Adaptive fusion of texture-based grading: Application to Alzheimer's disease detection

Kilian Hett[1,2], Vinh-Thong Ta[1,2,3], José V. Manjón[4], and 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] Bordeaux INP, LaBRI, UMR 5800, PICTURA, F-33600 Pessac, France
[4] Universitat Politècnia de València, ITACA, 46022 Valencia, Spain

Abstract. Alzheimer's disease is a neurodegenerative process leading to irreversible mental dysfunctions. The development of new biomarkers is crucial to perform an early detection of this disease. Among new biomarkers proposed during the last decades, patch-based grading framework demonstrated state-of-the-art results. In this paper, we study the potential using texture information based on Gabor filters to improve patch-based grading method performance, with a focus on the hippocampal structure. We also propose a novel fusion framework to efficiently combine multiple grading maps derived from a Gabor filters bank. Finally, we compare our new texture-based grading biomarker with the state-of-the-art approaches to demonstrate the high potential of the proposed method.

Keywords: Patch-based grading fusion, multi-features, Alzheimer's disease classification, Mild Cognitive Impairment

1 Introduction

Alzheimer's disease (AD) is the most prevalent dementia. AD is characterized by an irreversible neurodegeneration leading to mental dysfunctions. Subjects with Mild Cognitive Impairment (MCI) present higher risk to develop AD. To date, diagnosis of AD is established after advanced brain structure alterations motivating the crucial need to develop new imaging biomarkers able to detect the early stages of the disease. Furthermore, the early detection of AD can accelerate the development of new therapies by making easier the design of clinical trials. During the last decades, new biomarkers with competitive performances were developed to detect AD by taking advantage of the improvement of medical imaging like magnetic resonance imaging (MRI) [1].

Most of the proposed methods have been based on specific regions of interest (ROI). Among structures impacted by AD, previous investigations mainly focused on medial temporal lobe and especially on hippocampus (HC). Alterations on this structure are usually estimated using volume, shape or cortical thickness measurements [11]. Besides ROI-based methods, whole brain analyses performed on structural MRI (s-MRI) have also been proposed to detect areas impacted by AD. These methods are usually based on voxel-based morphometry (VBM) or tensor based morphometry (TBM) frameworks. It is interesting to note that both VBM and ROI-based studies confirmed that medial temporal lobe is a key area to detect the first signs of AD [11]. In the medial temporal lobe, the HC is one of the earliest region altered by AD. Recently, advanced methods were proposed to capture structural alterations of HC. Those techniques demonstrated their efficiency to detect the different stages of AD [8]. Among them, patch-based methods obtained competitive results to detect the earliest stages of AD [7, 2, 9]. Therefore, such advanced image analysis methods seem promising candidates to perform AD tracking. Recently, [4] demonstrated the efficiency of using edge detection filters to improve of patch-based segmentation. This result highlights that patches comparison can be improved by estimating patterns similarity on derivative image features. Moreover, it has been recently showed that HC texture plays a crucial role for the detection of early stages of AD [8]. Therefore, we propose to perform patch-based grading on multiple texture maps obtained with Gabor filters. Gabor filters are designed to detect salient features at specific resolution and direction. These filters were widely used for texture classification [13]. Consequently, the proposed strategy enables at the same time to improve patches comparison and to capture HC texture modifications occurring at the first stages of the pathology.

Contributions: The first contribution of this work is intended to develop a new texture-based grading framework to better capture structural alterations caused by AD. Secondly, in order to combine all the grading maps estimated on texture maps, we propose an innovate adaptive fusion strategy based on local confidence criterion. This fusion framework can be applied to any patch-based processing to combine different features or modalities. Moreover, contrary to usual grading-based methods, we propose a classification step involving weak classifiers distribution to better discriminate pathologies stages. Finally, to highlight the improvement of classification performances provided by our new framework, we compare our new biomarker with the state-of-the-art biomarkers and demonstrate its efficiency.

2   Materials and Methods

2.1   Dataset

Data used in this work were obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset[1]. ADNI is a North American campaign launched in

---

[1] http://adni.loni.ucla.edu

2003 with aims to provide MRI, positron emission tomography scans, clinical neurological measures and other biomarkers. The data used in this study are all the baseline T1-weighted (T1-w) MRI of the ADNI1 phase. This dataset includes AD patients, MCI and cognitive normal (CN) subjects. The group of MCI is composed of subjects who have abnormal memory dysfunctions and embed two groups, the first one is composed with patients having stable MCI (sMCI) and the second one is composed with patients with progressive MCI (pMCI). The information of the dataset used in our work is summarized in Table 1.

Table 1. Description of the dataset used in this work. Data are provided by ADNI.

