Multivariate Analysis of Structural MRI and PET (FDG and 18F-AV-45) for Alzheimer's Disease and Its Prodromal Stages

Qi Zhou, Mohammed Goryawala, Mercedes Cabrerizo, Warren Barker, David Loewenstein, Ranjan Duara and Malek Adjouadi*

***Abstract*—A multivariate analysis method, orthogonal partial least squares to latent structures (OPLS), was used to discriminate Alzheimer's disease (AD), early and late mild cognitive impairment (EMCI and LMCI) from cognitively normal control (CN) using MRI and PET measures. FreeSurfer 5.1 generated 271 MRI features including 49 subcortical volumes, 68 cortical volumes, 68 cortical thicknesses, 70 surface areas and 16 hippocampus subfields. Subjects with all aforementioned MRI measures passing quality control and valid Fludeoxyglucose (18F) (FDG) and Florbetapir (18F) PET scans were selected from ADNI database, resulting in a total of 524 participants (137 CN, 214 EMCI, 103 LMCI and 70 AD) for the study. Altogether 286 features including 15 significant PET uptake features (7 for FDG and 8 for AV-45) were utilized for OPLS analysis. Predictive power was evaluated by $Q_2(Y)$, a quantifier of the statistical significance for class separation. The results show that MRI features ($Q_2(Y)$ =0.645), and PET features ($Q_2(Y)$ = 0.636) has comparable predictive power in separating AD from CN, and MRI features are better predictor of LMCI ($Q_2(Y)$ = 0.282) than PET ($Q_2(Y)$ = 0.294). Combination of PET and MRI has the most predictive power for LMCI and AD with $Q_2(Y)$ of 0.294 and 0.721, respectively. While for EMCI, cortical thickness was found to be the best predictor with a $Q_2(Y)$ of 0.108, suggesting cortical thickness may be the first structural change ahead of others and should be prioritized in prediction of very mild cognitive impairment.**

I. Introduction

Characterized as a neurodegenerative disease, Alzheimer's disease is thought to be the cause of the majority of dementia cases[1]. Early and reliable diagnosis of AD and its prodromal stages (i.e. mild cognitive impairment (MCI)) is seen as an essential step in search of prospective early intervention and treatments. Structural MRI capturing regional brain atrophy [2-4] and PET scan, such as Fludeoxyglucose (FDG) and Florbetapir (AV-45), characterizing amyloid deposition in brain [5] are the two most popular imaging biomarkers for early detection and diagnosis of AD.

Orthogonal partial least square to latent structures (OPLS) is a supervised multivariate data analysis method that has shown its efficiency in analyzing complex biological data [6-9]. Westman et al. combined manual hippocampal volume measurements with automated regional and global volume measures to discriminate AD and MCI from controls and compared their discriminative powers[8]. They also compared and combined MRI data from European ADDNeuroMed with ADNI using OPLS, and the results indicated that the two cohorts showed similar pattern of atrophy and predictive power (between 80 and 90%)[7].

In this study, we have used additional MRI measures (i.e. subcortical volume, cortical volume, cortical thickness, surface area and hippocampus subfield), and features extracted from FDG and AV-45 PET scans as inputs to OPLS. The aim was to find the individual and combined discriminative power of these features in separating EMCI, LMCI and AD from normal control (CN) and the best models for predicting each stage of disease. We also aimed to investigate efficiency of cross testing by performing cross-model validation with an external dataset.

II. Materials and Methods

A. Dataset

The data used were downloaded from Alzheimer's disease Neuroimaging Initiative (ADNI) database which is publicly accessible (https://ida.loni.usc.edu/login.jsp), and were used in this study for the goal of detecting AD at its earliest stage and identifying ways to track the disease through biomarkers including brain-imaging techniques, such as FDG and AV-45 PET and structural MRI, among others.

B. Inclusion and Diagnostic Criteria

As this study aimed to find the discriminative power of MRI and PET, only subjects with all MRI measures passing quality control and valid FDG and AV-45PET scans at baseline were included. This has yielded a total of 524 subjects qualified for this study (CN = 137, EMCI = 214, LMCI = 103 and AD = 70) as of February 2014 on ADNI website, as shown in Table I.

