Available online at www.sciencerepository.org

Science Repository

Radiology and Medical Diagnostic Imaging
(ISSN 2613 – 7836)

Classification of Alzheimer's disease from MRI using the principle of consensus segmentation based atlas

Ayman Khlif, Max Mignotte

***Abstract*—In this paper, we develop an original and reliable detection and classification framework for Alzheimer's Disease (AD) in structural Magnetic Resonance Imaging (MRI). This work exploits recent advances made in segmentation and multimedia indexing and classification for Content Based Visual Information Retrieval (CBVIR). More precisely, our strategy exploits the concept of consensus segmentation to build two segmentation prototypes (Prototype Normal Control (Prototype_NC) and Prototype Alzheimer's Disease (Prototype_AD)) of the brain in terms of cerebral Grey Matter (GM), White Matter (WM) and Cerebro-Spinal Fluid (CSF) regions. A first classification is then performed by computing a specific distance between the three-class segmentation of each MR brain image w.r.t the two prototypes. Based on a threshold on this computed distance, brain images are then classified using either Minimal Distance (MD) or K-Nearest Neighbors (KNN) classifier. Our approach has been evaluated on the baseline data of MR images of 98 subjects from the Open Access Series of Imaging Studies (OASIS) database, which contains a large number of subjects compared to current reported studies. The used subsets consist of 49 subjects who have been diagnosed with very mild to mild AD and 49 non-demented individuals. The experimental results show that our classification of patients with AD versus NC subjects achieves accuracy of 90% by performing a Leave-One-Out Cross Validation (LOOCV). Results demonstrate the validity of the proposed method and especially its simplicity and high accuracy compared to the other state-of-the-art AD diagnosis approaches proposed in the literature.**

***Index Terms*—Alzheimer's disease, classification, consensual segmentation, Global Consistency Error (GCE), K-Nearest Neighbors (KNN), Minimal Distance Classifier (MDC), Magnetic Resonance Image (MRI).**

I. Introduction

Alzheimer's disease is an irreversible neurodegenerative dementia that occurs most frequently in older adults and that gradually destroys regions of the brain that are responsible for memory, learning, thinking, and behavior [1]. Current estimates indicate that 5.3 million Americans of all ages are afflicted with this illness and this number is expected to increase to 16 million people by 2050, unless a cure is found. The socio-economic consequences of this increase are cumbersome and makes early diagnosis of AD a public health emergency.

Medical information from structural Magnetic Resonance Imaging (sMRI) has long time been the most used neuroimaging modality to detect brain atrophy in AD studies [2], [3]. In fact, two main families of methods can be distinguished to extract features from MRI for AD classification which are the morphometric methods and volumetric methods. Several studies report the use of sophisticated measurement techniques that assess anatomical changes in areas compromised by AD such as the Hippocampal Volume (HV), the Lateral Ventricles Volume (LVV), CSF Volume (CSFV) [4]–[6]. These (so-called volumetric) methods are only based on form, size and/or shape derived features extracted from the brain structures. Indeed, such volumetric measurements, require the segmentation of these ROIs from the MR images, most often manually. Furthermore, a priori assumptions about the expectedly affected brain structures is needed to select the appropriate ROI [7]. Aside from volumetric approaches, morphometric methods have gained great interest among which we can distinguish: Voxel Based Morphometry (VBM) [8] which is a widely used whole-brain analysis method, which allows an exploration of the differences in local concentrations of grey matter and white matter. Tensor Based Morphometry (TBM) [9] was proposed to identify local structural changes from the gradients of deformations fields. Object Based Morphometry (OBM) [10] was introduced to perform shape analysis of anatomical structures and recently, Features Based Morphometry (FBM) [11] was proposed as a method for relevant brain features comparison using a probabilistic model on local image features in scale-space [12], [13].

The authors are with the [a]vision lab. of the Département d'Informatique et de Recherche Opérationnelle (DIRO), Université de Montréal, Faculté des Arts et des Sciences, Montréal, H3C 3J7, QC, Canada. E-mail: mignotte@iro.umontreal.ca

Voxel based methods work directly on the voxel grid and are computationally very efficient. An advantage of these approaches, compared to the ROI-based volumetric methods, is the fact that they do not require *a priori* assumptions about the location, the size or number of ROIs to be analyzed, since they provide voxel wise measures determined in the entire brain [7]. More, there is no evidence that other regions (except hippocampus and entorhinal cortex) did not provide any information for AD and NC [14]. Recent studies on AD diagnosis found that the quantity of CSF is a bio-marker of AD [15]. Indeed, smaller hippocampal volume is associated with greater CSF amount [5].

Nevertheless, these methods are less accurate due to the limited resolution of the voxel grid and less robust to noise, mainly because of the inherent inter subject anatomical variability and the effects of a brain pathology [7]. Indeed, some pathologies may affect not only a single anatomical structure or interconnected regions, but specific structures localized far away from each other. This kind of patterns are difficult to find and analyze with the standard morphometrical techniques [7]. To cope with this issue, we have developed an automatic content-based framework for recognition of Alzheimer's disease subjects using MRI segmented scans with the concept of consensus segmentation. In our case, the principle of *consensus* segmentation allows us to build two reliable segmentation-based prototypes, one corresponding to healthy individuals and the second one corresponding to unhealthy subjects (with AD) from a training database of previously segmented brain MR images. First, let us stress us that the segmentation is a representation of an image into something that is easier to analyze than the MR image itself. In addition, these two consensus segmentation-based prototypes have also several appealing advantages whose the first one is to suppress or remove undesired components in the brain image (to be classified) such as the anatomical variability existing between individuals which are not relevant for the detection and quantification of AD. The second advantage is that the segmentation-based prototypes allows us to not to take into account the noise of the MR imaging modality into the classification procedure. In addition and finally, the difference between these two prototypes also allows us to localize the regions in the brain that are affected by the AD or to identify ROIs with a specific textural pattern in the segmentation which could represent an interesting early bio-marker of this disease (cf. fig 2). These two segmentation-based prototypes (NC and AD), built from a training segmented database can be viewed and understood as being a mean and "denoised atlas" of an healthy or unhealthy subjects (with AD). Indeed, in estimation theory, an estimator based on the average (or the weighted average) operation generally yields to an optimal denoised solution (when the noise is uncorrelated). In the consensus theory, this average (or consensus or compromise) can be achieved according to different defined criteria (see Section II-C).

With the features estimated, either by volumetric or voxel-based methods, on training cases, a classifier can be trained and applied to predict the diagnosis of a testing case, whose features are extracted in the same way. Among the most popular classifiers: Linear Discriminant Analysis (LDA) [16], Neural Network [17], linear or non-linear Support Vector Machines (SVM) [4], [18]–[20] and KNN [21].

Many of the classification studies on the detection of AD were done with both men and women. However, it has been demonstrated that brains of women are different from men to the extent that it is possible to discriminate the gender *via* MRI analysis. Moreover, it has been shown that VBM is sensitive to the gender differences [22]. For these reasons, we have been very cautious in this study; therefore, as proposed in [1], [17], [23]–[31], we have selected a set of 98 MRI [1] women's brain volumes. This subsample from OASIS contains the same number of AD patients and controls, therefore is a well class balanced sample [23]. In addition, our study was also validated for the entire database to demonstrate the stability of the proposed method regardless of the set of data used. It must be noted that this is a large number of subjects compared with the other studies referred above as it has also been mentioned in [29].

In this work we propose an automatic content analysis based framework for recognition of AD using MRI scans. Image content analysis and classification methodologies are now more and more used for medical information mining and retrieval [32], [33] with the aim of Computer-Aided Diagnosis (CAD). The proposed method exploits recent advances made in segmentation and multimedia indexing and classification for content based visual information retrieval and, more precisely, the concept of consensus segmentation to define two segmentation prototypes (Prototype_NC and Prototype_AD) of the brain in terms of cerebral GM, WM, and CSF. These two prototypes are then used with a Minimal Distance Classifier (MDC) and cleverly combined with the KNN algorithm to increase the classification accuracy.

The rest of the paper is organized as follows: Section 2 presents the related work on the consensus of segmentation in imaging, describes the combination model and the consensus model, according to different criteria (or different metrics between two segmentations), to built two segmentation pro-

[1] http://www.ehu.es/ccwintco/index.php/GIC-experimental-databases

totypes Prototype_NC and Prototype_AD. Section 3 presents the proposed MDC-KNN classifier. Section 4 presents experimental results, and some comparisons with several state-of-the-art methods and finally, Section 5 concludes this work and outlines its further perspectives.

II. Consensus of segmentations in imaging

A. Introduction

Image segmentation is a frequent pre-processing step in computer vision whose goal is to simplify and transform the representation of an image with a set of coherent significant and spatial regions with similar attributes [34]. This low-level vision task, which changes the representation of an image into something that becomes easier to analyze, is often the preliminary and crucial step in the development of many image-understanding algorithms and computer vision systems. In our case, the segmentation of a brain MR image into three kinds of regions (classes), each one associated with a specific brain anatomical tissue (CSF, WM and GM) has the merit to efficiently reduce the information content of a brain MR image and to suppress undesired components such as noise or artifacts which are not relevant for the detection and classification of AD.