Characteristic / Group	CN	sMCI	pMCI	AD
Number of subjects	226	223	165	186
Ages (years)	76.0 ± 5.0	75.1 ± 7.5	74.5 ± 7.2	75.3 ± 7.4
Sex (M/F)	117/109	150/73	101/64	98/88
MMSE	29.05 ± 0.9	27.1 ± 2.5	26.3 ± 2.0	22.8 ± 2.9

2.2 Preprocessing

All the T1-w images were processed using the volBrain system [12][4]. This system is based on an advanced pipeline providing automatic segmentation of different brain structures from T1-w MRI. The preprocessing is based on: (a) a denoising step with an adaptive non-local means filter, (b) an affine registration in the MNI space, (c) a correction of the image inhomogeneities and (d) an intensity normalization.

2.3 Methods

Patch-based grading: Grading framework uses patch-based techniques to capture modifications related to anatomical degradations caused by AD [2]. To date, patch-based grading methods demonstrate state-of-the-art performances to detect the earliest stages of AD [6, 10]. To determine the pathological status of a subject, grading-based methods estimate at each voxel the state of cerebral tissues using anatomical patterns extracted from a training library T composed of two datasets, one with images from CN subjects and one with AD patients. Then, for each voxel of the considered subject, the patch-based grading method produces a weak classifier denoted g. This weak classifier is based on the similarity between the patch surrounding the voxel under study x_i and a set K_i of similar patches extracted from T. In this work, we used an approximative nearest neighbor method to drastically reduce the required computational time [5]. The grading value g at x_i is defined as:

$$g(x_i) = \frac{\sum_{x_{j,t} \in K_i} w(x_i, x_{j,t}) p_t}{\sum_{x_{j,t} \in K_i} w(x_i, x_{j,t})} \tag{1}$$

---

[4] http://volbrain.upv.es

where $x_{j,t}$ is the voxel j belonging to the training template $t \in T$. $w(x_i, x_{j,t})$ is the weight assigned to the pathological status p_t of t. We estimate w such as:

$$w(x_i, x_{j,t}) = e^{1 - \frac{(d(x_i, x_{j,t}))^2}{h^2 + \epsilon}} \qquad (2)$$

where $h = \min_{x_{j,t}} d(x_i, x_{j,t})$ with $\epsilon \to 0$, d is a distance between two patches surrounding the voxels x_i and $x_{j,t}$. p_t is set to -1 for patches extracted from AD patient and to 1 for those extracted from CN subject. The L2-norm is used to estimate the similarly between patches. Thus, our patch-based grading method provides at each voxel a score representing an estimation of the alterations caused by AD.

Texture maps estimation: The estimation of patch similarities could be improved by using texture representation instead of using raw intensities. Indeed, it was demonstrated that the use of edge detectors improves patch-based segmentation accuracy [4]. Moreover, it was also demonstrated that HC textural information plays an important role in AD detection [8]. Hence, we propose a new texture-based grading framework that simultaneously captures HC texture alterations and improves patches similarity estimation. In this work, texture information is extracted from MRI using a bank of 3D Gabor filters. We used Gabor filters since they are designed to detect texture patterns at different scales and directions [13]. In the proposed pipeline (see Fig. 1), the preprocessed MRI of the subject under study is filtered with a bank of Gabor filters to obtain multiple texture maps. It has to be noted that all the training library is also filtered with the same filters bank. Therefore, for each texture map, a texture-based grading map can be estimated.

Adaptive fusion: In this work, we propose an novel framework to fuse the multiple texture-based grading maps obtained from the estimated texture maps. Our fusion strategy is based on the fact that all the estimated grading maps may not have the same relevance, but more importantly all local weak classifiers in these maps do not have the same quality. Hence, at each location, we propose to combine weak classifiers derived from multiple texture maps according to a confidence criterion. Therefore, the grading value of a texture-based grading map m, denoted g_m, at voxel x_i, is weighted by $\alpha_m(x_i) = \sum_{x_{j,t} \in K_{i,m}} w_m(x_i, x_{j,t})$ that reflects the confidence of g_m. Thus, each texture-based grading map provides a weak classifier at each voxel that is weighted with its degree of confidence $\alpha_m(x_i)$. At the end, the final grading value, denoted g_M, resulting from our adaptive fusion strategy is given by;

$$g_M(x_i) = \frac{\sum_{m \in M} \alpha_m(x_i) g_m(x_i)}{\sum_{m \in M} \alpha_m(x_i)} \; . \qquad (3)$$

The proposed fusion framework is spatially adaptive and take advantage of having access to a local degree of confidence $\alpha_m(x_i)$ for each grading map m. Basically, the confidence $\alpha_m(x_i)$ gives more weight to a weak classifier estimated

with a well matched set of patches. Our adaptive fusion strategy can applied to any patch-based processing to combine multiple feature or modalities.

