917 for HC vs. AD, HC vs. MCI and MCI vs. AD respectively. The constructed biomarkers were relatively compact: from 5 to 11 features. Those multimodal biomarkers included several widely accepted univariate biomarkers and novel image and biochemical features. Multimodal biomarkers constructed from previously and non-previously AD associated features showed improved diagnostic performance when compared to those based solely on previously AD associated features.

Keywords: multimodal biomarker, Alzheimer’s disease, mild cognitive impairment, ADNI

1. INTRODUCTION

Alzheimer’s disease (AD) is the most common form of dementia, affecting more than five million people around the world. Its hallmark abnormalities are deposits of the protein fragment beta-amyloid and twisted strands of the protein tau. An established risk factor for AD is mild cognitive impairment (MCI), a condition in which a person has problems with language, memory, or another cognitive ability. MCI represents in some cases a transitional stage between normal aging and AD [1].

An early accurate detection of AD and MCI confers many benefits: allows prompt evaluation and treatment of reversible or treatable causes, allows potential management of symptoms with medication, enables potential inclusion in clinical trials, and allows physicians and caregivers to be aware of patients who may soon have difficulties in managing their own health, therefore allowing the affected person to plan ahead while they still have the capacity to make important decisions about their future care. The most used criteria for the clinical diagnosis of AD was defined almost 30 years ago and has been reported to be inaccurate even among experienced investigators in up to 20% of cases [2]. Because of this, there is a pressing need for imaging and biological biomarkers to improve the accuracy of diagnosis and to assess the efficacy of potential treatments [3].

Several biomarkers from distinct information modalities have been proposed for the diagnosis and prediction of AD and MCI: biological samples [2, 4-8], anatomical Magnetic Resonance Imaging (MRI) [9-11], perfusion MRI [12], Positron Emission Tomography (PET) [13-15], and family medical history [16]. Multimodal biomarkers have shown to improve the accuracy of AD and MCI diagnosis, and might also serve as indirect measures of disease severity [17-19]. However,
the modalities of information and features being used to construct such biomarkers have been reduced to only previously AD and MCI associated measurements, such as: brain regions volumes (obtained through MRI analysis), cerebral metabolic rate for glucose (CMRgl) in several brain regions (obtained through PET analysis) and cerebrospinal fluid proteins. This study extends these efforts by exploring other imaging information (i.e. cortical thicknesses and hypometabolic convergence index), and information from other modalities (e.g. genotyping and medical history).

The objective of this work was to determine whether by using a specific multivariate feature selection strategy within a multimodal database, including previously and non-previously AD associated variables, classification accuracy of subjects with AD, MCI and HC could be improved, and if novel AD and MCI associated features could be found. Biomarkers that have already been proven to be associated to MCI and AD were used to validate the methodology and the proposed biomarkers.

2. METHODS

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Available information from laboratory essays, genotyping, MRI and PET analyses, medical and family history, physical exams, and demographics, up to June 2012 was acquired. Neuropsychological information from questionnaires was disregarded because diagnosis, the variable of interest in this study, was based on some of these. The MRI analyses from which information was acquired were: the Stroke Summary analysis, reporting the number and location of strokes, and the white matter hyperintensity volume of the whole brain; the UA Gene Alexander Lab statistic parametric mapping voxel based morphometry analysis, reporting the mean gray matter value from 90 regions of interest (ROI); the UCSD Ander Dale Lab derived volumes analysis, reporting the volumes of 15 ROI; and the UCSF FreeSurfer analysis, reporting the volume, surface area, and cortical thickness of 139 ROI. The PET analyses from which information was obtained were: the Banner Alzheimer’s Institute NMRC summaries analysis, reporting the globally normalized cerebral metabolic rate for glucose (CMRgl) in 70 ROI; the UCB Jagust Lab PET ROI analysis, reporting the mean, median, mode, minimum, maximum, and standard deviation of FDG-PET from 5 different ROI; the UU PET analysis, reporting the average CMRgl normalized to the pons in 3 ROI, and the number of pixels with hypometabolic activity that are two and three standard deviations below normal mean; and the NYU FDG-PET hippocampus analysis, reporting the mean FDG-PET of the right and left hippocampus, normalized to the pons.