Recently, the notion of *consensus* segmentation has recently been proposed in the image processing community. A *consensus* segmentation is conceptually the compromise (in terms of level of details, contour accuracy, number of regions, etc.) exhibited by each segmentation map (or spatial clustering) belonging to a set of segmentations. In some sense, the *consensus* segmentation is the average of all the individual segmentations belonging to a (so-called) segmentation ensemble, according to a defined criterion.

This concept of *consensus* clustering has been first exploited in image processing as an interesting alternative for segmenting complex scenes. Indeed, in this case, a possible and reliable segmentation approach consists in combining (fusing or merging) multiple low-cost and rough image segmentation results of the same scene associated with simpler segmentation model[2] to provide a *consensus* solution which is a final improved segmentation map. In fact, this strategy can be viewed and easily understood as a (special type of) denoising problem in which each segmentation (to be merged) is in fact a noisy segmentation solution (or observation). In estimation theory, an estimator based on the average operation generally yields to an optimal denoised solution (when the noise is uncorrelated). In the same way, in the consensus theory, an interesting denoised segmentation solution also turns to be the average (in a statistical criterion sense) of all the individual segmentations to be combined.

[2]by varying the internal parameters of a given segmentation algorithm (or seeds for stochastic algorithms) and/or by using different features for an input image or simply by using different segmentation algorithms.

Let us mention that when the different anatomical classes can be easily identified from the spatial clustering result, as it is the case of a three-class segmentation of a brain MR image (since the CSF, grey and white matter exhibit regions with increasing mean value of grey level), the simplest *consensus* segmentation model can be based on the local majority vote criterion. This technique simply consists firstly in collecting, for each pixel the class labels of the different segmentation maps of the segmentation ensemble and secondly by assigning to that central pixel, the class label that has the majority vote. A spatially regularized variant of this method consists in gathering the class labels contained in a 3D window centered on this pixel. Nevertheless, many other consensus criteria in the segmentation field exist and which have also the appealing advantage of not requiring the number of classes and to previously identify the semantic class of each segmented region. For example, we can mention the consensus model proposed in [35] that averages or merges a set of segmentation maps in the sense of inertia or intra-cluster variance criterion (or its variant [36], in the weighted intra-class inertia sense), in the Probability Rand Index [37] (PRI) sense with different optimizers or strategies [38]–[42], in the Variation of Information (VoI) [43], [44], Global Consistency Error (GCE) [45], F-measure [46], in the sense of the Support Vector Machine (SVM) criterion [47] (*i.e.*, in the maximum margin hyper-plane sense between the classes) or finally in the sense of the Maximum A Posteriori (MAP) of the Logit distribution [48] to name a few.

It is also worth mentioning that the notion of *consensus* segmentation has been also recently exploited in CBVIR [49] in order to help a user to browse through a large database in an intuitive and efficient manner. For example, it is also used for searching a specific subset or class of images in terms of their segmentation-based descriptive content and, more precisely, according to the geometrical layout and shapes of the different objects detected and segmented within the image (and not based on low level color or texture visual features as it has been commonly and widely proposed in this domain and which is limited in usefulness).

B. Consensus segmentation model principle

Let us consider that we have at our disposal, a set of L segmentations $\left\{S_k\right\}_{k\leq L} = \left\{S_1, S_2, ..., S_L\right\}$ of size N pixels to be averaged (or combined) in order to obtain either a final improved segmentation result relatively to each member of $\left\{S_k\right\}_{k\leq L}$ if each segmentation map S_k results from the same image or either to build a prototype if each segmentation is related to an anatomical structure affected by a same disease. In this latter case, which is herein considered, it allows us to provide a kind of segmented anatomical atlas that captures the segmented anatomy of unhealthy anatomical structure without taking into account the inherent anatomical shape variability existing between individuals which are not relevant for the detection and quantification of Alzheimer disease.

Conceptually, most of the consensus criteria used to fuse or average a set of segmentations, is based upon the definition of a distance or metric between two segmentation maps (measuring the similarity between the two clusterings). This is straightforwardly the case for the VoI [43], GCE [45], PRI [41], F-measure [46] criterion and this distance can be easily found for the other criteria (*e.g.*, for the majority vote criterion this distance could be simply the sum (for each pixel) of the proportion of labels that are different to the label associated with the final majority vote).

Based on this distance, the estimation of the consensus or prototype of these L segmentations can be conceptually defined as the segmentation solution $\hat{S}$ which is at the center of the segmentation ensemble $\{S_k\}_{k\leq L}$ or equivalently as being the segmentation $\hat{S}$ that minimizes the considered average pairwise distance between the consensus segmentation $\hat{S}$ and all other segmentation S_k of the segmentation ensemble. Formally, this optimization problem called the *median partition* [50]) problem can be expressed as:

$$\hat{S}_{\text{Dist.}} = \arg\min_{S\in\mathcal{S}_n}\Big\{\frac{1}{L}\sum_{k=1}^{L}\text{Dist.}(S,S_k)\Big\} \tag{1}$$

where $\mathcal{S}_n$ represents the set of all possible segmentations of size N pixels and $\sum_{k=1}^{L}\text{Dist.}(S,S_k)$ is commonly called the consensus or cost function of this energy-based consensus model for the considered distance or criterion. This optimization problem can then be solved either with deterministic gradient procedures, algebraic optimization methods, dynamic programming or either with stochastic search or simulation procedures or machine Learning based optimization schemes [38]–[43], [45]–[48], [50].

C. Criteria or distance used in consensus segmentation field

1) Global Consistency Error (GCE): The Global Consistency Error (GCE) [45] measures the extent to which one segmentation map can be viewed as a refinement of another segmentation. Segmentations which are related in this manner are considered to be consistent, since they could represent the same image segmented at different scales.

Let $S_1 = \{C_1^1, C_2^1, \ldots, C_{R^1}^1\}$ and $S_2 = \{C_1^2, C_2^2, \ldots, C_{R^2}^2\}$ be respectively the first and second segmentation and R^1 and R^2 being respectively the number of clusters C (or segments[3]) in S_1 and S_2. For a given pixel p_i, let $C^1_{<p_i>}$ and $C^2_{<p_i>}$ be the cluster that contain that pixel. The local refinement error (LRE) is then defined at pixel p_i as:

$$\text{LRE}(S_1,S_2,p_i) = \frac{|C^1_{<p_i>}\backslash C^2_{<p_i>}|}{|C^1_{<p_i>}|} \tag{2}$$

where $\backslash$ denotes the set differencing operator and $|C|$ the cardinality of the set of pixels C. As noticed in [45], this clustering (or segmentation) error measure is not symmetric and encodes a measure of refinement in one direction only. A possible and natural way to combine the LRE at each pixel into a measure which is symmetric is to consider:

$$\text{GCE}_{(S_1,S_2)} = \frac{1}{2N}\Big\{\sum_{i=1}^{N}\text{LRE}_{(S_1,S_2,p_i)} + \sum_{i=1}^{N}\text{LRE}_{(S_2,S_1,p_i)}\Big\} \tag{3}$$

This segmentation error, based on the GCE, is a distance whose values lie in the range $[0,1]$. A value of 0 indicates that the two segmentations are identical (perfect match) and a value of 1 indicates maximum deviation between the two segmentations being compared.

2) Variation of Information (VoI): The variation of information (VoI) metric is a recent information theory based measure for comparing the similarity of two segmentation results. This metric quantifies the information shared between two segmentations by, more precisely, measuring the amount of information that is lost or gained in changing from one segmentation to another [49].

$$VoI(S_1,S_2) = H(S_1) + H(S_2) - 2\,I(S_1,S_2) \tag{4}$$

Where H and I are respectively the entropies of and mutual information between segmentation S_1 and S_2. The VoI is a true distance on the space of clusterings which is positive, symmetric and obeys the triangle inequality. It takes a value of 0 when two clusterings are identical and positive otherwise. It is bounded by $\log N$ (N is the number of pixels), and if S_1 and S_2 have at most $R^{\max}$ clusters, it is bounded by $2\log R^{\max}$ [51].

