Neurocomputing 220 (2017) 98–110

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Recognition of Alzheimer's disease and Mild Cognitive Impairment with multimodal image-derived biomarkers and Multiple Kernel Learning

Olfa Ben Ahmed [a,*], Jenny Benois-Pineau [b], Michelle Allard [c,d,e], Gwénaëlle Catheline [c,d,e], Chokri Ben Amar [f], For the Alzheimer's Disease Neuroimaging Initiative[1]

[a] XLIM-SIC, UMR CNRS 7252, Bvd Marie and Pierre Curie, 86962 Futuroscope Chasseneuil Cedex, France
[b] University of Bordeaux, Laboratoire Bordelais de Recherche en Informatique (LaBRI), France
[c] University of Bordeaux, INCIA, UMR 5287, F-33400 Talence, France
[d] CNRS, INCIA, UMR 5287, F-33400 Talence, France
[e] EPHE, Bordeaux, France
[f] REsearch Group on Intelligent Machines, Tunisia

ARTICLE INFO

Article history:
Received 3 January 2016
Received in revised form
6 August 2016
Accepted 7 August 2016
Available online 25 August 2016

Keywords:
Alzheimer's disease
Multiple Kernel Learning
Mutlimodal fusion
Diagnosis
Local features
DTI
MD maps
CHFs
Imaging biomarkers

ABSTRACT

Computer-Aided Diagnosis (CAD) of Alzheimer's disease (AD) has drawn the attention of computer vision research community over the last few years. Several attempts have been made to adapt pattern recognition approaches to specific neuroimaging data such as Structural MRI (sMRI) for early AD diagnosis. One strategy is to boost the discrimination power of such approaches by integrating complementary imaging modalities in a single learning framework. Diffusion Tensor Imaging (DTI) is a new and promising modality giving complementary information to the anatomical MRI. However, including relevant DTI information from such modality is a challenging problem. In this paper, we propose to extract local image-derived biomarkers from DTI and sMRI to construct multimodal AD signatures. To assess the relevance of such modalities as well as to optimize the classifier, we integrate complementary information using a Multiple Kernel Learning (MKL) framework for AD subjects recognition. To evaluate our method, we perform experiments on a subset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Both T1-weighted MRI and Mean Diffusivity (MD) maps from the DTI modality of 45 AD patients, 52 Normal Control (NC) and 58 Mild Cognitive Impairment (MCI) subjects have been used. The obtained results indicate that our multimodal approach yields significant improvement in accuracy over using each single modality independently. The classification accuracies obtained by the proposed method are 90.2%, 79.42% and 76.63% for respectively AD vs. NC, MCI vs. NC and AD vs. MCI binary classification problems. For the MCI classification problem, the proposed fusion framework leads to an average increase about at least 9% for the accuracy, 5% for the specificity and 15% for the sensitivity.

1. Introduction

Alzheimer's disease (AD) is the most common chronic neurodegenerative disorder and the first cause of dementia nowadays. Early detection of AD is of primary importance in biomedical research for providing a new therapeutics slowing its progression. Computer-Aided Diagnosis (CAD) tools for an automated and early AD detection are urgently needed to help clinician's decision. Medical information from structural Magnetic Resonance Imaging (sMRI) has long time been the most used neuroimaging modality to detect brain atrophy in AD studies [1–4]. Recent studies have shown that combining several information sources from multiple neuroimaging modalities may carry complementary atrophy information and thus

* Corresponding author.
E-mail addresses: olfa.ben-ahmed@unilim.fr (O.B. Ahmed), jenny.benois@labri.fr (J. Benois-Pineau), michelle.allard@chu-bordeaux.fr (M. Allard), gwenaelle.catheline@chu-bordeaux.fr (G. Catheline), chokri.benamar@ieee.org (C.B. Amar).

[1] Data used in preparation of this paper 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.

http://dx.doi.org/10.1016/j.neucom.2016.08.041
0925-2312/

may enhance the AD diagnosis accuracy [5]. In AD diagnosis research, usually two classic fusion strategies were applied to fuse features, namely early fusion and late fusion. The first strategy simply concatenates features extracted from different modalities and/or regions of interest (ROI) into a huge vector and then builds a classifier [6–9]. However, the performance of early fusion could be affected by features that have low contribution and by the curse of dimensionality. Meanwhile, in the late fusion scheme visual features are extracted first from each modality. Then a classifier is trained separately on each modality as well. The outputs of classifiers are then combined for a final decision [10–14]. However, such methods are unable to exploit the correlations between the different modalities, since each modality is treated independently. Recent works provide an alternative by using a kernel-based machine learning technique known as Multiple Kernel Learning (MKL) [15]. The MKL aims to combine kernels derived from several sources of information [16–22]. A kernel implicitly represents a notion of similarity for the features. For instance, [19] proposed an MKL method that represents PET and Cerebrospinal Fluid (CSF) modalities as one or more kernels. Indeed, SVM model parameters and kernel combination weights are simultaneously optimized using the SimpleMKL algorithm [15].

In [17], MKL with Fourier transform on the Gaussian kernels has been applied to AD classification using both structural MRI and functional MRI (fMRI). In [18], the authors proposed an SVM-based model to combine kernels from MRI, PET and CSF features. Their approach does not learn kernel coefficients. Instead, they use a grid-search method to select the optimal kernel weights which can be very time consuming. Ref. [23] integrates multimodal data (volumetric MRI, FDG-PET, CSF, and APOE genotype) with the classification process using of a mixed kernel to predict conversion of MCI patients to AD. In fact, the CSF features are biological measures obtained intrusively. In addition to PET, fMRI and CSF modalities, Diffusion Tensor Imaging modality has been used as a new MRI modality to detect micro-structural changes that remain not visible in anatomical scans. In DTI modality, the water diffusion in the brain is interpreted as MR signal loss. Effectively, the neurodegenerative process is accompanied by a loss of obstacles that restrict motion of water molecules [24]. Usually, DTI features are illustrated by scalar measurements calculated on two DTI-derived maps: the Fractional Anisotropy (FA) and the Mean Diffusivity (MD). The FA represents the degree of anisotropy of water diffusion while MD represents its magnitude. DTI is less expensive, safe and noninvasive unlike fMRI or PET modalities. Recent studies in AD diagnosis proved that combining information extracted from the DTI modality with other modalities such as sMRI [25–28] or fMRI [13] improved the diagnosis performance. It is worth noting that most of the cited-above multimodal methods resorted to direct volumetric features or voxel-wise features to analyze the brain atrophy. However, volumetric and voxel-wise methods rely on large-scale structural changes and thus present significant limitations such as difficulties to reflect localized details of the scan. Indeed, the analysis of the MRI signal may therefore bring additional information to the early diagnosis of AD. In fact, MRI signal varies across tissue characteristics and/or types meaning that, for example, locally shrunk brain structures will display a different amount of GM and CSF compared with when they are healthy. Therefore, quantification of the amount of brain cell loss in terms of signal variation across individual brains may provide information about the disease. Recently, a family of local features-based methods has demonstrated impressive level of performance for Alzheimer's disease related atrophy description [3,8,29–39]. Those methods focus mainly on characterizing visual properties that can be computed from pixels and image patterns.

Referring to the domain knowledge, the sMRI and the DTI modalities are used to assess respectively the micro- and macro-structural alteration of the hippocampus region [40,41]. However, local features extraction from DTI-derived maps remains a challenging problem since this modality does not contain any anatomical information about the brain structure. The second challenge is how to efficiently integrate DTI features with complementary sMRI information. To the best of our knowledge, there is the first work that investigated jointly local DTI and MRI derived features to deeply capture hippocampal atrophy for AD/MCI subjects discrimination. In this paper, we propose a multimodal CAD system that simultaneously considers and integrates local image-derived biomarkers from MD and sMRI using a Multiple Kernel Learning framework. Our premise is that T1-weighted MRI and DTI data provide complementary information about hippocampus atrophy. The rest of this paper is organized as follows. Section 2 presents the overall proposed fusion framework for multimodal AD subjects classification. Section 3 describes the MKL fusion method. Section 4 presents the data used to evaluate our method. Experiments and results are reported in Section 5. Finally, we conclude and provide perspectives of this research in Section 6.

2. Methodology

We design a Multiple Kernel Learning (MKL) framework to combine visual features derived from sMRI and DTI images of brain. The proposed method consists initially in brain image processing which helps to overlay the DTI features onto an individual's own anatomy and select the hippocampus ROI. The hippocampus ROI is then selected using and normalized brain template. After that, the AD imaging biomarkers are extracted from the hippocampus area from both sMRI and MD images. Two kinds of features are extracted from the T1 weighted MRI; structural signature from the hippocampus and the amount of CSF in the hippocampal area. From the DTI modality, we extract Means Diffusivity (MD) signal variation in the hippocampus ROI. Our method is a slice-based method and features are extracted in a 2D fashion. Fig. 1 presents the different steps of our method namely image preprocessing, hippocampus ROI selection, generation of visual signatures, and finally the MKL classification.

