HAL
open science

Adaptive fusion of texture-based grading for Alzheimer's disease classification

Kilian Hett, Vinh-Thong Ta, José V Manjón, Pierrick Coupé

To cite this version:

Kilian Hett, Vinh-Thong Ta, José V Manjón, Pierrick Coupé. Adaptive fusion of texture-based grading for Alzheimer's disease classification. Computerized Medical Imaging and Graphics, In press. hal-01859256

HAL Id: hal-01859256

https://hal.science/hal-01859256v1

Submitted on 21 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptive fusion of texture-based grading for Alzheimer's disease classification

Kilian Hett[a,b], Vinh-Thong TA[a,b,c], José V. Manjón[d], Pierrick Coupé[a,b], Alzheimer's Disease Neuroimaging Initiative[1]

[a] *Univ. Bordeaux, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France*
[b] *CNRS, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France*
[c] *Bordeaux INP, LaBRI, UMR 5800, PICTURA, F-33600 Pessac, France*
[d] *Universitat Politècnia de València, ITACA, 46022 Valencia, Spain*

Abstract

Alzheimer's disease is a neurodegenerative process leading to irreversible mental dysfunctions. To date, diagnosis is established after incurable brain structure alterations. The development of new biomarkers is crucial to perform an early detection of this disease. With the recent improvement of magnetic resonance imaging, numerous methods were proposed to improve computer-aided detection. Among these methods, patch-based grading framework demonstrated state-of-the-art performance. Usually, methods based on this framework use intensity or grey matter maps. However, it has been shown that texture filters improve classification performance in many cases. The aim of this work is to improve performance of patch-based grading framework with the development of a novel texture-based grading method. In this paper, we study the potential of multi-directional texture maps extracted with 3D Gabor filters to improve patch-based grading method. We also proposed a novel patch-based fusion scheme to efficiently combine multiple grading maps. To validate our approach, we study the optimal set of filters and compare the proposed method with different fusion schemes. In addition, we also compare our new texture-based grading biomarker with state-of-the-art methods. Experiments show an improvement of AD detection and prediction accuracy. Moreover, our method obtains competitive performance with 91.3% of accuracy and 94.6% of area under a curve for AD detection.

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

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`.

1. Introduction

Alzheimer's disease (AD) is the most prevalent dementia [27] that is characterized by an irreversible neurodegeneration leading to mental dysfunctions. Patients with AD have a memory lost with difficulty of remembering newly learned information that disrupts daily life [2]. Subjects with Mild Cognitive Impairment (MCI) present higher risk to develop AD [26]. However, before first clinical symptoms brain changes occur such as synaptic and neuronal loses. To date, diagnosis of AD is established after advanced brain structure alterations. This motivates the development of 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, the improvement of medical imaging like magnetic resonance imaging (MRI) led to the development of new biomarkers with competitive performance for AD diagnosis and prognosis [6]. 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 (HIPP). Alterations on this structure are usually estimated using volume [18], shape [1] or cortical thickness measurements [13] (see [36] for a review). Besides ROI-based methods, whole brain analysis performed on anatomical 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 [3]. 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 [36]. In the medial temporal lobe, the HIPP is one of the earliest region altered by AD [16]. Recently, advanced methods were proposed to capture structural alterations of the HIPP [6]. Those techniques demonstrated their efficiency to detect the different stages of AD [28]. Among them, patch-based methods obtained competitive results to detect the earliest stages of AD [21; 9; 32]. Moreover, it has been shown that patch-based methods could detect AD more than seven years before conversion to dementia [10] or can be used for differential diagnosis [19]. Therefore, such advanced image analysis methods seem promising candidates to perform AD prediction.

Patch-based methods are usually based on intensities [9; 33] or grey matter density maps [20]. However, it is has been demonstrated that HIPP texture improves the detection of early stages of AD [28]. A VBM method using several textural filter on medial temporal lobe area demonstrated the reliability of texture information in AD detection [7]. Moreover, HIPP texture enables to improve AD detection compared to HIPP volume [29]. This method could potentially capture MRI signal alterations related to neurofibrillary tangles and beta-amyloid plaque deposition, although such alterations are not directly detectable with MRI at current resolution. Besides, a recent study recently showed the efficiency of using edge detection filters to improve of patch-based segmentation [15]. This result highlights that patch-based grading methods could be

improved by estimating patterns similarity on derivative image features. 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 [25; 12; 14]. The proposed strategy enables to better capture texture modifications occurring at the first stages of the pathology by improving patch comparison.

The first contribution of this work is to propose a new texture-based grading framework to better capture structural alterations caused by AD. This new framework proposes multi-directional texture grading based on 3D Gabor filters. Secondly, in order to combine all the grading maps estimated on texture maps, we propose an innovative adaptive patch-based fusion strategy based on a 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 using the average grading values over the considered ROI, we propose a classification step based on a nonparametric grading values distribution representation to better discriminate pathologies stages. In our experiments, we first study the optimal number of Gabor filter directions for AD detection. In addition, we compare different texture filers such as local variance or entropy. We also compare our new adaptive fusion method with different fusion schemes. Finally, to highlight the improvement of classification performances provided by our new framework, we compare our new method with the state-of-the-art approaches and demonstrate its efficiency.