3) Probabilistic Rand Index (PRI): The Rand index [37] simply computes the proportion of pairs of pixels with compatible region (or cluster[3]) label relationships between the two segmentations to be compared. if $l_i^{S_1}$ designates the region or cluster labels associated to the segmentation maps S_1 at pixel location i and if $\mathcal{I}$ is the identity function, the Rand index is given by:

$$\text{PRand}(S_1,S_2) = \frac{1}{N(N-1)}\sum_{i,j\ i\neq j}\Big\{\mathcal{I}(l_i^{S_1}=l_j^{S_1}\text{ and }l_i^{S_2}=l_j^{S_2}) + \mathcal{I}(l_i^{S_1}\neq l_j^{S_1}\text{ and }l_i^{S_2}\neq l_j^{S_2})\Big\} \tag{5}$$

A value of 1 indicates that the two segmentations are identical and a value of 0 indicates that the two segmentations do not agree on any pair of points (e.g., when all the pixels are gathered in a single region in one segmentation whereas the other segmentation assigns each pixel to an individual region). When the number of labels in S_1 and S_2 are much smaller than the number of pixels N, a computationally inexpensive estimator of the RI can be found in [37]. Since the PRI is a measure of similarity lying in $[0,1]$, a distance based on the PRI criterion can be $[1-\text{PRand}(S_1,S_2)]$.

4) The F Measure: The global F measure (or harmonic mean of precision-recall measure) [52] provides a performance score, evaluating the agreement between region boundaries

[3]In fact, two different measures can be defined, one applying on clusters or class, another applying on regions or segments. For an image, a region is a set of connected pixels belonging to the same class and a class, a set of pixels possessing similar textural characteristics.

between two segmentations. This latter measure is, in fact, deduced from the well-known precision/recall values. Qualitatively, the precision measure (P) is defined as the fraction of detections that are true boundaries; this measure is low when there is significant over-segmentation, or when a large number of boundary pixels have poor localization. The Recall (R) measure gives the fraction of the true boundaries detected [46]. A particular (vision) application can define a relative cost α related to these two quantities, which controls a harmony between P and R [52] (or equivalently focuses attention at a specific point on the precision-recall curve). The F measure is defined as:

$$F_\alpha(S_1, S_2) = \frac{PR}{\alpha\, R + (1-\alpha)\, P} \tag{6}$$

The best F measure, for a given α reflects the optimal compromise between how much true signal is required and how much false alarms can be tolerated. Since the F measure of similarity lying in $[0, 1]$, a symmetric distance based on the F criterion can be $[1 - (1/2)\{F_\alpha(S_1, S_2) + F_\alpha(S_2, S_1)\}]$.

III. Proposed classification model

A. Data description

The data analyzed in this paper was obtained from the OASIS database[4]. The data set consists of a cross-sectional collection of 416 subjects aged 18 to 96 years, including 218 subjects aged 18 to 59 years and 198 subjects aged 60 to 96 years. For the older subjects, 98 subjects NC and 100 subjects who were diagnosed with very mild to moderate AD. The detailed statistics of the data set was described in the literature [53]. As proposed in [1], [17], [23]–[31], a subset of the OASIS database, *i.e.*, 98 right-handed women (aged $65-96$ years) is considered herein and was selected in order to evaluate the detection performance of the proposed classifier. More precisely, the used subset consists of 49 subjects who have been diagnosed with very mild to mild AD (class 1) and 49 non-demented (class 2). The designation of demented and non-demented is based on the Clinical Dementia Rate (CDR). The CDR is a dementia-staging factor that categorizes subjects for impairment in each of the six domains: memory, orientation, judgment and problem solving, function in community affairs, home and hobbies, and personal care. Absence of dementia is indicated by a CDR of 0, very mild, mild and moderate are respectively represented by CDRs of 0.5, 1 and 2. The demographic information of these subjects is summarized in Table I.

TABLE I: Demographic information of the subjects in the two classification classes. Education codes correspond to the following levels of education: 1: less than high school grad., 2: high school grad., 3: some college, 4: college grad., 5: beyond college

	Very mild to mild AD	Normal Control
No. of subjects	49	49
Age	78.08 (66-96)	77.77 (65-94)
Education	2.63 (1-5)	2.87 (1-5)
Socioeconomic status	2.94 (1-5)	2.88 (1-5)
CDR (0.5/1/2)	31/17/1	0
MMSE	24 (15-30)	28.96 (26-30)

B. MRI data preprocessing

In this work, we used T1-weighted magnetic resonance images of the brain, that were captured with a T Vision 1.5 scanner. As described in [53], all (axial section) MRI images were: (a) corrected for inter-scan head movement and rigidly aligned to the Talairach and Tournoux space [54], (b) transformed to a template with a 12-parameter affine registration and merged into a 1-mm isotropic image [53], (c) skull stripping with the Brain Extraction Tool (BET[5]) and corrected for intensity inhomogeneity [55], and finally (d) segmented into $K = 3$ classes corresponding to the three existing cerebral tissues (CSF, WM and GM). This 3-class segmentation is obtained by applying firstly, 10 times the K-means clustering algorithm with different seeds and different number of neighbors of the pixel to be classified and secondly by combining them with a simple majority vote scheme[6].

In our application, the majority vote is achieved with a spatial window (of size 3×3 pixels and centered on the pixel s to be classified) that collects the class labels of the 10 segmentation results obtained by each K-mean clustering and by finally assigning to that central pixel s, the class label that has the majority vote. This strategy ensures both an efficient spatial regularization of the final segmentation result and also a reliable decision fusion between results obtained by these K-mean clusterings. In this segmentation, the *CSF*, the *white matter* and the *grey matter* are represented by a dark, a grey, and a white region respectively, in order to visually express the activity level of the blood flow. The preprocessing steps are summarized in Fig. 1 for the T1 image of one subject.

C. Prototypes NC and AD

Our classification model is based on a two-step classification procedure whose the first step relies on a minimum distance based classifier using two consensus segmentation based prototypes of the brain, in terms of CSF, WM and GM regions and built from the entire OASIS basis (see Sections II-B & II). More precisely, one brain prototype corresponds to unhealthy individuals (with AD or class 1) and is obtained by combining all segmentations related to subjects who have been diagnosed with very mild to mild AD (from 100 subjects) and the second one (class 2), corresponding to healthy subjects, is

[4]http://www.oasis-brains.org

[5]https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BEText
For all the above-mentioned pre-treatments, we use the Statistical Parametric Mapping SPM software (VBM8 toolbox) with its parameters by default.

[6]We have noticed that this simple segmentation technique provides better classification results than a SPM software (Statistical Parametric Mapping) based segmentation technique (the classification accuracy is better by 2%).

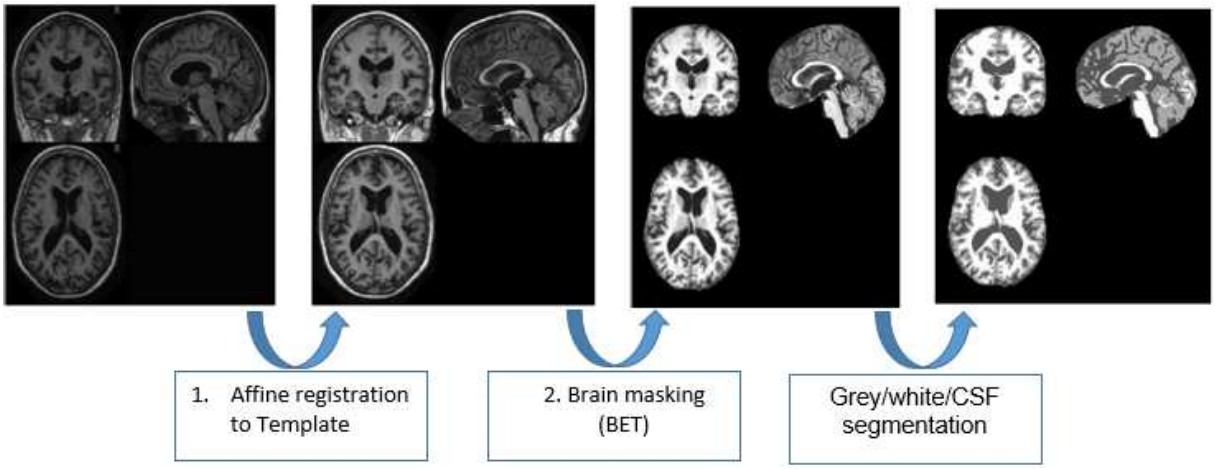

Fig. 1: Snapshot of a specific subject. (a) One original scan. (b) Atlas-registered image. (c) Brain masked version of Fig. 1 (b). (d) The GM/WM/CSF segmentation image

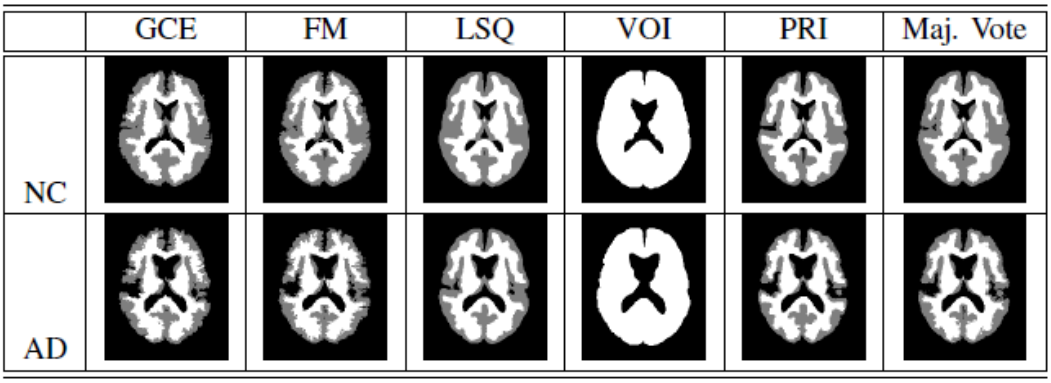

Fig. 2: Prototype AD (Alzheimer) and NC (Normal Control) according to different criteria of merging (GCE [45], FM [46], LSQ [35], VOI [43], PRI [41])

built from the set of non-demented subjects contained in the OASIS database (316 healthy subjects).