2.1. Image preprocessing

Preprocessing of DTI modality of brain imaging includes correction for eddy currents and head motion, skull stripping with the Brain Extraction Tool (BET) and fitting of diffusion tensors to the data with DTIfit module of the Software Library FSL.[2] Fitting step allows the generation of the MD and FA maps. In our work, we focus only on the MD maps. Indeed MD maps express the loss of internal structure and can be processed in the same way as sMR images, while FA maps express the anisotropy and specific analysis and feature extraction techniques still need to be designed. The MD images have to be processed to allow for ROI selection using a normalized anatomical atlas. To do this, MD images are affinely co-registered to the corresponding anatomical scans. Co-registration consists in superimposing MD maps on the subject's corresponding anatomical MRI. Indeed, co-registration consists in superimposing DTI-derived maps (MD images) on the subject's corresponding anatomical scan. This helps to overlay MD values onto an individual's own anatomy. We follow here [42,26] where co-registration was used to extract regional values of DTI parameters in some specific areas. Structural MRI was corrected to lower the

[2] http://www.fmrib.ox.ac.uk/fsl

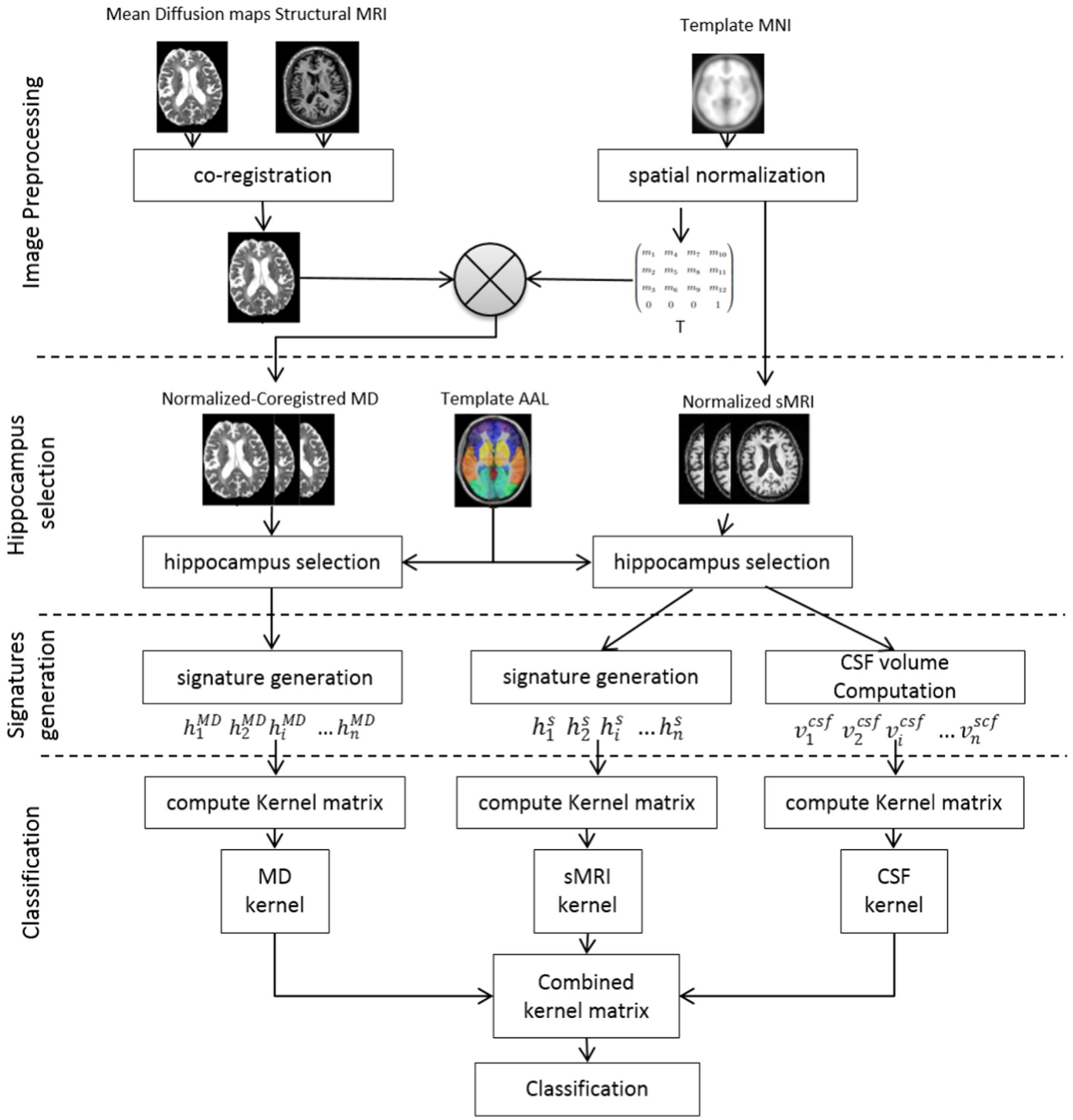

Fig. 1. Multimodal fusion pipeline for AD subjects recognition.

intensity inhomogeneity and spatially normalized onto the Montreal Neurological Institute (MNI) brain template [43]. The resulting transformation parameters were applied to the corresponding co-registered MD maps. It is to note that all performed transformations are affine in order to not deform the pattern of the features. Finally, the spatially normalized MD maps are smoothed with a Gaussian filter to improve signal to noise ratio. Transformed images were visually checked for co-registration errors. Preprocessing steps are presented in the top of Fig. 1. All image preprocessing steps were performed using Statistical Parametric Mapping software running on matlab (SPM8, Welcome Department of Imaging Neuroscience, London, UK).[3]

2.2. Hippocampus ROI selection

Since each modality is affinely normalized, we are able to identify a region of interest (ROI). To select the hippocampus ROI, we do not

[3] http://www.fil.ion.ucl.ac.uk/spm

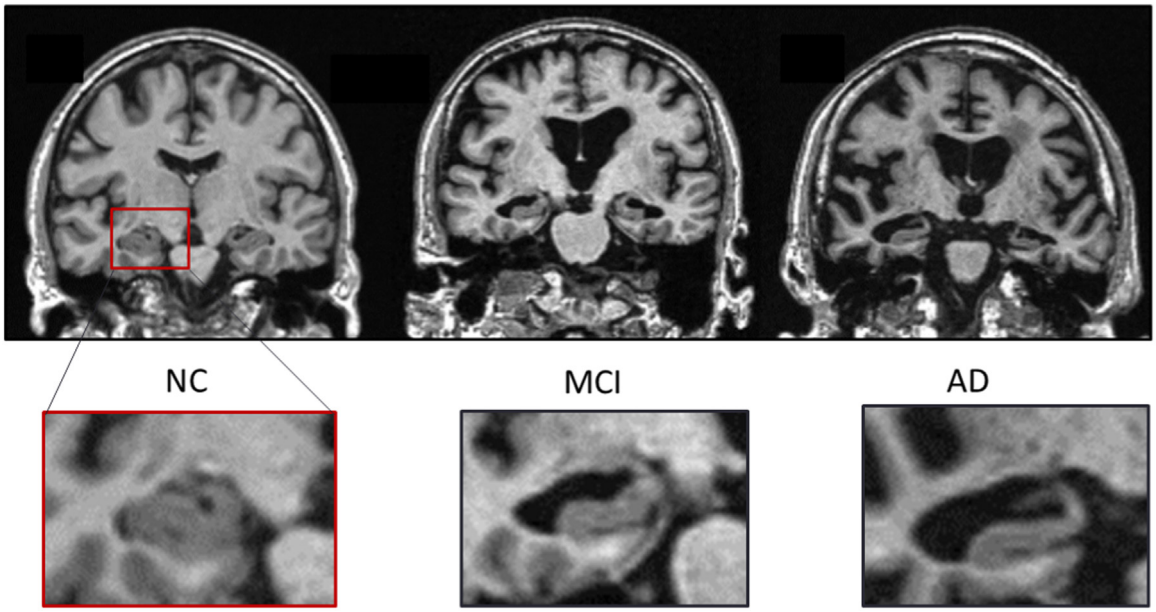

Fig. 2. A bounding box around the hippocampus region from coronal slices of the T1-weighted MRI of respectively NC, MCI and AD subjects.

use a segmentation step, but instead we follow an atlas-based selection method we previously proposed in [8]. The method allows for a rough extraction of the ROI. We use an MNI normalized brain atlas called Anatomic Automated Labeling Atlas (AAL) [44]. Furthermore, in order to limit the processing only to brain tissues, we also generated a mask to remove skull voxels. Both sMRI and MD images are mapped to the AAL to select the hippocampus ROI. The use of the atlas parcels helps to characterize brain abnormalities in terms of intra-ROI local pattern. In fact, the pattern overlapping with the extracted ROI mask in all modalities varies between healthy subjects and those exhibiting clinical signs of disease.

2.2.1. *Pattern of hippocampus shrinkage from sMRI*

Referring to domain knowledge, at the AD stage the hippocampus ROI undergoes a significant volume reduction and then shrinks. The liberated volume of the hippocampus is filled with CSF. An illustration of this phenomenon is given in Fig. 2. The dark areas in the image correspond to the CSF.

Hence, to quantify the hippocampus shrinkage in sMRI, we extract two complementary features: hippocampus structural alteration and the amount of CSF in the hippocampal area. First, we follow the method we previously proposed in [30] which is based on the use of local features such as SIFT, SURF and Gauss–Laguerre Circular Harmonic Functions descriptors (GL-CHFs) to quantify the structural hippocampus alteration. Then, an adaptive thresholding method based on the Otsu-method technique is used to compute the percentage of pixel corresponding to CSF in the hippocampus area.

2.2.2. *Pattern of water diffusion in hippocampus from MD maps*

The brain cells loss is accompanied by a loss of barriers that restrict motion of water molecules in brain tissues. Fig. 3, illustrates an example of the MD maps of both healthy and AD subjects. In AD case, the diffusion in ventricles for example is faster and the MD map is brighter due to the free motion of water molecules. While in white and gray matter regions, the diffusion is slower and the MD pixels are darker. In addition, the faster diffusion of water in the affected hippocampal area results in brighter pixels in the MD maps for this region as well.

Therefore, we assume that the fast diffusion of water in the hippocampal area is expressed by hyper-signal (brighter pixels) on the MD maps (Fig. 4). Hence, we propose to quantify the hippocampal atrophy by considering the pattern of water diffusion as an MD image signal variation inside the hippocampus ROI.

2.3. *Multimodal AD disease-related signature generation*

Once the hippocampus ROI is localized in both sMRI and MD images, we present now the process of generating a visual signature or the so-called “AD disease-related signature”. Indeed, an MD and sMRI signature per subject is generated to reflect the bimodal brain atrophy at the individual level. We represent signal variations inside the ROI anatomy by a set of local features. Here, we use a multi-resolution approach based on the Gauss–Laguerre Circular Harmonic Functions (GL-CHFs) descriptors [45], which is suitable for extracting the most relevant image features and even small and localized patterns. The local descriptor of the image is a vector of coefficients of development of image signal on the CHF basis [30]. CHFs proved their performance in capturing atrophy from sMRI and MD maps [30,46] since they give interesting approximations of blurred signal in MRI and DTI.