Diagnostic criteria are as follow: Cognitively Normal Subjects (CN): MMSE scores between 24 and 30 (inclusive), a CDR of 0, non-depressed, non-MCI, and non-demented. EMCI: MMSE scores between 24 and 30 (inclusive), a subjective memory concern reported by subject, informant, or clinician, objective memory loss measured by education adjusted scores on delayed recall of one paragraph from

Q. Zhou, M. Goryawala, M. Cabrerizo, and M. Adjouadi* are with the Department of Electrical Engineering at the Florida International University, Miami FL 33174 USA (email: qzhou003@fiu.edu; mgory001@fiu.edu; cabreriz@fiu.edu; jwang006@fiu.edu and adjouadi@fiu.edu).

W. Barker, D. A. Loewenstein and R. Duara are with the Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, FL 33140 USA (email: warren.barker@msmc.com; DLoewenstein@med.miami.edu; ranjan.duara@msmc.com). Asterisk indicates corresponding author.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

Wechsler Memory Scale Logical Memory II (WMSLM II)(⩾16 years: 9-11; 8-15 years: 5-9; 0-7 years: 3-6), a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia. LMCI: Same as EMCI with a difference only in objective memory loss measured by education adjusted scores on delayed recall of one paragraph from WMSLM II(⩾16 years: ⩽8; 8-15 years: ⩽4; 0-7 years: ⩽ 2). Mild AD: MMSE scores between 20-26 (inclusive), a CDR of 0.5 or 1.0, and meets NINCDS/ADRDA criteria for probable AD.

TABLE I. CHARACTERISTICS OF PARTICPANTS

	Training Set				Testing set			
	CN	EMCI	LMCI	AD	CN	EMCI	LMCI	p-value*
Number	70	70	70	70	67	144	33	-
Gender (F/M)	33/37/	29/41	33/37	31/39	38/29	69/75	16/17	ns†
Age	73.6 ± 6.6	70.6 ± 6.8	71.2 ± 7.4	75.0 ± 7.9	73.6 ± 5.4	69.6 ± 7.0	70.9 ± 8.5	**< 0.001**
Years of education	16.5 ± 2.4	15.5 ± 2.6	16.7 ± 2.7	15.9 ± 2.6	16.3 ± 2.6	16.3 ± 2.6	16.1 ± 2.8	0.192
MMSE	29.2 ± 1.0	28.5 ±1.4	27.5 ± 1.8	22.8 ± 2.1	29.0 ± 1.3	28.4 ± 1.6	27.9 ± 1.9	**< 0.001**
FDG	6.6 ± 0.5	6.4 ± 0.6	6.2 ± 0.8	5.3 ± 0.7	6.6 ± 0.6	6.5 ± 0.6	6.5 ± 0.6	**< 0.001**
AV45	1.09 ± 0.19	1.17 ± 0.22	1.29 ± 0.23	1.41 ± 0.18	1.11 ± 0.17	1.16 ± 0.20	1.3 ± 0.24	**< 0.001**

Data are represented as mean ± standard deviation where applicable. CN = cognitively normal, EMCI = early mild cognitive impairment, LMCI = late mild cognitive impairment, AD = Alzheimer's disease, MMSE = Mini Mental State Examination, FDG = PET with [18F]-fluorodeoxyglucose (FDG), AV45 = PET with 18F-AV-45 (florbetapir) † ns = not significant. Fisher's exact test was performed for AD vs. CN, LMCI vs. CN and EMCI vs. CN using the training set, and all the p values were larger than 0.05. *Student's t-tests were performed for the possible factors between CN and AD in the training set and p-values were shown with those smaller than 0.05 deemed significant as bolded.

C. MRI and PET

MRI scans were acquired from a variety of 3T scanners with protocols individualized for each scanner, as defined at (http://adni.loni.usc.edu/methods/documents/mri-protocols/). All scans were reviewed for quality by quality control (QC) team at the Mayo Clinic. A detailed description of PET protocols and acquisition procedures can be found at (http://adni.loni.usc.edu/methods/pet-analysis/pre-processig).