2. Materials

2.1. Dataset

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

Data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset[1]. ADNI is a North American campaign launched in 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 (T1w) MRI of the ADNI1 phase. This dataset includes AD patients, MCI and cognitive normal (CN) subjects.

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

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), such patients converted to AD during the following 48 months from the baseline [36]. The information of the dataset used in our work is summarized in Table 1.

2.2. Preprocessing

All the T1w images were processed using the volBrain system [22][4]. This system is based on an advanced pipeline providing automatic segmentation of different brain structures from T1w MRI. However, in this work, only hippocampus segmentations were used. The preprocessing is based on: (a) a denoising step with an adaptive non-local means filter [23], (b) an affine registration in the MNI space [4], (c) a correction of the image inhomogeneities [34] and (d) an intensity normalization [24]. Afterwards, MRI were segmented in the MNI space using non-local patch-based multi-atlas methods [8]. The obtained hippocampus were segmented according to the EADC protocol [5] designed for AD studies.

3. Methods

In this section, we describe the different steps of the proposed texture-based grading framework as illustrated in Figure 1. First, we use multi-directional Gabor filters to extract texture in different directions. Second, a patch-based grading method is applied within each texture map computed. Afterwards, all the texture grading maps are merged with our novel adaptive fusion method. Finally, the input of classification method is a histogram representation of texture grading values in each hippocampus.

3.1. 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 [15]. Moreover, it was also demonstrated that HIPP textural information plays an important role in AD detection [28]. Hence, we propose a new texture-based grading framework that simultaneously captures HIPP texture alterations and improves patches similarity estimation. In this work, texture information is extracted from MRI using a bank of 3D Gabor filters (see Figure 2). We used Gabor filters since they are designed to detect texture patterns at different

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

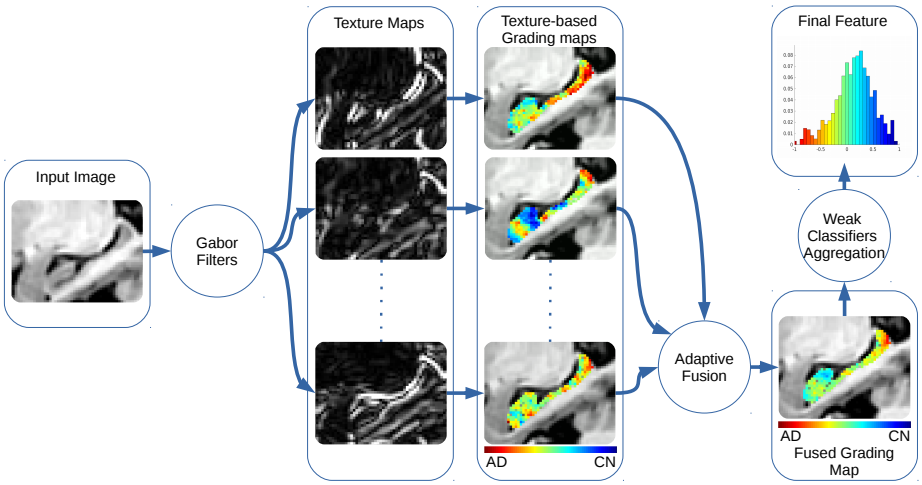

Figure 1: Proposed adaptive fusion of texture-based grading framework: from left to right, the T1w input data, the texture maps for different directions, the intermediate texture-based grading maps, the final fused grading map and the histogram-based weak classifiers aggregation.

scales and directions [25]. Impulse response of the 3D Gabor filter is given by the following equation:

$$h(X,\sigma,F,\theta,\phi) = \frac{1}{(2\pi)^{3/2}\sigma^3}\exp\big(-\frac{x^2+y^2+z^2}{2\sigma^2}\big)\hat{h}(X,F,\theta,\phi) \tag{1}$$

where $X=(x,y,z)$, σ represents the standard deviation of the Gaussian function. (θ,ϕ) are the orientation angles. F represents the central frequency of the frequency response. Finally, $\hat{h}(X,F,\theta,\phi)$ is given by:

$$\hat{h}(X,F,\theta,\phi) = \exp\big(j2\pi F(x\sin\theta\cos\phi + y\sin\theta\sin\phi + z\cos\theta)\big) \tag{2}$$

In our method, the texture maps are the magnitude of the signal resulting from the convolution of our Gabor filters bank with the MRI. 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. 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.

