

Early Prediction of Alzheimer’s Disease with Non-local Patch-based Longitudinal Descriptors

Gerard Sanroma¹, Víctor Andrea¹, Oualid M. Benkarim¹, José V. Manjón², Pierrick Coupé^{3,4}, Oscar Camara¹, Gemma Piella¹, and Miguel A. González Ballester^{1,5}

¹ DTIC, Universitat Pompeu Fabra, Barcelona, Spain

² ITACA, Universitat Politècnica de València, Spain

³ Univ. Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France

⁴ CNRS, LaBRI, UMR 5800, F-33400 Talence, France

⁵ ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

`gerard.sanroma@upf.edu`

Abstract. Alzheimer’s disease (AD) is characterized by a progressive decline in the cognitive functions accompanied by an atrophic process which can already be observed in the early stages using magnetic resonance images (MRI). Individualized prediction of future progression to AD, when patients are still in the mild cognitive impairment (MCI) stage, has potential impact for preventive treatment. Atrophy patterns extracted from longitudinal MRI sequences provide valuable information to identify MCI patients at higher risk of developing AD in the future. We present a novel descriptor that uses the similarity between local image patches to encode local displacements due to atrophy between a pair of longitudinal MRI scans. Using a conventional logistic regression classifier, our descriptor achieves 76% accuracy in predicting which MCI patients will progress to AD up to 3 years before conversion.

Keywords: early AD prediction, non-local patch-based label fusion, longitudinal analysis

1 Introduction

Alzheimer’s disease (AD) is characterized by a progressive decline of the cognitive abilities. Before being diagnosed as probable AD, patients usually go through a mild cognitive impairment (MCI) stage. The earliest signs of neurodegeneration can be observed using magnetic resonance images (MRI) already at the MCI stage [7]. Machine learning techniques have taken advantage of this fact to characterize individuals at different stages of the disease. Cuingnet *et al.* [6] presented a comparison of 10 methods for discrimination of healthy controls (HC) and AD patients with different degrees of neurodegeneration. The most common MRI-based features used for discrimination include tissue probability maps, cortical thickness, hippocampal morphometry or a combination of them [14].

Among the personalized medicine approaches related to AD, the discrimination between patients that will remain stable in the MCI stage (i.e., s-MCI)

and the ones that will progress to AD in the future (i.e., p-MCI) is possibly the one with most potential impact. Successful early identification of p-MCI patients opens up the possibility for improving clinical trials aimed at assessing preventive care treatments. Moradi *et al.* [9] specifically focused on the discrimination of p-MCI vs. s-MCI patients up to 3 years prior to conversion. Tong *et al.* [12] also identified p-MCI subjects depending on their similarity with a pre-defined dictionary containing both HC and AD subjects. This latter work was inspired by an hippocampal grading method by Coupé *et al.* [4] (i.e., SNIPE) that assessed hippocampal abnormality based on local similarities to a pre-defined training library. The grade produced by SNIPE could discriminate s-MCI vs. p-MCI with high accuracy.

The structures in the medial temporal lobe (MTL), including the hippocampus, are among the first ones to be atrophied during the early stages of AD [11]. Strictly speaking, atrophy can only be measured using repeated acquisitions from the same subject over time (rather than using a single MRI, as in the above methods). Several approaches agree in finding atrophy rates in the MTL structures following the trend $AD > MCI > HC$ [2, 3]. However, these approaches are not designed for personalized predictions at the individual level and can only reveal the general trends in the population.

We propose a novel method to describe, with a high level of detail, the atrophy patterns across a pair of MRI scans from the same subject at different time points. The proposed descriptor is suitable for being used by machine learning techniques for personalized medicine. Inspired by patch-based label fusion in multi-atlas segmentation [5], our descriptor computes local patch-wise similarities between baseline and follow-up images. Therefore, one-to-many correspondences are used to encode local displacements. For the early prediction of AD, we feed the proposed high-dimensional descriptors extracted from the hippocampal region to a conventional logistic regression classifier.

