Anatomically Constrained Weak Classifier Fusion for Early Detection of Alzheimer's Disease

Mawulawoé Komlagan, Vinh-Thong Ta, Xingyu Pan, Jean-Philippe Domenger, D. Louis Collins, Pierrick Coupé

HAL Id: hal-01062759

https://hal.science/hal-01062759

Submitted on 10 Sep 2014

Anatomically Constrained Weak Classifier Fusion for Early Detection of Alzheimer's Disease

Mawulawoé Komlagan[1,2], Vinh-Thong Ta[1,2,3], Xingyu Pan[1,2],
Jean-Philippe Domenger[1,2], D. Louis Collins[4], Pierrick Coupé[1,2], and
the Alzheimer's Disease Neuroimaging Initiative [⋆]

[1] Univ. Bordeaux, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.
[2] CNRS, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.
[3] IPB, LaBRI, UMR 5800, PICTURA, F-33600 Pessac, France.
[4] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill
University, Montreal, Canada.

Abstract. The early detection of Alzheimer's disease (AD) is a key
step to accelerate the development of new therapies and to diminish the
associated socio-economic burden. To address this challenging problem,
several biomarkers based on MRI have been proposed. Although numer-
ous efforts have been devoted to improve MRI-based feature quality or
to increase machine learning methods accuracy, the current AD prog-
nosis accuracy remains limited. In this paper, we propose to combine
both high quality biomarkers and advanced learning method. Our ap-
proach is based on a robust ensemble learning strategy using gray matter
grading. The estimated weak classifiers are then fused into high infor-
mative anatomical sub-ensembles. Through a sparse logistic regression,
the most relevant anatomical sub-ensembles are selected, weighted and
used as input to a global classifier. Validation on the full ADNI1 dataset
demonstrates that the proposed method obtains competitive results of
prediction of conversion to AD in the Mild Cognitive Impairment group
with an accuracy of 75.6%.

Keywords: Ensemble learning, Weak classifier, Sparse logistic regres-
sion

1 Introduction

Alzheimer's disease (AD) and its prodromal phase, Mild Cognitive Impairment
(MCI), are the most common neurodegenerative diseases affecting elderly peo-
ple. In the early stage of the disease, neural degeneration is subtle making it
difficult to predict which MCI subjects will progress to AD (pMCI) and which

MCI subjects will remain stable (sMCI) during the follow up. Hereon, AD prediction, *i.e.*, AD early detection will address the classification of MCI subjects into pMCI and sMCI subjects.

Several biomarkers have been proposed to achieve early AD diagnosis [1]. Among them, it has been established that measurements of brain atrophy extracted from structural MRI are valid markers of early stages of AD [2]. Therefore, automatic frameworks using MRI-based features have been developed to achieve computer-aided prognosis [3–5]. One part of these works focused on advanced machine learning techniques [6] while another part aimed to enhance the biomarker quality [4, 7]. Among them, patch-based methods [8, 9] demonstrated competitive AD prediction results. Despite these efforts, the current AD prognosis accuracy remains around 70%, that suggests the limitation of using (i) traditional features with advanced learning processes or (ii) high quality features with basic machine learning methods. In this paper, we propose to combine high quality biomarkers with advanced learning method to improve AD prediction accuracy.

To this end, we first propose to extend the patch-based scoring method proposed in [8]. In this approach, the anatomical pattern similarity is estimated between the MCI test subject and two training populations (*i.e.*, Cognitively Normal (CN) and AD) using a non-local patch-based scoring method. For each voxel, a score (*i.e.*, a grade) that measures the proximity to both training populations is computed. In [8], the *a priori* ROI-based strategy focused mainly on hippocampus and may discard other possible informative anatomical regions. To overcome this limitation, we propose to score the whole gray matter (GM). Moreover, to be more robust to intensity normalization discrepancies between MRI, probabilities are used in place of intensities during patch comparison. Finally, while a local patch-based strategy is used in [9], a non-local approach is privileged to better handle inter-subject variability and registration error [8].

Afterwards, an ensemble learning method [10] is considered to efficiently use the estimated advanced biomarkers. Since the scoring value, assigned to each voxel of the GM, estimates the proximity to AD and CN, it can be viewed as the posterior probability of a weak classifier. Combined together, these weak classifiers form an ensemble that can be used to classify subjects [11]. As noticed in [2], it appears that AD-related brain alterations are mainly a region-by-region process. Hence, we propose to further use this clinical knowledge to create atlas-based anatomical sub-ensembles of weak classifiers before fusing them into intermediate classifiers. Finally, to discard brain areas that may not be related to AD, we propose to select the most relevant anatomical sub-ensembles using a Sparse Logistic Regression (SLR).