3.2. Patch-based grading

Grading framework uses patch-based techniques to capture signal modifications related to anatomical degradations caused by AD [9]. To date, patch-based grading methods demonstrate state-of-the-art performances to detect the earliest stages of AD [20; 33]. To determine the pathological status of a subject,

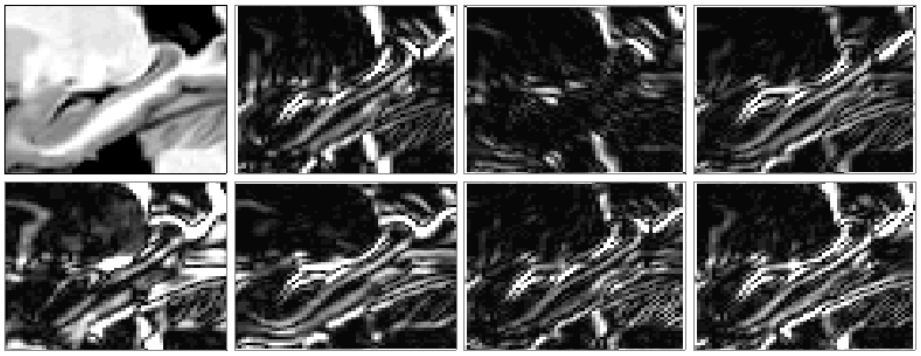

Figure 2: Sample of absolute value of complex number computed by Gabor filters. The hippocampal area is filtered with gabor at different directions. From left: Original image, textural maps provided by our Gabor filters bank given by the parameter set $\{(\theta = 0; \phi = 0), (\pi/2; \pi/2), (0; \pi/2), (0; \pi/4), (0; -\pi/4), (-\pi/4, \pi/4), (\pi/4, \pi/4)\}$. It can be noticed that for each direction, different textures are extracted.

.

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. Next, for each voxel of the considered subject, the patch-based grading method produces weak classifiers 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 find similar patches in the training library in order to drastically reduce the required computational time [17]. The grading value g at x_i is defined as:

$$g_{x_i} = \frac{\sum_{t_j \in K_i} w(P_{x_i}, P_{t_j}) p_t}{\sum_{t_j \in K_i} w(P_{x_i}, P_{t_j})}, \tag{3}$$

where t_j is the voxel j belonging to the training template $t \in T$, P_{x_i} and P_{t_j} are the patches surrounding x_i and t_j, respectively. The function $w(P_{x_i}, P_{t_j})$ is the weight assigned to the pathological status p_t of t. We estimate w such as:

$$w(P_{x_i}, P_{t_j}) = \exp\Big(- \frac{||P_{x_i} - P_{t_j}||_2^2}{h^2 + \epsilon} \Big) \tag{4}$$

where $h = \min_{t_j} ||P_{x_i} - P_{t_j}||_2^2$ with $\epsilon \to 0$, 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.

3.3. Adaptive fusion

In this work, we propose a novel framework to fuse the multiple texture-based grading maps obtained from the estimated texture maps. Our fusion strategy assumes that all the estimated grading maps may not have the same relevance, but more importantly all local weak classifiers g_{x_i} 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 $g_{x_{i,n}}$ at voxel $x_{i,n}$ of a texture-based grading map n , is weighted by:

$$\alpha_{x_{i,n}} = \sum_{t_j \in K_{i,n}} w(P_{x_{i,n}}, P_{t_{j,n}}) \tag{5}$$

that reflects the confidence of $g_{x_{i,n}}$ for the texture-based grading map n, here t_j is the voxel j belonging to the training template $t \in T$ and $K_{i,n}$ is the set of similar patches extracted from the training library T at the voxel x_i of the texture map n. Thus, each texture-based grading map provides a weak classifier at each voxel that is weighted with its degree of confidence $\alpha_{x_{i,n}}$. The final grading value g_{x_i} is given by:

$$g_{x_i} = \frac{\sum_{n \in N} \alpha_{x_{i,n}} g_{x_{i,n}}}{\sum_{n \in N} \alpha_{x_{i,n}}} \ . \tag{6}$$

The proposed fusion framework is spatially adaptive and take advantage of having access to a local degree of confidence $\alpha_{x_{i,n}}$ for each grading map n. Basically, the confidence $\alpha_{x_{i,n}}$ gives more weight to a weak classifier estimated with a well-matched set of patches. This adaptive fusion strategy can applied to any patch-based processing to combine multiple feature or modalities.

3.4. Weak classifiers aggregation

In previous works on patch-based grading, the weak classifier aggregation was performed using a simple averaging [9; 17]. 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 and concatenate into a single feature vector. The number of bins are set following Sturge's formula that intends to find the optimum number of bin for a certain size of sample [30]. Finally, 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 [11]. This correction is done by removing the test CN subjects into a cross-validation procedure.

