Extracting Salient Brain Patterns for Imaging-Based Classification of Neurodegenerative Diseases

Andrea Rueda, Fabio A. González, *Senior Member, IEEE*, and Eduardo Romero*

***Abstract*—Neurodegenerative diseases comprise a wide variety of mental symptoms whose evolution is not directly related to the visual analysis made by radiologists, who can hardly quantify systematic differences. Moreover, automatic brain morphometric analyses, that do perform this quantification, contribute very little to the comprehension of the disease, i.e., many of these methods classify but they do not produce useful anatomo-functional correlations. This paper presents a new fully automatic image analysis method that reveals discriminative brain patterns associated to the presence of neurodegenerative diseases, mining systematic differences and therefore grading objectively any neurological disorder. This is accomplished by a fusion strategy that mixes together bottom-up and top-down information flows. Bottom-up information comes from a multiscale analysis of different image features, while the top-down stage includes learning and fusion strategies formulated as a max-margin multiple-kernel optimization problem. The capacity of finding discriminative anatomic patterns was evaluated using the Alzheimer's disease (AD) as the use case. The classification performance was assessed under different configurations of the proposed approach in two public brain magnetic resonance datasets (OASIS-MIRIAD) with patients diagnosed with AD, showing an improvement varying from 6.2% to 13% in the equal error rate measure, with respect to what has been reported by the feature-based morphometry strategy. In terms of the anatomical analysis, discriminant regions found by the proposed approach highly correlates to what has been reported in clinical studies of AD.**

***Index Terms*—Alzheimer's disease (AD), automated pattern recognition, computer-assisted image analysis, magnetic resonance imaging (MRI), support vector machines (SVMs).**

I. Introduction

EXISTING studies suggest neuroimaging may become a valuable tool in the early diagnosis of neurodegenerative diseases by extracting anatomical patterns and revealing hidden relations from structural magnetic resonance (MR) images. The value of neuroimaging against clinical, neuropsychological, and biochemical analysis remains to be demonstrated in large representative populations, yet there exists sufficient evidence in small series of patients with different states of neurodegenerative disorders. The usual examination workflow is performed by expert neurologists or radiologists, who are able to figure out complex anatomical patterns and subtle changes with clinical meaning. The process that an expert follows when examining a particular case involves two different kinds of tasks: those related with image perception, such as visual search or exploration paths, and others associated with cognitive skills, mainly related to diagnostic reasoning and decision making [1]. An expert structures a diagnosis by using contextual knowledge and fusing information from different sources, a process that has been recently under study [1].

At analyzing structural brain MR images, a main aim is to find anatomical changes, either local or global, related to functional disturbances. In particular, radiologists examine images by looking at distinctively regions and compare them by searching differences [1]. In the computational attempt of emulating the human vision process—a synchronized collaborative work between the brain and low level visual mechanisms—the concept of visual attention has introduced a generation of techniques that are able to transform an image into a hierarchy of relevant regions, known as salient regions. Relevant regions in radiological terms may be defined as those image areas that are visually altered and are entailed with a certain degree of clinical interpretability. Nevertheless, most methods used to compare brains establish local rather than regional (salient) differences.

Currently, a morphometric brain analysis consists of a set of strategies aimed to extract and quantify anatomical differences between groups of subjects. Commonly, this analysis comprises two main processes: first, all images are warped or registered together to a common reference frame or template, and second, a quantification of the estimated local deformation required to register is computed, producing specific measurements of interest. voxel-based morphometry (VBM) [2] and deformation-based morphometry (DBM) [3] are currently the most used techniques to compare populations. In VBM, local differences, found in brain tissue segmentations, are voxel-by-voxel statistically analyzed, while DBM statistically compares information coming from the deformations fields obtained after registration to the template. With these methods, one-to-one correspondences between subjects are assumed and statistics are computed for the same voxel across all subjects. However, conclusions are limited when the same structure may be partially present, or when a single anatomical region may exhibit multiple shapes across the population. On the other hand, some pathologies may affect not only a unique anatomical structure

Manuscript received December 18, 2013; accepted February 20, 2014. Date of publication February 27, 2014; date of current version May 29, 2014. This work was supported by project "Visual Attention Models and Sparse Representations for Morphometrical Image Analysis" (number 12108) funded by Universidad Nacional de Colombia through "Apoyo de la DIB a tesis de investigación en posgrados" and in part by projects "Anotación Automática y Recuperación por Contenido de Imágenes Radiológicas usando Semántica Latente" (number 110152128803) and "Sistema para la Recuperación de Imágenes Médicas utilizando Indexación Multimodal" (number 110152128767) by Convocatoria Colciencias 521 de 2010. *Asterisk indicates corresponding author.*

A. Rueda is with the Computer Imaging and Medical Applications Laboratory—CIM@LAB, Universidad Nacional de Colombia, Bogotá, Colombia.

F. A. González is with the Machine Learning, Perception and Discovery Laboratory—MindLab, Universidad Nacional de Colombia, Bogotá, Colombia (e-mail: fagonzalezo@unal.edu.co).

*E. Romero is with the Computer Imaging and Medical Applications Laboratory—CIM@LAB, Universidad Nacional de Colombia, Bogotá, Colombia (e-mail: edromero@unal.edu.co).

Digital Object Identifier 10.1109/TMI.2014.2308999

or even contiguous regions, but localized structures separated from each other. These kinds of patterns are difficult to find and analyze with these classic morphometric techniques. A recent proposal, the feature-based morphometry (FBM) [4], copes with these issues by modeling the image as a collage of local scale-invariant features and by learning, from them, a probabilistic model that reflects group-related anatomical characteristics. However, these approaches disregard the local statistical dependences, and then subtle changes are hardly detected; exactly the opposite strategy used by the radiologists, who analyze regions rather than pixels [5].

In recent years, there has been an increasing interest in using analytical methods to improve inferences using a small set of individuals. These methods include, among others, supervised machine learning techniques and supervised pattern recognition algorithms, which are able to automatically extract information from data. Most of them are used for classification, while they attempt to automatically discover data patterns. The most popular technique has been by far the support vector machine (SVM), which has been applied to classifying individuals with several neurological disorders. A complete review and comparison of about 40 SVM-based approaches for classification of neurological and psychiatric diseases using neuroimaging data can be found in [6], using structural MR brain images. Classifications of patients with depressive disorders, psychosis, schizophrenia, Alzheimer's disease, and Parkinson's disease, among others, have been presented in this review. Useful information for classification can be extracted from the whole brain volume or from specific regions of interest (ROIs), manually delineated or registered to a parcellated atlas. The SVM classifier is usually fed with features such as intensity [7], textural and statistical information [8], [9], binary tissue segmentations [10] or cortical thickness estimations [11]. Overall, a dimensionality reduction technique helps to decrease the computational time and the presence of irrelevant and noisy features. Recently, it has been shown that the analysis using only ROIs [12] outperforms any of the other methods. This statement points out the fact that a biased analysis, performed specifically on the known anatomical disease locations, systematically leads to stronger and more significant conclusions. However, most neurological disorders have a very variable clinical and pathological presentation, whereby subtle patterns or the atypical disease entities can be easily missed whilst, on the other hand, previous ROIs delineations are highly time-consuming and expert-dependent.

Among the neurodegenerative diseases studied using structural MR image information, the most known are the Alzheimer's disease (AD), the mild cognitive impairment (MCI), and the schizophrenia. In particular, AD is the most common type of dementia, worldwide affecting over 20 millions of people. In the clinical practice, a probable diagnosis is reached with specific neuropsychological tests, clinical examinations and particular conditions of relatives. Complementary sources of information, relevant for accurate AD diagnosis, come from different medical imaging techniques, such as structural and functional magnetic resonance imaging and positron emission tomography (PET) [13]. At present, brain atrophy, hypometabolism and quantification of specific proteins measured using these techniques have been proved to be sensitive to AD. In the particular case of structural MR images, early diagnosis of AD turns out to be a challenging task [14], [15], basically because the atrophy patterns associated with aging can be confounded with complex patterns associated with the pathology. The necessity of arriving to accurate diagnoses, given these difficulties, has increased the interest in structural neuroimaging.

Several studies, by taking structural MRI volumetric measurements of specific brain regions, have demonstrated significant differences between patients with probable AD and normal controls [11], [16]–[20]. These studies have reported several signs of the disease progression by using exclusively anatomical relationships, i.e., from MCI to AD. In particular, it has been classically described that AD early stages are characterized by premature atrophic changes along the perforant hippocampal pathway, including the entorhinal cortex (a portion of the anterior parahippocampal gyrus), the hippocampus, and the posterior cingulate cortex [21], and, in some cases, along the banks of the superior temporal sulcus [14], [22]. Later on, larger regions such as the temporal, parietal and frontal neocortices may also be compromised and exhibit neuronal loss. In terms of the disease progression, Frisoni *et al.* [21] have established the morphometric protocol when measuring the whole-brain, paying particularly attention to the entorhinal cortex, hippocampus, and temporal lobe volume estimations, as well as to the associated ventricular enlargement percentages. From the diverse automatic classification approaches that have been applied to the study of AD, the relevant anatomical regions usually correspond to those clinically described: hippocampus, amygdala, entorhinal cortex, temporal gyrus, and parahippocampal gyrus [4], [11].