Only subjects with an AD, MCI, or HC diagnosis, as defined by ADNI, were taken into account for this study. Additionally, subjects without apolipoprotein E (ApoE), cerebrospinal fluids (CSF), hippocampal volume, or CMRgl information, widely accepted AD and MCI biomarkers [20]; or without, TOMM40 gene information, a novel potential genetic biomarker [21], were excluded. This resulted in a study group of 47 HC, 83 MCI and 43 AD subjects. The demographics of this study group are shown in Table I. Using this subjects, 3 binary classification subsets were made: HC vs. AD, HC vs. MCI, and MCI vs. AD.
For each subset, variables with missing data in more than 20% of the subjects were excluded, resulting in subsets with 983, 1,029, and 1,124 variables. All variables were z-standardized, as defined by:

\[ z_{ij} = \frac{x_{ij} - \hat{\mu}_j}{\hat{\sigma}_j} \]  

where \( z_{ij} \) and \( x_{ij} \) are the standardized score and the raw measurement of the \( i^{th} \) subject for the \( j^{th} \) variable, and \( \hat{\mu}_j \) and \( \hat{\sigma}_j \) are the mean and the standard deviation of the entire ADNI population for the \( j^{th} \) variable.

The three subsets were analyzed, using the same methodology, to obtain an accurate and compact set of classifying features. For the HC vs. AD subset, two thirds of the subjects of each class were randomly selected to generate and train the biomarker, and the rest were set apart to use as a blind test. For the other two subsets, two thirds of the subjects from the class with the least subjects, and an equal number of subjects from the other class, were selected for training. This was done because the feature selection methodology required a balance between classes. Galgo [22], a powerful multivariate feature selection strategy based on genetic algorithms, was used to generate 300 five-variable nearest centroid classifiers. Each one of these models evolved from a set of random models throughout 200 generations, where the fitness of the model, defined as the accuracy using a k-folds strategy, was optimized.

Nearest centroid classifiers were used because of their simplicity and their ability to work with datasets having missing data. A centroid is defined as the mean of a given set of samples of the same class. For every fold (five for the HC vs. AD and the MCI vs. AD subsets, and four for the HC vs. MCI subset), \( n - n/k \) subjects were used to estimate a centroid for each class and each feature of the model, where \( n \) represents the number of subjects in the subset, and \( k \), the number of folds. The remaining \( n/k \) subjects were predicted to pertain to the class whose centroid had the minimal Euclidean distance to the subject, as defined by:

\[ class_i = \min_k \left( \sqrt{\sum_j (z_{ij} - c_{kj} - th)^2} \right) \]  

where \( class_i \) is the predicted class of the \( i^{th} \) subject, \( z_{ij} \) is the standard score of the \( i^{th} \) subject for the \( j^{th} \) variable, \( c_{kj} \) is the centroid of the class \( k \) for the \( j^{th} \) feature, and \( th \) is a threshold defined as zero in this case.

After the 300 models were generated, the variables were ranked according to their frequency, and the ranking was used in a derive mode forward selection (DMFS) strategy. Auxiliary models were generated, one with the top 2 ranked variables, one with the top 3, and so on, up to a model with the top 30 ranked variables. The fitness of the auxiliary models was then evaluated, from the one with the least variables to the one with the most. If the additional variable between two auxiliary models resulted in an increased fitness, that variable was selected to be part of the DMFS model. The auxiliary models in the DMFS methodology were evaluated as follows: the receiver operating characteristic (ROC) was obtained for every fold; an optimal cutoff value was then generated, consisting of the mean value of the overlapping best cutoff ranges of each fold (the ones resulting in the largest accuracy); if there was no overlapping, the mean of the median bottom and median top cutoff values were selected; the class of each validation subject in each fold was assessed as described by the nearest centroid equation (2), defining \( th \) as the optimal cutoff value; fitness was finally defined as the mean accuracy using these predictions.