As already said (see Section II), these two consensus segmentation based prototypes can be achieved according to different criteria (each criterion being, in fact, conceptually based on the definition of a specific distance between two segmentation maps). In this work, we will test and compare the efficiency of different consensus criteria for the two prototypes used in our classification scheme. The methodology for the creation of the two consensus segmentation based prototypes is illustrated in Fig. 3.

D. Two-step classification

The proposed hybrid classification technique named MDC-KNN (for Minimum Consensus Distance-KNN), combining the previously estimated prototypes (Prototype_NC and Prototype_AD) and the KNN (weighted KNN as proposed in [56]) algorithm consists of three stages:

1) Calculate the two Consensus distances between the segmentation related to an input MRI image and respectively the Prototype Prototype_NC (D_NC) and the Prototype Prototype_AD (D_AD).
2) Choose the classifier KNN or (prototype-based) Minimal distance (MD). The choice of classifier (KNN or MD) depends on the difference between D_NC and D_AD. If the difference is large (greater than a threshold T which was set to 1.5%[7]), then, in this case (case 1 or 2 of Fig. 4), we choose the MD classifier since in this case we are certain that the segmentation is very close to one of the two prototypes and consequently we can thus rely on this classification procedure. Otherwise, we apply the weighted KNN classifier (case 3 of Fig. 4) in with the GCE distance (see Section II-C1) between the segmentation related to an input MRI image and the 3-class segmentations obtained for each image of the training data set[8].
3) Classification.

[7]We have computed than the distance (in percentage of pixel difference) between Prototype_NC & Prototype_AD is 15%. Based on this, we have set this threshold T to an order of magnitude ($\div 10$) lower (*i.e.*, $T = 1.5\%$).

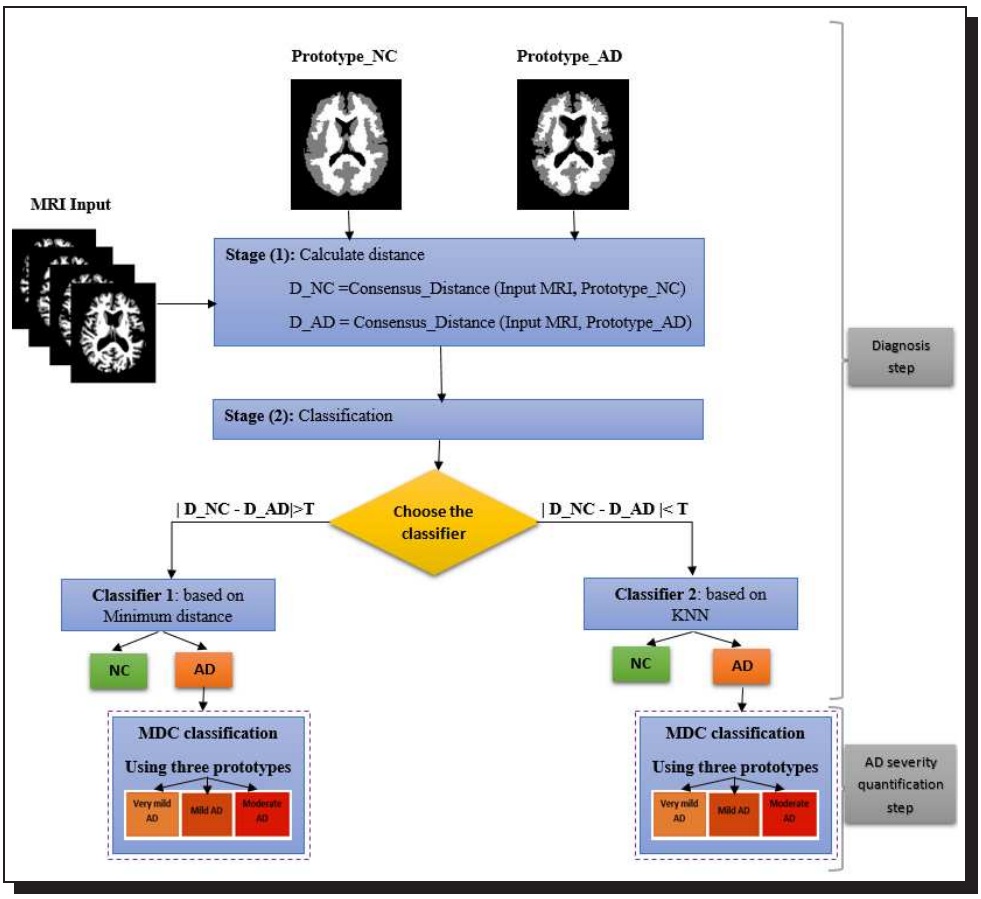

Fig. 5: The methodology of the proposed two-step classification method (see Section III-D)

Let us note that our two-step classification procedure which combines, in the first step, the MD classifier based on two consensus segmentation based prototypes followed by the second step, which uses a classical weighted KNN classifier, can be also viewed as a single hierarchical KNN classifier. Indeed the MD classifier based on our two prototypes can be seen as a simple 1-nearest neighbor in which the prototype summarizes the set of segmented brains belonging to a same pathology class (healthy or AD). The schematic diagram for the proposed two-step classification methodology is outlined in Fig. 5.

IV. Results and Discussion

In this section we describe our experiments and report the results of the proposed method. In order to investigate the detection performance of the proposed MDC-KNN classifier, a set of appropriate experiments were conducted. For the experimental purposes, specific software was developed in C++. All experiments were executed in an Intel i7-3.3 GHz PC with 16 GB RAM.

[8]To examine the performance of the classifier, a leave-one-out cross-validation approach was taken, and every subject was selected once as the test data, with the remaining subjects forming the training data.

A. Evaluation criteria

We evaluate the performance of the proposed classification method, against recent leading classification methods, in terms of sensitivity=TP/(TP + FN), specificity=TN/(TN + FP) and accuracy=(TN + TP)/(TN+TP+FN+FP) where True Positives (TP) are AD patients correctly identified as AD, True Negatives (TN) are controls correctly classified as controls, False Negatives (FN) are AD patients incorrectly identified as controls and False Positives (FP) are controls incorrectly identified as AD. Sensitivity is the proportion of AD subjects correctly classified, and the specificity is the proportion of correctly classified controls [57], [58] (see Table II).

In addition, we also compare the performance of our classification method with different consensus criteria for the creation of the two prototypes (see Section III-C and Figures 3, 2 and 5), namely; GCE, FM, Least Square (LSQ), VoI, PRI and Majority vote (see Tables III and VI).

The validation was performed using a leave-one-out cross-validation. In this scheme, a single MRI is classified using the classifier trained with the remaining observations.

The validation of the whole methodology was performed for two experimental groups, the first one consisted of a well-balanced subset of the original OASIS dataset, including 98 right-handed women (aged 65-96 years). More precisely, 49