This paper proposes an automatic image analysis method inspired by the radiologist visual perception. The method builds on a visual saliency model and extends it to involve a learning process that mimics the adaptation of a radiologist visual perception. The method performs a multiscale analysis of saliency maps that are optimally combined. This method is able to map any brain to a set of visual patterns that previously have been learned as associated to the pathological or normal condition. This is not about certain salient points but salient regions, whereby the whole brain structure results classified either as pathological or normal while no elastic registration among brains is required. The proposed method has been validated by accurately classifying patients from two public brain MR datasets (OASIS [23] and MIRIAD [24]) as probable AD subjects or normal controls.

A main contribution of this paper is a fusion strategy that learns, from training data, the discriminant structural patterns of neurological disorders, in particular, the Alzheimer's disease. Another important contribution is the model interpretability since the learned patterns can be mapped to the original brain and used to quantitatively estimate the importance of each region, normal or pathological, for the final classification. Technical contributions include: the use of a 3-D multiscale analysis of the brain saliency inspired by what radiologists do when examining cases, the use of low-level features that sparsify data, the formulation of a model and fusion strategies as a max-margin multiple-kernel optimization problem, and a

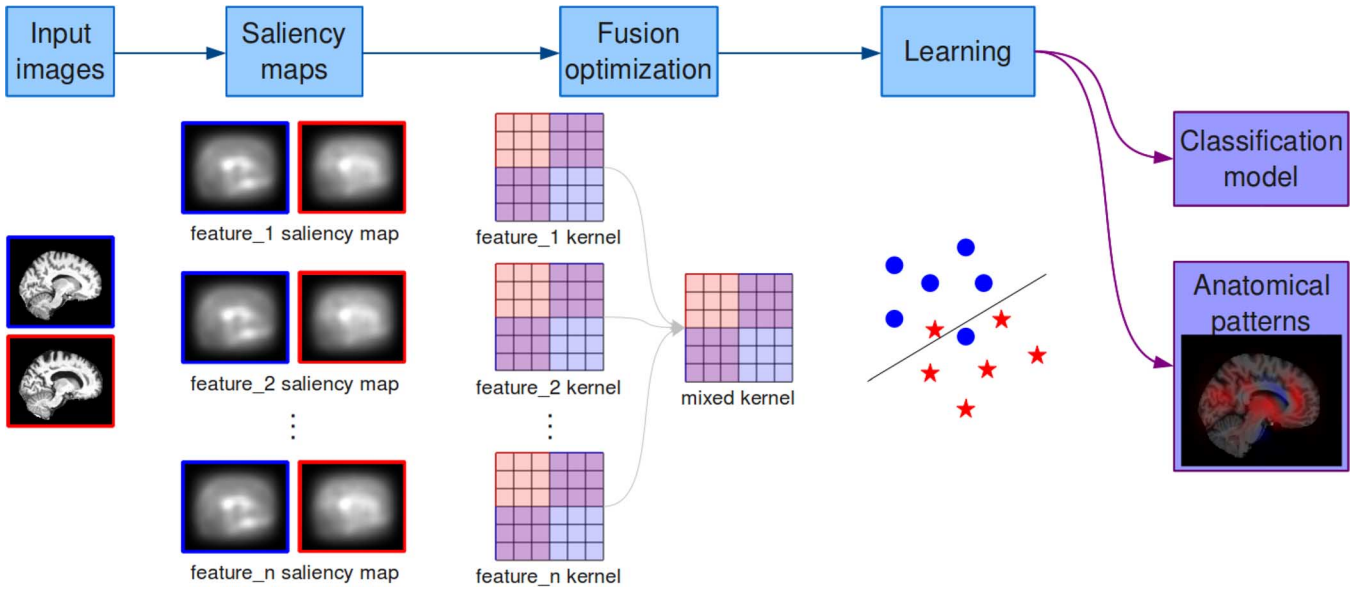

Fig. 1. Graphical overview of the proposed method. Different feature-scale saliency maps are extracted, then a learning algorithm fuses optimally this information to feed a SVM classifier, to produce both a classification model as well as maps of relevant anatomical regions.

regional analysis method that completely avoids any nonrigid preregistration step, which at the end constitutes another important variability source. An extensive parameter analysis of the influence of the image features as discriminative factors is also carried out. The classification accuracy between normal controls and probable AD subjects is improved by applying this approach, outperforming a recently proposed technique (FBM). To the best of our knowledge, this kind of visual-saliency-based pattern extraction approach has not been previously investigated for AD characterization and classification in structural MR images.

The rest of the paper is organized as follows. The description of the proposed saliency-based classification framework is presented in Section II, together with extensive experiments on two different datasets and comparisons with a state-of-the-art approach presented in Section III. Finally, the conclusions and future work are discussed in Section IV.

II. Proposed Approach

The proposed method is based on a two-phase visual saliency model that combines bottom-up and top-down approaches to achieve accurate classification of brain MR images as normal controls or probable AD subjects. The bottom-up phase performs a multiscale analysis of different basic image characteristics, similar to what was previously described [9], and fuse them to feed a discriminative model. The top-down phase uses high level knowledge, represented by the labels assigned to the training brain MR volumes, to adapt the parameters of the bottom-up saliency model using a multiple kernel learning strategy. A graphical overview of the proposed approach is presented in Fig. 1, with the respective algorithm (Algorithm 1).

Algorithm 1 Saliency-based Pattern Extraction

Require: N structural brain volumes (with $N = N_{AD} + N_{NC}$, N_{AD}: number of pathological subjects, N_{NC}: number of normal subjects)
Step 1: calculation of saliency maps (Section II-A, Algorithm 2)
Step 2: bottom-up saliency fusion (Section II-B, Algorithm 3)
Step 3: top-down learning (Section II-C, Algorithm 4)
Step 4: anatomical interpretation (Section II-D, Algorithm 5)

A. Calculation of Saliency Maps

The search of particular patterns among the anatomical areas in structural brain MR images can be considered as equivalent to figure out a preferential information flux through a net of nodes belonging to a fully-connected graph, being each node a particular anatomical region and each edge a similarity (or dissimilarity) measure between regions. If an image is partitioned and its parts are somehow related together as a fully-connected graph, the interaction process between an user and the image can be modeled as a privileged path of that graph. Some approaches have taken into account the possibility of using graphs and induced Markov chains to model saliency and attentional fixations in natural images. The graph-based visual saliency (GBVS) approach, proposed by Harel *et al.* [25], encode dissimilarities between feature pixels modulated by a closeness measure in the graph connections, i.e., salient pixels are the most dissimilar in a local context, a closer approximation to the visual analysis made by radiologist when studying a medical image and herein adapted.

Calculation of saliency information starts by extracting a set of feature maps from a given image volume x, $\{T^{\phi}(x)\}_{\phi}$, where $\phi \in \Phi$ indicates a particular combination of scale and visual feature. Commonly used features include intensity, orientation, and contrast. Projection of the original data into these different scale spaces attempt to sparsify the raw brain data, facilitating dimensionality reduction [26], a crucial factor for the success

of any classification strategy. Furthermore, the features aim to approximate the sparsity of the Human Visual System, a concept recently illustrated by showing that sparse coding of images produces Gabor-like oriented filters that resemble the receptive fields of simple cells in the visual cortex [27]. On the other hand, since it has been considered that the relevance of visual information is proportional to its coherence through different scales [28], the feature maps are also calculated at different scales by subsampling the original volume.

Subsequently, a per-slice-fully-connected graph G_A^ϕ is associated to each feature map $T^\phi(x)$. The graph vertices correspond to the image pixels while the edges stand for the regional dissimilarity between nodes. As proposed in [25], the edge weight between graph nodes $g_{i,j}^A$ and $g_{p,q}^A$ is calculated as

$$w_A\left(g_{i,j}^A, g_{p,q}^A\right) = d\left(g_{i,j}^A, g_{p,q}^A\right) \cdot F(i-p, j-q) \qquad (1)$$

where $d(g_{i,j}^A, g_{p,q}^A)$ encodes the dissimilarity (in terms of the respective feature information) and $F(i-p, j-q)$ represents the spatial closeness between nodes. Dissimilarity is calculated as

$$d\left(g_{i,j}^A, g_{p,q}^A\right) = \left|\log \frac{T^\phi(v)_{i,j}}{T^\phi(v)_{p,q}}\right|$$

where the inclusion of a logarithmic metric guarantees that larger feature dissimilarities pop out easily while similar features have little impact in the edge weight. On the other hand, the closeness is measured with

$$F(a,b) = \exp\left(-\frac{a^2+b^2}{2\sigma^2}\right) \qquad (2)$$

where σ is a free parameter of the GBVS algorithm [25]. This means that feature dissimilarity information is modulated by the spatial distance between nodes, thus encoding regional dissimilarity information at the graph edges.