The final biomarker was obtained after employing a variable-elimination strategy, intended to avoid redundant information, in which the fitness of the DMFS model, evaluated as in the DMFS strategy, was assessed after removing all the variables of the model, one at a time. For each cycle, the variable whose elimination from the model resulted in the largest fitness was permanently removed. This process was carried on until no variable removal resulted in an equal

---

### Table I. Demographics of the study group.

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
<th>Age</th>
<th>apoE4+</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC</td>
<td>30</td>
<td>17</td>
<td>74.4 (±5.2)</td>
<td>29.8%</td>
</tr>
<tr>
<td>MCI</td>
<td>54</td>
<td>29</td>
<td>74 (±7.1)</td>
<td>53.0%</td>
</tr>
<tr>
<td>AD</td>
<td>27</td>
<td>16</td>
<td>74.3 (±7.4)</td>
<td>79.1%</td>
</tr>
<tr>
<td>Age</td>
<td>75.1 (±6.5)</td>
<td>72.7 (±7.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>apoE4+</td>
<td>51.4%</td>
<td>56.5%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
or better fitness than the parent model. The resulting biomarker was then evaluated for its sensitivity, specificity, accuracy and AUROC, using the DMFS methodology, and the binomial proportion confidence intervals were also calculated. Sensitivity for the HC vs. AD and the MCI vs. AD subsets refers to the ratio of accurately predicted AD subjects to the total diagnosed AD subjects, and similarly for the HC vs. MCI subset, substituting AD with MCI. Subjects left apart for a blind test were evaluated in a similar way, but computing the centroids using all the train subjects at once and using the optimal cutoff value obtained when evaluating the train subjects. Figure 1 shows the employed methodology, from the data acquisition to the generation of results.

3. RESULTS

The feature selection algorithm generated 300 models for every subset, each of which evolved for 200 generations from initial random models, optimizing for the fitness of the model. The evolution process for the HC vs. AD subset is shown in Figure 2, where it can be seen that the fitness curve reaches stability. The variables were then ranked according to their frequency in those models. In the negative y-axis of Figure 3, the number of occurrences of the top 30 ranked variables from the HC vs. AD subset can be seen, several of which are previously proven AD associated variables, such
as the volumes of the entorhinal cortex and of the hippocampus, ApoE4, amyloid-beta and the mean glucose intake of the temporal left gyrus. However, the DMFS model did not include all these variables. In the positive y-axis of Figure 3, the selected variables, the ones increasing the fitness of the previous model, are indicated with a vertical bold gray line. Finally, after the variable-elimination strategy was implemented, a final biomarker was obtained for every subset.

The eight variables that were selected as part of the HC vs. AD biomarker were: the presence or absence of an ApoE4 allele, the volume of the left entorhinal cortex, the volume of the left hippocampus, the volume of the left inferior temporal gyrus, the globally normalized CMRgl in the right inferior temporal gyrus, the globally normalized CMRgl in the left angular gyrus, the maximum CMRgl in the bilateral posterior cingulate, and the median CMRgl in the bilateral posterior cingulate. For the HC vs. MCI subset, the biomarker consisted of eleven variables: the presence or absence of an ApoE2 allele, the volume of the right supramarginal gyrus, the volume of the left inferior temporal gyrus, the average cortical thickness of the left medial orbitofrontal cortex, the average cortical thickness of the left rostral anterior cingulate, the volume of the left supramarginal gyrus, the surface area of the left temporal pole, the volume of the right hippocampus, the maximum CMRgl in the right angular gyrus, the mean CMRgl in the right angular gyrus, and the maximum CMRgl in the bilateral posterior cingulate. The MCI vs. AD biomarker was generated using five variables: the
plasma concentration of α1-antitrypsin (A1AT), the plasma concentration of the complement component 3 protein (C3), the plasma concentration of the monocyte chemotactic protein 3 (MCP3), the maximum CMRgI in the right angular gyrus, and the mean CMRgl in the left temporal gyrus.