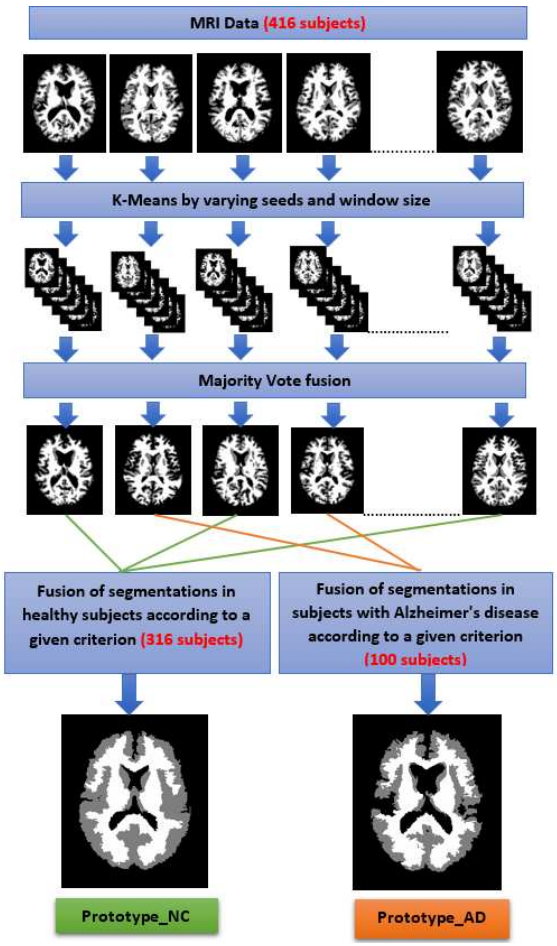

Fig. 3: Methodology for the NC and AD prototype creation

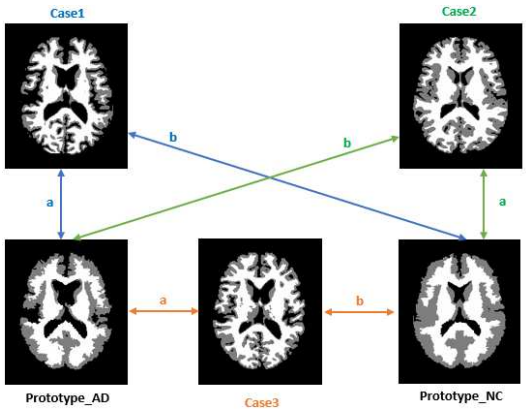

Fig. 4: $|D_{NC} - D_{AD}|$ large (case 1 or 2) or low (case 3)

control subjects and 49 patients diagnosed with AD, The demographic information about these subjects is summarized in Table I. The second one was the whole OASIS dataset, consisted of 416 subjects aged 18 to 96. The subjects are all right-handed and include both men and women subjects (156 males and 260 females). 100 of the included subjects over the age of 60 have been clinically diagnosed with very mild to moderate AD. More precisely, 316 subjects with CDR=0, 70 subjects with CDR=0.5, 28 subjects with CDR=1 and 2 subjects with CDR=2.

At first we will make the classification into two classes (NC *vs.* AD). Nevertheless, when we include the subjects with CDR = 0.5, CDR=1 and CDR=2, we have 4 classes of the disease (NC, very mild AD, mild AD and moderate AD) that are not easily separable, to distinguish between these classes we divided the task in two steps as shown in Figure 5.

B. Performance measures & comparison with state-of-the-art methods

1) Classification results into two classes (NC vs. AD): The actual feature dataset (98 subjects: 49 AD and 49 NC) has been used in several works in the literature, hence results obtained with a variety of classifier models are publicly available for comparison, for example, [1] applied a K-NN classifier defined in the LC (Lattice Computing) context, on the 98 female subjects. He obtained 80% accuracy, 79% specificity and 80% sensitivity, also, authors in [24] studied the feature-extraction process with VBM analysis and achieved the best results with 85% accuracy for the rbf-DAB-SVM classifier (see Table II). The AD detection statistics, accuracy, sensitivity and specificity, of the proposed MDC-KNN classifier by performing a leave-one-out cross validation[9] test, in comparison with the state-of-the-art models [1], [23], [24], [28], [30] are summarized in Table II regarding an OASIS subset of 98 subjects (see Section III-A).

TABLE II: Classification results comparison between our method and the morphometric approaches proposed in [1] and several other approaches (using a leave-one-out cross validation test and the same dataset)

Classifier type	Accuracy	Sensitivity	Specificity
MDC-KNN (GCE meas.)	**0.90**	**0.90**	**0.91**
rbf-DAB-SVM	0.85	0.78	0.92
PCA+Multi-Kernel SVM	0.84	0.85	0.86
PCA+Linear SVM	0.83	0.83	0.83
LVQ2	0.83	0.74	0.92
LVQ1	0.81	0.72	0.90
MDC (GCE meas.)	**0.80**	**0.81**	**0.79**
LC-KNN	0.80	0.80	0.79
rbf-AB-SVM	0.79	0.78	0.80
PCA+SRAN	0.79	0.79	0.79
MLP-BP	0.78	0.69	0.88
PNN	0.78	0.62	0.94
Linear SVM	0.78	0.72	0.88
Indep-rbf-SVM	0.75	0.56	0.95
Kernel-LICA-DC	0.74	0.96	0.52
Indep-Linear-SVM	0.74	0.51	0.97
KNN (GCE meas.)	**0.73**	**0.71**	**0.75**
Linear-AB-SVM	0.71	0.54	0.88
RBF	0.66	0.65	0.68

[9]This procedure iteratively leaves out the information on each subject and trains the model on the remaining subjects for subsequent class assignation of the person that was not included in the training procedure.

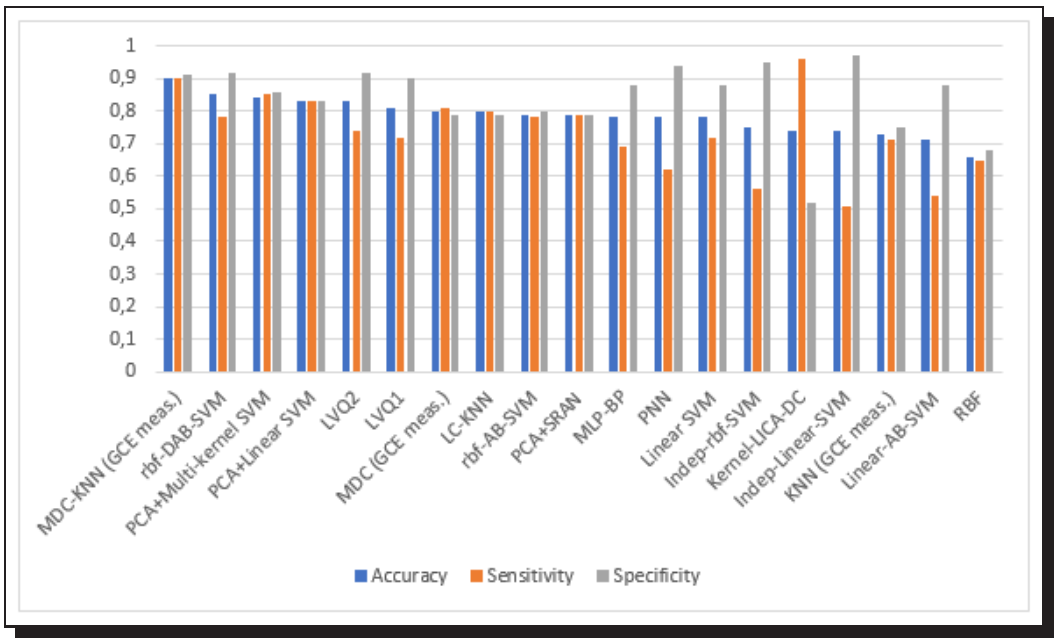

Fig. 6: Performance comparison between our method and the morphometric approaches proposed in [1] for AD *vs.* NC

TABLE III: Comparison of the classification results of our approach with different consensus criteria for the creation of prototypes (98 subjects)

Criterion	Accuracy	Sensitivity	Specificity
GCE	0.90	0.90	0.91
FM	0.88	0.81	0.95
Least Square	0.88	0.86	0.90
VoI	0.79	0.61	0.97
PRI	0.85	0.85	0.86
Majority Vote	0.89	0.88	0.90

TABLE IV: Confusion matrix of classification accuracy for classifier MDC-KNN (GCE measure)

2 class problem	NC subjects	AD subjects
NC subjects	45	4
AD subjects	5	44

In order to evaluate the performance of our classification pipeline to distinguish Alzheimer's disease (AD) patients from normal control subjects, we can use confusion matrix (See Table IV):

- Altogether, the classifier made 98 predictions (98 subjects were classified in AD or NC class).
- Out of 98 subjects, our model correctly classified 89 subjects: 44 were correctly classified as AD, and 45 of them were correctly classified as NC. This result to 90% accuracy.
- Further, 5 out of 98 subjects were classified falsely: 5 subjects, which were actual NC, were not predicted as NC (False Negative). And 4 subjects, which were actual AD, were not predicted as AD (False Positive).

By computing additional measures such as Matthews Correlation Coefficient (MCC) and F1 Score from the classification matrix, we can get additional insights about our approach (See Table V).

By examining the results of Table II & Fig. 6, it follows that the proposed classifier MDC-KNN with GCE measure (for the MD classifier) and $K = 3$ (this number was empirically tested) demonstrated superior performance than some conventional classifiers such as RBF, MLP-BP, PNN, Linear SVM (12% higher accuracy) and some advanced classification models (Indep-linear-SVM, Indep-rbf-SVM, linear-AB-SVM, rbf-AB-SVM, Kernel-LICA-DC, LVQ1, LVQ2, rbf-DAB-SVM, PCA+Linear SVM, PCA+SRAN, PCA+Multi-Kernel SVM). Let us note that the single KNN classifier (without being combined with our prototype based MD classifier) with the GCE distance reaches a classification value of only 73%. Conversely, the single MD classifier (without being combined with our KNN classifier) with the GCE distance reaches a classification value of only 80% (See Table II). In addition, we have processed a total of 416 MR images from 100 AD and 316 control subjects and noticed that our classification results is stable over the entire database and thus reliable (see Table VI).