3.5. Implementation

During our experiments, texture maps were obtained using one scale and 11 different 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 [17] was performed using $K = 50$. The required computational time was 3s per texture maps, thus the global grading step required 15 seconds with our setup. A support vector machine (SVM) with a linear kernel was used to classify each test subject. We used the Matlab function provided by the Statistics and Machine Learning Toolbox. In our experiments, the soft margin parameter C was optimized with a Bayesian optimization method. The results of each experiment were compared in terms of accuracy (ACC) and area under the ROC curve (AUC), specificity (SPE), and sensitivity (SEN). The AUC is estimated with the *a posteriori* probabilities provided by the SVM classifier. We carried out several experiments: CN versus AD, CN versus pMCI, AD versus sMCI and sMCI versus pMCI. A t-test were performed to study the significance of the results provided by adaptive fusion scheme compared to mean of textural maps and late fusion into SVM classifier. Finally, our new texture-based grading framework was validated within a repeated stratified 10-fold cross-validation procedure iterated 50 times for CN versus AD, CN versus pMCI and AD versus sMCI comparisons. The mean ACC, AUC, SPE, and SEN over these 50 iterations are provided as results. As demonstrated in [33], training the classifier with CN and AD enables to discriminate sMCI and pMCI subjects better. Moreover, it enables to perform classification without cross-validation procedure and to limit bias and over-fitting problem. Therefore, only one run was performed for sMCI versus pMCI comparison.

4. Results

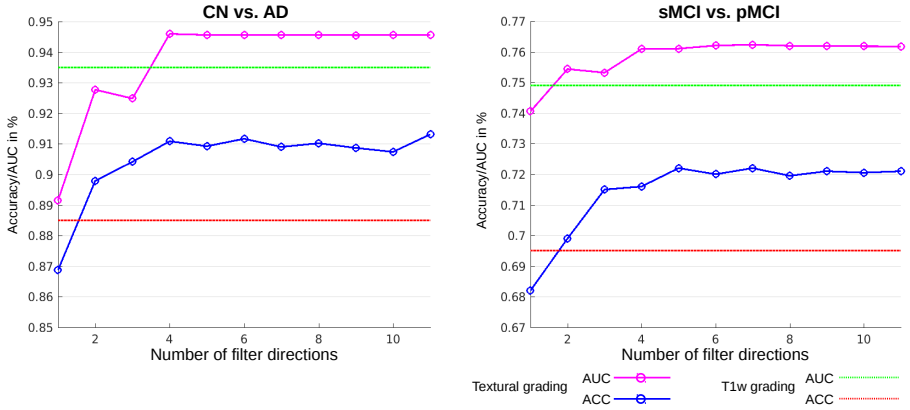

Figure 3: Evolution of the accuracy and AUC for CN versus AD and sMCI versus pMCI comparisons according to the number of Gabor filters. Red and green doted line represented the accuracy and the AUC of intensity-based grading.

Table 2: Comparison of grading features based on histogram representation of the probability distribution of grading values. This table presents a comparison of intensity-based grading and texture-based grading. These results show that texture-based grading improves AUC of all comparisons. Moreover, these results show that histogram representation provides similar or better results for all comparisons than using average value (see Table 3). All the results are expressed in percentage of AUC, SEN, and SPE.

	Intensity-based grading histo. (AUC/SEN/SPE in %)	Texture-based grading histo. (AUC/SEN/SPE in %)
CN vs. AD	93.5/**95.5**/82.7	**94.6**/94.2/**86.6**
CN vs. pMCI	90.0/81.8/**81.4**	**92.0**/**92.5**/81.2
AD vs. sMCI	81.1/**78.5**/68.3	**82.6**/77.6/**72.6**
sMCI vs. pMCI	74.9/**77.6**/67.2	**76.1**/74.9/**70.2**

4.1. Optimal number of directions

First, the optimal number of filter directions were investigated. Figure 3 shows the evolution of accuracy related to the number of directions. This experiment demonstrates that 5 different directions are enough to obtain the best results for CN versus AD comparison. Indeed, the accuracy does not increase using more directions. The best accuracy is reached with 5 different directions for sMCI versus pMCI. A fusion of Gabor filters at different scales was also performed. However, this experiment shown that filters at the full image resolution is enough to obtain the best results. The experiments showed that the optimal set of filters directions is $\{(\theta = 0; \phi = 0), (\pi/2; \pi/2), (0; \pi/2), (\pi/4; \pi/4), (\pi/4; -\pi/4)\}$. Therefore, in the rest of the experiments, comparisons were performed with Gabor filters in these 5 different directions and at the full images resolution.

Table 3: Comparison of grading features based on mean of the grading values within the HIPP structure. This table presents a comparison of intensity-based grading and texture-based grading. These results show that texture-based grading improves AUC of all comparisons. All the results are expressed in percentage of AUC, SEN, and SPE.

	Intensity-based grading mean (AUC/SEN/SPE in %)	Texture-based grading mean (AUC/SEN/SPE in %)
CN vs. AD	92.6/86.7/83.3	**94.7**/**93.4**/**87.6**
CN vs. pMCI	89.9/78.2/**85.4**	**92.3**/**91.6**/83.0
AD vs. sMCI	80.8/76.2/69.9	**82.2**/**77.6**/**71.0**
sMCI vs. pMCI	73.2/76.4/**65.0**	**75.1**/**77.0**/64.1

4.2. Comparison grading based on intensity vs. texture

To estimate the improvement provided by texture-based approach, we compare results obtained with our framework using intensities of the images in the MNI space (*i.e.*, intensity-based grading) and texture maps. For this comparisons, intensity 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 intensity-based grading and the proposed texture-based grading obtained with 5 Gabor filters. Results are expressed with area under the curve (AUC), sensibility (SEN) and specificity (SPE) measures.