D. Image Analysis and Data Preprocessing

FreeSurfer 5.1 was used to generate aforementioned 271 MRI features defined by UCSFFSX51_DICT_02_04_14.csv on ADNI website (https://ida.loni.usc.edu/login.jsp) excluding 5th ventricle (due to missing values) and ICV. All MRIs passed QC for the 271 features. Averaged florbetapir uptakes of the cortical and reference regions and a florbetapir composite (SUVR) across the brain were the 8 features extracted from AV-45 PET as shown in Table II, which also showed the 7 features of FDG PET including FDG uptakes in 5 previously identified regions of interests (ROI)[10] plus their sum and weighted average. Altogether this yielded a total of 286 features for analysis.

TABLE II. VARIABLES OF PET FEATURES INCLUDED IN OPLS ANALYSIS

FDG PET	AV-45 PET
Cingulum Post	Cerebellum GM
Left Temporal	Whole cerebellum
Left Angular	Brainstem
Sum of FDG	Frontal
Right Temporal	Cingulate
Right Angular	Parietal
Weighted Average	Temporal
	Summary of SUVR

GM = grey matter, SUVR = standardized uptake value ratio
FDG = PET with [18F]-fluorodeoxyglucose (FDG), AV45 = PET with 18F-AV-45 (florbetapir)

Since ICV and age were found to be significant factors as demonstrated in Table 1, all MRI measures but for ICV were adjusted for ICV and age as per Eq. 1 if p value of the linear regressions were less than 0.05.

$$V_a = V_{ua} - G_{ICV} \cdot (V_{sICV} - V_{mICV}) - G_{AGE} \cdot (A_s - A_m) \quad (1)$$

Where V_a is the adjusted measure, V_{ua} is the unadjusted measure, V_{sICV} and A_s are the subject ICV and age (years), respectively; V_{mICV} and A_m are the corresponding means for all the control subjects. The gradients G_{ICV} and G_{AGE} were derived by a region specific linear regression against subject ICV and age of all the participants. As per Chiang et al.[11], the above regression also has the advantage that the regressing order doesn't affect the results. And then all data was processed by mean centering and unit variance scaling.

E. Multivariate Data Analysis

The aforementioned 286 features were used as input to OPLS [6, 9], a supervised multivariate data analysis method comes with the software package SIMCA (Umetrics AB, Umea, Sweden). OPLS removes variation from descriptor variables that is not related to group separation and the information related to class separation is found in the predictive component[6]. There are 7 single models and 3 hierarchical models that were created for AD vs. CN, LMCI vs. CN and EMCI vs. CN for separation purpose. Single models used one of the 5 groups of MRI measures (i.e. subcortical volumes, cortical volumes, cortical thickness average, surface area and hippocampus subfields) or one of the PET (AV-45 or FDG). Three hierarchical models included one with all 5 groups of MRI measures, one with two PET measures and one with all of them combined.

The predictive power of OPLS model for separating two groups is found in $Q_2(Y)$ and is defined as follows:

$$Q_2(Y) = 1 - PRESS/SSY_{tot.corr}$$

Where PRESS (predictive residual sum of squares) $= \sum(y_{actual} - y_{predicted})^2$, is the squared differences between observed and predicted Y-values, and $SSY_{tot.corr}$ represents the total variation of the Y variable (diagnosis)

after scaling and mean centering [12]. $Q_2(Y)$ is the results of 7-fold (by default) cross validation with some subjects kept out of the model development and prediction of the kept out observations by the developed model are compared with the actual classes. This procedure is repeated until every observation has been kept out once and only once. A model with a $Q_2(Y)$ value larger than 0.5 is regarded as good [12].

III. Results

A. Predictive Power of OPLS Model

The predicative power for the aforementioned models was summarized in Table III with the corresponding$Q_2(Y)$. The hierarchical models are italicized and highest $Q_2(Y)$ are bolded for predicting different stages of the disease. Table III showed that combining all MRI and PET features had the highest predictive power of LMCI and AD. While using cortical thickness alone yielded a higher $Q_2(Y)$ for EMCI than using any measures even when all features combined.