In this work, the contributions are threefold: (i) unlike ROI-based approach, non-local scoring values are estimated over the whole GM and considered as weak classifiers; (ii) an advanced ensemble learning technique is used to fuse these weak classifiers into anatomical sub-ensembles; and (iii) a sparse approximation is used to efficiently select and weight the most relevant anatomical sub-ensembles.

Table 1. Demographic information about the considered study subjects.

Pathological group	Size	Gender (% Female)	Age ± SD	MMSE± SD
AD	192	48%	75.7±7.6	22.9±3.0
CN	220	49%	76.1±4.9	29.1±0.9
pMCI	166	39%	74.5±7.2	26.4±2.0
sMCI	236	33%	74.9±7.8	27.2±2.5

2 Materials and Methods

2.1 The ADNI Dataset and Image Processing

To evaluate the performance of the proposed method, all the subjects with an available baseline ADNI preprocessed 1.5T MRI scan are used. The considered dataset is composed of 814 subjects divided into 4 groups AD, CN, pMCI and sMCI. AD and CN groups are used exclusively as training population during GM grading step (see 2.3). The size, the genders, the average ages and the average MMSE (Minimal Mental State Examination) are summarized in Table 1. These groups are similar to the ones used in [4,7–9]. All 814 MRI were first segmented, normalized, modulated (correction of volume changes due to the normalization), and registered into a common space. These processing steps were performed with the VBM8 toolbox[5] added to the SPM8 software[6]. The resulting images correspond to tissue-class probability maps in the MNI space. The obtained GM probability maps are then used as inputs of our GM grading process.

2.2 Method Overview

The framework of the proposed method is summarized here and in Fig. 1. First, the grading method is applied to all the MCI subjects GM maps using the AD and CN populations. Second, the grading values obtained over the whole GM are fused into anatomical sub-ensembles to form intermediate classifiers. Third, the age-effect is corrected using a control population. Afterwards, SLR feature selection is applied to select and weight the most relevant intermediate classifiers. Finally, the selected intermediate classifiers are used to train a global linear SVM classifier. The methods are detailed in the following sections.

2.3 Weak Classifier Estimation via Whole GM Grading

This work is based on the Scoring by Non-local Image Patch Estimator (SNIPE) method [8] where a non-local patch-based estimator is used to perform anatomical structure grading. The patch surrounding each voxel of a test subject is involved to estimate the anatomical pattern similarity between the considered patch and the most similar patches extracted from AD and CN training populations. This pattern similarity is quantified with L_2-norm between the patch

---

[5] http://dbm.neuro.uni-jena.de/vbm.html
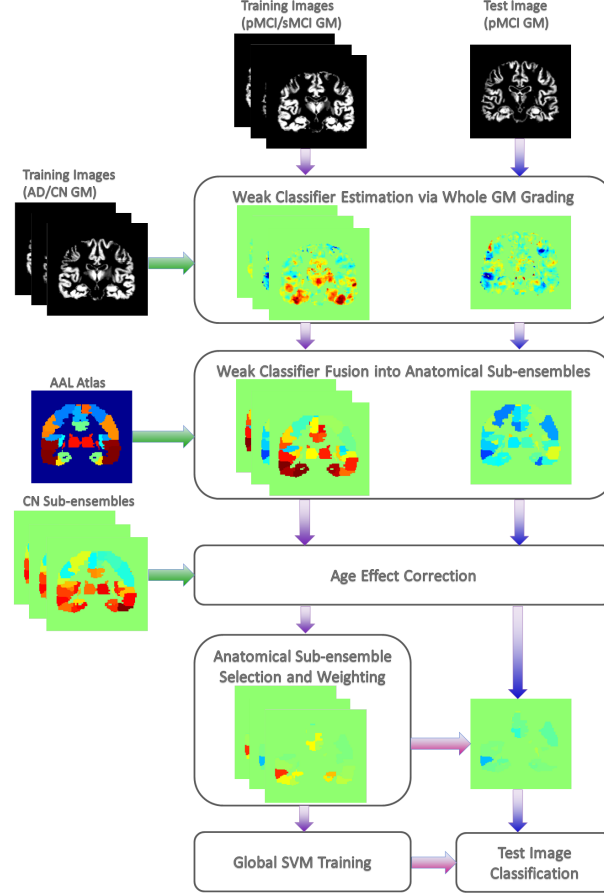[6] http://www.fil.ion.ucl.ac.uk/spm

Fig. 1. Overview of the proposed method.

intensities. The resulting grading value indicates if the considered anatomical pattern is typical of AD (AD-like) or CN (CN-like) populations. Such values can be viewed as the posterior probabilities of a weak classifier. In the proposed method, the grade of each GM voxel is calculated using probability of GM tissue instead of voxel intensities. By using GM tissue probability, our method is more robust to multi-site MR image acquisition. In addition, in our approach the grading is performed on the whole GM, and not only in the hippocampal area. This prevents discarding any relevant information that could be found in other brain regions.