As it is shown, texture-based grading improves classification performances in all experiments using mean or histogram-based grading. Indeed, the comparisons conducted with histogram-based representation show that texture-based grading obtains 94.6% of AUC for CN versus AD, 92.0% of AUC for CN versus pMCI, and 82.6% of AUC for AD versus sMCI comparisons while intensity-based grading obtains 93.5% of AUC for CN versus AD, 90.0% of AUC for CN versus pMCI, and 81.1% of AUC for AD versus sMCI comparisons. Finally, with histogram representation, texture-based grading obtains 76.1% of AUC for sMCI versus pMCI comparisons and intensity-based grading obtains 74.9%. As results based on histogram representation, the average grading aggregation follows the same tendency. These results demonstrate that texture maps enable to better capture structural alterations.

4.3. Comparison average grading vs. histogram-based grading

In this section, we compare our proposed histogram-based weak-classifier aggregation of texture-based grading values with a straightforward average that is usually used in patch-based grading framework. As presented in Table 3 and 2, the results show that histogram representation of weak classifiers distribution provides similar or better classifications results for all comparisons. Histogram-based aggregation obtains 94.6% and 92.0% of AUC while the average obtains 94.7% and 92.3% of AUC for CN versus AD and CN versus pMCI comparisons, respectively. Moreover, histogram-based aggregation obtains better results for AD versus sMCI and sMCI versus pMCI comparisons with 82.6%, and 76.1% of AUC compare to the average that obtains 82.2% and 75.1% of AUC for the same comparisons, respectively.

4.4. Comparison of different fusion schemes

Our fusion scheme was compared with a fusion based on the mean of texture-based grading maps (*i.e.*, fusion of the different grading maps provided at each direction with a straightforward average) and a SVM fusion of our texture-based grading features (*i.e.*, concatenation of the histogram features at the different considered directions into the SVM classifier). Results are summarized in Figure 4. During the experiments, adaptive fusion obtained an accuracy of 91.3% for CN versus AD comparison, the fusion using SVM classifier obtained 90.1% and the mean fusion obtained 89.1%. Moreover, for sMCI versus pMCI comparison adaptive fusion obtains 72.2% of accuracy while SVM fusion obtains 68.3% and mean fusion obtains 69.1%. Thus, adaptive fusion obtained the best results. Indeed, the results obtained by adaptive fusion is 1.2 percentage point higher than SVM fusion and 2.1 percentage point higher than mean fusion for CN versus AD comparison and 3.8 percentage point higher than SVM fusion and 3.1 percentage point higher than mean fusion. In order to study the significance of the accuracy differences between each fusion method, p-values were