2.3.1. *Gauss–Laguerre Circular Harmonic Functions (GL-CHFs)*

CHFs were first introduced in the pattern recognition domain [45]. They have several advantages over other descriptors particularly for MRI. CHFs present a decomposition of image signal on the orthonormal functional basis. They allow for capturing local direction of image signal and also capture intermediate frequencies in the signal similar to Fourier decomposition. This propriety is more convenient for sMRI/MD images with smooth contrasts.

Mathematically speaking, Gauss–Laguerre Harmonic Functions are complex-valued radial profile functions multiplied by complex exponent:

$$\Psi(r,\theta;\sigma) = \Psi_n^{|\alpha|}\left(\frac{r^2}{\sigma}\right)e^{i\alpha\theta} \tag{1}$$

Their radial profiles are Laguerre functions:

$$\Psi_n^{|\alpha|}(x) = \frac{1}{\sqrt{n!\Gamma(n+\alpha+1)}}x^{\frac{\alpha}{2}}e^{\frac{-x}{2}}L_n^{\alpha}(x) \tag{2}$$

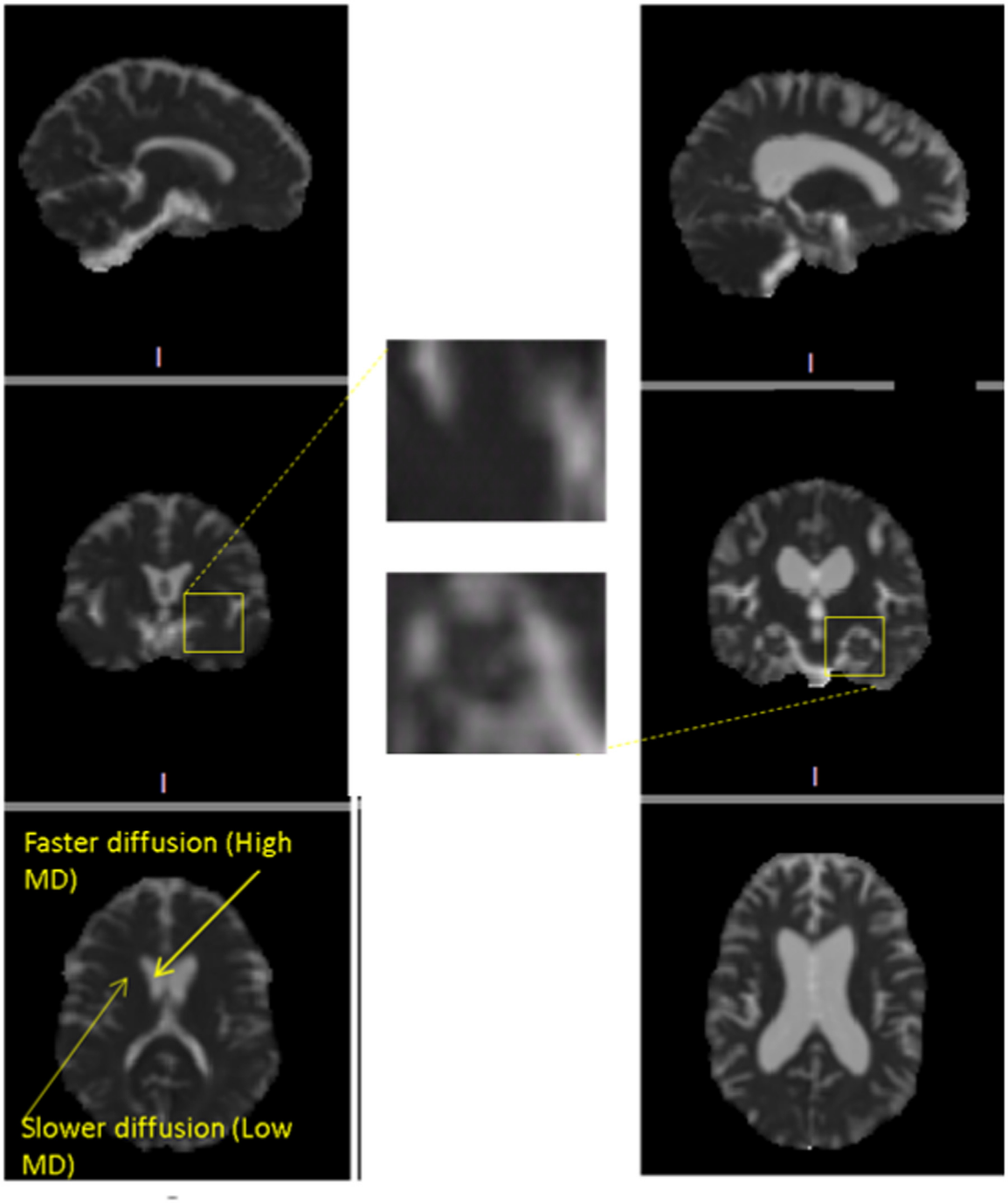

Fig. 3. Water diffusion in MD maps from normal and AD subjects. The faster the diffusion is the higher the MD values are.

where $n = 0, 1, \ldots$; $\alpha \pm 1, \pm 2\ldots$ and $L_n^{\alpha}(x)$ are Laguerre polynomials. r, θ are polar coordinates, σ is a scale parameter and Γ is a gamma function.

$$L_n^{\alpha}(x) = (-1)^n x^{-\alpha} \exp^x \frac{d}{dx^n}(x^{n+\alpha} e^{-x}) \tag{3}$$

LG-CHF is a complete orthogonal set of functions in the real plane. Thus, each brain slice $S_m(a, b)$ per modality m can be expanded in the analysis point (a_0, b_0) for fixed scale σ in Cartesian system. The coefficients of the partial expansion of local neighborhood of (a_0, b_0) (patch) can be used as a feature descriptor. The advantages of these features are such that they capture both the direction and smooth variations of slice signal which is not the case of conventional SIFT and SURF features computed on the basis of high-pass filtering by gradient computation at different scales as we have shown in [47]. The number of coefficients retained defines the dimensionality of the descriptor. The reasonable dimensionality of 150 coefficients (see [30]) was used in the present work. More mathematical details about the CHF descriptors can be found in [45].

To extract features from slices, we used a dense sampling scheme with a regular grid of circular patterns of the same radius. Each grid disk is then considered as a patch of fixed size and forms the descriptor support, that is the area where the signal will be approximated by CHFs. Such sampling may provide a rough model of clinician vision. It results in a good coverage of the entire hippocampus ROI and in a constant amount of features per image area. In addition, regions with less contrast contribute equally to

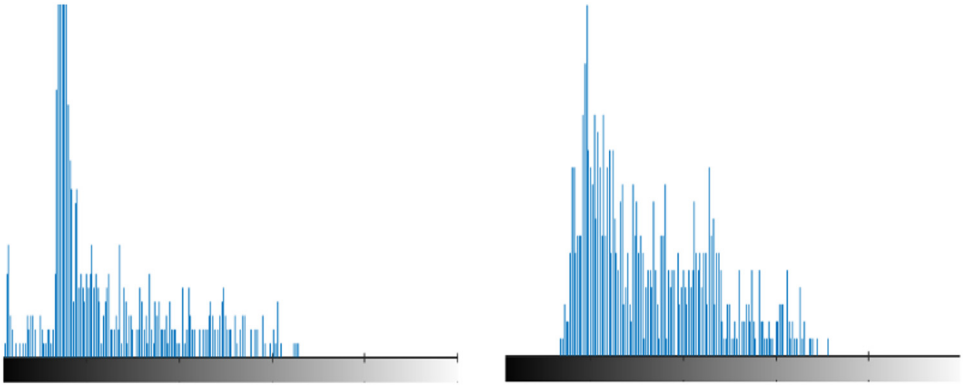

Fig. 4. Histograms of intensity distribution in the hippocampal ROI from the MD map of (from left to right) NC and AD patients.

the overall slice representation. Fig. 5 presents CHF features placement in the hippocamups of both sMRI and MD coronal slice of an ADNI subject. The descriptor support areas are selected by simply scanning the hippocampus mask line by line and by placing the feature centers in masked pixels of each slice.

2.3.2. Features quantization and signatures generation

To build a compact visual signature per modality, we use the Bag-of-visual-Words (BoVW) approach which is popular in computer vision [48]. To capture the maximum of atrophy information, we build a bag of words per projection and then per modality. First, all features $f_{p,i}^{m,sagittal}$, where p and i stand respectively for slice and feature indexes, are extracted from the modality m ($m \in \{smri, md\}$) on all slices from the sagittal projection then the features are quantized by the k-means clustering algorithm. The centers $c_k^{m,sagittal}$, $k \in [1, K]$, are then calculated to form the codebook, that is the set of cluster centers of cardinal k. The latter is the parameter given to the k-means algorithm. The same is done for axial and coronal projections. All features $f_{p,i}^{m,sagittal}$, $f_{p,i}^{m,axial}$, $f_{p,i}^{m,coronal}$ and centers $c_k^{m,sagital}$, $c_k^{m,axial}$, $c_k^{m,coronal}$ are 150-dimensional descriptors. Once the cluster centers have been determined, the image signature per projection is generated. Each feature is "encoded", i.e., it is assigned to the closest center using the Euclidean distance $d\left(f_{p,i}^{m,sagital}, c_k^{m,sagital}\right)$ metric. Then each projection is represented by a normalized histogram $h^{m,projection}$ of visual words occurrence. The image signature per modality m is acquired by the concatenation of the histograms from all projections: $h^m = [h^{m,sagital}, h^{m,axial}, h^{m,coronal}]$. The obtained visual signatures are used to compute the bi-modal representation of one subject. The size of the resulting visual signature for the modality m is $N = 3 * size\ of\ codebook_m$.