Other learning-based methods use longitudinal data to predict AD in the first stage of the pathology. Zhu *et al.* [15] proposed a constrained SVM specifically designed for longitudinal data. Jie *et al.* [8] proposed a constrained regression for the prediction of the evolution of cognitive scores in AD patients. The main difference between these methods and the proposed one is that the former ones propose longitudinally-aware classifiers that use conventional MRI images, whereas we propose longitudinally-specific descriptors that can be used by conventional classifiers.

2 Method

We present an atrophy descriptor between a pair of baseline B_i and follow-up F_i images for the i -th patient, aimed at capturing the subtle atrophy patterns discriminating s-MCI and p-MCI patients. In order to bring the pair of images into correspondence while still preserving the local differences due to atrophy, we affinely register the follow-up image to its baseline. We use the notation $F_i^{\rightarrow B}$ to denote that the follow-up image has been registered to its corresponding

baseline. In the case that we have more than one follow-up image per patient, we can divide the entire sequence into a set of pairs baseline / follow-up and treat them independently. The proposed method is divided in the following steps: (i) defining the region-of-interest (ROI), (ii) computing the patch-based similarity maps, (iii) building the atrophy descriptor and (iv) learning the classifier.

2.1 Region of Interest

We extract the high-dimensional atrophy descriptors from a ROI around the hippocampus. As shown in the literature, the hippocampus is among the first regions to be atrophied due to AD [11] and therefore it is a reasonable choice as ROI for early prediction of AD [5, 6]. We propagate the hippocampal ROI, denoted as Ω , from a template image T onto each baseline B_i using spatial warpings $\mathcal{T}_{T \rightarrow B_i}$ obtained via non-rigid image registration. The hippocampal ROI in the template was computed by dilating (with a structuring element of $3 \times 3 \times 3$) the hippocampal segmentation obtained through multi-atlas segmentation [5]. Fig. 1 shows the hippocampal ROI (in red) overlaid onto the template.



Fig. 1. Hippocampal ROI (in red) overlaid onto a template image.

Finally, let us denote as $(B_i(x), F_i^{\rightarrow B}(x))$, $x \in \Omega_i$, the pair of voxel intensities at corresponding location x within the ROI Ω_i in the baseline and follow-up images of a given subject.

2.2 Patch-based Similarity Maps

We encode the atrophy patterns as one-to-many correspondences between each point in the baseline $x \in \Omega_i$ and the neighboring points in the follow-up $x' \in \mathcal{N}_x$, where \mathcal{N}_x is a cubic neighborhood of size s^3 around point x . This gives high-dimensional information about local displacements undergone by each point between the two scans. We use the similarity between image patches to compute the local correspondences, where the patches centered at x and x' in the baseline and follow-up images are defined respectively as $\mathcal{P}_i^B(x)$, $\mathcal{P}_i^F(x')$. In our experiments we use a patch size of $3 \times 3 \times 3$.

For each subject, we compute a set of similarity maps $W_i^{(j)}$, $j = 1 \dots s^3$, one for each offset in the cubic neighborhood \mathcal{N}_x , as follows:

$$W_i^{(j)}(x) = \exp\left(\frac{-\|\mathcal{P}_i^B(x) - \mathcal{P}_i^F(\mathcal{N}_x(j))\|_2^2}{h^2}\right) \quad (1)$$

where $\mathcal{N}_x(j)$ is the j -th offset in the cubic neighborhood and we use the exponential of the negative sum of squared differences as measure of patch similarity with a normalization constant $h = \sum_j \|\mathcal{P}_i^B(x) - \mathcal{P}_i^F(\mathcal{N}_x(j))\|_2$. Fig. 2 shows examples of similarity maps across each neighbor offset in a cubic 27-neighborhood.

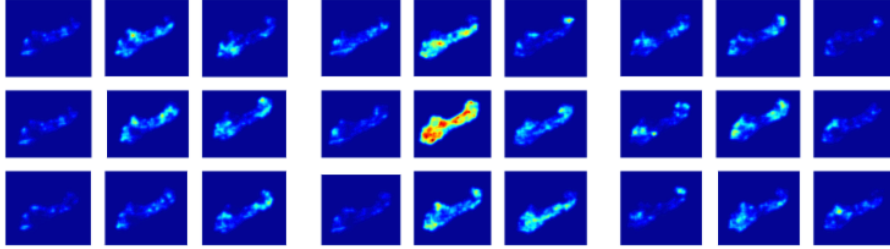


Fig. 2. Each of the 27 tiles shows the similarity map for a different neighbor, with red and blue denoting higher and lower similarities, respectively. We have used a cubic neighborhood of size $s^3 = 3 \times 3 \times 3 = 27$. Tiles in each group of 9 are coherently placed according to their neighborhood offset within the sagittal plane. The three groups correspond to neighbors along the sagittal axis.