Afterward, activation maps A^ϕ that point out connected regions of high dissimilarity, are estimated by constructing a Markov Chain on each G_A^ϕ and estimating its equilibrium distribution as the principal eigenvector of the stochastic matrix, using the Power Iteration Method [29]. Once activation maps are computed, a normalization step is required to guarantee that these maps concentrate the activation (saliency) only in few key locations. To do so, the same Markovian approach is applied to each activation map A^ϕ, using a new graph G_N^ϕ with image pixels as vertices, but edges now storing information about regional activation

$$w_N\left(g_{i,j}^N, g_{p,q}^N\right) = A^\phi(p,q) \cdot F(i-p, j-q)$$

so the equilibrium distribution of a new Markov chain on each G_N^ϕ highlights pixels with high activation (saliency). Finally, the feature saliency map $S^\phi(x)$ for the whole volume x is constructed by stacking the 2D-per-slice saliency maps. The complete process is summarized in Algorithm 2.

Algorithm 2 - Step 1: calculation of saliency maps

Require: N structural brain volumes
  for each structural brain volume **do**
    multi-scale decomposition
    feature extraction (per each scale)
    for each combination of feature and scale **do**
      construction of fully-connected graph, using dissimilarity and closeness given by Equation 1
      construction of Markov chain upon graph
      calculation of equilibrium distribution of the Markov chain
      distribution normalization
    end for
  end for
  return N sets of saliency maps (one set per each input volume, set size: $S = n_f \times n_s$, n_f: number of features, n_s: number of scales)

B. Bottom-Up Saliency Fusion

The model described in the previous section provides a set of saliency maps for a volume x, $\{S^\phi(x)\}_\phi$, that encompasses different scales and different types of basic visual features. Each saliency map is calculated by a function $S^\phi : I \rightarrow [0,1]^{m\times n\times l}$, where $\phi \in \Phi$ indicates a particular combination of scale and visual feature, and I indicates the volume representation space which is usually $\{0,\ldots,255\}^{m\times n\times l}$, with (m,n,l) the volume size.

Computational visual saliency models use different strategies to fuse information from saliency corresponding to different visual features. A common strategy is to weight the maps and then sum them up to calculate an overall saliency map [30]

$$S^*(x) = \sum_{\sigma,\phi} \omega_{\sigma,\phi} S_\sigma^\phi(x).$$

The problem with this strategy is that important information from the individual features may be lost when somehow linearly filtering out the maps. The proposed model uses a different strategy that keeps all the information from the different saliency maps and only fuse them when a decision needs to be made. Specifically, the saliency maps are used as input to a discriminant function, $g_W(x)$, that indicates to which extend a particular volume x corresponds to a probable AD case

$$g_W(x) = \langle W, (S^1(x), \ldots, S^{|\Phi|}(x))\rangle$$

where $W \in \mathbb{R}^{|\Phi|\times m\times n\times l}$ is the vector of parameters that indicates the relative importance of each voxel extracted from each saliency map $S^\phi(x)$. The value of $g_W(x)$ is expected to satisfy $g(x) \geq 0$ if x corresponds to an AD case and $g_W(x) < 0$ if x corresponds to a healthy subject. This model can be extended by introducing a saliency mapping function as follows:

$$g_{W,\Gamma}(x) = \langle W, (\gamma_1 \Psi(S^1(x)), \ldots, \gamma_{|\Phi|}\Psi(S^{|\Phi|}(x)))\rangle \qquad (3)$$

where $\Psi : [0,1]^{m\times n} \rightarrow F$ is a function that maps each saliency map $S^\phi(x)$ to a feature space F, $W \in \mathbb{R}^{\dim(F)\times|\Phi|}$ and the parameters $\Gamma = \{\gamma_\phi\}$ indicate the relative importance of each saliency map $S^\phi(x)$. This formulation has the advantage of improving the flexibility by allowing the model to account for potentially complex nonlinear interactions between original single saliency values corresponding to different features and scales.

An important parameter of the model in (3) is the function Ψ, which maps a saliency map to a new feature space. The main

Algorithm 3 - Step 2: bottom-up saliency fusion

Require: N sets of saliency maps, $S = n_f \times n_s$
  for each combination of feature and scale **do**
    normalize all saliency maps
    calculate kernel function, i.e. measure similarity among pairs of volumes using Equation 4
  end for
  return S feature-scale kernel matrices (one matrix per each combination of feature and scale, matrix size: $N \times N$)

aim of such representation transformation is that complex non-linear patterns in the original space become linear in the new, potentially high-dimensional, feature space. This is a well-known strategy used in kernel methods, the *kernel trick*, where the mapping Ψ is implicitly induced by a kernel function. A kernel is a function $k : X \times X \rightarrow \mathbb{R}$ associated to a mapping $\Psi : X \rightarrow F$ such that $\forall x, y \in X, k(x, y) = \langle \Psi(x), \Psi(y) \rangle_F$, i.e., k calculates the dot product in F. Intuitively, a kernel may be seem as a function that measures the similarity between two objects from the input space. In the proposed model, the input space is the space of saliency maps, so a kernel function measures the similarity between saliency maps, as detailed in Algorithm 3.

Regarding the function Ψ, in this work we used the histogram intersection kernel defined as

$$k_{hi}(s_p, s_q) = \sum_i \sum_j \sum_k \min(s_p(i,j,k), s_q(i,j,k)). \quad (4)$$

The histogram intersection kernel requires the input saliency maps to be normalized in such a way that all their values add up to 1. This is inspired by the fact that a saliency map may be seen as a saliency probability distribution over the voxels of a volume.

C. Top-Down Learning

The goal of the top-down model is to adapt the parameters of the bottom-up model in such a way that the most discriminative image features, represented in the different saliency maps, receive a higher weight value. Specifically, the top-down model uses domain knowledge codified as a set of labeled training volumes, to find optimal values for the parameters W and Γ that maximize the discriminative ability of the model. This is formulated as the following min-max-margin-discrimination optimization problem:

$$\begin{aligned} \min_{W,\Gamma} \quad & C \sum_{i=1}^{N} \max(0, 1 - y_i g_{W,\Gamma}(x_i)) + \|W\|_2^2 + \|\Gamma\|_1 \\ \text{s.t.} \quad & \Gamma \geq 0 \end{aligned} \quad (5)$$

where x_i represents a training volume, $y_i \in \{-1, 1\}$ represents the corresponding label, N is the number of training samples, and C controls the regularization of the model. The first term of the objective function in (5) is a loss function that penalizes the wrong classification of training samples, the second and the third terms are regularizers of the W and Γ parameters respectively. It is important to notice the difference between the regularizer of W (l_2 norm) and Γ (l_1 norm). The regularization of W is associated to finding a max-margin classifier as done for support vector classification [31], while the regularization of Γ aims to find the sparsest set of feature weights. In this context, sparsity is motivated by the goal of finding a reduced set of saliency maps that better codify visual patterns to discriminate probable AD cases from normal controls.

Algorithm 4 - Step 3: top-down learning

Require: S kernel matrices, N class labels (one label per each brain volume)
  solve the min-max-margin-discrimination problem in Equation 5, to obtain a feature-scale kernel weighting vector (relevance of each feature-scale kernel in the classification between the classes)
  generate the optimal kernel using a linear combination of the feature-scale kernels and the learned weights
  solve the min-max-margin-discrimination problem in Equation 5, now with the single optimal kernel, to obtain the classification model
  return trained classification model M

A saliency map kernel k may be extended to a volume kernel as follows:

$$k_\phi(x_p, x_q) = k(S^\phi(x_p), S^\phi(x_q)).$$

These kernels can be combined in a single kernel $k^* = \sum_\phi \gamma_\phi k_\phi$. It is not difficult to show that the kernel k^* is associated to the mapping

$$\begin{aligned} \Psi^* : & I \rightarrow F^{|\Phi| \times |\Sigma|} \\ & x \mapsto (\gamma_1 \Psi(S^1(x)), \ldots, \gamma_{|\Phi|} \Psi(S^{|\Phi|}(x))) \end{aligned}$$

that maps a volume to the feature space where the discriminant function in (3) is defined. In other words, the proposed model fuses the information from the different saliency maps by combining the respective kernels instead of directly adding the saliency maps themselves. This means that the optimization problem in (5) can be seen as a multiple kernel learning (MKL) problem, where in addition to finding a good discriminant hyperplane determined by the parameters W, the contribution of each kernel, determined by coefficients in Γ, must be also found. Several MKL formulations have been proposed so far, the approach herein used was proposed by [32], and describes a generalized multiple kernel learning (GMKL), suitable for learning different combinations of kernels. Learning of the classification model is presented in Algorithm 4.

D. Anatomical Interpretation

A main contribution of this work is that the method allows not only to classify structural MRI brain images but rather to highlight the anatomical areas related with a particular pattern, thereby achieving clinical interpretability. The method actually finds a quantitative estimate of the found brain differences, an important issue in terms of the clinical management of the Alzheimer's disease.