3. Multiple Kernel Learning for multimodal visual signatures fusion

A global fusion framework is presented in this section to combine visual signatures computed on sMRI and MD images to distinguish between, NC, MCI and AD subjects. The features to be combined are: MD visual signature h^{md}, sMRI visual signature h^{smri} and the amount of CFS in hippocampus area v^{csf} computed with the method proposed in [30]. Therefore, three kernels are computed (i.e., one kernel per feature) and a set of weights are estimated for the kernel combination. The assumption behind Multiple Kernel Learning is to create a weighted linear combination of the kernels from each feature space, and to adapt these weights in order to achieve the best performance [49].

In our framework, we use the MD signatures, sMRI signatures and CSF values to spawn a set of kernels k_{md}, k_{smri} and k_{csf}. The overall kernel is constructed by summing those kernels and weighting them as specified. The optimization problem then reduces to finding their weights while simultaneously maximizing the margin for the training data. A kernel is effectively a kind of similarity measure that explicits feature maps and properties, so the choice of a suitable kernel for each kind of features is crucial here. Actually, the Chi-square kernel has proven to be a powerful measure of similarity between histograms in the area of image classification [50] and specially with BoVW-based MKL applications [51]. In our case, the h^{md} and h^{smri} visual signatures are histogram-based features. Hence, we resort to the use of a χ^2 chi-square kernel given by:

$$K_{\chi^2}(h_1, h_2) = \exp(-\beta\, d_{\chi^2}(h_1, h_2)) \text{ with } d_{\chi^2}(h_1, h_2) = \sum_1^k \frac{(h_1^i - h_2^i)^2}{(h_1^i + h_2^i)} \quad (4)$$

where ${h_1}^i$ and ${h_2}^i$ are the corresponding bins from the histograms h_1 and h_2 and k denotes the codebook size. Here, β is the kernel width, this parameter was fixed to be the mean of all chi-square distances between all training features. Meanwhile, for the v^{csf} features, which are not histograms but simple quantities of pixels classified as CSF (see [30] for the Bayesian classification approach), we choose a Gaussian kernel $\exp(-\| v_1 - v_2 \|^2/2\sigma^2)$. All kernels are then normalized to the unit trace through the formula $\bar{k}_m(x, y) = k_m(x, x)/(sqrt(k_m(x, x)_* k_m(y, y)))$.

We set w_{md}, w_{smri} and w_{csf} the weights accordingly to respectively the MD, sMRI and CSF features. The combined kernel is presented as follows:

$$K(x, y) = w_{md}K_{md}(x, y) + w_{smri}K_{smri}(x, y) + w_{csf}K_{csf}(x, y) \quad \text{with } w_{md} + w_{smri} + w_{csf} = 1 \quad (5)$$

Then, we use the simpleMKL [15] solver to search a combination of kernels that maximizes the classifier performance (Accuracy). To this end, in the training step, both SVM parameters and MKL weights are simultaneously estimated within the same optimization problem. The best classification accuracy is obtained using an alternating method performed between the optimization of the SVM classifier and the optimization of the kernel weights. In each step, given the current solution of kernel weights, MKL solves a standard SVM optimization problem with the combined kernels. Suppose that we have a set of training samples $\{(u_i, v_i)\}_{i=1}^l$ where u_i is a N-dimensional descriptor and $v_i \in \{-1, +1\}$ is its corresponding label. Classification performances depend strongly on the data representation. In kernel methods, the data representation is implicitly chosen through the so-called kernel $K(u_i, u_j)$ which intuitively computes similarity between u_i and u_j. The decision function is defined as flow:

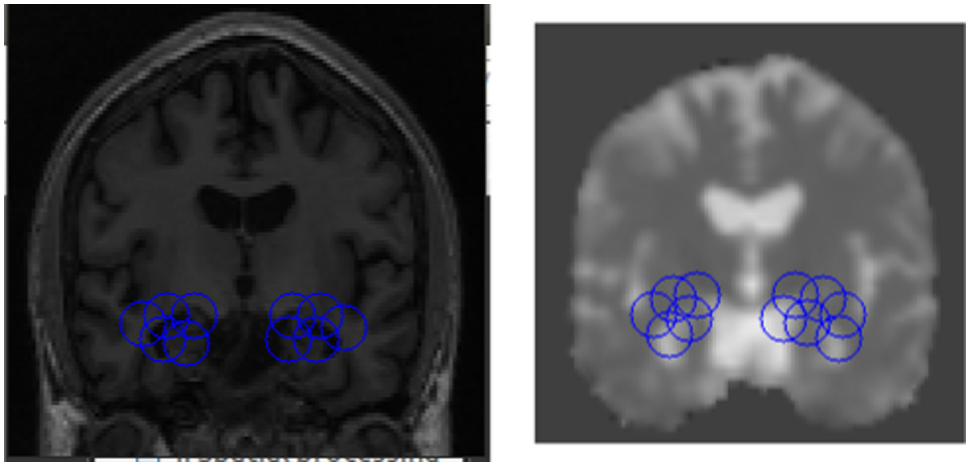

Fig. 5. Example of CHF features placement in, respectively from the left to the right, T1-weighted MRI and MD coronal brain slices (AD subject from the ADNI dataset). Circles represent the locations of features "support area" (i.e., where the descriptors are computed).

$$f(u) = \sum_{i=1}^{l} \alpha_i^* K(u, u_i) + b^* \tag{6}$$

where α_i^* and b^* are coefficients to be learned from data. $K(.,.)$ is a positive definite Gram matrix. The learning task here involves heterogeneous features extracted from multiple data sources. The MKL aims in general to learn kernel from training data, this kernel is defined as linear combination of a series of basic kernels:

$$K(u, u') = \sum_{m=1}^{M} w_m k_m(u, u')$$
$$\text{with } w_m \geq 0, \quad \sum_{m=1}^{M} w_m = 1 \tag{7}$$

With M is the number of kernels, kernels k_m are the MD, CSF and sMRI kernels with different parameters. The additionally introduced parameters w_m define the weight for each modality m. MKL aims at simultaneously optimizing the α_i and the w_m subject to $w_m \geq 0$, $\sum_{m=1}^{M} w_m = 1$.

In the current work, we use the SimpleMKL algorithm proposed by [15] to solve the classification problem. SimpleMKL iteratively determines the combination of kernels by gradient descent wrapping a standard SVM solver.

$$\min_w J(w) \text{ such that } \sum_{m=1}^{M} w_m = 1\,, \quad w_m \geq 0 \tag{8}$$

where

$$J(w) = \begin{cases} \min\limits_{\{f\},b,\epsilon} & \frac{1}{2}\sum_m \frac{1}{w_m}\|f_m\|^2_{H_m} + C\sum_i \epsilon_i \\ \text{s. t.} & y_i \sum_m^M f_m(u_i) + y_i b \geq 1 - \epsilon_i \\ & \epsilon_i \geq 0 \end{cases}$$

where each function f_m belongs to a reproducing kernel Hilbert space (RKHS H_m) associated with a kernel k_m. The primal MKL problem is addressed by solving the following convex problem which can be transformed into the following dual form using Lagrange multipliers:

$$\max_{\alpha} \; -\frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_m w_m K_m(x_i, x_j) + \sum_i \alpha_i$$
$$\text{with } \sum_i \alpha_i y_i = 0$$
$$C \geq \alpha_i \geq 0 \quad \forall\, i \tag{9}$$

the weights w_m are updated with the gradient descent. The update scheme for the gradient descent is $w \longleftarrow w + \gamma D$, where γ is the optimal gain and D is the gradient of the objective function. This procedure is repeated until convergence (the objective function value stops decreasing).

4. Imaging data

Data used in the experiments comes from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. It is to note that the ADNI recently added at the second phase ADNI2/ADNIGO Diffusion Tensor Imaging (DTI), among several other new imaging modalities, in an effort to identify sensitive biomarkers of Alzheimer's disease (AD). In our work, we selected both T1-weighted MR imaging and DTI imaging data of 155 subjects including 52 NC, 45 AD and 58 MCI. Table 1 presents a summary of the demographic characteristics of the selected groups (including the number, age, gender and Mini Mental State Examination (MMSE) score of cognitive function of the subjects). The number of scans available for this work was limited by the availability of DTI data. In fact, the DTI is a relatively new MR modality, and not all subjects have both DTI and sMRI scans. Anatomical images will serve also as a spatial reference for the DTI data. Details about the acquisition protocol and the initial processing steps can be found in the ADNI2 protocol (http://adni-info.org/Scientists/Pdfs/ADNI2_Protocol_FINAL_20100917.pdf).

5. Experiments and results

5.1. Experiments

In order to fix the optimal codebook size, we plot the variation of classification accuracies (AD vs. NC, NC vs. MCI and MCI vs. AD) function to the codebook size changes. Fig. 6 illustrates that the performance of the three binary classification problems gradually increases with the increase of the codebook size but it can also decrease in certain cases. In general, the accuracy does not change significantly with codebook size. Hence, we fix the optimal codebook sizes to 250 and 150 respectively for sMRI and MD images.

All experiments were conducted for three different classification tasks: (AD vs. NC), (NC vs. MCI) and (AD vs. MCI). A 10-fold cross validation strategy was used to assess the classification performance. We repeated the 10 fold cross-validation 10 times to avoid any bias introduced by the random partitioning of the dataset during cross

Table 1
Demographic description of the ADNI group. Values are denoted as mean ± standard deviation.

Diagnosis	Number	Age	Gender (M/F)	MMSE
AD	45	75 ± 4.5 ($16 \in [55-65]$, $18 \in [65-75]$ and $11 > 75$)	28/17	23.3 ± 1.8
NC	52	73 ± 7 ($13 \in [55-65]$, $18 \in [65-75]$ and $21 > 75$)	19/33	29.2 ± 1.1
MCI	58	$73{,}8 \pm 7$ ($19 \in [55-65]$, $20 \in [65-75]$ and $19 > 75$)	35/23	27.2 ± 0.9

Table 2
AD vs. NC classification results.