Let us note that our method (MDC-KNN) achieves the best result with the GCE measure or criterion and the second best result with the Majority_Vote criterion (see Tables III & VI). This indicates that the GCE is an interesting distance measure between segmentations with which useful consensus-based segmentations can be efficiently exploited in a classification framework.

2) Classification results into four classes (NC, very mild AD, mild AD and moderate AD): Previous classification aims

TABLE V: Computation of basic measures from the confusion matrix of Table IV

Measure	Value	Derivations
Sensitivity	0.900	TPR = TP / (TP + FN)
Specificity	0.917	SPC = TN / (FP + TN)
Precision	0.918	PPV = TP / (TP + FP)
Accuracy	0.908	ACC = (TP + TN) / (TP + TN + FP + FN)
F1 Score	0.909	F1 = 2TP / (2TP + FP + FN)
Matthews Correlation Coefficient	0.817	$\frac{TP*TN-FP*FN}{\sqrt{((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))}}$

TABLE VI: Comparison of the classification results of our approach with different consensus criteria for the creation of prototypes (for the entire database : 416 subjects)

Criterion	Accuracy	Sensitivity	Specificity
GCE	0.90	0.88	0.91
FM	0.89	0.88	0.90
Least Square	0.88	0.86	0.89
VOI	0.81	0.62	0.88
PRI	0.86	0.85	0.87
Majority Vote	0.90	0.87	0.91

TABLE VII: Confusion matrix of the classification accuracy for the classifier MDC-KNN (GCE measure) for a 4-class classification task (416 subjects)

4 class problem	NC	Very mild AD	Mild AD	Moderate AD	Accuracy
NC	274	26	16	0	86.7%
Very mild AD	8	58	4	0	82.9%
Mild AD	0	2	24	2	85.7%
Moderate AD	0	0	0	2	100%

to distinguish between two separated classes (AD *vs.* NC). Nevertheless, when we include the subjects with CDR = 0.5, CDR=1 and CDR=2, we have 4 classes of the disease (NC, very mild AD, mild AD and moderate AD), and it is interesting to distinguish between these different classes. To this end, we can divide the task into two stages:

- Step I: in this step, all the subjects are taken and classified as NC or AD as described in Fig. 5, the AD class includes patients with very mild to moderate AD, this first step can be viewed as a diagnosis step.
- Step II: in this step, we use the MDC classifier based on three new prototypes, *i.e.*, respectively built from brain segmented images having (in the training base) a CDR=0.5, CDR=1 and CDR=2 (Fig. 7) to classify them into the very mild AD, mild AD and moderate AD classes, this second step allows us to quantify the severity of the disease.

The classification performance with the second experimental group achieved 86% with the GCE measure for the entire database (416 subjects) and 87% for the subset of database (98 subjects).

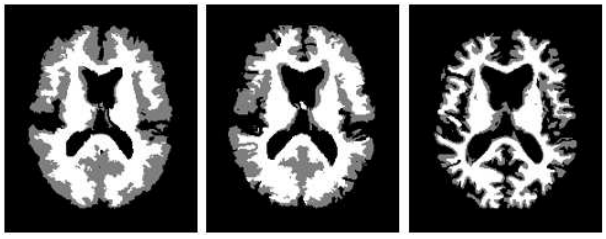

Fig. 7: From the lexicographic older: prototypes using the GCE distance and built from (a): CDR=0.5 (very mild AD), (b) CDR=1 (mild AD), (c) CDR=2 (moderate AD)

C. Computation time

The averaged computation time of every stage is listed in Table VIII. The preprocessing (registration) costs the most time, namely 35 minutes for the total dataset (48 Gigabytes representing 416 subjects). The 3-class segmentation of the dataset costs about 20 minutes. Afterwards, the creation of the two prototypes cost approximately 5 minutes. Each additional registration, for a subject outside the dataset, takes approximately 5 sec. Finally, the classification of MDC-KNN takes 0.7 sec. In practice, let us note that the construction of the prototype with the registration and the segmentation of the initial dataset is off-line. Consequently, it costs at most 5 sec. to perform a new registration, 3 sec. for the segmentation step and about 0.7 sec. to get the computer-aided diagnosis for each new patient for a non-optimized C++ code under Linux for a Intel i7 3.3 GHz PC.

TABLE VIII: Averaged computation time

Stage	Time
Registration of the dataset	35 min.
Segmentation of the dataset	20 min.
Prototypes construction	5 min.
Additional registration and segmentation	8 sec.
MDC-KNN classification	0.7 sec.

D. Discussion

First, we would like to recall that the use of consensus segmentation-based prototypes is the core idea of our proposed detection and classification approach and emphasize why these prototypes have several appealing characteristics for our classification task.

Since these two consensus models are built from segmentation maps, it allows us to reduce the information content of a brain MR image and to suppress undesired components such

as noise which may degrade the classification performance. Second, as a consensus model, it also allows us to suppress undesired components in the brain image such as the anatomical variability existing between individuals (of the same group) which are not relevant for the detection and quantification of AD (and consequently to improve the classification scheme).

We have compared the performance of our classification method with different segmentation consensus criteria, for the creation of the two prototypes namely; GCE, F-Measure, Least Square (LSQ), VoI, PRI and Majority vote. It turns out that the GCE seems the most relevant criterion for modeling a healthy or unhealthy segmented brain consensus, for the specific application of detection and quantification of AD. This is not surprising and can be explained by the fact that the GCE use a (purely) geometrical criterion for building the consensus segmentation-based prototypes. Conceptually speaking, the GCE searches the segmentation solution that minimizes the geometrical difference or overlap (see Equation 2) between its segmented regions (in terms of CSF, WM and GM) and the segmented regions of the different segmentation maps to be combined. It is meaningful since recent studies on AD diagnosis found that the excess of CSF or locally the lack of GM (relative to the WM region), at certain specific parts of the brain, is a good bio-marker of AD.
Another explanation may be due to the nature of this GCE criterion which can take into account the inherent multiscale nature of any segmentation by measuring the level of refinement existing between two segmentations.

Let us also add that the proposed approach yields higher classification accuracy compared to the other state-of-the-art methods. In addition, the proposed method produces small classification errors with the appealing property to be well balanced across the predictions, as indicated in the confusion matrices.

Combining MDC and KNN seems especially interesting in our classification method maybe because our two-step procedure can be considered as a hierarchical, multiscale or a coarse-to-fine KNN classifier in which the MD classifier can be regarded, at the coarsest scale, as a 1-nearest neighbour in which each prototype is assumed to summarize (with a well-suited geometrical criterion, as proposed by GCE) the set of segmented brains belonging to a same pathology class.

Finally, it is worth mentioning that, in the clinical practice, the results presented in our paper should be used in conjunction with other data (Neuropsychological Tests, Early Diagnosis, etc.) to reach a classification performance that may turn out useful in practice. Indeed, we will be able to see, in the following section, how mixed the two-class distributions are (*i.e.*, the two prototypes are closed to each other in term of GCE similarity distance, see Fig. 8).

E. Visualization of image databases

In order to appreciate all the relevancy of the GCE distance, used in our classification procedure, it may be interesting to find a strategy to provide a quick overview of how are distributed the 49 healthy and 49 unhealthy brains of our subset of the OASIS database with our two prototypes according to the GCE distance. This can be done with the Multi-Dimensional Scaling (MDS) and the technique described in [49] with, in our application the GCE distance between pair of segmentations (instead of the VoI distance as applied in [49]). To this end, it consists in computing the distance matrix describing the dissimilarities between each existing pair of segmentations, in term of GCE distance, and used this distance matrix with a technique which will attempt to find an embedding, in a (for example) 2D space, such that pairwise distances between these segmentations are preserved as much as possible. This embedding method, aiming at preserving the original relationships of these images (in term of a given distance), is done with the MDS visualization technique in the least square sense. It is now used in navigation systems, browsing or image database in which it may be interesting to arrange the images of the database according to their descriptive content extracted by a segmentation process, or otherwise said, based on the spatial arrangement of the different objects detected or segmented in the image (instead of simply their own color or texture) [49].

This 2D MDS visualization map of the MRIs based on the GCE distance is shown in Fig. 8. The MDS estimates a (possible) 2D mapping with 17% of error (or loss of information if we consider the MDS technique as a dimensionality reduction scheme). In spite of this error, we can see two clusters; the blue brains are related to the healthy brains and are in a cluster located at the bottom and left of the image and conversely the red brains, corresponding to unhealthy brains are in a cluster located at the top and right of the image.