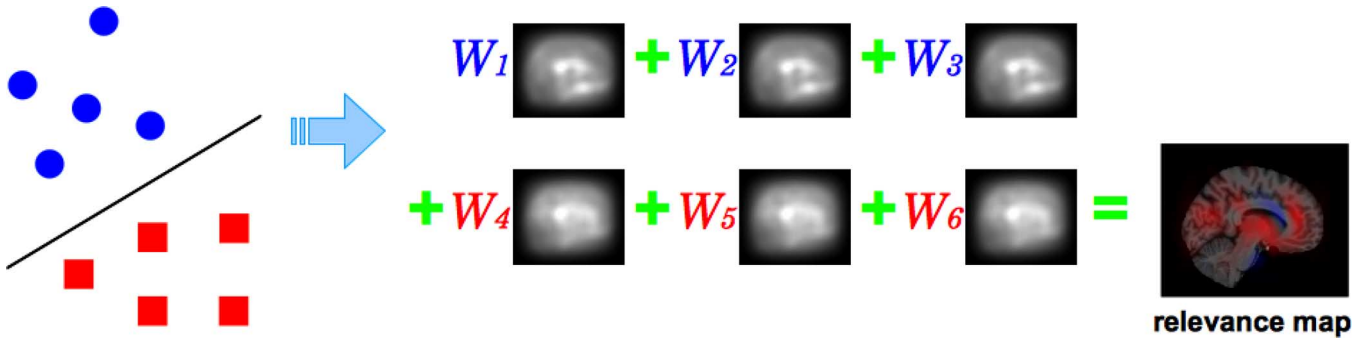

Fig. 2. Schematic description of the relevance weight map construction, using the found optimal hyperplane (5). Voxels with the highest positive or negative coefficients are colored according to the class: red for AD and blue for NC.

In the proposed approach, the pre-defined kernels convert the input image into individual feature saliency maps, whose voxels correspond to dimensions of the saliency map space. The vector Γ allows to construct a master saliency map that incorporates the most relevant feature information, by combining the individual feature saliency maps according with the learned kernel weights. Likewise, the vector W defining the separating hyperplane allows to identify the most relevant regions for AD discrimination, i.e., to select a particular set of saliency map voxels as important for discriminating AD class from the NC class, by using the coefficient value (positive or negative) associated to each master saliency map voxel. Those relevant regions can then be visualized in an overall discrimination relevance map by performing a linear combination of the master saliency maps and their corresponding coefficients, as illustrated in Fig. 2.

By aligning these relevance maps to some previously delineated brains in standard atlases, it is possible to accurately identify the specific anatomical brain areas involved in the identification of normal controls and probable AD patients. In the present investigation, 96 cortical and 21 subcortical structural areas obtained from the Harvard–Oxford atlas [33], have been used to label the Regions of Interest. For each anatomical region, the maximum value of its discrimination relevance map is stored, allowing to compare relevance values of the different anatomical regions. This scale allows to set different disease patterns at quantifying the importance of each of the anatomical areas. The whole process is presented in Algorithm 5.

Algorithm 5 - Step 4: anatomical interpretation

Require: trained classification model M, N sets of saliency maps
  for each structural brain volume (subject) **do**
    using the learned feature-scale kernel weighting vector, construct a master saliency map using a linear combination of the individual feature-scale saliency maps
  end for
  using the trained classification model, construct an overall discrimination relevance map using a linear combination of the master saliency maps and their corresponding classification coefficients
  align the discrimination relevance map with the Harvard-Oxford atlas to extract the maximum relevance values of specific anatomical regions
  return discrimination relevance map R_M (red regions associated to pathology and blue regions associated to normality), anatomical discrimination relevance map A_M

III. Experimental Results

The proposed method was assessed with respect to its capability to discriminate different experimental groups of the dataset as well as its ability to determine the important regions and to weight them in diagnostic terms.

A. Implementation Details

The saliency-based pattern extraction strategy has been implemented in MATLAB R14, running on a Linux PC with 2 Intel Quad Core i7 at 3.07 GHz and 24 GB of RAM. In all experiments, we used the GMKL source code[1] as the implementation of the Multiple Kernel Learning strategy, while the SVM classifier was implemented using the LIBSVM toolbox [34], with precomputed kernel matrices. With this configuration, the complete computational performance of the proposed approach has been measured as follows: per each volume, the calculation of the set of saliency maps takes about 15 min, the calculation of the set of kernel matrices can be completed in about 1 h/matrix, the training using the MKL approach (with the parameter selection stage) takes about 20 min, the classification stage with SVM (training and test with parameter selection) takes about 15 min, and finally, the calculation of the relevance classification maps takes about 10 min. Most of the time involved in these stages can be associated to the opening and reading of the brain volumes, process that takes a considerable time.

For calculating the saliency information from brain MR images, the selected features include intensity, orientation and edges. Intensity information corresponds to the individual gray value of each voxel in the volume; orientation information is

[1]http://research.microsoft.com/en-us/um/people/manik/code/gmkl/download.html

calculated using a bank of Gabor filters with four different orientations (0°, 45°, 90°, and 135°); and finally, edge information is extracted by applying a Sobel operator in the three different directions, with a kernel size of $5 \times 5 \times 5$. The feature maps are also calculated at different scales, in this case by subsampling the volume to 1/4, 1/8, and 1/16 of the original size. In (2), the σ value was fixed to 0.15, which can be related to the definition of a neighborhood of interest of about 1/6 of the original image size. In summary, a set of 18 different 3-D feature maps at various scales (three for intensity, 12 for orientation, and three for edges) is finally collected.

The proposed method uses a group of 18 individual kernels that comprise different intensity, orientation (each orientation angle evaluated separately) and Sobel edge information evaluated at three different image scales (1/4, 1/8, 1/16). Then, saliency maps per feature and scale were individually compared using the histogram intersection, while the model parameters (W and Γ) were learned by solving the optimization problem in (2) for a set of labeled training volumes. Cross-validation over a subset of training images (10-fold cross-validation) was used to find an optimal value for the regularization parameter C and with the optimal C, the final classification of test subjects was performed. As the baseline, the original GBVS method [25] was applied to the whole volume in a slice-per-slice basis, i.e., a single saliency map is calculated per each subject, and directly used as the input of the histogram intersection. The obtained kernel feeds a conventional SVM classifier, as previously described. Essentially, this baseline amounts to a similar characteristic but with no learning phase.

B. Brain MR Data Sets

Two different public data sets were used to evaluate the performance of the proposed method: the OASIS (Open Access Series of Imaging Studies) data set [23], and the MIRIAD (Minimal Interval Resonance Imaging in Alzheimer's Disease) data set [24]. Both datasets include brain MR volumes from healthy subjects and probable Alzheimer's disease patients.

- *OASIS (Open Access Series of Imaging Studies) database* [23]: This data set comprises 198 brain MR images from healthy (98) and pathological (100) subjects. Each subject has been previously analyzed with a Mini-Mental State Examination (MMSE) and a Clinical Dementia Rating (CDR), and diagnosed as normal controls (NC) or with probable Alzheimer's disease (AD) using the scores obtained in the MMSE and CDR tests. As described in [23], a set of 3–4 images were acquired per each subject on a 1.5T Vision scanner (Siemens, Erlangen, Germany), using a T1-weighted magnetization prepared rapid gradient-echo (MP-RAGE) sequence, on a single imaging session. Images were first spatially warped into the 1988 atlas space of Talairach and Tournoux, using a 12-parameter affine transformation as described by Buckner *et al.* [35]. Images per each subject were then averaged, obtaining a single, high-contrast MP-RAGE image in atlas space per subject, and skull-stripped by application of a loose-fitting atlas mask, as described in [23]. Finally, images were gain-field corrected, based on a fitted quadratic inhomogeneity model introduced in [36]. For the sake of a comparison with the FBM technique, which has been tested on the same dataset [4], results are reported for four different groups.
  - **Group 1 (OASIS-G1)**: 86 subjects, aged between 60 to 80 years: includes 66 healthy controls and 20 patients suffering only mild AD $(\mathrm{CDR} = 1)$.
  - **Group 2 (OASIS-G2)**: 126 subjects, aged between 60 to 96 years: includes 98 healthy controls and 28 patients suffering only mild AD $(\mathrm{CDR} = 1)$. This group results after including in the Group 1 elderly subjects (>80 years) to increase the difficulty level, given that aging patterns can be confounded with the disease patterns.
  - **Group 3 (OASIS-G3)**: 136 subjects, aged between 60 to 80 years: includes 66 healthy controls and 70 patients suffering both very mild and mild AD $(\mathrm{CDR} = 0.5, 1)$: This group results after including in Group 1, patients with very mild AD to increase the difficulty level, given that in this case the very mild AD patients could not necessarily present visual differentiating patterns.
  - **Group 4 (OASIS-G4)**: 198 subjects, aged between 60 to 96 years: includes 98 healthy controls and 100 patients suffering very mild, mild and moderate AD $(\mathrm{CDR} = 0.5, 1, 2)$. This group includes all subjects available in the OASIS data set, and can be categorized as the most difficult to classify, given that it mixes both elderly subjects and different stages of the disease.
- *MIRIAD (Minimal Interval Resonance Imaging in Alzheimer's Disease) database* [24]: Comprises 69 brain MR images from healthy (23) and pathological (46) subjects. Subjects were previously analyzed with a MMSE, and the score obtained was used to classify them as normal controls (NC) or probable Alzheimer's disease patients (AD). As described in [24], images were acquired on a 1.5T Signa MRI scanner (GE Medical systems, Milwaukee, WI, USA), using a T1-weighted IR-FSPGR (Inversion Recovery Prepared Fast Spoiled Gradient Recalled) sequence. Images were spatially warped into the 1988 atlas space of Talairach and Tournoux, using a 12-parameter affine transformation, and skull-stripped using the FLIRT and BET tools in the FSL package. It is worthy to strengthen out that this database is not categorized and different degrees of the disease are mixed up.