2.4 Weak Classifier Fusion into Anatomical Sub-ensembles

After the grading step, the dimensionality of the weak classifiers space is too high to be directly used for classification. A straightforward solution is to fuse

the weak classifiers into a global classifier [11]. However, this may lead to a sub-optimal result since local relevant information may be lost in a high level global fusion. Additionally, as noticed in [2], AD affects specific regions of the brain in a typical progressive manner. Therefore, we propose to group the weak classifiers into anatomical sub-ensembles using an atlas-based strategy. An ensemble learning principle fuses them into intermediate classifiers [10]. The intermediate classifiers, $c_i = \frac{1}{K} \sum_k c_w(k)$, are constructed by an un-weighted vote of the K weak classifiers c_w included in each anatomical sub-ensemble. In this work, the whole GM is divided into the 116 segmented anatomical regions corresponding to the Automatic Anatomical Labeling (AAL) atlas [12]. Thus, the grades are averaged within each anatomical structure and their mean values considered as predictor values of the 116 c_i intermediate classifiers. Since the grades estimate the AD-related brain anatomical changes, it could be interesting to remove the normal aging effect from the features used. Moreover, it has been shown that SNIPE grades are correlated to age [8]. Therefore, as in [13], we used the CN population to correct the age effect on the MCI populations. For each intermediate classifier, we estimated the age-related effect on the CN population using linear regression. Intermediate classifiers in the MCI populations were then corrected using the estimated linear regression coefficients (see [14]).

2.5 Anatomical Sub-ensemble Selection and Weighting

As shown in [1], anatomical regions may not be similarly impacted by the progression from MCI stage to the moderate stage of AD. Therefore, using all the intermediate classifiers could be suboptimal. Moreover, beyond classification efficiency reasons and for clinical considerations, it could also be interesting to know the most impacted brain regions. In this work, we selected the most relevant anatomical sub-ensembles by using SLR with L_1/L_2-norm regularization [15, 16]. It has been established that combining the two norms take into account possible inter-feature correlation while imposing sparsity [17]. Additionally, SLR provides a coefficient for each intermediate classifier that represents its relative importance in the sparse approximation. In our method, these coefficients are used to weight each corresponding intermediate classifier before global classification. We used the SLEP package[7] to solve SLR. The selected weighted intermediate classifiers are then used to train a linear SVM as implemented in LIBSVM[8].

2.6 Validation Framework

As is done in [7–9], the classification process is performed using a leave-one-out cross-validation procedure to avoid bias. To validate the efficiency of our framework, we conducted several experiments. First, to highlight the relevance of using

---

[7] http://www.public.asu.edu/~jye02/Software/SLEP/

[8] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Table 2. Methods comparison. The used features were corrected for age-effect.

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
GM Volume	59.7	47.6	68.2
GM Volume + SLR	70.1	56.0	80.0
GM Grading	67.7	57.8	74.8
GM Grading + SLR	75.6	61.5	85.6

high quality features, we compared the efficiency of our framework using volume-based and grading-based features. For the volume-based approach, we computed the volumes of each AAL region performing the sum of its corresponding GM probability values. Second, to measure the contribution of SLR sub-ensemble selection, we tested our framework while removing this step for both volume-based and grading-based approaches. For the grading step, we used the default parameters proposed in [8]. In each experiment, the L_1/L_2 regularization parameters for solving SLR were set by searching their optimal values while the penalization parameter of the SVM was estimated by a grid search and a nested 10-fold cross validation over the training set.

3 Results and Discussion

The results are summarized in Table 2. First, we notice that using grading-based features improves the result of the classification compared to volume-based features with an increase of about 5pp (percentage points). This confirms the relevance of using high quality features in our method. Second, we observe an improved accuracy of at least 8pp when performing an SLR feature selection with both volume and grading. Moreover, compared to hippocampal scoring [8], we improve the accuracy of 4.6pp using our framework (see Table 3). It is interesting to note that directly using all the anatomical sub-ensembles (*i.e.*, without SLR) provided worst results than using only hippocampal grading. However, when selecting the most relevant anatomical sub-ensembles an important increase is observed. This indicates that areas other than hippocampus seem to be impacted at MCI stage. Thus, automatic *a posteriori* selection of these areas instead of using predefined ROIs leads to higher accuracy.