When analyzing the HC vs. AD biomarker, the accuracy for the train subjects and for the blind test subjects was 0.983 and 0.9, respectively. The area under receiver operating characteristic (AUROC) was also measured, obtaining values of 0.994 and 0.955 for the train subjects and for the blind test subjects, respectively. The HC vs. MCI biomarker achieved train and blind test accuracies of 0.871 and 0.779, and AUROC values of 0.882 and 0.8, respectively. The MCI vs. AD classifier obtained an accuracy of 0.917 and an AUROC of 0.922 for the train subjects, and an accuracy of 0.848 and AUROC of 0.808 for the blind test subjects. Results for the three subsets, including their accuracies, sensitivities, specificities, AUROC, and confidence intervals, are shown in Table II and Table III. The ROC curves for the three subsets are shown in Figure 4.

Table II. Results obtained with the train subjects, using the proposed biomarkers.

<table>
<thead>
<tr>
<th>TRAIN (K-FOLDS)</th>
<th>Accuracy 0.05% Mean 0.95% Sensitivity 0.05% Mean 0.95% Specificity 0.05% Mean 0.95% AUROC 0.05% Mean 0.95%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>HC vs AD</strong></td>
</tr>
<tr>
<td></td>
<td><strong>HC vs MCI</strong></td>
</tr>
<tr>
<td></td>
<td><strong>MCI vs AD</strong></td>
</tr>
</tbody>
</table>

Table III. Results obtained with the blind test subjects, using the proposed biomarkers.

<table>
<thead>
<tr>
<th>BLIND TEST</th>
<th>Accuracy 0.05% Mean 0.95% Sensitivity 0.05% Mean 0.95% Specificity 0.05% Mean 0.95% AUROC 0.05% Mean 0.95%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>HC vs AD</strong></td>
</tr>
<tr>
<td></td>
<td><strong>HC vs MCI</strong></td>
</tr>
<tr>
<td></td>
<td><strong>MCI vs AD</strong></td>
</tr>
</tbody>
</table>

Figure 4. ROC curves generated for the train subjects (a, b, c) and for the blind test subjects (d, e, f) using the proposed biomarkers.
4. DISCUSSION

The present work showed that some features not previously proven as univariate biomarkers, add value for the diagnosis of AD and MCI in a multivariate setting. To achieve this conclusion a set of advanced bioinformatics tools were used to explore the complex multi-dimensional space of multimodal variables composed of laboratory essays, genotype, MRI analyses, PET analyses, medical history, physical exams, family history, and subject demographics. The study has some limitations that may affect the generalization of the results to other AD-MCI cohorts. First, the relatively small number of included subjects may not represent a reliable sample of the entire population, and could have biased the training phase of feature selection algorithm. Second, the nearest centroid classifiers are affected by the data normalization strategy: changes in normalization algorithm may yield a different set of classifiers. Finally, subject diagnostic was based on medical history and neuropsychological tests, and not on biopsies. Therefore, the AD diagnostic reliability has to be considered when interpreting the results. Regardless of these limitations, the proposed method created a HC vs. AD model that included some of the most widely accepted AD biomarkers: ApoE4, the best established risk factor for sporadic AD; the volume of the hippocampus, the best established structural biomarker for AD; the volume of the entorhinal cortex, a very promising anatomic structure for the early diagnosis of AD; and the CMRgl in the inferior temporal gyrus and the posterior cingulum, regions where a typical pattern of reduced cortical uptake has been demonstrated [20]. This result shows the concurrent validity of the employed feature selection methodology. Other widely accepted AD biomarkers, such as CSF amyloid-beta, total tau and hyperphosphorylated cortical uptake did not end up being added to the model, suggesting that they do not provide complementary information. Additional model features were the volume of the inferior temporal gyrus and the CMRgl of the angular gyrus. The inferior temporal gyrus is involved in higher order visual processing (i.e. assessment of faces) and has already been suggested to have AD predictive power [23]. Likewise, the angular gyrus, having a role in the central process for spelling [24], has also been associated to AD [25]. But, to our knowledge, both of these variables have not been used as an end point in neither clinical trials nor any kind of research, thus they have been considered as novel features.