As pointed out by [4], analysis of the classification performance must take into account the clinical and demographic information of subjects in the dataset, given that it is more difficult to discriminate between elderly normal and pathological subjects, or between healthy subjects and patients with very mild AD. The dataset groups are proposed to illustrate the influence of these aspects.

C. Performance Evaluation

Classification of each group is performed in a leave-one-out manner, as classically described [4], for which one subject at a time is set aside during the training phase and then classified using the SVM model trained with the remaining subjects. The use of a leave-one-out strategy for classification evaluation was

TABLE I
COMPARISON OF CLASSIFICATION PERFORMANCE (USING THE EER) OF THE PROPOSED APPROACH (SINGLE KERNEL AGAINST COMBINATION OF FEATURE-SCALE KERNELS) WITH RESPECT TO THE FBM APPROACH [4]. BEST CLASSIFICATION PERFORMANCE PER EACH DATASET GROUP IS HIGHLIGHTED IN BOLD

	FBM	Our approach	
		Baseline	Kernel Group
OASIS-G1	0.80	0.80	**0.86**
OASIS-G2	0.70	0.71	**0.79**
OASIS-G3	0.71	0.71	**0.76**
OASIS-G4	0.65	0.68	**0.69**
MIRIAD	-	0.76	**0.83**

chosen for two reasons: 1) to produce the same scenario proposed in [4], where FBM is introduced and, therefore, to perform a fair comparison of the two approaches, and 2) given the reduced number of available subjects and the composition of the different groups in terms of age and disease stages, it was difficult to obtain significant training/validation/test groups.

Classification performance was validated using the following metrics.

- Accuracy $(\mathrm{Acc}) = (\mathrm{TP}+\mathrm{TN})/(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN})$.
- Sensitivity $(\mathrm{Sens}) = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$.
- Specificity $(\mathrm{Spec}) = \mathrm{TN}/(\mathrm{FP}+\mathrm{TN})$.
- Balanced Accuracy $(\mathrm{BAC}) = (\mathrm{Sens}+\mathrm{Spec})/2$.
- F-measure $(\mathrm{F1}) = (2\cdot\mathrm{TP})/(2\cdot\mathrm{TP}+\mathrm{FP}+\mathrm{FN})$.
- Equal error rate (EER): the point on a receiving operating characteristic (ROC) curve where the false positive rate and false reject rate (1—true positive rate) are equal.

where TP stands for true positives (AD individuals correctly classified), TN for true negatives (NC individuals correctly classified), FP for false positives (NC individuals misclassified), and FN for false negatives (AD individuals misclassified).

To avoid the possible inflated performance estimation on the unbalanced datasets, the balanced classification accuracy was also computed, a simple arithmetic mean of the sensitivity and specificity. The balanced accuracy (BAC) removes the bias that may arise by imbalanced datasets [37]. In a binary classification problem, if the classifier performs equally well on either class, BAC reduces to the ordinary accuracy. If, however, the classifier has taken advantage of an imbalanced dataset, then the ordinary accuracy will be inflated, whereas the BAC will drop to chance (50%), as desired. Significance of each metric was assessed using a two-tailed t-test. The F-measure [38] trades off precision $(\mathrm{TP}/\mathrm{TP}+\mathrm{FP}))$ and recall $(\mathrm{TP}/\mathrm{TP}+\mathrm{FN})$ by averaging FP and FN, that is to say $\mathrm{F}-\mathrm{measure} = \mathrm{TP}/(\mathrm{TP}+\mathrm{avg}(\mathrm{FP},\mathrm{FN})) = 2\cdot\mathrm{TP}/(2\cdot\mathrm{TP}+\mathrm{FP}+\mathrm{FN})$. This measure is insensitive to how the incorrect predictions are distributed over the classes, and has been recently used to measure the performance of multiclass classification algorithms [39].

D. Classification Results

First of all, the obtained performance was compared with a state-of-the-art method, the FBM framework [4], that reports classification results over the same OASIS groups using the EER. Table I shows the EER values obtained with the kernel group and the values reported in [4] for the same classification groups.

TABLE II
CLASSIFICATION PERFORMANCE MEASURES FOR THE 18 FEATURE-SCALE KERNEL GROUP, IN THE DATASET SUBJECT GROUPS, USING BOTH l_1 AND l_2 REGULARIZATIONS. BEST CLASSIFICATION PERFORMANCE PER GROUP IS HIGHLIGHTED IN BOLD

		Acc	Sens	Spec	BAC	F1	EER
OASIS-G1	l_1	**86.05**	**85**	**86.36**	**85.68**	**73.91**	**0.86**
	l_2	82.56	80	83.33	81.67	68.09	0.80
OASIS-G2	l_1	**80.16**	**75**	**81.63**	**78.32**	**62.29**	**0.79**
	l_2	76.19	**75**	76.53	75.77	58.33	0.76
OASIS-G3	l_1	75.74	**87.14**	63.64	75.39	**78.71**	**0.76**
	l_2	**76.47**	82.86	**69.70**	**76.28**	78.38	0.74
OASIS-G4	l_1	**70.20**	67	**73.47**	**70.23**	69.43	**0.69**
	l_2	69.20	**70**	68.37	69.18	**69.65**	0.68
MIRIAD	l_1	**84.06**	84.78	**82.61**	**83.70**	87.64	**0.83**
	l_2	**84.06**	**91.30**	69.57	80.43	**88.42**	0.78

The presented results show the performance of the kernel group (18 feature-scale combinations) and the reported classification values of the FBM framework, available for the OASIS and MIRIAD datasets. The baseline reaches a similar performance to what has been reported by [4] $(p = 0.25)$. This performance improves when introducing the learning phase. For classification of OASIS G1, the kernel group with 18 dimensions reports 0.86 and outperforms the baseline. Overall, this trend is observed through the whole experimental setup in the OASIS dataset: the performance increases in about 7.5% for G1, 13% for G2, 7% for G3, and 6.2% for G4 with respect to the baseline $(p < 0.05)$. For the MIRIAD dataset, only the performance increase can be reported: 9.2% with respect to the baseline.

Evaluation has been extended to other type of measurements, not only to the reported EER, as introduced in previous Section, but also to the full set of biomedical measures: accuracy, sensitivity, specificity, F-measure and balanced accuracy, aiming to obtain a more precise experimental description. In addition, we have tested the influence of the particular type of metric (l_1 or l_2) used by the kernel regularizer. Overall, one expects that l_1 norm would prefer sparsest solutions for the optimization problem. The classification performance obtained with the proposed method for the kernel group is presented in Table II.

The OASIS G1 and G3 groups are better separated by the l_1 norm in terms of sensitivity while the G1, G2, G4, and MIRIAD groups are better classified according to the specificity with the same norm, indicating that the discriminant capability of this kernel is easily adaptable, yet this ability decreases when groups are mixed and the classifier is not class-specific anymore. In fact, the l_1 regularizer has been recently introduced to find sparsest solutions in a search space [40]. Specificity is nevertheless better for the l_2 norm in the OASIS G3, but this is reversed for the G4, showing that when groups are very different, the anatomical and pathological variabilities end up by hiding the characteristic pattern of each group. As long as the variability increases, (G3 and G4 have older subjects) also the number of discriminative areas grow and the method is unable to find a minimum number of discriminative patterns (G3 and G4). By contrast, when the group is composed of subjects with less variability, the l_1 norm outperforms the l_2 norm and is able to highlight a reduced number of discriminative patterns. This findings strongly suggest that the disease patterns are very different for different age groups.

TABLE III
KERNEL WEIGHTS LEARNED FOR THE KERNEL GROUP (18 FEATURE-SCALE KERNELS, EACH IN ONE ROW) ON THE DIFFERENT DATASET SUBJECT GROUPS, USING BOTH l_1 AND l_2 REGULARIZATIONS

		OASIS-G1		OASIS-G2		OASIS-G3		OASIS-G4		MIRIAD	
Feature	**Scale**	l_1	l_2	l_1	l_2	l_1	l_2	l_1	l_2	l_1	l_2
Intensity	1/4	0	1.03	0	1.05	0	1.03	0	1.08	0	1.03
	1/8	0	1.04	0	1.05	0	1.03	0	1.09	0	1.03
	1/16	0	1.04	0	1.07	0	1.04	0	1.12	0	1.03
Orientation 0°	1/4	0	1.06	0	1.09	0	1.05	0	1.19	0	1.04
	1/8	0.22	1.06	0.27	1.09	0	1.05	0	1.15	1.37	1.06
	1/16	1.97	1.08	0.57	1.10	2.66	1.09	2.46	1.18	0.54	1.05
Orientation 45°	1/4	4.67	1.08	6.46	1.13	2.26	1.06	1.50	1.20	3.22	1.06
	1/8	0	1.06	0.62	1.10	0	1.06	1.65	1.19	0	1.05
	1/16	0	1.04	0	1.08	0	1.04	0	1.14	0	1.04
Orientation 90°	1/4	0	1.07	0	1.10	0	1.05	0	1.18	0.21	1.05
	1/8	0	1.06	0	1.09	0	1.05	0	1.16	0	1.05
	1/16	0	1.05	0	1.08	0	1.05	0	1.14	0	1.04
Orientation 135°	1/4	4.30	1.08	6.68	1.13	10.68	1.08	4.63	1.23	5.37	1.06
	1/8	0.60	1.06	0.20	1.09	1.37	1.07	1.17	1.18	0.22	1.05
	1/16	0	1.05	0	1.07	0	1.05	0	1.14	0	1.04
Sobel edges	1/4	0	1.04	0	1.06	0	1.04	0	1.09	0	1.03
	1/8	0	1.03	0	1.04	0	1.03	0	1.08	0	1.02
	1/16	0	1.03	0	1.03	0	1.02	0	1.07	0	1.02

This set of measurements shows in general that the multiscale analysis improves the classification rates for every group and that the learning strategy effectively chooses those characteristics with more discriminant power. This was evidenced in the classification results since best values were obtained using the l_1 norm regularizer for the whole set of experimental groups. Nevertheless, as long as the group size increases and the experimental group includes both very mild versions of the disease and elderly subjects, the difference between the results obtained with both norms decreases. This sparsity condition emerges when increasing the number of analysis sources and introducing the l_1 norm regularizer, thereby facilitating the search of the particular anatomical areas with more important differences. The sparsity of the found solutions can be verified by looking at the learned kernel weights in Table III.