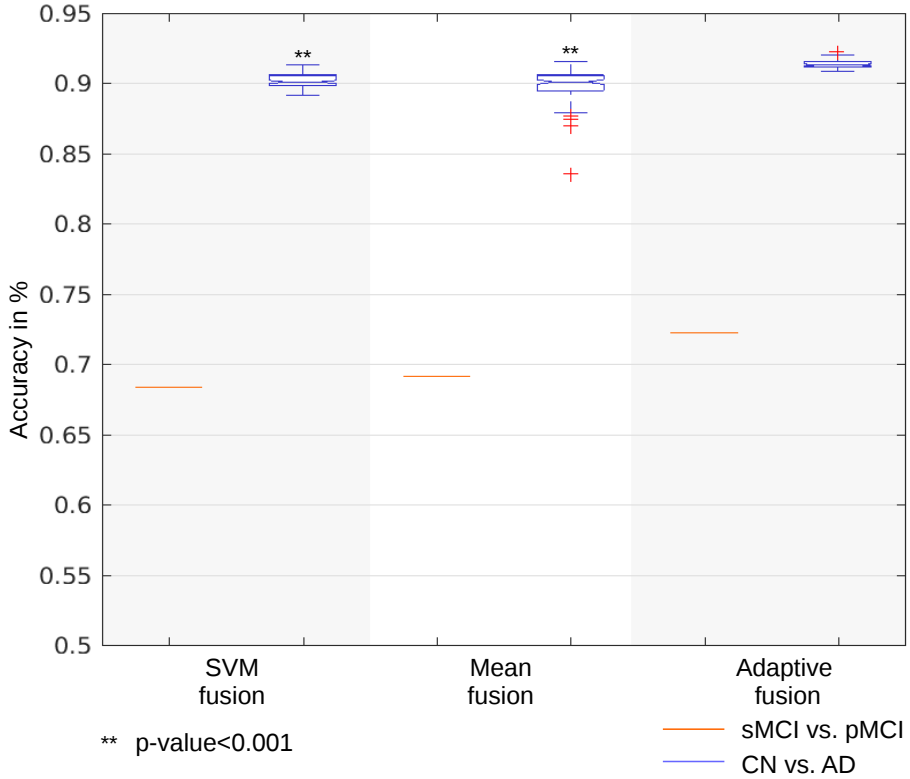

Figure 4: Comparison of different fusion schemes: mean of textural-grading maps, adaptive fusion of textural-grading maps and fusion into a SVM classifier (*i.e.*, concatenation of the histogram representing the grading of each texture direction as the input of SVM method). This comparison shows that adaptive fusion provides best results for both AD detection and prediction (*i.e.*, CN versus AD and sMCI versus pMCI). P-values were estimated with a t-test to compare adaptive fusion with other fusion methods.

estimated with a t-test. These experiments showed that adaptive fusion provides significantly better results for CN versus AD comparisons.

4.5. Comparison of different texture-based filters

Figure 5 shows comparison of classification results for different texture filters. This comparison was conducted with standard deviation (STD), entropy, gradient, and Gabor filters. Entropy and STD filters were computed into a window size of 5×5×5 voxels. Patch-based grading maps from gradient and Gabor filters were merged with our novel adaptive fusion scheme. STD filters obtained 89.1% of accuracy for CN versus AD comparison and 67.1% for sMCI versus pMCI comparison. Entropy filter obtained an accuracy of 90.5% for CN versus AD comparison and 68.1% for sMCI versus pMCI comparison. Gradient filter obtained 90.8% for CN versus AD and 70.1% for sMCI versus pMCI. This exper-

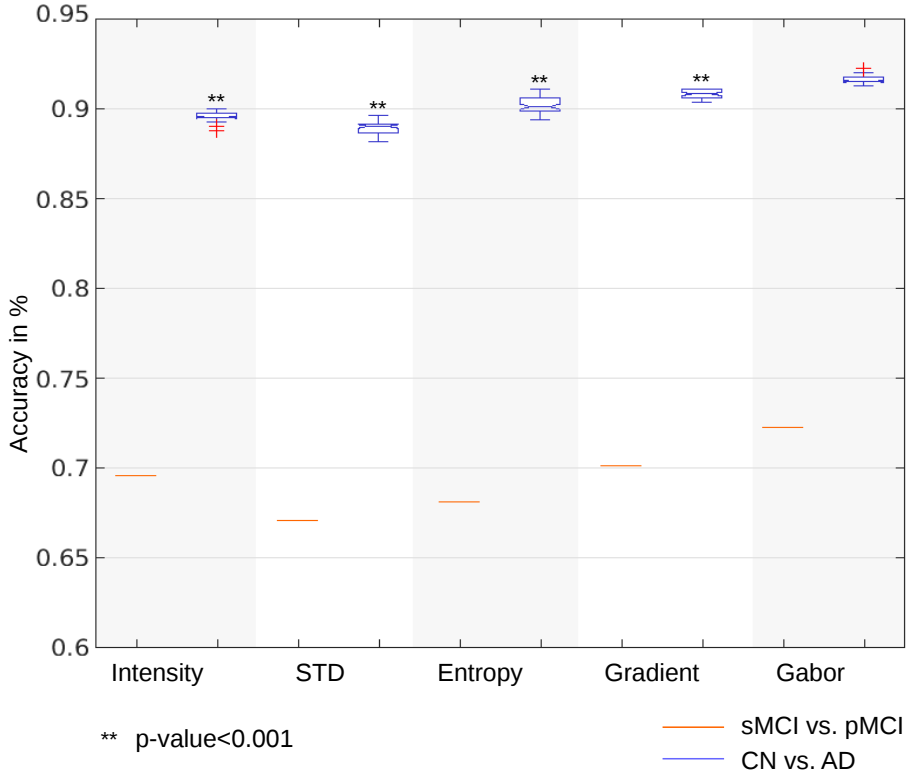

Figure 5: Comparison of classification performances for different texture filters. From left to right, we compare standard deviation (STD), entropy, gradient and Gabor filters. Entropy and STD filters were computed with a window size of 5×5×5 voxels. The results shows that only gradient and Gabor filters enable to improve performances for both AD detection and prediction. P-values were estimated with a t-test to compare Gabor filters to other texture filters.

iment shows that only gradient and Gabor filters obtain better results for both, AD detection and prediction compared to intensity-based grading with 91.3% and 72.2% of ACC for CN versus AD and sMCI versus pMCI comparisons, respectively. Thus, patch-based grading applied on an optimal set of Gabor filters provides better results than others texture filters for both considered comparison. Gabor filters improve by 3.2 and 2.1 percent points of accuracy for CN versus AD and sMCI versus pMCI, respectively, compared to gradient filter.