Weak classifiers aggregation: First, to prevent bias introduced by structure alterations related to aging, all the grading values are age corrected with a linear regression based on the CN group [3]. In previous works on patch-based grading [2, 5], the weak classifier aggregation was performed using a simple averaging. While using a strategy based on averaging enables to be robust to noise, this may remove relevant information on weak classifiers distribution. Therefore, in this paper we propose to approximate weak classifiers distribution using histogram. Consequently, we classify histogram bins instead of classifying mean grading value over the segmentation mask. Here, histograms were separately estimated for right and left hippocampus.

Validation: During our experiments, texture maps were obtained using one scale and 3 orthogonal directions. The texture-based grading maps were estimated using patches of $5 \times 5 \times 5$ voxels. The grading step based on an optimized PatchMatch [5] was performed using $K = 50$. The required computational time was 3s per texture maps, thus the global grading step required 10s with our setup. Our new texture-based grading framework was validated with a leave-one-out cross validation procedure. A support vector machine (SVM) was used to classify each test subject. The results of each experiment were compared in terms of accuracy (ACC) and area under the ROC curve (AUC). The AUC is estimated with the *a posteriori* probabilities provided by the SVM classifier. We carried out several experiments: CN vs. AD, CN vs. pMCI, AD vs. sMCI and sMCI vs. pMCI.

3 Results

Table 2. Comparison of different features HC-based, all results are expressed in AUC.

Features	CN vs. AD (AUC in %)	CN vs. pMCI (AUC in %)	AD vs. sMCI (AUC in %)	sMCI vs. pMCI (AUC in %)
T1-w grading	93.5	90.0	81.1	73.6
Proposed method	**94.2**	**90.9**	**81.3**	**75.4**

Firstly, in order to validate the improvement provided by our method, we compare results obtained with our framework using raw intensities (T1-w grading) and texture maps. T1-w and texture-based grading were estimated using exactly the same pipeline involving adaptive fusion and histogram-based weak classifiers aggregation. Table 2 summarizes the results of T1-w grading and our proposed method. Results are expressed with area under the curve (AUC) measure. As it is shown, texture-based grading improves classification performances
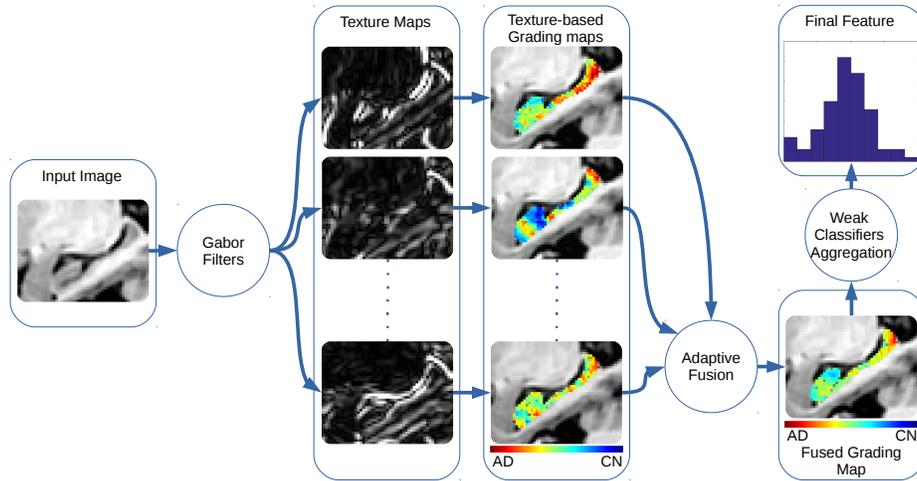
Fig. 1. Proposed adaptive fusion of texture-based grading framework.: from left to right, the T1-w input data, the texture maps for different directions, the intermediate texture-based grading maps, the final fused grading map and the histogram-based weak classifiers aggregation.

in all experiments especially MCI classification with 94.2% of AUC in CN vs. AD, 90.9% of AUC in the CN vs. pMCI, 81.3% of AUC in AD vs. sMCI and 75.4% of AUC in sMCI vs. pMCI comparisons. During our experiments, weak classifiers aggregation based on histogram did not provide improvement in CN vs. AD comparison. That could be explained by the fact that CN and AD distributions are separated. However, in sMCI vs. pMCI case, the two distributions are less separable and histogram representation yielded to better classification performances. The experiments carried out showed that the use of only one scale is enough. Moreover, using more than 3 directions did not improve the results while increasing computational time.