This table shows that the relevant information for discriminating AD patients from NC subjects is mainly due to the 0°, 45°, and 135° orientations at the three different scales. Interestingly, intensity information, as well as Sobel edges and 90° orientations are always disregarded. Basically, this amounts to consider horizontal and diagonal changes of a bi-dimensional brain view in a multi-scale analysis, a topic on which we will come back later in this paper.

E. Anatomical Patterns and Saliency

The anatomical analysis carried out with the discrimination relevance maps, as described in Section II-D, remarkably agrees with these very known anatomical findings [21], [22]. By aligning each of the maps to the Harvard–Oxford cortical and subcortical atlases, the discrimination values per region can be better identified. As described in Section II-D, the relevance maps contain both positive and negative values, associated to the AD and NC classes, respectively. Then, using the atlases, the largest positive and negative relevance values enclosed within each anatomical region are then selected for the two different classes, producing a quantitative indicator that allows to identify those regions that enclose the most relevant patterns for discrimination of the AD and NC classes. Thus, this procedure can deliver high relevance values for both classes simultaneously, leading to find the same anatomical region as relevant for classification for both AD and NC classes.

Separation between cortical and subcortical regions is performed with these two atlas to facilitate the analysis and visualization of the relevance values. Fig. 3 presents selected sagittal slices of the corresponding discrimination relevance maps, constructed upon the classification model trained using the kernel group and the l_1 norm as regularizer, assigned to each cortical and subcortical region for all subject groups. Those regions, usually similar in case of pathological condition, are drawn in red, while those areas that remain similar for the normal brains, are colored in blue, all of them overlaid upon a structural brain MR image. Notice that some regions appear systematically as AD discriminant for the different experimental groups, namely the anterior division of the left and right parahippocampal gyrus (left and right entorhinal cortices), the left precuneous cortex, the left and right amygdalas and the left hippocampus. The discriminant level is obviously not the same but this difference can be used to characterize different levels of AD, a side result of the presented method. In contrast, normal subjects are mainly discriminated with very different regions, including the temporooccipital part of the right middle temporal gyrus, the inferior division of the right lateral occipital cortex, the right angular gyrus and the right lateral ventricle. The differences in the relevant patterns found among the different experimentation groups may be attributed to two different factors: the presence or absence of elderly subjects and the different stages of the disease.

In the first case, it is possible to identify, for example, certain correlations between the patterns of G1 and G3, those ones that do not include very elderly subjects, in particular for those regions relevant for NC discrimination. Interestingly, OASIS G1 and G2, those with only mild AD subjects, show a slight trend of reduced differences in regions such as the cingulate gyrus, the middle frontal gyrus, the lateral ventricles and the caudate nucleus, as expected since the elderly subjects should decrease the signal-to-noise ratio of the relevance values. Likewise, relevance differences between G3 and G4 are also diminished, probably because of the experimental group enlargement

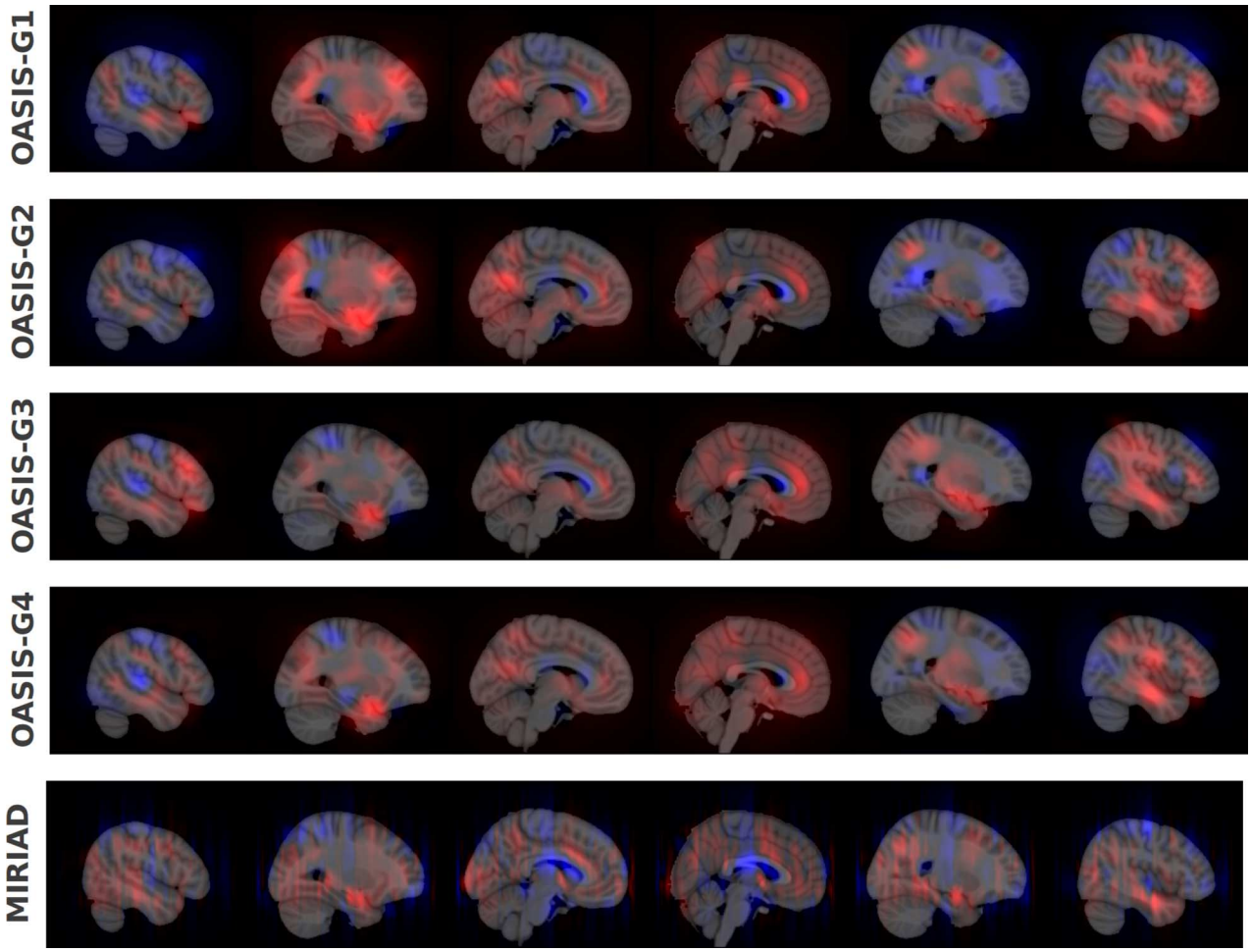

Fig. 3. Anatomical relevance maps for AD classification. Rows: database subject groups, Columns: selected sagittal slices (same slice for all groups). Blue regions are associated to NC class and red regions are associated to AD class.

with elderly subjects. Specifically, the following regions were found: the superior, middle and inferior frontal gyrus, the cingulate gyrus, the lateral ventricles, and the right amygdala, caudate and hippocampus.

In the second case, the relevance values of specific regions decay as long as different stages of the disease are included in the classification groups. For example, when comparing the AD relevant regions in OASIS G1 and G3 (those that have only subjects between 60–80 years), it can be noticed the strong relevance reduction in areas such as the anterior division of the cingulate gyrus, the left entorhinal cortex, the intracalcarine and supracalcarine cortices, the caudate nucleus and the left amygdala, putamen and hippocampus. Moreover, the same tendency is observed in OASIS G2 and G4, case in which the differences come from the fact that G2 includes exclusively the mild stage of the disease, while G4 includes all the illness phases.

Given that the volume saliency maps are generated following the acquisition plane, namely the coronal plane for subjects in the MIRIAD dataset and the sagittal plane for subjects in the OASIS dataset, one would expect that the relevance map obtained for the MIRIAD group differ in appearance with the ones of the different OASIS groups. However, it can be noticed that the relevant regions are visually very alike, as illustrated when comparing the maps obtained from the OASIS G4 and the MIRIAD group in Fig. 3, suggesting that the discriminant regions do not depend on the particular acquisition plane.