2.3 Atrophy Descriptors

The proposed atrophy descriptors are built by encapsulating the similarity maps $W_i^{(j)}$ into feature vectors according to the following steps:

1. We spatially align to a reference space the similarity maps, denoted as $\tilde{W}_i^{(j)}$, using the inverse non-rigid transformations between template and baselines $\mathcal{T}_{T \rightarrow B_i}^{-1}$ (recall that similarity maps originally lie in the space of their baseline images).
2. To compensate for moderate registration errors, we smooth the warped similarities using a Gaussian kernel of width σ .
3. We build the longitudinal atrophy descriptor for i -th subject, denoted as \mathbf{z}_i , by concatenating the similarities across ROI locations and neighbors, i.e., $\{\tilde{W}_i^{(j)}(x) | x \in \Omega, j = 1 \dots s^3\}$ (in practice, we subsample the locations with a step size ρ along each dimension in order to reduce redundancy and decrease the vector's length).

The length of the final vector is approximately $|\Omega| \cdot s^3 / \rho^3$. In our experiments we set $\sigma = 1.0$, $s^3 = 27$ and $\rho = 2$.

2.4 Learning

Given the atrophy descriptors computed in the previous section in a population of training subjects $\{\mathbf{z}_i, i = 1 \dots n\}$, we learn a logistic regression classifier to predict the future outcome of each patient, denoted as $y_i = \{-1, 1\}$, for s-MCI and p-MCI, respectively. Prior to learning, we select the most important features by training a random forest classifier with 1000 trees on the future outcome of each patient. As input to the logistic regression classifier, we only use the features with an importance above $0.5 \cdot \mu$, where the importance is computed according how much a feature decreases the average impurity on the forest and μ is the average importance across features⁶. After the feature selection step, we train a logistic regression classifier by minimizing the empirical loss over our training data subject to some regularization constraint. We define the optimization as:

$$\min_{\mathbf{v}, b} \sum_i^n \frac{1}{1 + \exp(-y_i (\mathbf{v}^\top \mathbf{z}'_i + b))} + \lambda \|\mathbf{v}\|_1, \quad (2)$$

where \mathbf{v} and b are the parameters of the logistic regression classifier and \mathbf{z}'_i is the vector of selected features from the i -th subject. The first term penalizes the classification errors and the second term, modulated by the scalar λ , enforces the sparseness of the coefficients-vector \mathbf{v} through the L_1 -norm. The sparsity regularization is suitable when the number of features is much larger than the number of training samples, as in our case. Given the atrophy descriptor extracted from a new testing subject \mathbf{z} , first we obtain \mathbf{z}' by picking the most important features as determined during training and then we classify it as p-MCI or s-MCI according to the output of the function: $\text{sign}(f(\mathbf{z}'; \mathbf{v}, b))$, where $f(\cdot)$ is the learned logistic regression classifier.

3 Experiments

We evaluate our method in classification experiments between MCI patients that remain stable (i.e., s-MCI) and MCI patients that will progress to AD in the following 3 years (i.e., p-MCI). We use the same subset of ADNI⁷ as in [9, 12], containing 164 p-MCI and 100 s-MCI subjects⁸.

We use the first scan (i.e., baseline, B_i) and second scan (i.e., follow-up F_i) of each subject in order to compute the atrophy descriptors. Images are corrected for inhomogeneities with the N4 algorithm [13] and their histograms linearly matched to a reference template [10]. Follow-up images F_i are affinely registered

⁶ This is implemented in the `feature_importance_` attribute of the random forest classifier in `scikit-learn` package in Python

⁷ <http://www.adni-info.org/>

⁸ More details at: <https://sites.google.com/site/machinelearning4mcitoad/>

to their respective baselines B_i with ANTs [1]. Subsequently, also with ANTs, we compute non-rigid spatial transformations from the MNI152 template to each of the baselines $\mathcal{T}_{T \rightarrow B_i}$.