V. Conclusion

In this paper, we have presented a novel and reliable prototype-based classification framework, in structural MRI, for the early detection and classification of the Alzheimer's disease. The proposed framework relies on two prototypes, based on the recent concept of consensus segmentation, to define two average models of segmented brain, corresponding to healthy subjects or unhealthy individuals affected by dementia. These two consensus segmentation-based prototypes have several appealing characteristics for our classification task. First, since these two consensus models are built from segmentation maps, it allows us to reduce the information content of a brain MR image and to suppress undesired components such as noise which may degrade the classification performance. Second, as a consensus model, it also allows us to suppress undesired components in the brain image such as the anatomical variability existing between individuals (of the same group) which are not relevant for the detection and quantification of AD. In this study, different consensus criteria of segmentations have been tested and it turns out that the Global Consistency Error seems the most relevant criterion for modeling a healthy or unhealthy brains consensus which could be subsequently used for the early detection of AD in structural MRI. This can be due to the inherent nature of this criterion

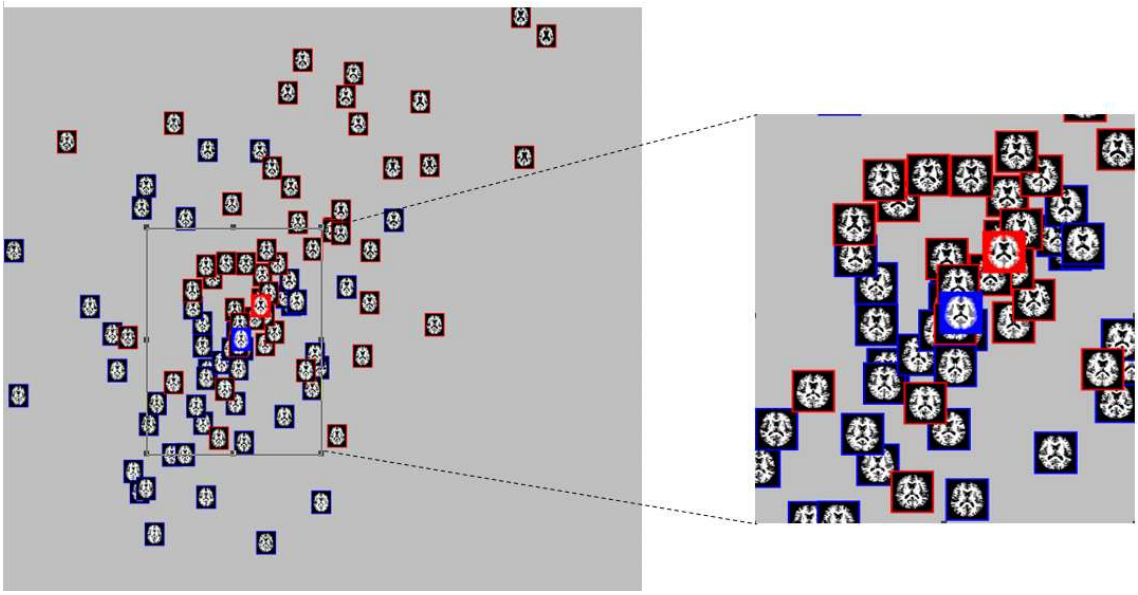

Fig. 8: A MDS visualization map of the considered subset of the OASIS database (49 healthy and 49 unhealthy subjects) based on the GCE distance between (pair of) segmentations according to their similarity. MRIs framed in blue are NC and MRIs framed in red are AD. the large red and blue frame is respectively the AD and NC prototypes

which can take into account the inherent multiscale nature of an image segmentation by measuring the level of refinement existing between two segmentations. Experiments, conducted on 98 subjects show the validity of the proposed method and especially its simplicity and high accuracy compared to the other state-of-the-art AD diagnosis approaches proposed in the literature. In addition,

- Our approach is less time-consuming compared to the state of the arts computer-based Volumetric methods, and is fully automatic (it does not require the intervention of an expert during the classification/retrieval phase).
- It is extensible to other diseases that can be diagnosed by brain MRI such as Schizophrenia.
- The method could be extended by combining axial, coronal, and sagittal MRI data for improving the classification accuracy.
- The method is fast and reliable, so it can be used for clinical applications.
- As the proposed approach rely on the training phase, so increasing the number of the data in the training set can improve the performance of the approach and leads to a better detection in subjects with/without AD.

ACKNOWLEDGMENT

Authors would like to thank FESP-UDeM (Faculty of Graduate and Postdoctoral studies -University of Montreal) for having supported this work.

REFERENCES

[1] G. Papakostas, A. Savio, M. Graña, and V. Kaburlasos, "A lattice computing approach to alzheimer's disease computer assisted diagnosis based on MRI data," *Neurocomputing*, vol. 150, pp. 35–42, 2015.

[2] N. Villain, B. Desgranges, F. Viader, V. D. L. Sayette, F. Mézenge, B. Landeau, J.-C. Baron, F. Eustache, and G. Chételat, "Relationships between hippocampal atrophy, white matter disruption, and gray matter hypometabolism in alzheimer's disease," *Journal of Neuroscience*, vol. 28, no. 24, pp. 6174–6181, 2008.

[3] O. B. Ahmed, J. Benois-Pineau, M. Allard, G. Catheline, and C. B. Amar, "Recognition of alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning," *Neurocomputing*, vol. 220, pp. 98–110, 2017.

[4] P. de Magalhães Oliveira Jr, R. Nitrini, G. Busatto, C. Buchpiguel, J. Sato, R. João, and E. A. Jr, "Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 19, no. 4, pp. 1263–1272, 2010.

[5] R. Brian, R. Cohen, A. Gongvatana, O. Okonkwo, C. Johanson, E. Stopa, J. Donahue, G. Silverberg, A. D. N. Initiative *et al.*, "Brain ventricular volume and cerebrospinal fluid biomarkers of alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 20, no. 2, pp. 647–657, 2010.

[6] S. Lu, Y. Xia, W. Cai, M. Fulham, D. Feng, D. Dagan, A. D. N. Initiative *et al.*, "Early identification of mild cognitive impairment using incomplete random forest-robust support vector machine and fdg-pet imaging," *Computerized Medical Imaging and Graphics*, 2017.

[7] O. B. Ahmed, "Features-based MRI brain classfi
cation with domain knowledge : Application to alzheimer's disease diagnosis," Ph.D. dissertation, University of bordeaux; University of Sfax, 2015.

[8] J. Ashburner and K. Friston, "Voxel-based morphometry the methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.

[9] R. Wolz, V. Julkunen, J. Koikkalainen, E. Niskanen, D. Zhang, D. Rueckert, and A. D. N. Initiative, "Multi-method analysis of MRI images in early diagnostics of alzheimer's disease," *PloS one*, vol. 6, no. 10, p. e25446, 2011.

[10] J.-F. Mangin, D. Rivière, A. Cachia, D. Papadopoulos-Orfanos, D. L. Collins, A. C. Evans, and J. Régis, "Object-based strategy for morphometry of the cerebral cortex," in *Biennial International Conference on Information Processing in Medical Imaging*. IPMI, July 2003, pp. 160–171.

[11] M. Toews, W. Wells, D. Collins, and T. Arbel, "Feature-based morphometry: Discovering group-related anatomical patterns," *NeuroImage*, vol. 49, no. 3, pp. 2318–2327, 2010.

[12] O. Ahmed, J. Benois-Pineau, M. Allard, C. B. Amar, and G. Catheline, "Classification of alzheimer's disease subjects from MRI using hippocampal visual features," *Multimedia Tools and Applications*, vol. 74, no. 4, pp. 1249–1266, 2015.

[13] O. Ahmed, M. Mizotin, J. Benois-Pineau, M. Allard, G. Catheline, C. C.B. Amar, A. D. N. Initiative *et al.*, "Alzheimer's disease diagnosis on structural mr images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex," *Computerized Medical Imaging and Graphics*, vol. 44, pp. 13–25, 2015.

[14] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, and T.-F. Yuan, "Detection of subjects and brain regions related to alzheimer's disease using 3d MRI scans based on eigenbrain and machine learning," *Frontiers in Computational Neuroscience*, vol. 9, p. 66, 2015.

[15] L. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. Clark, P. Aisen, R. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk *et al.*, "Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects," *Annals of neurology*, vol. 65, no. 4, pp. 403–413, 2009.

[16] B. M. French, M. R. Dawson, and A. R. Dobbs, "Classification and staging of dementia of the alzheimer type: a comparison between neural networks and linear discriminant analysis," *Archives of neurology*, vol. 54, no. 8, pp. 1001–1009, 1997.

[17] A. Savio, M. García-Sebastián, C. Hernández, M. Graña, and J. Villanúa, "Classification results of artificial neural networks for alzheimer's disease detection," in *International Conference on Intelligent Data Engineering and Automated Learning*, September 2009, pp. 641–648.

[18] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, and H. Benali, "Support vector machine-based classification of alzheimer's disease from whole-brain anatomical MRI," *Neuroradiology*, vol. 51, no. 2, pp. 73–83, 2009.