The tessellation of the relevance maps allows to quantify directly the importance of each anatomical region in the classification. Different patterns can be obtained by setting the number of regions responsible for a given percentage of the found differences, value which defines also the discrimination degree of these regions. The regions with higher relevance values are then identified per each class and compared through the different configurations tested. In this case, up to the eight most relevant cortical structures and the six most relevant subcortical structures, responsible for about the 60% of the found differences, are selected. However, in some cases the number of relevant subcortical structures are less than six, revealing that only few regions encode the largest percentage of differences. Tables IV and V present the cortical and subcortical regions identified as relevant per each dataset group.

Relevant cortical and subcortical regions, shown in Tables IV and V, are consistent with what has been described in clinical studies. In this case, the right insular cortex, the left and right hippocampus, the left and right amygdalas and the right putamen always appear as discriminant for AD, while the right precuneous cortex, the left and right lateral ventricles and the right caudate always pop out as discriminant for NC

TABLE IV
LIST OF RELEVANT CORTICAL REGIONS HIGHLIGHTED BY THE CLASSIFICATION PROCESS USING THE 18 FEATURE-SCALE KERNEL GROUP, DISCRIMINATED FOR THE DATASET SUBJECT GROUPS. RED CELLS: REGIONS RELEVANT FOR AD CLASS. BLUE ROWS: REGIONS RELEVANT FOR NC CLASS

Relevant cortical regions	Gr. 1		Gr. 2		Gr. 3		Gr. 4		MIRIAD	
	AD	NC	AD	NC	AD	NC	AD	NC	AD	NC
L Middle Frontal Gyrus										
L Inferior Frontal Gyrus, pars trian.										
L Inferior Frontal Gyrus, pars oper.										
L Temporal Pole										
L Postcentral Gyrus										
L Superior Parietal Lob.										
L Intracalcarine Cortex										
L Subcallosal Cortex										
L Cingulate Gyrus, ant. div.										
L Cingulate Gyrus, post. div.										
L Frontal Orbital Cortex										
L Parahippocampal Gyrus, ant. div.										
L Supracalcarine Cortex										
L Temporal Fusiform Cortex, post. div.										
R Frontal Pole										
R Insular Cortex										
R Superior Frontal Gyr.										
R Middle Frontal Gyrus										
R Temporal Pole										
R Superior Temporal Gyrus, post. div.										
R Inferior Temporal Gyrus, post. div.										
R Supramarginal Gyrus, post. div.										
R Intracalcarine Cortex										
R Paracingulate Gyrus										
R Cingulate Gyrus, ant. div.										
R Cingulate Gyrus, post. div.										
R Precuneous Cortex										
R Parahippocampal Gyrus, ant. div.										
R Lingual Gyrus										
R Temporal Fusiform Cortex, ant. div.										
R Temporal Fusiform Cortex, post. div.										
R Parietal Operculum Cortex										
R Planum Polare										
R Planum Temporale										
R Supracalcarine Cortex										
R Precentral Gyrus										
R Postcentral Gyrus										
R Subcallosal Cortex										

TABLE V
LIST OF RELEVANT SUBCORTICAL REGIONS HIGHLIGHTED BY THE CLASSIFICATION PROCESS USING THE 18 FEATURE-SCALE KERNEL GROUP, DISCRIMINATED FOR THE DATASET SUBJECT GROUPS. RED CELLS: REGIONS RELEVANT FOR AD CLASS. BLUE ROWS: REGIONS RELEVANT FOR NC CLASS

Relevant subcortical regions	Gr. 1		Gr. 2		Gr. 3		Gr. 4		MIRIAD	
	AD	NC	AD	NC	AD	NC	AD	NC	AD	NC
Left Lateral Ventricle										
Left Caudate										
Left Putamen										
Left Thalamus										
Brain-Stem										
Left Hippocampus										
Left Amygdala										
Left Accumbens										
Right Lateral Ventricle										
Right Caudate										
Right Putamen										
Right Thalamus										
Right Hippocampus										
Right Amygdala										

in all subject groups. Regions such as the left temporal pole, the left intracalcarine cortex and the anterior division of the left parahippocampal gyrus (left entorhinal cortex) can be closely related with presence of mild AD, given that they are mainly present for OASIS G1 and G2, which enclose only pathological subjects with mild AD ($\mathrm{CDR} = 1$). In contrast, the posterior division of the right temporal fusiform cortex and the right planum polare emerge as significant for OASIS G3, G4 and MIRIAD groups, where mixed stages of AD (very mild, mild, moderate) are present, suggesting that changes in these structures can generally indicate presence of the pathology. The right hippocampus is the only region in Table V present in OASIS G3 and G4 as relevant for both AD and NC classes. These findings indicate that frontal and temporal cortex and global subcortical structures are fundamentally altered in AD patients, presenting a complex composition of different levels of local alterations which are therefore difficult to characterize and manage.

IV. DISCUSSION

This paper has introduced a fully automatic strategy that reveals structural brain patterns associated to the presence of the Alzheimer's disease in a public dataset of brain MR images. The underlying idea behind this proposal is that it is possible to find the discriminant patterns that an expert clinician might discover in similar images. This is accomplished using a fusion strategy that mixes together bottom-up and top-down information flows, achieving accurate classifications of probable AD patients or healthy controls. The bottom-up representation is given by a visual saliency method that automatically highlights relevant regions correlated with the AD diagnosis, using contributions from different multi-scale visual features. On the other hand, the top-down scheme allows to adaptively select the meaningful part of the representation, identifying patterns associated to pathological stages. The whole strategy allows to find anatomical regions with clinical meaning that can be quantitatively related to the diagnosis, and therefore, may be suitable for an objective graduation and understanding of the different AD stages.

Morphometrical analysis of groups of subjects, for identifying discriminant patterns associated to diverse pathologies (mainly neurodegenerative diseases), is currently a wide and active research area [41]. As mentioned in the Introduction (Section I), information coming from voxel intensities (VBM) or deformation fields (DBM) is commonly used for statistical identification of anatomical between-group-differences. However, so far such analyses are only able to establish very localized differences that can not be systematically found at exactly the same place along an experimental group, whereby their clinical meaning is still limited. In addition to their anatomical inconsistency, these analyses can hardly identify complex relations between these local differences. In the search of morphometrical methods, more robust to the anatomical variability and to the lack of one-to-one correspondence between all subjects, the analysis paradigm has evolved from the very local approaches to the identification of distinctive and

reproducible patterns, represented by scale-invariant salient features, along with a probabilistic framework that together permit to evaluate the significance and differentiation degree of salient features. This is the main idea behind the FBM approach [4], which automatically discovers sets of anatomical features which are consistent with clinically established differences between normal controls and probable AD patients. These sets of features, with their appearance and geometric information, are then considered as group-related anatomical patterns, suitable to be used as image biomarkers [4]. Nevertheless, this analysis is really far from a usual diagnostic analysis, in which case anatomo-physiological correlations are required to determine the physiopathology of a particular disease. In contrast, the approach herein described can be seen as an improvement to the state-of-the-art analysis of localized salient features, much closer to the clinical interpretation of a pathological finding, and therefore more suitable to support any diagnostic decision. By combining local and global visual analyses, the presented method allows to extract anatomic relevant regions and weight their contribution to the differentiation of pathological stages.

The present investigation has included an extensive validation and parameter study, evaluating both its accuracy for discriminating different experimental groups and its capacity of determining the relevant anatomical regions together with their weights. Regarding discriminative power, different parameters involved in the top-down and bottom-up information flows, were assessed in terms of classification accuracy, allowing to identify the influence of the different visual features and image scales in the final discrimination between AD and NC classes. The simpler version of our proposal (combining a single saliency-based kernel with a SVM learning) has reached an equivalent performance to a state-of-the-art approach (FBM proposed by [4]). Comparisons between the kernel group and the baseline, have shown that the segregation of information into different feature-scale kernels, improves the classification performance in all subject groups (OASIS-MIRIAD), reaching an average increment of 8.8% in the EER measure, with respect to what was reported by the FBM technique. It should be strengthen out that the method was evaluated in two different databases and different acquisition planes, and results were completely consistent since approximately the same patterns were found in the different populations. The presented approach allows to identify in addition that the most relevant information for AD classification comes from the orientation feature, specially at 0°, 45°, and 135°, and at the three different scales. This result illustrates that the learning technique herein used is able to separately explore the parameter space and to optimally combine or fuse each part.

One of the main contributions of this work has been the design of a fully automatic method which is also completely interpretable and consistent with what radiologists diagnose. The optimal fusion of different features and their learned discriminative power, facilitates an objective understanding and localization of pathological differences. The optimal fusion of the relevant features, through the l_1 or l_2 norm regularizers, brings out at the end those anatomical areas with systematically important differences, with respect to the available feature information. When the information is more segregated and there exists a large number of features, the sparse l_1 norm regularizer selects only a small set of relevant sources, disregarding those that may provide redundant or divergent information for the classification.