For each subject, we build the atrophy descriptors as follows:

1. The hippocampal ROI is propagated from the template to each of the baselines B_i , as described in Sec. 2.1.
2. Similarity maps $W_i^{(j)}$ are computed using baseline and registered follow-up scans, as described in Sec. 2.2.
3. Atrophy descriptors \mathbf{z}_i are built after smoothing and subsampling the warped similarity maps $\tilde{W}_i^{(j)}$, as described in Sec. 2.3.

Alternatively, we also compute more compact representations by decomposing the similarities at each point through PCA. We took the first 10 components explaining $> 90\%$ of the variance of the data.

Table 1 shows the average classification accuracy obtained by logistic regression with the proposed atrophy descriptors in 10-fold cross-validation experiments (with and without PCA decomposition) for a range of regularization strengths λ .

	$\lambda = 0.1$	$\lambda = 1.0$	$\lambda = 10$	$\lambda = 50$	$\lambda = 100$	$\lambda = 200$	$\lambda = 300$	$\lambda = 500$
original	0.709	0.744	0.745	0.745	0.766	0.757	0.754	0.753
PCA	0.715	0.732	0.742	0.737	0.733	0.737	0.713	0.741

Table 1. Accuracy of the proposed method in 10-fold cross-validation classification of s-MCI and p-MCI subjects for increasing regularization strengths. First row corresponds to the original proposed descriptor. Second row corresponds to the proposed descriptor with an additional PCA decomposition step.

As we can see in Table 1, the additional PCA decomposition step degrades the discrimination accuracy, thus suggesting that some important information may be lost after the linear decomposition in the present application. Note that, even we only sacrifice 10% of the variance of the data in the PCA decomposition, we may also lose some structure imposed by the normalization of the similarities (i.e., similarities may not add up to one after the reconstruction).

For comparison, in Table 2 we show the results reported by state-of-the-art methods in s-MCI vs. p-MCI classification using MRI features (including some methods using the same dataset as ours).

Comparing results in Table 2, we can see that our method achieves state-of-the-art performance. Our results are directly comparable to [9, 12], since we use the same dataset. It is worth noting that [9, 12] use the whole brain whereas we only use the hippocampal ROI. On the other hand, we use a pair of baseline and follow-up scans whereas [9, 12] only use single baseline scan for classification. Coupé *et al.* [4] also focused on the hippocampal ROI (including enthorinal cortex), suggesting that this area might convey important information for early

Method	Cross-sectional				Longitudinal		
	Moradi [9] *	Tong [12] *	Coupé [4]	Wolz [14]	Zhu [15]	Jie [8]	proposed
Perform.	0.747	0.789	0.74	0.68	0.76–0.84	0.757	0.766

Table 2. Performance in s-MCI vs. p-MCI classification of state-of-the-art methods using only MRI features. The former 4 methods in the table use only a single baseline MRI for classification whereas the latter 3 use at least one longitudinal follow-up as well. See the main text for details about performances reported as an interval. Methods with an asterisk (*) have been evaluated in the same dataset as the proposed method.

AD classification [11]. As another difference, results of [4, 14] correspond to early prediction of up to 4 years before conversion, whereas our results (as well as those in [9, 12]) correspond to prediction up to 3 years before conversion. The rest of longitudinal methods (i.e., [15, 8]) use at least 4 follow-up scans for each subject, whereas we use only 1 baseline and 1 follow-up scan. Zhu *et al.* [15] discriminate between progression to AD at intervals 18, 12, 6 and 0 months with accuracies 0.76, 0.81, 0.83 and 0.84, respectively, hence the interval 0.76 – 0.84 in the table.