AD vs. NC Features	Acc[95%CI]	Spe[95%CI]	Sen[95%CI]	BAC%
Hippo-SMRI	81.5[81.28 81.71]	82.63[82.38 82.88]	80.16[79.79 80.53]	81.40
Hippo-MD	79.66[79.36 79.95]	84.04[83.61 84.46]	74.48[74.05 74.91]	79.26
Concat (MD,SMRI)	83.09[82.8 83.38]	85.21[84.82 85.6]	80.58[80.14 81.02]	82.9
MKL(MD,SMRI)	88.16[86.22 90.11]	95.6[93.26 97.94]	80.42[77.58 83.26]	88.01
MKL(MD,SMRI,CSF)	90.2[88.7 91.5]	97.20 [94.6 98.8]	82.92 [81.5 84.3]	90.06
p-Value:				
MKL(MD,SMR,CSF) vs. sMRI	<0.001	<0.001	<0.001	
MKL(MD,SMR,CSF) vs. MD	<0.001	<0.001	<0.001	

validation. The whole set of subject samples was equally partitioned into 10 subsets, and each time the subject samples within one subset were selected as the testing samples and all remaining subject samples in the 9 other subsets were used for training the multiple kernel classifier. We report the mean average 95% confidence interval of $Accuracy\ (Acc) = (TP + TN)/(TP + TN + FN + FP)$, $Sensitivity\ (Sen) = TP/(TP + FN)$, $Specificity\ (Spe) = TN/(TN + FP)$ and $Balanced\ Accuracy\ (BAC) = 0.5_{*}(Sensitivity + Specificity)$. Here True Positives (TP) are AD patients correctly identified as AD, True Negatives (TN) are NC correctly classified as NC, False Negatives (FN) are AD patients incorrectly identified as NC and False Positives (FP) are NC incorrectly identified as AD. Similar definition holds for other binary classification problems NC vs. MCI and AD vs. MCI.

In a first part of experiments, we perform a direct feature concatenation as baseline early fusion method (Concat) that combines visual signatures h^{smri} and h^{md} into a global feature vector. The obtained 400-dimensional signature is normalized and reduced using Principle Component Analysis technique (PCA) [52]. When preserving 95% of the total energy, we obtained a 68-dimensional feature vector. We use an SVM classifier with an RBF kernel to do classification as this scheme has shown good results in our previous work [8]. To select the spread parameter σ for RBF kernel and the regularization parameter C, we performed a grid search on the training dataset and selected the values (σ, C) which gave the best classification accuracy to build the classifier. We used the LiBSVM software.[4] To evaluate performance of the baseline concatenation method we classify subjects using modality alone (sMRI-Hippo and MD-Hippo). Classification of single modality is performed in the same way as the Concat methods (an SVM classifier is trained for each modality).

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

In a second part of experiments, we compute features kernels and then we conduct the MKL classification. First, we start by combing h_{smri} and h_{md} (MKL(sMRI,MD). At the next step we add the v^{csf} features (MKL(sMRI,MD,CSF)). The SimpleMKL solver is then used to perform classification.

5.2. Classification results

Tables 2, 3 and 4 represent the classification results among different tests (Hippo-MRI, Hippo-DTI, Concat, MKL(SMRI,MD) and MKL(SMRI,MD,CSF)) in terms of mean accuracy, specificity, sensitivity with [95% Confidence Interval] for binary classification tasks (i.e., AD vs. NC, NC vs. MCI and AD vs. MCI).

Furthermore, we performed a paired Student *t*-test to assess the statistical differences in classification accuracy, sensitivity and specificity between the single modality use and the MKL fusion. We formulated the null hypothesis ($H0$) as "there is no significant improvement in performance when we combine MD and sMRI imaging-biomarkers via the MKL fusion". All statistical tests were considered significant at the $p < 0.001$ level. This means that we can confidently reject the null hypothesis and declare that the multimodal MKL-based fusion method has shown a statistically significant improvement in the experiment compared to the use of a single modality (MD or sMRI). This suggests that integrating multi-modal imaging biomarkers offers optimal results for AD subjects classification.

The experimental results show that the MKL method achieves better classification scores compared to single modality of features or a simple concatenation of multiple features. For the AD vs. NC classification problem, using the MKL fusion method we obtained 90.2% of accuracy, 97.2% of specificity and 82.92% of sensitivity in average with cross validation. However, using for example the structural MRI alone we have respectively only 81.5%, 82.63% and 80.16% of accuracy, specificity and sensitivity. A very close figure for the MD modality alone are observed (see Table 3 Hippo-MD). Adding the CSF modality in the whole MKL fusion scheme also increases the performance metrics up to 2%. From Tables 2–4, we can see that the MKL fusion of sMRI and MD improves classification performance by an average about 7%, 6% and 10% respectively for AD vs. NC, MCI vs. NC and AD vs. MCI classification problems compared to the use of a single modality. However, adding CSF features to the MKL (MD, SMR) slightly increases the performance. This can be explained by the fact that both sMRI and CSF features are derived from the same modalities(sMRI) and may have redundant information. Indeed, the correlation between features from different modalities is more informative than features computed from the same modality.

In conclusion, compared with the single modality use, the MKL fusion leads to an increase about 9% for the accuracy and 15% for the specificity and 2.8% for sensitivity (see Table 2). In addition, the baseline concatenation method gives better results than the single modality approach but the proposed MKL fusion method still performs better.

For the NC vs. MCI classification task, we obtained respectively 79.42%, 86.05% and 71.58% of accuracy, specificity and sensitivity. The MKL fusion of all imaging biomarkers yields to an increase in the classification performance by in average about 8.5% for the accuracy, 6% for the specificity and 15% for sensitivity. Additionally, combining the two visual signatures into a single vector (early fusion) improves the classification results but it is still less efficient than the MKL method.

Even for the most challenging discrimination problem (AD vs. MCI), because of the heterogeneous nature of the MCI group, MKL is still giving good classification rates (76.63% accuracy, 81.33% specificity and 65.62% sensitivity) and outperforms the single modality method by about 9% for the accuracy, 6.56% for the

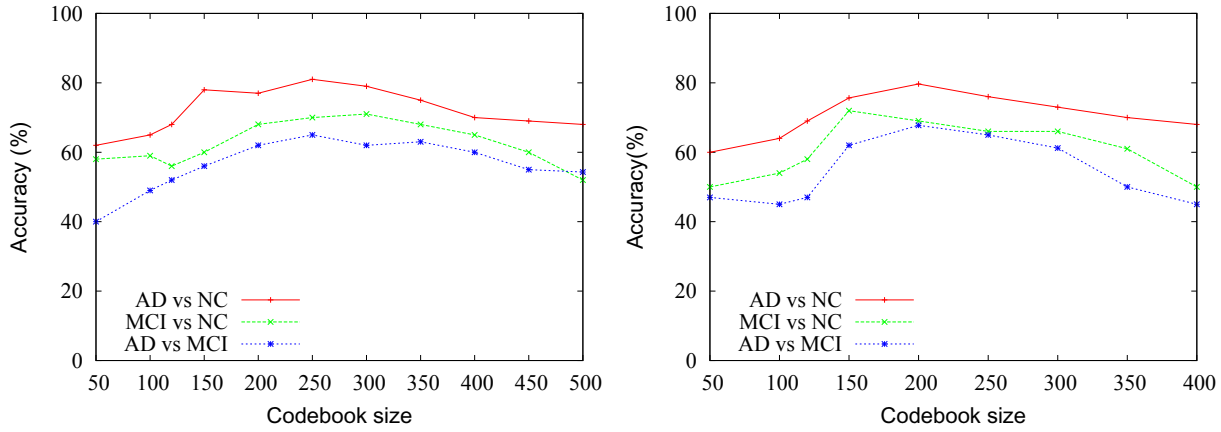

Fig. 6. Codebook variation or respectively structural MRI and MD maps.

Table 3
NC vs. MCI classification results.

MCI vs. NC Features	Acc[95%CI]	Spe[95%CI]	Sen[95%CI]	BAC%
Hippo-SMRI	71.03[70,01 72.05]	81.98[80,18 83,78]	54.16[52,96 56.36]	68.07
Hippo-MD	71.96[70,66 73.26]	80[79,19 80.81]	56.75[55.51 57,99]	68.37
Concat(MD,SMRI)	72.61[71.95 73.27]	82.40[81.19 83.61]	60.95[59.18 62.72]	71.68
MKL(MD,SMRI)	76.92[74,62 79,22]	83.95[81,15 86,75]	68.61[66,71 70,51]	76.28
MKL(MD,SMRI,CSF)	79.42[75,62 81,22]	86.05[84,94 87.16]	71.58[70.16 72]	78.81
p-Value:				
MKL(MD,SMR,CSF) vs. sMRI	<0.001	<0.001	<0.001	
MKL(MD,SMR,CSF) vs. MD	<0.001	<0.001	<0.001	

Table 4
AD vs. MCI classification results.

MCI vs. AD Features	Acc[95%CI]	Spe[95%CI]	Sen[95%CI]	BAC%
Hippo-SMRI	65.02[64,29 65.75]	74.77[73.56 75.98]	50 [49.1 50.9]	62.38
Hippo-MD	67.75[65,95 69,55]	72.41[70.31 74.52]	55.55[53,85 57,25]	65.61
Concat (MD,SMRI)	71.23[70.72 71.74]	78.68[78.27 79.10]	56.08[55.28 56.87]	67.38
MKL(MD,SMRI)	75.25[74.4 76.29]	79.84[78,97 80.71]	64.01[62,68 65,33]	71.92
MKL(MD,SMRI,CSF)	76.63[75.42 77.84]	81.33[80,42 82,24]	65.62[64,9 66,34]	73.48
p-Value:				
MKL(MD,SMR,CSF) vs. sMRI	<0.001	<0.001	<0.001	
MKL(MD,SMR,CSF) vs. MD	<0.001	<0.001	<0.001	

specificity and about 10% for the sensitivity. Fig. 7 illustrates classification performance on the most challenging subjects (MCI) group. We can see that all metrics improve when combining the MD and sMRI features using the MKL method (Fig. 7).