The problem of overfitting is a major concern for any machine learning approach. To prevent overfitting we limited the complexity of the model by finding an optimal regularization parameter C through cross validation. In addition, using the current GBVS implementation, the size of the final saliency maps per brain volume corresponds to about 1/8 of the original slice size, a considerable dimensionality reduction. Yet the parameter size of the model in (5) seems to be quite large, it is important to highlight that the actual parameter space has a lower dimensionality thanks to the fact that the optimization problem in (5) is solved using a dual formulation, which is standard in SVM and that reduces the number of parameters to the number of samples in the training data set. On the other hand, the model has two regularization terms which control the complexity of the model by restricting the l_2 and l_1 norm of W and Γ, respectively. In the case of Γ, this promotes sparse solutions, which amounts to use a restricted number of kernels. Finally, we want to highlight that the quality of the model is not only given by the quantitative performance measures, but by its aptness to automatically detect highly discriminative brain regions, consistent with those regions that have been described as important in the progression of the disease.

In terms of the anatomical analysis, the regions found with the proposed approach as systematically relevant for discrimination of AD patients, include the cingulate gyrus, the anterior division of the parahippocampal gyrus (entorhinal cortex), and subcortical structures such as the putamen, amygdala, and hippocampus; results completely coherent to what has been reported by clinical studies of AD [21], [22]. Likewise, this analysis has determined that main discriminative features are orientations, in particular, systematic changes were mainly detected at the horizontal and diagonal directions. Yet the number of cases is not enough as to statistically conclude that this finding is disease-related, similar outcomes have been described in other global neurological disorders. It has been observed for instance that patients with schizophrenia exhibit faster volume decline in regions like the right frontal gray matter and the bilateral posterior superior temporal gray matter, that is to say global changes in anterior or posterior brain regions [42]. In this study we have found that main changes are located in horizontal and diagonal directions, an indirect evidence that changes occur very likely in oriented areas but not precisely located in a particular region. From the clinical perspective, the proposed strategy follows the visual analysis made by radiologists when diagnosing medical images, allowing in addition a quantitative determination of the brain anatomical regions which are different between experimental groups. With an adequate and exhaustive evaluation in larger data sets, containing sufficient examples of the different AD stages, this method can be also used as a second diagnostic opinion in the current clinical practice.

ACKNOWLEDGMENT

Data used in the preparation of this article were obtained from the Open Access Series of Imaging Studies (OASIS) project and the MIRIAD database (http://miriad.drc.ion.ucl.ac.uk).

The MIRIAD investigators did not participate in analysis or writing of this report. The MIRIAD dataset is made available through the support of the U.K. Alzheimer's Society (Grant RF116), while the original data collection was funded through an unrestricted educational grant from GlaxoSmithKline (Grant 6GKC).

REFERENCES

[1] J. Beutel, H. Kundel, and R. Van Metter, *Handbook of Medical Imaging*. Bellingham, WA: SPIE Press, 2000, vol. 1, Phys. Psychophys..

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

[3] J. Ashburner *et al.*, "Identifying global anatomical differences: Deformation-based morphometry," *Hum. Brain Mapp.*, vol. 6, no. 5–6, pp. 348–357, 1998.

[4] 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, Feb. 2010.

[5] H. Kundel, C. Nodine, D. Thickman, and L. Toto, "Searching for lung nodules a comparison of human performance with random and systematic scanning models," *Invest. Radiol.*, vol. 22, no. 5, pp. 417–422, May 1987.

[6] G. Orrù, W. Pettersson-Yeo, A. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review," *Neurosci. Biobehav. Rev.*, vol. 36, no. 4, pp. 1140–1152, Apr. 2012.

[7] P. Padilla, M. López, J. Górriz, J. Ramirez, D. Salas-Gonzalez, and I. Álvarez, "NMF-SVM based cad tool applied to functional brain images for the diagnosis of Alzheimer's disease," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 207–216, Feb. 2012.

[8] M. García-Sebastián, A. Savio, M. Graña, and J. Villanúa, "On the use of morphometry based features for Alzheimer's disease detection on MRI," in *Bio-Inspired Systems: Computational and Ambient Intelligence*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2009, vol. 5517, pp. 957–964.

[9] N. Doan, B. van Lew, B. Lelieveldt, M. van Buchem, J. Reiber, and J. Milles, "Deformation texture-based features for classification in Alzheimer's disease," *SPIE Med. Imag.*, 2013.

[10] M. Liu, D. Zhang, P. Yap, and D. Shen, "Hierarchical ensemble of multi-level classifiers for diagnosis of Alzheimer's disease," in *Machine Learning in Medical Imaging*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2012, vol. 7588, pp. 27–35.

[11] E. Westman *et al.*, "Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion," *NeuroImage*, vol. 62, no. 1, pp. 229–238, Aug. 2012.

[12] B. Magnin *et al.*, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," *Neuroradiology*, vol. 51, no. 2, pp. 73–83, Feb. 2009.

[13] J. Masdeu, J. Zubieta, and J. Arbizu, "Neuroimaging as a marker of the onset and progression of Alzheimer's disease," *J. Neurol. Sci.*, vol. 236, no. 1:2, pp. 55–64, 2005.

[14] B. Dubois *et al.*, "Revising the definition of Alzheimer's disease: A new lexicon," *Lancet Neurol.*, vol. 9, no. 11, pp. 1118–1127, Nov. 2010.

[15] R. Sperling *et al.*, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the national institute on aging—Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers Dement*, vol. 7, no. 3, pp. 280–292, May 2011.

[16] N. Fox and J. Schott, "Imaging cerebral atrophy: Normal ageing to Alzheimer's disease," *Lancet*, vol. 363, no. 9406, pp. 392–394, Jan. 2004.

[17] B. Ridha, J. Barnes, J. Bartlett, A. Godbolt, T. Pepple, M. Rossor, and N. Fox, "Tracking atrophy progression in familial Alzheimer's disease: A serial fMRI study," *Lancet Neurol.*, vol. 5, no. 10, pp. 828–834, Oct. 2006.

[18] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. Resnick, "Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging," *Neurobiol. Aging*, vol. 29, no. 4, pp. 514–523, 2008.

[19] C. Plant *et al.*, "Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease," *NeuroImage*, vol. 50, no. 1, pp. 162–174, Mar. 2010.

[20] A. Chincarini *et al.*, "Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease," *NeuroImage*, vol. 58, no. 2, pp. 469–480, Sep. 2011.

[21] G. Frisoni, N. Fox, C. Jack, P. Scheltens, and P. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nat. Rev. Neurol.*, vol. 6, no. 2, pp. 67–77, Feb. 2010.

[22] R. Killiany *et al.*, "Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease," *Ann. Neurol.*, vol. 47, no. 4, pp. 430–439, Apr. 2000.

[23] 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," *J. Cognitive Neurosci.*, vol. 19, no. 9, pp. 1498–1507, Sep. 2007.

[24] I. Malone, D. Cash, G. Ridgway, D. MacManus, S. Ourselin, N. Fox, and J. Schott, "MIRIAD—Public release of a multiple time point Alzheimer's MR imaging dataset," *NeuroImage*, 2012.

[25] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Adv. Neural Inf. Process. Syst.*, 2007, vol. 19, p. 545.

[26] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Jul. 2001.

[27] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.

[28] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, 1998.

[29] L. Page, S. Brin, R. Motwani, and T. Winograd, The PageRank citation ranking: Bringing order to the web 1999.

[30] S. Frintrop, G. Backer, and E. Rome, "Goal-directed search with a top-down modulated computational attention system," *Pattern Recognit.*, pp. 117–124, 2005.

[31] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[32] M. Varma and B. Babu, "More generality in efficient multiple kernel learning," in *Proc. ACM 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 1065–1072.

[33] R. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, Jul. 2006.

[34] C. Chang and C. Lin, "LIBSVM, a library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, Apr. 2011.

[35] R. Buckner, D. Head, J. Parker, A. Fotenos, D. Marcus, J. Morris, and A. Snyder, "A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume," *NeuroImage*, vol. 23, no. 2, pp. 724–738, Oct. 2004.

[36] M. Styner, C. Brechbuhler, G. Szckely, and G. Gerig, "Parametric estimate of intensity inhomogeneities applied to MRI," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 153–165, Mar. 2000.

[37] K. Brodersen, C. Ong, K. Stephan, and J. Buhmann, "The balanced accuracy and its posterior distribution," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3121–3124.

[38] C. Van Rijsbergen, *Information Retrieval*. London, U.K.: Butterworths, 1979.

[39] W. Cheng, K. Dembczyński, E. Hüllermeier, A. Jaroszewicz, and W. Waegeman, "F-measure maximization in topical classification," in *Rough Sets and Current Trends in Computing*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2012, vol. 7413, pp. 439–446.

[40] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York: Springer, 2010.

[41] J. Ashburner and S. Klöppel, "Multivariate models of inter-subject anatomical variability," *NeuroImage*, vol. 56, no. 2, pp. 422–439, 2011.

[42] D. Mathalon, E. Sullivan, K. Lim, and A. Pfefferbaum, "Progressive brain volume changes and the clinical course of schizophrenia in men, a longitudinal magnetic resonance imaging study," *Arch. Gen. Psychiat.*, vol. 58, no. 2, p. 148, Feb. 2001.