Separating efficiency of the best models found could be better visualized with scatter plots as given in Fig. 1. Perfect separation of AD using all features was shown as Fig. 1A, with a high $Q_2(Y)$ of 0.721. Fig. 1B-1C showed high efficiency of separating LMCI and EMCI from CN as well.

TABLE III. SUMMARY OF $Q_2(Y)$ FOR ALL MODELS

Models		EMCI*	LMCI	AD
MRI	Subcortical volume	-	0.188	0.585
	Cortical volume	-	0.257	0.528
	Cortical thickness	**0.108**	0.154	0.538
	Surface area	-	0.040	0.202
	Hippocampal Subfields	0.029	0.277	0.547
	Combined	-	*0.282*	*0.645*
PET	18F-AV-45	0.076	0.227	0.518
	FDG	0.038	0.055	0.512
	Combined	*0.093*	*0.229*	*0.636*
MRI+PET		*0.008*	***0.294***	***0.721***

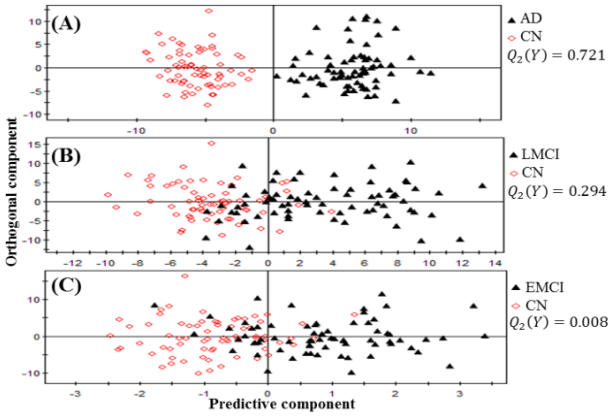

Figure 1. Scatter plots under best models as indicated in Table III for separation between (A) AD and CN. (B) LMCI and CN (C) EMCI and AD.

B. Model Cross-Model Validation with External Set

External testing set was used to validate the best models as indicated in Table III, which were also cross tested with best AD model to see how the AD model classify MCI subjects. Results were shown with scatter plots as Fig. 2A-2D. Generally, the EMCI and LMCI best model has a better sensitivity and accuracy in predicting EMCI and LMCI than using AD model, however best AD model exhibited higher specificity as the model classified almost all controls correctly. The results also showed that the best LMCI model was almost as efficient in classifying LMCI in the testing set as it did in the training set indicating high model efficiency. However, the same is not seen for EMCI.

IV. Discussion

A. Model Efficiency with OPLS

This study aimed to investigate the predictive power of MRI measures, AV-45 and FDG PET in discriminating AD, LMCI, and EMCI from control using OPLS as a multivariate analysis tool. Models using part or all of the features were created for EMCI, LMCI and AD. Cross-validated predictive power $Q_2(Y)$ was used to evaluate these models, which were also validated with external testing dataset. The results as shown in Table III indicated that models generally had much more predictive power of AD than LMCI and EMCI as expected. The best model for predicting stages of AD is deemed the one with highest $Q_2(Y)$ value. For AD and LMCI prediction, the best model was obtained by combining all MRI and PET features. The best model of EMCI was average cortical thickness with a $Q_2(Y)$ value of 0.108, which was even slightly higher than combinational power of the two PET scans. This was very interesting that inclusion of more features did not strengthen the predictive power.

With European AddNeuroMed project data, Westman et al. [8] utilized automated regional volumes and manual outlining of hippocampus as inputs to OPLS and found a $Q_2(Y)$of 0.64 when discriminating AD from CN, which was consistent with the $Q_2(Y)$ of 0.645 achieved using MRI only in this study, though more features were being used in this study. We have also showed PET scans has comparable discriminating power with MRI ($Q_2(Y)$ = 0.636). Westman et al. also built a MCI model with the same features and obtained a $Q_2(Y)$of 0.22. Comparing with LMCI model using all MRI features in this study, the predictive power of our model ($Q_2(Y)$ = 0.282) is higher which could be due to the difference in diagnostic criteria between MCI in AddNeuroMed dataset and LMCI in ADNI, or perhaps due to the MRI features that were included in this study contributing more complementary information to the model.