4.6. Comparison to state-of-the-art methods

In addition, a comparison with state-of-the-art methods are provided in Table 4. The results of this comparison are expressed in accuracy. On one hand, to compare classification results using the same structure, the proposed framework is compared with grading methods based on HIPP (see the upper part of Table

4). Thus, our proposed texture-based grading method is compared with the original patch-based grading method [9], a grading based on multiple instance learning method [32], and a patch-based grading based on a sparse representation using two different registration strategies [33]. This comparison shows that our method provides best results among HIPP-based grading methods. It reaches 91.3% of accuracy for CN versus AD, and 72.2% of accuracy for sMCI versus pMCI comparisons. On the other hand, our proposed method applied into hippocampus is compared with methods based on a whole brain analysis using similar dataset (see the lower part of Table 4). Indeed, we compare our texture-based grading approach applied on HIPP with an ensemble grading that proposed to extend the original patch-based grading to a whole brain analysis based on grey matter maps [20], the patch-based grading method based on a sparse representation applied on the whole brain [33], a sparse ensemble grading method that analyzes the whole brain [21], and a Deep Learning (DL) method based a whole brain analysis [31]. The results show that our method obtains the best accuracy for AD versus CN. This result is similar to classification results obtained with a DL and sparse ensemble grading method [31; 21]. However, methods based on a whole brain analysis and using non linear registration obtain more accurate classification results for sMCI versus pMCI.

5. Discussion

In this work, to improve patch-based grading framework, we proposed to capture texture information with a bank of Gabor filters. Our experiments showed that using more than 5 directions does not improve the results while increasing computational time (see Figure 3). Moreover, we also investigated a multi-scale texture approach. However, the experiments carried out showed that only one scale, at the full image resolution, is enough and the use of multi-scale texture did not improve classification performances. Therefore, we propose a multi-directional texture-based grading framework based on 1 scale and 5 directions.

A new grading values aggregation method based on histogram was also proposed. During our experiments, histogram representation of grading values distribution did not provide improvement for CN versus AD comparison compared to use a simple average value. That could be explained by the fact that CN and AD distributions are well separated and a parametric representation of their distributions is enough to discriminate these two groups. However, for sMCI versus pMCI case, the two distributions are less separable and histogram representation lead to better classification performances with in average a gain of 1 percentage point of AUC compared to a simple average value.

In order to fuse efficiently the different texture-based grading maps, we proposed a novel patch-based grading fusion scheme. This method is based on a confidence value estimated at each voxel. The comparison with different fusion schemes demonstrated the efficiency of our method. Indeed, compared to a straight average of the texture-based grading maps and a SVM fusion of the final histogram features with the classifier, our proposed method obtained best

Table 4: Comparison with state-of-the-art methods, all the results are expressed in accuracy. The upper part of this table presents results of results of methods applied on the HIPP structure. Texture-based grading improves CN versus AD classification by 2.3 percent points and sMCI versus pMCI classification by 1.2 percent points of accuracy compared to grading approaches based on the same HIPP structure. The lower part presents results of methods that propose a whole brain analysis. Compared to these approaches, our method obtains similar results than advanced methods based on whole brain analysis for CN versus AD classification. However, such methods obtain better performances for sMCI versus pMCI classification. In this table, we provide the type of registration (Reg.) involved in the methods and the type of features (Feat.).

Methods	Reg.	Feat.	CN vs. AD (ACC in %)	sMCI vs. pMCI (ACC in %)
Hippocampus				
Original Grading [9]	Affine	Intensity	88.0	71.0
Multiple Instance Grading [32]	Affine	Intensity	89.0	70.0
Sparse-based Grading [33]	Affine	Intensity	–	66.0
Sparse-based Grading [33]	NL	Intensity	–	69.0
Proposed Method	Affine	Texture	**91.3**	**72.2**
Whole Brain				
Ensemble Grading [20]	NL	GM	–	**75.6**
Sparse-based Grading [33]	Affine	Intensity	–	66.7
Sparse-based Grading [33]	NL	Intensity	–	75.0
Sparse Ensemble Grading [21]	NL	GM	90.8	–
Deep Ensemble Learning [31]	NL	GM	91.0	74.8

GM = Grey matter
NL = Non linear

accuracy for AD detection and prediction. Moreover, the obtained improvement was significant. This improvement can be explained by the fact that our proposed adaptive fusion method weights the grading values of each texture map according to their relevances while the fusion into SVM classifier and the average of texture maps considers each grading value as having the same importance.

Our work hypothesis is also that directional texture filters enable to improve patch comparison, and thus increase AD detection and prediction accuracy. To validate this hypothesis, our novel texture-based grading using an optimal set of Gabor filters were compared with others texture filters as done for segmentation in [35] (see Figure 5). STD, entropy, gradient and Gabor filters were compared for AD detection and prediction. This experiment showed that STD and entropy does not enable to improve patch comparison compare to intensity. The limitation of these filters might be to perform feature estimation within a window. Thus, only gradient and Gabor filters improve classification performances for AD detection and prediction. Moreover, Gabor filters obtain best results for both comparisons. This improvement is related to the use of additional texture directions compared to the three texture directions provided by gradient filter.