4 Conclusions

We have presented a high-dimensional atrophy descriptor for early AD prediction using longitudinal MRI data. We achieve state-of-the-art performance by feeding our proposed descriptor to a conventional logistic regression classifier. Results suggest that our descriptor is suitable for capturing subtle atrophic patterns distinguishing s-MCI and p-MCI patients up to 3 years before conversion. Indeed, the hippocampal ROI is a suitable region for prediction in the early stages because it is among the first areas revealing atrophy due to AD [11]. Other methods have also focused in this ROI, achieving comparable performance to methods using the whole brain [4]. Effective ways to reduce the dimensionality should be explored in order to extend the use of the proposed descriptor to larger areas of the brain. Results in [15] suggest room for improvement in using the full sequence of follow-up scans (instead of the first 2 ones). Possible lines of future work include combining our descriptor extracted from full follow-up sequences with longitudinally-aware classifiers.

Acknowledgements The first author is co-financed by the Marie Curie FP7-PEOPLE-2012-COFUND 462 Action. Grant agreement no: 600387

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* 12(1), 26–41 (2008)

2. Cash, D.M., Frost, C., Iheme, L.O., Ünay, D., Kandemir, M., Fripp, J., Salvado, O., Bourgeat, P., Reuter, M., Fischl, B., Lorenzi, M., Frisoni, G.B., Pennek, X., Peirson, R.K., Gunter, J.L., Senjem, M.L., Jack, C.R., Guizard, N., Fonov, V.S., Collins, D.L., Modat, M., Cardoso, M.J., Leung, K.K., Wang, H., Das, S.R., Yushkevich, P.A., Malone, I.B., Fox, N.C., Schott, J.M., Ourselin, S.: Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge. *NeuroImage* 123, 149–164 (2015)
3. Chincarini, A., Sensi, F., Rei, L., Gemme, G., Squarcia, S., Longo, R., Brun, F., Tangaro, S., Bellotti, R., Amoroso, N., Bocchetta, M., Redolfi, A., Bosco, P., Boccardi, M., Frisoni, G.B., Nobili, F.: Integrating longitudinal information in hippocampal volume measurements for the early detection of Alzheimer’s disease. *NeuroImage* 125, 834–847 (2016)
4. Coupé, P., Eskildsen, S.F., Manjón, J.V., Fonov, V.S., Pruessner, J.C., Allard, M., Collins, D.L.: Scoring by nonlocal image patch estimator for early detection of Alzheimer’s disease. *NeuroImage: Clinical* 1, 141–152 (2012)
5. Coupé, P., Manjón, J.V., Fonov, V.S., Pruessner, J., Robles, M., Collins, D.L.: Patch-based Segmentation using Expert Priors: Application to Hippocampus and Ventricle Segmentation. *NeuroImage* 54(2), 940–954 (2011)
6. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O.: Automatic classification of patients with Alzheimer’s disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage* 56, 766–781 (2011)
7. Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurol.* 6(2), 67–77 (2010)
8. Jie, B., Liu, M., Zhang, D., Shen, D.: Temporally-Constrained Group Sparse Learning for Longitudinal Data Analysis in Alzheimer’s Disease. *IEEE Transactions on Biomedical Engineering* 64(1), 238–249 (2015)
9. Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J.: Machine learning framework for early MRI-based Alzheimer’s conversion prediction in MCI subjects. *NeuroImage* 104, 398–412 (2015)
10. Nyúl, L.G., Udupa, J.K.: On standardizing the mr image instensity scale. *Magnetic Resonance in Medicine* 42(6), 1072–1081 (1999)
11. Thompson, P.M., Hayashi, K.M., de Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., Toga, A.W.: Dynamics of gray matter loss in Alzheimer’s disease. *Journal of Neuroscience* 23(3), 994–1005 (2003)
12. Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D.: A Novel Grading Biomarker for the Prediction of Conversion From Mild Cognitive Impairment to Alzheimer’s Disease. *IEEE Transactions on Biomedical Engineering* 64(1), 155–165 (2017)
13. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging* 29(6), 1310–1320 (2010)
14. Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D.P., Rueckert, D., Soininen, H., Lötjönen, J.: Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer’s Disease. *PLOS ONE* 6(10) (2011)
15. Zhu, Y., Zhu, X., Kim, M., Shen, D., Wu, G.: Early Diagnosis of Alzheimer’s Disease by Joint Feature Selection and Classification on Temporally Structured Support Vector Machine. In: *MICCAI* (2016)