To the best of our best knowledge of the literature, we were the first to include the EMCI group into the OPLS model. Interestingly, cortical thickness was found to have the most discriminative power than all other MRI and PET features, even higher than when combining all of them. The same trend was not seen for LCMI and AD models, which showed that volumetric measures had more power. This could be due to more reliability of cortical thickness than cortical volumes and other MRI measures at the very early

stage of AD [13]. This is consistent with some researches in literature that cortical thickness is recommended for MCI prediction [14]. Also it could imply that cortical thinning is the very first anatomical change that occurs before any other volumetric change or effective response to PET scan, and as disease processes volumetric change starts to dominate.

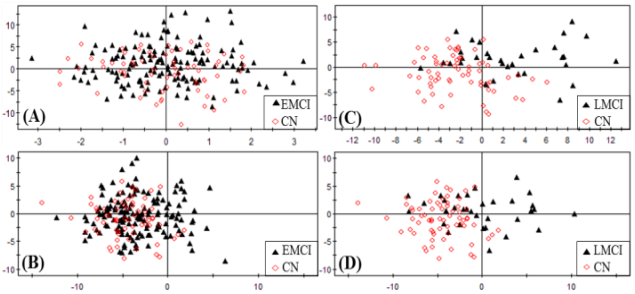

Figure 2. Scatter plots of model validation with external testing set. (A) Testing best EMCI model with external EMCI data. (B) Testing best AD model with external EMCI data. (C) Testing best LMCI model with external LMCI data. (D) Testing best AD model with external LMCI data.

Since the discriminative power of the best EMCI model is still weak, it can be argued that this could be due to partitioning errors embedded in the cross validation process in that the data distribution may slightly favor cortical thickness than others as reflected in $Q_2(Y)$. However, since $Q_2(Y)$ of 0.108 is higher than the threshold for significance of an OPLS model (0.05) [8], the model is still considered significant and the results are considered reliable.

B. Model and Cross-Model Validation

External EMCI and LMCI dataset was used to validate previously found best models. Fig. 2A-2B showed that a model trained using EMCI did a better job in classifying EMCI yielding higher sensitivity and accuracy than using AD model. Fig. 2C-2D showed similar results for LMCI. From Fig. 2, it could be seen that AD model did classify most of the CN subjects correctly indicating higher specificity than EMCI and LMCI models. This was because AD-CN trained model was very robust to recognize its training groups as manifested by Fig. 1A with no misclassification, while EMCI-CN and LMCI-CN trained models were not that efficient in defining the separating boundary due to limited distinctive features existing.

Validation results of best EMCI and LMCI models as shown in Fig. 2A and 2C, showed that LMCI model was more robust, as LMCI model classified the testing LMCI subjects as well as it did for the training set (see Fig. 1B) and EMCI showed poorer separation than it did for the training set (see Fig. 1C). However, EMCI model showed some degree of generalization as validated with external dataset shown in Fig. 2A, indicating structural change of brain (cortical thickness in particular) already showed up at this stage. Besides, combining two PET scans received a predictive power ($Q_2(Y) = 0.093$) close to that of the cortical thickness indicating that amyloid deposition was also shown to accumulate at the EMCI stage.

ACKNOWLEDGEMENTS

The work was supported by the National Science Foundation under grants CNS-0959985, HRD-0833093, CNS-1042341, IIP-1338922 and IIP-1230661. The philanthropic support provided by the Ware Foundation is also greatly appreciated.

REFERENCES

[1] S. Duchesne, A. Caroli, C. Geroldi, C. Barillot, G. B. Frisoni, and D. L. Collins, "MRI-based automated computer classification of probable AD versus normal controls," *Medical Imaging, IEEE Transactions on,* vol. 27, no. 4, pp. 509-520, 2008.