5.2.1. MKL fusion vs. baseline concatenation scheme

MKL fusion is a heavier computational approach than concatenation of signatures followed by PCA. Nevertheless, as we can see from the results (Tables 2 and 4) it gives an average improvement about 6% in BAC metric. The effectiveness of the MKL fusion with regard to the baseline concatenation approach can be explained by the way that concatenation method represents an equal confidence in each feature type, MKL successfully handles discrepancies in the discriminative power of different features by assigning lower weights to less discriminative feature kernels.

5.2.2. Effect of kernel weights for MKL-based multimodal imaging-biomarkers fusion

To investigate how the combining kernel weights w_{md}, w_{smri} and w_{csf} affect the classification performance of our fusion method, we set kernel weights from 0 to 1 at a step size of 0.1, under the constraint of $w_{md} + w_{smri} + w_{csf} = 1$. Figs. 8 and 9 show the accuracy and balanced accuracy values with respect to different combining weights of MRI, MD and CSF for respectively AD vs. MCI and MCI vs. NC classification problems. Note that in each plot of Figs. 8 and 9, only values in the squares of the upper triangular part are valid because of the constraint $w_{md} + w_{smri} + w_{csf} = 1$. Moreover, in each plot, the three vertices of the upper triangle (the top left, top right, and bottom left squares) present the individual-imaging biomarker based binary classification results, respectively CSF, MD and sMRI. Also, for each plot, the three edges of the upper triangle (excluding the three vertices of the upper triangle) denote two-imaging biomarkers based binary classification results using $MRI + CSF\,(w_{md} = 0)$, $MRI + DTI\,(w_{csf} = 0)$, and $CSF + DTI\,(w_{md} = 0)$. From Figs. 8 and 9, most of ACC and BAC values within squares of the upper triangle are larger than those on the three vertices and edges. These results further validate that combining imaging-biomarkers derived from different modalities can achieve better classification results than combining only two imaging biomarkers or using only one.

5.2.3. Comparison with the state of the arts

In this work, we showed that using complementary visual information derived from MD maps and structural MRI in the hippocampus ROI increases the AD/MCI subjects classification performance. The proposed fusion method exhibited a superior classification performance compared with the single modality approaches and yielded comparable accuracy with previous multimodal classification studies. However, direct comparison with existing work could not be fair due to the different subjects and modalities used, as well as the different methods for feature extraction and classification protocol.

In Table 5, we reported some work from the literature that used the MKL concept to combine different neuroimaging data for AD

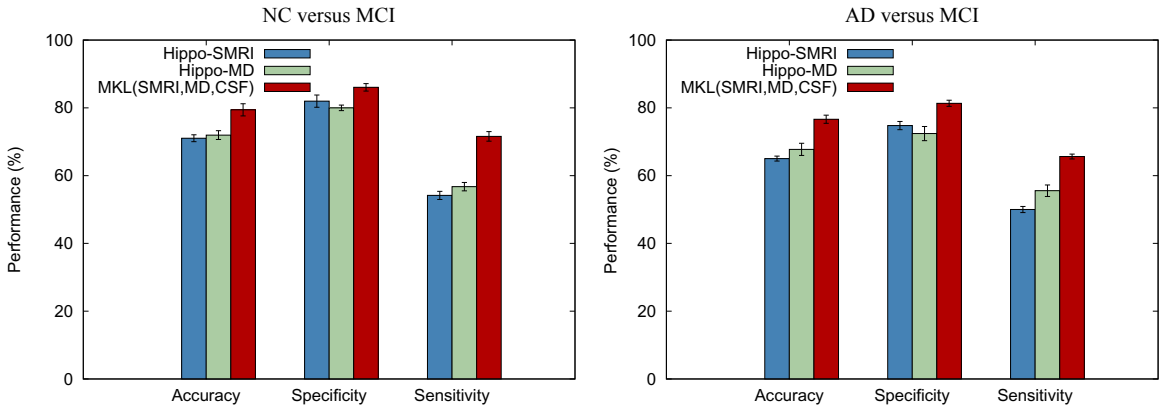

Fig. 7. Performance classification on the MCI group.

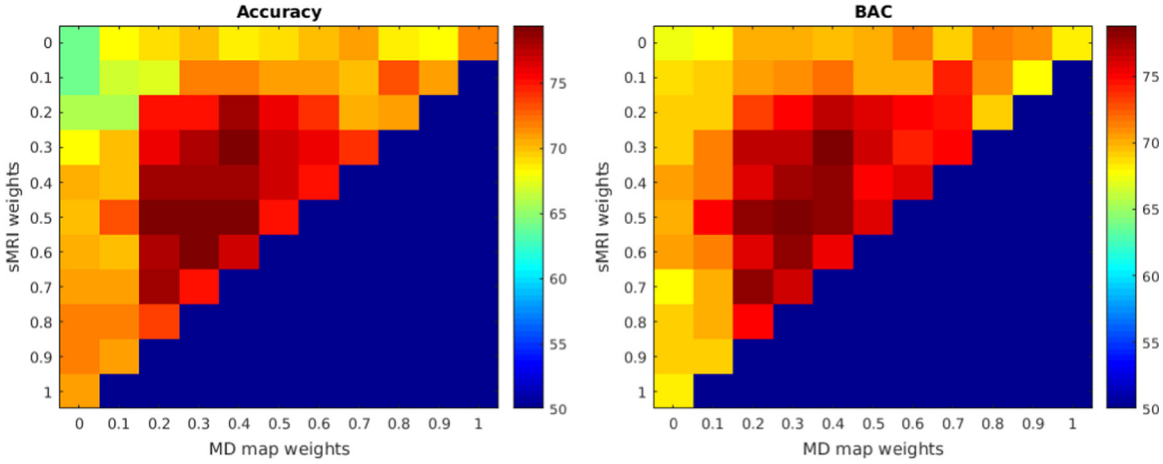

Fig. 8. MCI vs. NC classification accuracy and BAC metrics with respect to MD, sMRI, CSF imaging-biomarkers weights.

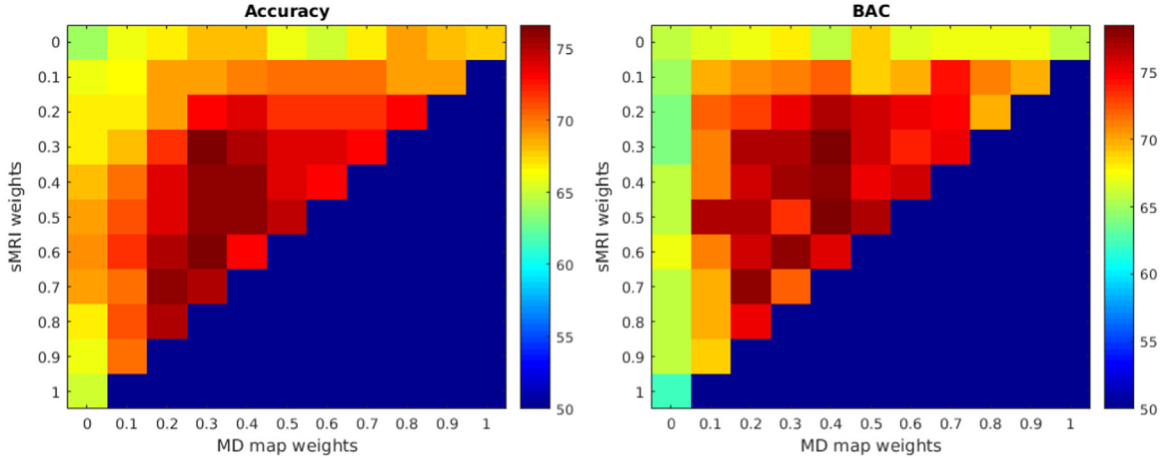

Fig. 9. MCI vs. AD classification accuracy and BAC metrics with respect to MD, sMRI, CSF imaging-biomarkers weights.

subjects discrimination. We perform comparison between methods based on BAC metric because this provides a more meaningful performance metric for diagnostic groups of unequal sizes. Ref. [18] reported an accuracy of 90.6% for AD vs. NC classification by using MRI and PET and an accuracy of 93.20% for AD classification by using MRI, PET and CSF. Besides, they reported an accuracy of 76.40% for MCI vs. NC classification by using three modalities, and obtained an accuracy of 73.79% when using only PET and MRI. In [55], the authors combined regional subcortical volumes and cortical thickness measure from MRI and CSF data. Their method gives, for distinguishing between AD and NC, an accuracy of 91.8% and 77.6% when distinguishing between MCI vs. NC. The obtained BAC in MCI vs. NC classification is lower than ours (78.81% vs. 78.75%). Finally, [54] proposed a multi-modal method to combine PET, MRI and CSF features, they obtained for the MCI vs. AD classification an accuracy of 73.21% which is lower than ours (76.63%).

Few are the works that combined the DTI modality with other modalities and specifically the structural MRI modality with the

Table 5
Comparison of classification performance with the state of the arts methods using multimodal neuroimaging data.