[19] L. Mesrob, B. Magnin, O. Colliot, M. Sarazin, V. Hahn-Barma, B. Dubois, and H. Benali, "Identification of atrophy patterns in alzheimer's disease based on SVM feature selection and anatomical parcellation," in *International Workshop on Medical Imaging and Virtual Reality*, August 2008, pp. 124–132.

[20] A. Savio, M. Grańa, and J. Villanúa, "Deformation based features for alzheimer's disease detection with linear SVM," in *International Conference on Hybrid Artificial Intelligence Systems*, May 2011, pp. 336–343.

[21] E. El-Dahshan, T. Hosny, and A. Salem, "Hybrid intelligent techniques for MRI brain images classification," *Digital Signal Processing*, vol. 20, no. 2, pp. 433–441, 2010.

[22] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. Resnick, and C. Davatzikos, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," *Neuroimage*, vol. 21, no. 1, pp. 46–57, 2004.

[23] D. Chyzhyk, M. Graña, A. Savio, and J. Maiora, "Hybrid dendritic computing with kernel-lica applied to alzheimer's disease detection in MRI," *Neurocomputing*, vol. 75, no. 1, pp. 72–77, 2012.

[24] A. Savio, M. T. García-Sebastián, D. Chyzyk, C. Hernández, M. Graña, A. Sistiaga, A. D. Munain, and J. Villanúa, "Neurocognitive disorder detection based on feature vectors extracted from vbm analysis of structural MRI," *Computers in biology and medicine*, vol. 41, no. 8, pp. 600–610, 2011.

[25] N. Aggarwal, B. Rana, and R. Agrawal, "Detection of alzheimer's disease via statistical features from brain slices," in *FLAIRS 2013 - Proceedings of the 26th International Florida Artificial Intelligence Research Society Conference*, Jan. 2013, pp. 172–175.

[26] D. Chyzhyk and A. Savio, "Feature from structural MRI images based on VBM : data from oasis database," *University of the Basque Country, Internal research publication*, 2010.

[27] D. Chyzyk, M. Termenon, and A. Savio, "A comparison of VBM results by SPM, ICA and LICA," *Hybrid Artificial Intelligence Systems*, pp. 429–435, 2010.

[28] A. S, M. Kang, S. Shin, J.-Y. Pyun, and G.-R. Kwon, "Classification of alzheimer's disease using combination of each recognition method in mr scans," in *Biomedical Engineering and Science BIOENG*, 2015.

[29] A. Savio, M. García-Sebastián, M. Graña, and J. Villanúa, "Results of an adaboost approach on alzheimer's disease detection on MRI," in *Bioinspired Applications in Artificial and Natural Computation*, 2009, pp. 114–123.

[30] S. Alam, M. Kang, J.-Y. Pyun, and G.-R. Kwon, "Performance of classification based on pca, linear SVM, and multi-kernel SVM," in *Ubiquitous and Future Networks (ICUFN), 2016 Eighth International Conference on*. IEEE, 2016, pp. 987–989.

[31] S. Aruna and S. Chitra, "Machine learning approach for identifying dementia from mri images," *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 9, no. 3, pp. 881–888, 2016.

[32] H. Muller and T. M. Deserno, "Content-based medical image retrieval," *Biomedical Image Processing*, pp. 471–494, 2010.

[33] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data," *Journal of digital imaging*, vol. 26, no. 6, pp. 1025–1039, 2013.

[34] P. Zhang, G. Wu, Y. Gao, P.-T. Yap, and D. Shen, "A dynamic tree-based registration could handle possible large deformations among mr brain images," *Computerized Medical Imaging and Graphics*, vol. 52, pp. 1–7, 2016.

[35] M. Mignotte, "Segmentation by fusion of histogram-based k-means clusters in different color spaces," *IEEE Transactions on Image Processing*, vol. 17, pp. 780–787, 2008.

[36] R. Harrabi and E. B. Braiek, "Color image segmentation using multi-level thresholding approach and data fusion techniques: application in the breast cancer cells images," *EURASIP Journal on Image and Video Processing*, vol. 2012, 2012.

[37] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.

[38] A. Fred and A. Jain, "Data clustering using evidence accumulation," in *In Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02)*, August 2002, pp. 276–280.

[39] P. Wattuya, K. Rothaus, J.-S. Prassni, and X. Jiang, "A random walker based approach to combining multiple segmentations," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR'08)*, Tampa, Florida, USA, December 2008, pp. 1–4.

[40] S. Ghosh, J. Pfeiffer, and J. Mulligan, "A general framework for reconciling multiple weak segmentations of an image," in *Proceedings of the Workshop on Applications of Computer Vision, (WACV'09)*, Snowbird, Utah, USA, December 2009, pp. 1–8.

[41] M. Mignotte, "A label field fusion Bayesian model and its penalized maximum Rand estimator for image segmentation," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1610–1624, June 2010.

[42] A. Alush and J. Goldberger, "Ensemble segmentation using efficient integer linear programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1966–1977, 2012.

[43] M. Mignotte, "A label field fusion model with a variation of information estimator for image segmentation," *Information Fusion*, vol. 20, pp. 7–20, 2014.

[44] D. Nguyen, S. Benameur, M. Mignotte, and F. Lavoie, "Superpixel and entropy-based multi-atlas fusion framework for the segmentation of x-ray images," in *18th International Conference on Image Analysis and Processing, ICIAP'15*,, Genova, Italy, September 2015, pp. 151–161.

[45] L. Khelifi and M. Mignotte, "A novel fusion approach based on the global consistency criterion to fusing multiple segmentations," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2016.

[46] C. Helou and M. Mignotte, "A precision-recall criterion based consensus model for fusing multiple segmentations," *International Journal of Signal and Image Processing, IJSIP*, vol. 7, pp. 61–82, 2014.

[47] X. Ceamanos, B. Waske, J. A. Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson, "A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data," *International Journal of Image and Data Fusion*, vol. 1, no. 4, pp. 293–307, 2010.

[48] B. Song and P. Li, "A novel decision fusion method based on weights of evidence model," *International Journal of Image and Data Fusion*, vol. 5, no. 2, pp. 123–137, 2014.

[49] A. Khlif and M. Mignotte, "Segmentation data visualizing and clustering," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 1531–1552, 2017.

[50] S. Vega-Pons and J. Ruiz-Shulcloper, "A survey of clustering ensemble algorithms," *International Journal of Pattern Recognition and Artificial Intelligence, IJPRAI*, vol. 25, no. 3, pp. 337–372, May 2011.

[51] M. Meila, "Comparing clustering an information based distance," *Journal of multivariate analysis*, vol. 98, no. 5, pp. 873–895, 2007.

[52] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 5, pp. 530–549, 2004.
[53] D. Marcus, T. Wang, J. Parker, J. Csernansky, J. Morris, and R. Buckner, "Open access series of imaging studies (oasis): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of cognitive neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.
[54] J. Talairach and P. Tournoux, *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*. G. Thieme, 1988, 1988.
[55] S. Smith, "Fast robust automated brain extraction," *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.
[56] T. Bailey and A. Jain, "A note on distance-weighted k-nearest neighbor rules," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 4, pp. 311–313, 1978.
[57] M. López, J. Ramírez, J. Górriz, D. Salas-González, I. Illan, F. Segovia, and C. Puntonet, "Principal component analysis-based techniques and supervised classification schemes for the early detection of alzheimer's disease," *Neurocomputing*, vol. 74, no. 8, pp. 1260–1271, 2011.
[58] L. Khedher, J. Ramírez, J. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, pp. 139–150, 2015.

Ayman Khlif received the engineering diploma in computer science from the Tunisian engineering ENIS-SFAX in June 2012 He is currently a PhD student in computer science at the University of Montreal under the supervision of Pr. Max Mignotte. in the Department of Computer Science and Operational Research (DIRO). His research interests include web browsing, multimedia, image segmentation, fusion and change detection.

Max Mignotte received the DEA (Postgraduate degree) in Digital Signal, Image and Speech processing from the INPG University, France (Grenoble), in 1993 and the Ph.D. degree in electronics and computer engineering from the University of Bretagne Occidentale (UBO) and the digital signal laboratory (GTS) of the French Naval academy, France, in 1998. He was an INRIA post-doctoral fellow at University of Montreal (DIRO), Canada (Quebec), from 1998 to 1999. He is currently with DIRO at the Computer Vision & Geometric Modeling Lab as Professor at the University of Montreal. He is also a member of LIO (Laboratoire de recherche en imagerie et orthopedie, Centre de recherche du CHUM, Hopital Notre-Dame) and researcher at CHUM. His current research interests include statistical methods, Bayesian inference and energy-based models (especially encoding non-local pairwise pixel interactions) for solving diverse large-scale high-dimensional ill-posed inverse problems in imaging.