[2] N. C. Fox, and J. M. Schott, "Imaging cerebral atrophy: normal ageing to AD," *Lancet,* vol. 363, no. 9406, pp. 392-4, Jan 31, 2004.

[3] Q. Zhou, M. Goryawala, M. Cabrerizo, W. Barker, R. Duara, and M. Adjouadi, "Significance of normalization on anatomical MRI measures in predicting Alzheimer's disease," *ScientificWorldJournal,* vol. 2014, pp. 541802, 2014.

[4] Q. Zhou, M. Goryawala, M. Cabrerizo,J. Wang, W. Barker, R. Duara, and M. Adjouadi, et al. "An Optimal Decisional Space for the Classification of Alzheimer's Disease and Mild Cognitive Impairment " IEEE Trans. on Biomedical Engineering, DOI: 10.1109/TBME.2014.2310709, 8 pages, 2014

[5] A. Nordberg, J. O. Rinne, A. Kadir, and B. Långström, "The use of PET in AD," *Nature Reviews Neurology,* vol. 6, no. 2, pp. 78-87, 2010.

[6] J. Trygg, and S. Wold, "Orthogonal projections to latent structures (O-PLS)," *J. of Chemometrics,* vol. 16, no. 3, pp. 119-128, Mar, 2002.

[7] E. Westman, A. Simmons, J. S. Muehlboeck, P. Mecocci, B. Vellas, M. Tsolaki, I. Kloszewska, H. Soininen, M. W. Weiner, S. Lovestone, C. Spenger, L. O. Wahlund, A. Consortium, and A. s. D. N. Initi, "AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America," *Neuroimage,* vol. 58, no. 3, pp. 818-828, Oct 1, 2011.

[8] E. Westman, A. Simmons, Y. Zhang, J. S. Muehlboeck, C. Tunnard, Y. Liu, L. Collins, A. Evans, P. Mecocci, B. Vellas, M. Tsolaki, I. Kloszewska, H. Soininen, S. Lovestone, C. Spenger, and L. O. Wahlund, "Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls," *Neuroimage,* vol. 54, no. 2, pp. 1178-87, Jan 15, 2011.

[9] M. Bylesjo, M. Rantalainen, O. Cloarec, J. K. Nicholson, E. Holmes, and J. Trygg, "OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification," *Journal of Chemometrics,* vol. 20, no. 8-10, pp. 341-351, Aug-Oct, 2006.

[10] S. M. Landau, D. Harvey, C. M. Madison, R. A. Koeppe, E. M. Reiman, N. L. Foster, M. W. Weiner, W. J. Jagust, and A. s. D. N. Initi, "Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI," *Neurobiology of Aging,* vol. 32, no. 7, pp. 1207-1218, Jul, 2011.

[11] G. C. Chiang, P. S. Insel, D. Tosun, N. Schuff, D. Truran-Sacrey, S. Raptentsetsang, C. R. Jack, M. W. Weiner, and A. D. Neuroimaging, "Identifying Cognitively Healthy Elderly Individuals with Subsequent Memory Decline by Using Automated MR Temporoparietal Volumes," *Radiology,* vol. 259, no. 3, pp. 844-851, Jun, 2011.

[12] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström, and S. Wold, "Multi-and Megavariate Data Analysis: Part II: Advanced Applications and Method Extensions," 2006.

[13] O. Querbes, F. Aubry, J. Pariente, J.-A. Lotterie, J.-F. Démonet, V. Duret, M. Puel, I. Berry, J.-C. Fort, and P. Celsis, "Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve," *Brain,* vol. 132, no. 8, pp. 2036-2047, 2009.

[14] L. Wang, F. C. Goldstein, E. Veledar, A. I. Levey, J. J. Lah, C. C. Meltzer, C. A. Holder, and H. Mao, "Alterations in Cortical Thickness and White Matter Integrity in Mild Cognitive Impairment Measured by Whole-Brain Cortical Thickness Mapping and Diffusion Tensor Imaging," *American Journal of Neuroradiology,* vol. 30, no. 5, pp. 893-899, May, 2009.