As shown in Table 3, our method achieves better accuracy than other state-of-the-art methods validated on the same ADNI database and with the same unbiased leave-one-out cross-validation process [7–9]. This establishes the robustness and the efficiency of the proposed framework that combines high quality features with an advanced learning method, *i.e.*, sub-ensemble learning based on constrained weak-classifier-fusion combined with SLR. Additionally, it should also be noted that even using usual GM volumes as features in our framework leads to similar or even competitive accuracy as compared to other methods.

Finally, we can note that even though our method is based on one imaging modality it performs similarly or even better than recent multi-modality methods [6, 18, 19]. Furthermore, for clinical reasons, it could be interesting to analyze

Table 3. Comparison with recently published methods using similar dataset.

Method	Acc. (%)	Sen. (%)	Spe. (%)
GM Grading + SLR	75.6	61.5	85.6
GM Volume + SLR	70.1	56.0	80.0
ROI-based SNIPE (hippocampal grading) [8]	71	70	71
Multi-instance learning [9]	70.4	66.5	73.1
Multi-methods [4]	68	67	69
Cortical thickness [7]	67.8	64.6	70.0

Table 4. 10 first AAL regions selected by SLR and ordered by decreasing weight.

AAL-based gyrus:	
1. Right middle temporal	6. Left cerebelum
2. Left hippocampus	7. Right inferior frontal
3. Left superior frontal	8. Left medial orbital frontal
4. Right middle cingulum	9. Right hippocampus
5. Left posterior cingulum	10. Left para-hippocampal

the anatomical regions selected via SLR. Table 4 presents, on average, the first selected AAL regions at each run of the leave-one-out cross-validation process. It appears that some anatomical regions like middle temporal gyrus, hippocampus and parahippocampal gyrus are included in the presented list. Such structures are known to be impacted by AD [20]. They are also among the most selected regions in [7] using cortical thickness features.

4   Conclusion

In this study, we proposed an anatomically constrained weak classifier fusion classification procedure extending the grading technique presented by [8]. This work aimed to combine high quality biomarkers with advanced learning method to improve AD detection at its prodromal stage. We demonstrated through our experiments that the contributions made to the method proposed by [8] lead to high classification accuracy for the early detection of AD. Compared to recently proposed MRI-based prediction techniques, we obtained a very competitive accuracy result of 75.6% for the prediction of AD.

Acknowledgments

References

1. Jack Jr, C.R., et al: Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet Neurology **12**(2) (2013) 207–216
2. Frisoni, G.B., et al: The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology **6**(2) (2010) 67–77
3. Cuingnet, R., et al: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage **56**(2) (2011) 766–781
4. Wolz, R., et al: Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PloS one **6**(10) (2011) e25446
5. Davatzikos, C., et al: Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiology of aging **32**(12) (2011) 2322–e19
6. Suk, H.I., et al: Deep learning-based feature representation for AD/MCI classification. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer (2013) 583–590
7. Eskildsen, S.F., et al: Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. NeuroImage **65** (2013) 511–521
8. Coupé, P., et al: Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. NeuroImage: clinical **1**(1) (2012) 141–152
9. Tong, T., et al: Multiple instance learning for classification of dementia in brain MRI. Medical Image Analysis **18**(5) (2014) 808 – 818
10. Dietterich, T.G.: Ensemble methods in machine learning. In: Multiple classifier systems. Springer (2000) 1–15
11. Liu, M., et al: Ensemble sparse classification of Alzheimer's disease. NeuroImage **60**(2) (2012) 1106–1116
12. Tzourio-Mazoyer, N., et al: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage **15**(1) (2002) 273–289
13. Koikkalainen, J., et al: Improved classification of Alzheimer's disease data via removal of nuisance variability. PloS one **7**(2) (2012) e31112
14. Dukart, J., et al: Age correction in dementia – matching to a healthy brain. PLoS ONE **6**(7) (07 2011)
15. Dubey, R., et al: Analysis of sampling techniques for imbalanced data: An n= 648 ADNI study. NeuroImage **87** (2014) 220–241
16. Ye, J., et al: Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. BMC neurology **12**(1) (2012) 46
17. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B **67**(2) (2005) 301–320
18. Zhang, D., Shen, D.: Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. Neuroimage **59**(2) (2012) 895–907
19. Cheng, B., et al: Domain transfer learning for MCI conversion prediction. In: MICCAI 2012. Springer (2012) 82–90
20. Braak, H., Braak, E.: Neuropathological staging of Alzheimer-related changes. Acta neuropathologica **82**(4) (1991) 239–259