Table 4 summarizes the comparison of our proposed method with other grading methods proposed in the literature. These results demonstrate that Gabor filters enable to better capture structural alterations than method based on intensity or grey matter data. Indeed, texture maps provide enhance information leading to a better grading process. Thus, our method outperforms other grading methods using intensity when applied on the same structure [9; 32; 33]. At the lower part of Table 4, we compare the performance of our HIPP-based grading method with methods using the whole brain. First, for AD versus CN, the proposed method obtained similar or better results than methods applied over the whole brain. It is important to note that these methods require non linear registration [21; 20; 33; 31] while our method only requires affine registration and proposes a fast grading step. Second, for sMCI versus pMCI, our method obtained better results than all the methods involving a simple affine registration, including whole brain method proposed in [33]. On the other hand, the best results for sMCI versus pMCI are produced by whole brain grading [20; 33] using non linear registration. The improvement when using non linear registration is observed for HIPP-based and whole brain methods [33]. However, this improvement is obtained at the expense of using non linear registration, which is subject to failure and requires high computational time. Our method also demonstrated competitive performances for AD versus CN classification compared to advanced DL methods using whole brain and non linear registration [31]. Finally, this comparison shows that patch-based grading methods [20; 33] obtain similar or better results than recent deep learning methods [31] when applied over the entire brain after non linear registration.

6. Conclusion

In this work, we have proposed a new texture-based grading framework to better capture structural alterations caused by AD. Moreover, to combine grading maps estimated on texture maps, we have presented a new adaptive fusion scheme. We also have proposed an histogram-based weak classifiers aggregation step to better discriminate early stages of AD. We have studied the optimal set of texture directions and compare to others fusion schemes. Experiments conducted in this work demonstrated the relevance of using textural information in combination with with our novel locally adaptive fusion method. Finally, we have demonstrated the competitive performances of our new texture-based grading framework compared to several state-of-the-art approaches. In future works, we will investigate the extension of our texture-based grading framework to a whole brain analysis.

7. Acknowledgement

This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the

Investments for the future Program IdEx Bordeaux (HL-MRI ANR-10-IDEX-03-02), Cluster of excellence CPU and TRAIL (BigDataBrain ANR-10-LABX-57).

References

[1] Achterberg et al. Hippocampal shape is predictive for the development of dementia in a normal, elderly population. *Human brain mapping*, 35(5):2359–2371, 2014.

[2] Association Alzheimer's. 2015 Alzheimer's disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 11(3):332, 2015.

[3] Ashburner and Friston. Voxel-based morphometry—the methods. *Neuroimage*, 11(6):805–821, 2000.

[4] Avants et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.

[5] Boccardi et al. Training labels for hippocampal segmentation based on the eadc-ADNI harmonized hippocampal protocol. A*lzheimer's & Dementia*, 11(2):175–183, 2015.

[6] 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.

[7] Chincarini et al. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *Neuroimage*, 58(2):469–480, 2011.

[8] Coupé et al. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, 54(2):940–954, 2011.

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

[10] Coupé et al. Detection of Alzheimer's disease signature in MR images seven years before conversion to dementia: Toward an early individual prognosis. *Human brain mapping*, 36(12):4758–4770, 2015.

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

[12] Simona E et al. Comparison of texture features based on Gabor filters. *IEEE Transactions on Image processing*, 11(10):1160–1167, 2002.

[13] Eskildsen et al. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *Neuroimage*, 65:511–521, 2013.

[14] Farhan et al. Texture classification using rotation-and scale-invariant Gabor texture features. *IEEE Signal Processing Letters*, 20(6):607–610, 2013.

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

[16] Braak H. and Braak E. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of aging*, 16(3):271–278, 1995.

[17] 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.

[18] Jack et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

[19] Koikkalainen et al. Differential diagnosis of neurodegenerative diseases using structural MRI data. *NeuroImage: Clinical*, 11:435–449, 2016.

[20] 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.

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

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

[23] Manjón et al. Adaptive non-local means denoising of MR images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging*, 31(1):192–203, 2010.

[24] Manjón et al. NICE: non-local intracranial cavity extraction. *International Journal of Biomedical Imaging*, 2014.

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

[26] Petersen et al. Current concepts in mild cognitive impairment. *Archives of neurology*, 58(12):1985–1992, 2001.

[27] Petrella et al. Neuroimaging and early diagnosis of Alzheimer disease: a look to the future. *Radiology*, 226(2):315–336, 2003.

[28] 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.

[29] Sørensen et al. Early detection of Alzheimer's disease using MRI hippocampal texture. *Human brain mapping*, 37(3):1148–1161, 2016.

[30] Sturges. The choice of a class interval. *Journal of the american statistical association*, 21(153):65–66, 1926.

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

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

[33] 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.

[34] Tustison et al. N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*, 29(6):1310–1320, 2010.

[35] Wachinger et al. Efficient descriptor-based segmentation of parotid glands with nonlocal means. *IEEE Transactions on Biomedical Engineering*, 64(7):1492–1502, 2017.

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