Works	Modalities	Data(AD/MCI/NC)	acc/sep/sen(AD vs. NC)	acc/sep/sen(MCI vs. NC)	acc/sep/sen(MCI vs. AD)
[25]	SMRI, DTI	–/79/204	NA	71.09/78.4/51.96	NA
[13]	fMRI, DTI	–/10/17	NA	96.30/100/94.12	NA
[19]	MRI, PET	48/–/66	87.6/78.9/93.8	NA	NA
[18]	MRI, PET	51/99/52	90.6/90.5/90.7	73.79/NA/NA	NA
[53]	MRI, PET	51/99/52	94.37/94.71/94.04	78.8/84.8/67.06	NA
[18]	MRI,PET,CSF	51/99/52	93.2/93/93.3	76.4/81.8/66	NA
[54]	MRI, PET, CSF	51/99/52	96.18/NA/NA	81.45/NA/NA	73.21/NA/NA
[55]	MRI, PET	96/162/111	91.8/94.6/88.5	77.6/84.7/72.8	NA
Our	sMRI,MD maps + CSF amount computed on the SMRI	45/58/52	90.2/97.2/82.92	79.42/86.05/71.58	76.63/81.33/65.62

goal of AD detection. For instance, [13] proposed to integrate connectivity information from Diffusion Tensor Imaging (DTI) and resting-state functional Magnetic Resonance Imaging (rs-fMRI) for improving AD subjects classification performance with multiple-kernel SVM. The authors used the AAL atlas to extract regional features and they achieved an accuracy increase of 7.4% from the single modality-based methods and the direct data fusion. However, this work used a relatively small sample size (10 MCI and 17 NC subjects) and did not distinguish between MCI subjects and AD. In [25], the authors combined spatial atrophy, derived from T1-weighted images, and white matter alterations assessed with DTI to MCI discrimination. They obtained 71.09% accuracy, 78.40% specificity and 51.96% sensitivity when distinguishing between MCI and NC which are lower than our results (respectively 79.42%, 86.05% and 71.58%). [26] proposed a new multimodal measure that combines anatomical and diffusivity measures at the voxel level. They extract multimodal characteristics from 73 anatomical brain regions using the AAL atlas. They obtained 72.4% accuracy, 82% specificity and 62.4% sensitivity to classify only 15 AD from 16 NC subjects. To the best of our knowledge the present work is the first work to use a combined T1-weighted MRI and MD maps in an MKL framework for the automated detection of MCI/AD subjects.

We compare our classification results with tow recently published works mainly based on both PET and MRI data to classify AD subjects [19,53]. Although the images modalities are not exactly the same, we used the same data partition (subject number) with a closed subjects demographic characteristic (MMSE,gender). First, in [19], the authors combined MRI and PET data to achieve an accuracy of 87.6% for AD vs. NC classification with 86.35% of BAC which is lower than the results obtained using our method on the same subjects (91% accuracy and 88.6% BAC). It is worth noting that, in [19], both baseline and longitudinal data are used for MRI and PET modalities, while our proposed method uses only the baseline data of DTI and sMRI data. In addition, the MCI/AD classification problem is not addressed in their work. Second, we also applied our method on the same subjects as [53]. We obtained respectively for MCI vs. NC classification an accuracy of 87.60%, compared to 78.8% reported in [53].

Most of the proposed MKL fusion frameworks in the literature, use the PET modality in addition to the anatomical MRI or CSF modality. However, PET modality involves the injection of radioactive contrast agents making PET very expensive and difficult to be performed in non-specialized centers. Conversely, DTI scans are less expensive, safer, take only several minutes and are widely available. Moreover, extraction of CSF directly from the brain is intrusive. Hence, computing the CSF amount from imaging modality, i.e., structural MRI seems to be a good idea to avoid intrusive diagnostics. The proposed method is extensible to other MRI-based brain diseases that can be diagnosed by brain MRI such as Schizophrenia and brain tumors. Indeed, the proposed features-based approach extract and quantify the local brain tissue abnormality (neurodegeneration, spot, etc.) whatever the type of brain atrophy. We need just to localize the ROI involved in the diagnosis of such disease (referring to the domain knowledge) and then apply the method to quantify atrophy and classify subjects.

6. Conclusion

In this paper, we proposed an AD/MCI recognition approach based on image-derived biomarkers and Multiple Kernel Learning method. We extract visual features from structural MRI and DTI MD maps. Extracted features are combined in a Multiple Kernel Learning framework to discriminate between AD/MCI subjects. The obtained results showed that combining imaging biomarkers from those modalities improves the diagnosis accuracies compared to the single modality use. Our proposed approach is general and extensible and can potentially be used for integration of other neuroimaging biomarkers. We intend to further evaluate our approach performance using other datasets (MCI converters and MCI non-converters) in the aim of predicting subject conversion to AD rather than recognizing subject's category.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimers Association; Alzheimers Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles.

References

[1] E. Gerardin, G. Chételat, M. Chupin, et al., Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging, NeuroImage 47 (4) (2009) 1476–1486.

[2] R.e. Cuingnet, E. Gerardin, J.e. Tessieras, 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.

[3] M. Toews, W. Wells, D.L. Collins, T. Arbel, Feature-based morphometry: discovering group-related anatomical patterns, NeuroImage (2010) 2318–2327.

[4] Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis, Neurocomputing 175, Part A (2016) 132–145.

[5] S. Landau, D. Harvey, C. Madison, E. Reiman, N. Foster, P. Aisen, R. Petersen, L. Shaw, J. Trojanowski, C. Jack, et al., Comparing predictors of conversion and decline in mild cognitive impairment, Neurology 75 (3) (2010) 230–238.

[6] C. Davatzikos, P. Bhatt, L.M. Shaw, K.N. Batmanghelich, J.Q. Trojanowski, Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification, Neurobiol. Aging 32 (12) (2011) (2322-e19).

[7] J. Ye, K. Chen, T. Wu, J. Li, Z. Zhao, R. Patel, M. Bae, R. Janardan, H. Liu, G. Alexander, et al., Heterogeneous Data Fusion for Alzheimer's Disease Study, 2008.

[8] O. Ben-Ahmed, M. Mizotin, J. Benois-Pineau, M. Allard, G. Catheline, C.B. Amar, Alzheimer's disease diagnosis on structural {MR} images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex, Comput. Med. Imaging Graph. 44 (2015) 13–25.

[9] C. Eckerström, U. Andreasson, E. Olsson, S. Rolstad, K. Blennow, H. Zetterberg, H. Malmgren, Å. Edman, A. Wallin, Combination of hippocampal volume and cerebrospinal fluid biomarkers improves predictive value in mild cognitive impairment, Dement. Geriatr. Cogn. Disord. 29 (4) (2010) 294–300.

[10] M. Liu, D. Zhang, P.-T. Yap, D. Shen, Hierarchical ensemble of multi-level classifiers for diagnosis of Alzheimer's disease, in: Machine Learning in Medical Imaging, Springer, Nice, France, 2012, pp. 27–35.

[11] R. Polikar, C. Tilley, B. Hillis, C. Clark, Multimodal EEG, MRI and PET data fusion for Alzheimer's disease diagnosis, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2010, pp. 6058–6061.

[12] Z. Dai, C. Yan, Z. Wang, J. Wang, M. Xia, K. Li, Y. He, Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (m3), NeuroImage 59 (3) (2012) 2187–2195.

[13] C.-Y. Wee, P.-T. Yap, D. Zhang, K. Denny, J.N. Browndyke, G.G. Potter, K. A. Welsh-Bohmer, L. Wang, D. Shen, Identification of MCI individuals using structural and functional connectivity networks, Neuroimage 59 (3) (2012) 2045–2056.

[14] S. Farhan, M.A. Fahiem, H. Tauseef, An ensemble-of-classifiers based approach for early diagnosis of Alzheimer's disease: classification using structural features of brain images, Comput. Math. Methods Med. 2014 (2014) 1–11.

[15] A. Rakotomamonjy, F. Bach, S. Canu, Y. Grandvalet, et al., SimpleMKL, J. Mach. Learn. Res. 9 (2008) 2491–2521.

[16] C. Hinrichs, V. Singh, G. Xu, S. Johnson, MKL for robust multi-modality ad classification, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009, Springer, Imperial College London, UK, 2009, pp. 786–794.

[17] F. Liu, L. Zhou, C. Shen, J. Yin, Multiple kernel learning in the primal for multimodal Alzheimers disease classification, IEEE J. Biomed. Health Inf. 18 (3) (2014) 984–990.

[18] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, Multimodal classification of Alzheimer's disease and mild cognitive impairment, NeuroImage 55 (3) (2011) 856–867.

[19] C. Hinrichs, V. Singh, G. Xu, S.C. Johnson, Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population, NeuroImage 55 (2) (2011) 574–589.

[20] F. Liu, C.-Y. Wee, H. Chen, D. Shen, Inter-modality relationship constrained multi-task feature selection for AD/MCI classification, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, Springer, NAGOYA JAPAN, 2013, pp. 308–315.

[21] W. Luo, J. Yang, W. Xu, J. Li, J. Zhang, Higher-level feature combination via multiple kernel learning for image classification, Neurocomputing 167 (2015) 209–217.

[22] H. Wu, L. He, Combining visual and textual features for medical image modality classification with pnorm multiple kernel learning, Neurocomputing 147 (2015) 387–394 (Advances in Self-Organizing Maps Subtitle of the special issue: Selected Papers from the Workshop on Self-Organizing Maps 2012, WSOM 2012)..

[23] J. Young, M. Modat, M.J. Cardoso, A. Mendelson, D. Cash, S. Ourselin, Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment, NeuroImage: Clin. 2 (2013) 735–745.

[24] B. Zhang, Y. Xu, B. Zhu, K. Kantarci, The role of diffusion tensor imaging in detecting microstructural changes in prodromal Alzheimer's disease, CNS Neurosci. Therapeut. 20 (1), 2013.

[25] Y. Cui, W. Wen, D.M. Lipnicki, M.F. Beg, J.S. Jin, S. Luo, W. Zhu, N.A. Kochan, S. Reppermund, L. Zhuang, P.R. Raamana, T. Liu, J.N. Trollor, L. Wang, H. Brodaty, P.S. Sachdev, Automated detection of amnestic mild cognitive impairment in community-dwelling elderly adults: a combined spatial atrophy and white matter alteration approach, NeuroImage 59 (2) (2012) 1209–1217.