Table 3 summarizes the comparison of our proposed method with other grading methods proposed in the literature. In addition, classification results obtained with Deep Learning (DL) [14] ensemble are provided for comparison with last advanced methods. The results on Table 3 are expressed in accuracy (ACC). First, to compare classification results using the same structure, the proposed framework is compared with grading methods based on HC (see the upper part of Table 3). This comparison shows that our method provides the best results among HC-based grading methods. It reaches 91.3% of ACC for CN vs. AD, and 71.1% of ACC for sMCI vs. pMCI comparisons. These results demonstrate that texture maps provide valuable information during the grading process. At the lower part of Table 3, we compare the performance of our HC-based grading method with those using the whole brain.

First, for AD vs. CN, the proposed method obtained similar or better results

than those using whole brain and requiring non linear registration [7] while our method only requires affine registration and proposes a fast grading step (i.e., 10s).

Second, for sMCI vs. pMCI, our method obtained better results than all the methods involving a simple affine registration [10]. On the other hand, the best results for sMCI vs. pMCI are produced by whole brain grading [6, 10] using non linear registration. The improvement when using non linear registration is observed for HC-based and whole brain methods [10]. However, this improvement is obtained at the expense of using non linear registration, which is subject to failure and requires high computational time.

Finally, our method also demonstrated competitive performances for AD vs. CN classification compared to the most advanced DL methods using whole brain and non linear registration. In addition, this comparison shows that patch-based grading methods [6, 10] obtain similar or better results than recent DL methods [14] when applied with similar settings.

To conclude, according to our comparison, whole brain methods enable a better classification of sMCI vs. pMCI. Hence, in further works, we will investigate the extension of our texture-based grading framework to whole brain analysis.

Table 3. Comparison with state-of-the-art methods, all the results are expressed in accuracy.

Methods	Registration	Features	CN vs. AD (ACC in %)	sMCI vs. pMCI (ACC in %)
Hippocampus				
Original Grading [2]	Affine	Intensity	88.0	71.0
Multiple Instance Grading [9]	Affine	Intensity	89.0	70.0
Sparse-based Grading [10]	Affine	Intensity	—	66.0
Sparse-based Grading [10]	Non Linear	Intensity	—	69.0
Proposed Method	Affine	Texture	**<u>91.3</u>**	**71.1**
Whole Brain				
Ensemble Grading [6]	Non Linear	GM Map	—	**<u>75.6</u>**
Sparse-based Grading [10]	Affine	Intensity	—	66.7
Sparse-based Grading [10]	Non Linear	Intensity	—	75.0
Sparse Ensemble Grading [7]	Non Linear	GM Map	90.8	—
Deep Ensemble Learning [14]	Non Linear	GM Map	91.0	74.8

4   Conclusion

In this work we propose a new texture-based grading framework to better capture structural alterations caused by AD. Moreover, to combine grading maps estimated on texture maps, we present a new adaptive fusion scheme. We also propose an histogram-based weak classifiers aggregation step to better discriminate early stages of AD. Finally, we demonstrate the competitive performances of

our new texture-based grading framework compared to several state-of-the-art biomarkers. In future works, we will investigate the extension of our texture-based grading framework to whole brain analysis.

5 Acknowledgement

References

[1] Bron et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge. *NeuroImage*, 111:562–579, 2015.

[2] Coupé et al. Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. *NeuroImage: clinical*, 1(1):141–152, 2012.

[3] Dukart et al. Age correction in dementia–matching to a healthy brain. *PloS one*, 6(7):e22193, 2011.

[4] Giraud et al. An optimized patchmatch for multi-scale and multi-feature label fusion. *NeuroImage*, 124:770–782, 2016.

[5] Hett et al. Patch-based DTI grading: Application to Alzheimer's disease classification. In *International Workshop on Patch-based Techniques in Medical Imaging*, pages 76–83. Springer, 2016.

[6] Komlagan et al. Anatomically constrained weak classifier fusion for early detection of Alzheimer's disease. In *International Workshop on Machine Learning in Medical Imaging*, pages 141–148. Springer, 2014.

[7] Liu et al. Ensemble sparse classification of Alzheimer's disease. *NeuroImage*, 60(2):1106–1116, 2012.

[8] Sørensen et al. Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *NeuroImage: Clinical*, 2016.

[9] Tong et al. Multiple instance learning for classification of dementia in brain MRI. *Medical image analysis*, 18(5):808–818, 2014.

[10] Tong et al. 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.

[11] Wolz et al. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. *PloS one*, 6(10):e25446, 2011.

[12] Manjón and Coupé. volbrain: An online MRI brain volumetry system. *Frontiers in neuroinformatics*, 10, 2016.

[13] Bangalore S et al. Manjunath. Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8):837–842, 1996.

[14] Heung-Il et al. Suk. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Medical image analysis*, 37:101–113, 2017.