The results obtained with the proposed 8-feature biomarker were better than those reported by Hinrichs [18] and Walhovd [19], who achieved an accuracy of 0.810 and 0.888, and an AUROC of 0.885 and 0.961, respectively. When compared with the work shown by Zhang [17], where a 189-feature biomarker was used, the results are almost identical, since they obtained an accuracy of 0.932 (0.890 – 0.965) and an AUROC of 0.976. Using a more compact biomarker decreases overfitting issues, increasing the robustness of the biomarker, and makes it easier for the biomarker to be measured. The robustness of the proposed biomarker was made clear by the blind test accuracy and AUROC results, which overlap with the results obtained for the train subjects.

For the HC vs. MCI 11-variable model, several novel and relevant features were found: the cortical thickness of the medial orbitofrontal cortex and of the rostral anterior cingulate, the surface area of the temporal pole, ApoE2, the volume of the supramarginal and the inferior temporal gyri, and the CMRgl of the angular gyrus. Association between cortical thickness and MCI and AD has already been studied [20, 26], where a reduced cortical thickness has shown to be related to cognitive decline. Surface area measurements have recently started to be analyzed with respect to MCI and AD, showing that brain atrophy is due either to surface area reduction or to cortical thickening [27, 28]. The relevance of ApoE2 has been debated, with research proving it to have a protective effect against AD [29], and research proving it to be associated with an increased risk of early-onset AD [30]. The relevance of the inferior temporal and the angular gyri were previously discussed, and the supramarginal gyrus has not shown to be relevant in current literature. Additionally, the hippocampal volume and the CMRgl in the posterior cingulum, widely accepted biomarkers, were part of the proposed model.

The results for the train subjects with the HC vs. MCI biomarker were better to the ones obtained by Zhang [17] and Walhovd [19]. The former published an accuracy of 0.764 (0.735 – 0.797) and an AUROC of 0.809, and the latter obtained an accuracy of 0.791 and reported no AUROC. On the other hand, the results for the blind test subjects were similar.

The MCI vs. AD biomarker included novel features: A1AT, C3 and MCP3. A1AT may be functionally involved in the pathogenesis of the lesions of AD [31], but has not been suggested as a biomarker. It has been demonstrated that C3, a factor contained in amyloid-beta plaques, has an active role in the genesis of the senile plaques [32], but the expression of C3 did not differ significantly between AD cases and age-matched controls in the work presented by Fischer [33]. Although MCP3 has been analyzed in several studies [34, 35], it has not shown significant difference in expression
levels between AD subjects and controls. Also, the volume of the angular gyrus, whose role was previously discussed, and the CMRgl of the temporal gyrus, a widely accepted biomarker, were part of the the proposed model.

The three proposed biomarkers are formed with both widely accepted biomarkers and novel features, indicating that models constructed from previously and non-previously AD associated features outperform those based solely on previously AD associated features. Also, the CMRgl in the angular gyrus showed up in the three models, which may indicate that this feature plays an important role in assessing the progression of the disease.

As future work, it is our intention to process the images made available by ADNI, so that our analysis can be made without any missing data; and to generate predictive biomarkers, capable of classifying between MCI subjects that will convert to AD and subjects that will not.

5. CONCLUSION

A wide association study was carried out and a feature selection strategy was used to find multimodal biomarkers to classify HC vs. AD, HC vs. MCI, and MCI vs. AD subjects. The automatically generated compact-nearest-centroid-classifiers included a short list of imaging and non-imaging features. The models included proven univariate biomarkers and novel variables that may help to improve the discrimination between HC, MCI and AD subjects. In the near future, this methodology is intended to be used to find a multimodal biomarker that accurately predicts which subjects will have a MCI to AD conversion.

6. ACKNOWLEDGMENTS

This work was partially supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study Rev October 16, 2012 at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

REFERENCES