[26] L. Mesrob, M. Sarazin, V. Hahn-Barma, et al., DTI and structural MRI classification in Alzheimer's disease, Adv. Mol. Imaging 2 (2012) 12–20.

[27] M. Dyrba, M. Ewers, M. Wegrzyn, I. Kilimann, et al., Robust automated detection of microstructural white matter degeneration in Alzheimer's disease using machine learning classification of multicenter DTI data, PLoS ONE 8 (5) (2013) e64925.

[28] A. Cherubini, P. Péran, I. Spoletini, M. Di Paola, F. Di Iulio, G.E. Hagberg, G. Sancesario, W. Gianni, P. Bossu, C. Caltagirone, et al., Combined volumetry and DTI in subcortical structures of mild cognitive impairment and Alzheimer's disease patients, J. Alzheimer's Dis.: JAD 19 (4) (2009) 1273–1282.

[29] M.R. Daliri, Automated diagnosis of Alzheimer disease using the scale-invariant feature transforms in magnetic resonance images, J. Med. Syst. 36 (2) (2012) 995–1000.

[30] O. Ben Ahmed, J. Benois-Pineau, M. Allard, C. Ben Amar, G. Catheline, Classification of Alzheimer's disease subjects from MRI using hippocampal visual features, Multimed. Tools Appl. (2014) 1–18.

[31] D. Unay, Augmenting clinical observations with visual features from longitudinal MRI data for improved dementia diagnosis, in: Proceedings of the International Conference on Multimedia Information Retrieval, MIR'10, ACM, New York, NY, USA, 2010, pp. 193–200.

[32] M. Agarwal, J. Mostafa, Image retrieval for Alzheimer disease detection, in: Proceedings of the First MICCAI international conference on Medical Content-Based Retrieval for Clinical Decision Support, MCBR-CDS'09, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 49–60.

[33] C.B. Akgul, D. Unay, A. Ekin, Automated diagnosis of Alzheimer's disease using image similarity and user feedback, in: Proceedings of the ACM International Conference on Image and Video Retrieval, 2009.

[34] A. Rueda, J.E. Arevalo, A. Cruz-Roa, E. Romero, F.A. González, Bag of features for automatic classification of Alzheimer's disease in magnetic resonance images, in: CIARP, 2012, pp. 559–566.

[35] B.H. Ridha, J. Barnes, L.A. Van de Pol, et al., Application of automated medial temporal lobe atrophy scale to Alzheimer disease, Arch. Neurol. 64 (6) (2007) 849–854.

[36] J.C. Felipe, A.J.M. Traina, C. Traina, Jr., Retrieval by Content of Medical Images Using Texture for Tissue Identification, 2003, pp. 175–180.

[37] Y.-Y. Qin, J.T. Hsu, S. Yoshida, A.V. Faria, K. Oishi, P.G. Unschuld, G.W. Redgrave, S.H. Ying, C.A. Ross, P.C. van Zijl, A.E. Hillis, M.S. Albert, C.G. Lyketsos, M. I. Miller, S. Mori, K. Oishi, Gross feature recognition of anatomical images based on atlas grid (gaia): incorporating the local discrepancy between an atlas and a target image to capture the features of anatomic brain MRI, NeuroImage: Clin. 3 (0) (2013) 202–211.

[38] Y. Chen, J. Storrs, L. Tan, L.J. Mazlack, J.-H. Lee, L.J. Lu, Detecting brain structural changes as biomarker from magnetic resonance images using a local feature based SVM approach, J. Neurosci. Methods 221 (0) (2014) 22–31.

[39] N. Guizard, P. Coupé, V.S. Fonov, J.V. Manjn, D.L. Arnold, D.L. Collins, Rotation-invariant multi-contrast non-local means for {MS} lesion segmentation, NeuroImage: Clin. 8 (2015) 376–389.

[40] T. den Heijer, F. van der Lijn, et al., Structural and diffusion MRI measures of the hippocampus and memory performance, NeuroImage 63 (4) (2012) 1782–1789.

[41] A. Pelletier, O. Periot, B. Dilharreguy, B. Hiba, M. Bordessoules, K. Pérès, H. Amieva, J.-F. Dartigues, M. Allard, G. Catheline, Structural hippocampal network alterations during healthy aging: a multi-modal MRI study, Front. Aging Neurosci. 5.

[42] A. Cherubini, P. Péran, I. Spoletini, M.D. Paola, F.D. Iulio, et al., Combined volumetry and DTI in subcortical structures of mild cognitive impairment and Alzheimer's disease patients, J. Alzheimer's Dis. 19 (4) (2010) 47–55.

[43] G.B. Frisoni, C. Testa, F. Sabattoli, A. Beltramello, H. Soininen, M.P. Laakso, Structural correlates of early and late onset Alzheimer's disease: voxel based morphometric study, J. Neurol. Neurosurg. Psychiatry 76 (1) (2005) 112–114.

[44] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, 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.

[45] D.V. Sorokin, M. Mizotin, A.S. Krylov, Gauss–Laguerre keypoints extraction using fast hermite projection method, in: Proceedings of the 8th International Conference on Image Analysis and Recognition – Volume Part I, ICIAR'11, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 284–293.

[46] O. Ben Ahmed, J. Benois-Pineau, M. Allard, G. Catheline, C. Ben Amar, Diffusion tensor imaging retrieval for Alzheimer's disease diagnosis, in: 2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI), 2014, pp. 1–6.

[47] M. Mizotin, J. Benois-Pineau, M. Allard, G. Catheline, Feature-based brain MRI retrieval for Alzheimer disease diagnosis, in: 19th IEEE International Conference on Image Processing (ICIP), 2012, pp. 1241–1244.

[48] J. Sivic, A. Zisserman, Efficient visual search of videos cast as text retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 31 (4) (2009) 591–606.

[49] G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L.E. Ghaoui, M.I. Jordan, Learning the kernel matrix with semidefinite programming, J. Mach. Learn. Res. 5 (2004) 27–72.

[50] J. Zhang, S. Lazebnik, C. Schmid, Local features and kernels for classification of texture and object categories: a comprehensive study, Int. J. Comput. Vis. 73 (2007) 2007.

[51] M. Mühling, R. Ewerth, J. Zhou, B. Freisleben, Multimodal video concept detection via bag of auditory words and multiple kernel learning, in: K. Schoeffmann, B. Merialdo, A. Hauptmann, C.-W. Ngo, Y. Andreopoulos, C. Breiteneder (Eds.), Advances in Multimedia Modeling, Lecture Notes in Computer Science, vol. 7131, Springer, Berlin, Heidelberg, 2012, pp. 40–50.

[52] I.T. Jolliffe, Principal Component Analysis, 2nd ed., John Wiley & Sons, Ltd, 2002.

[53] M. Liu, D. Zhang, D. Shen, The Alzheimer's disease neuroimaging initiative, Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis, Hum. Brain Mapp. 35 (4) (2014) 1305–1319.

[54] H.-I. Suk, S.-W. Lee, D. Shen, A.D.N. Initiative, et al., Subclass-based multi-task learning for Alzheimer's disease diagnosis, Front. Aging Neurosci. 6.

[55] E. Westman, J.-S. Muehlboeck, A. Simmons, Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion, NeuroImage 62 (1) (2012) 229–238.

Olfa Ben Ahmed is a research and teaching assistant at the University of Limoges and the ICONES team from XLIM research center. She received the Ph.D. degree in computer science from the University of Bordeaux in January 2015. She obtained the master degree in the New Technologies of the Dedicated Computer Systems (NTSID) in 2011 from the National School of Sfax (ENIS). Her topics of interest include content-based image classification and retrieval, medical image retrieval and indexing for computer aided diagnosis and image and video processing.

Jenny Benois-Pineau is a full professor of Computer science at the University Bordeaux and chair of Video Analysis and Indexing research group in Image and Sound Department of LABRI UMR 58000 Université Bordeaux/CNRS/IPB[HYPHEN]ENSEIRB. She has been a deputy scientific director of theme B of French national research unity GDR CNRS ISIS (2008 [HYPHEN] 2015), and is now a chair od international relations at College of Sciences and Technologies at University Bordeaux. She obtained her PhD degree in Signals and Systems in Moscou and her Habilitation à Diriger la Recherche in Computer Science and Image Processing from University of Nantes France. Her topics of interest include image and video analysis and indexing, motion analysis and content description for content[HYPHEN]based multimedia retrieval.

Michelle Allard is a professor – hospital practitioner – M.D.–Ph.D., CHU Pellegrin – Service de Mèdecine nucléaire diagnostique; CHU Pellegrin Bordeaux; EPHE – Ecole Pratique des Hautes Etudes. Her research program is based on the dual approach developed in her team of clinical psychology and multimodal imaging, essential elements for the understanding of cognitive impairment mechanisms. Her research team is a member of the Labex TRAIL. Her research interests include human brain imaging, neurodegenerative and neurodevelopmental disorders, and autism.

Gwenaëlle Catheline is an assistant professor at the University of Bordeaux. A member of the team "Neuroimaging and Human Cognition". Integrative Neurobiology Lab and Adaptive EPHE, Incia UMR CNRS 5287-Université Segalen Bordeaux. She received the Ph.D. degree in Neuroscience in 1999. Her research of interest include neurodegenerative disease (Alzheimer's Disease) functional and anatomical brain networks, neuroimaging, morphological MRI (3D, DTI) and neuroanatomy.

Chokri Ben Amar received the B.S. degree in electrical engineering from the National Engineering School of Sfax (ENIS) in 1989, the M.S. and Ph.D. degrees in computer engineering from the National Institute of Applied Sciences in Lyon, France, in 1990 and 1994 respectively. He spent one year at the University of "Haute Savoie" (France) as a teaching assistant and researcher before joining the Higher School of Sciences and Techniques of Tunis (ESSTT) as assistant professor in 1995. In 1999, he joined the Sfax University (USS), where he is currently a full professor in the Department of Electrical Engineering of the National Engineering School of Sfax (ENIS). His research interests include computer vision and image and video analysis.