

Dissimilarity-Based Detection of Schizophrenia

Aydin Ulaş,¹ Robert P.W. Duin,² Umberto Castellani,¹ Marco Loog,²
Pasquale Mirtuono,¹ Manuele Bicego,^{1,3} Vittorio Murino,^{1,3}
Marcella Bellani,⁴ Stefania Cerruti,⁴ Michele Tansella,⁴ Paolo Brambilla^{5,6}

¹ Department of Computer Science, University of Verona, 37134, Verona, Italy

² Pattern Recognition Laboratory, Delft University of Technology, The Netherlands

³ Istituto Italiano di Tecnologia (IIT), Genova, Italy

⁴ Department of Public Health and Community Medicine, Section of Psychiatry and Clinical Psychology, Inter-University Centre for Behavioural Neurosciences, University of Verona, Verona, Italy

⁵ IRCCS “E. Medea” Scientific Institute, Udine, Italy

⁶ DISM, Inter-University Centre for Behavioural Neurosciences, University of Udine, Udine, Italy

Received 31 December 2010; accepted 17 February 2011

ABSTRACT: In this article, a novel approach to schizophrenia classification using magnetic resonance images (MRI) is proposed. The presented method is based on dissimilarity-based classification techniques applied to morphological MRIs and diffusion-weighted images (DWI). Instead of working with features directly, pairwise dissimilarities between expert delineated regions of interest (ROIs) are considered as representations based on which learning and classification can be performed. Experiments are carried out on a set of 59 patients and 55 controls and several pairwise dissimilarity measurements are analyzed. We demonstrate that significant improvements can be obtained when combining over different ROIs and different dissimilarity measures. We show that combining ROIs using the dissimilarity-based representation, we achieve higher accuracies. The dissimilarity-based representation outperforms the feature-based representation in all cases. Best results are obtained by combining the two modalities. In summary, our contribution is threefold: (i) We introduce the usage of dissimilarity-based classification to schizophrenia detection and show that dissimilarity-based classification achieves better results than normal features, (ii) We use dissimilarity combination to achieve better accuracies when carefully selected ROIs and dissimilarity measures are considered, and (iii) We show that by combining multiple modalities we can achieve even better results. © 2011 Wiley Periodicals, Inc. *Int J Imaging Syst Technol*, 21, 179–192, 2011; Published online in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/ima.20279

Key words: schizophrenia detection; dissimilarity-based classification; structural MRI; diffusion-weighted imaging

Correspondence to: Aydin Ulaş; e-mail: mehmetaydin.ulas@univr.it
Grant sponsor: FET programme within the EU FP7, under the SIMBAD project (contract 213250)

I. INTRODUCTION

Computational neuroanatomy using magnetic resonance imaging (MRI) is a growing research field that uses image analysis methods to quantify morphological characteristics of different brains (Giuliani et al., 2005). The ultimate goal is to identify structural brain abnormalities by comparing normal subjects with patients affected by a certain disease. Advanced computer vision and pattern recognition techniques may deeply help the understanding of brain characteristics and functionalities, and there are several studies where these techniques are applied (Davatzikos, 2004; Fan et al., 2007; Cheng et al., 2009a). In this sense, MRI techniques are crucial for effective automatic brain analysis, especially in the context of mental health research. Current approaches consist of identifying structural or functional brain abnormalities by comparing normal subjects (controls) with patients affected by a certain disease. A recent article by Agarwal et al. (2010) is an excellent review on the use of MR for psychiatric diseases.

In this work, we try to address one of the unsolved fundamental questions: “Is it possible to identify mental illnesses just by analyzing brain images?” Here, we focus on schizophrenia, which is a heterogeneous psychiatric disorder characterized by several symptoms such as hallucinations, delusions, cognitive, and thought disorders (Bellani et al., 2010). Although genetic and environmental factors play a role in the disorder, their etiology remains unknown, and a substantial body of research demonstrates numerous structural and functional brain abnormalities in patients (Shenton et al., 2001; Rujescu and Collier, 2009). In the context of schizophrenia research, several works have been proposed for human brain classification (Gerig et al., 2001; Fan et al., 2007; Yoon et al., 2007). Besides standard volumetric methods (Ashburner and Friston, 2000; Baiano

et al., 2008), the most promising approaches focus on: (i) shape characterization (Gerig et al., 2001), (ii) surface computation (Yoon et al., 2007), and (iii) high-dimensional pattern classification (Fan et al., 2007). Gerig et al. (2001) introduced a region of interest (ROI)-based morphometric analysis by defining spherical harmonics and a 3D skeleton as shape descriptors.

Yoon et al. (2007) proposed a support vector machine (SVM) to classify cortical thickness using the Euclidean distance between linked vertices on the inner and outer cortical surfaces. Fan et al., (2007) defined a new morphological signature by combining deformation-based morphometry with SVM. Pohl and Sabuncu (2009) used an anatomical parametrization of spatial warps to reveal structural differences between healthy and schizophrenic subjects. They use four ROIs: superior temporal gyrus (STG), amygdala, hippocampus, and parahippocampal gyrus. The parameters are estimated by minimizing a Kullback–Leibler divergence measure and are classified using SVMs.

Encouraged by our previous studies (Cheng et al., 2009a,c; Ulaş et al., 2010), we extend our earlier work and go beyond volumetric measurements by proposing an alternative approach to discriminate between subjects affected by schizophrenia and healthy controls on the basis of MRI imaging. We adopt a recent and promising framework, called dissimilarity-based paradigm (Pekalska and Duin, 2005), exploiting a particular selection of ROIs in the brain.

The dissimilarity-based paradigm pursued in this work differs from typical pattern recognition approaches where objects to be classified are represented by feature vectors. In the dissimilarity-based approach, objects are described using pairwise (dis)similarities (Pekalska and Duin, 2005). This offers the analyst a different way to express background knowledge when compared with features. In a second step, the dissimilarity representation is transformed into a vector space in which traditional statistical classifiers can be used. Unlike the related kernel approach, whose application is often constrained by technicalities such as fulfilling Mercer's condition, basically any dissimilarity measure can be used. We investigate the discriminativeness of such space by using different dissimilarity measures computed between subjects; in particular, dissimilarities are computed between histograms (or pdfs) derived from the MRI images. Using different dissimilarity measures between histograms and using the dissimilarity space, we transform the problem into a different space, where in our case shows improvements in accuracy compared to the original feature space.

Our main contribution is the application of this dissimilarity-based classification paradigm to the detection of schizophrenia in MR images and the demonstration of its accuracy using dissimilarity combination using multiple ROIs, dissimilarity measures, and modalities. An extensive experimental evaluation on a rather large dataset (114 subjects) shows the appropriateness of the paradigm to this scenario. Moreover, we see that by combining multiple modalities, one can achieve even better results than combining the dissimilarity matrices of the same modality. The rest of the article is organized as follows: In Section II, we introduce the population used in this study, and in Sections III and IV, we show how we preprocess the raw images and how the data set is formed. In Section V, we define the dissimilarity measures used in this study, and the classification methodology applied into the dissimilarity space is detailed in Section VI. Extensive experiments are illustrated in Section VII, and Section VIII concludes the work with a final discussion and future perspectives.

II. MATERIALS

Quantitative data collection and processing in MRI-based research implies facing several methodological issues to minimize biases

and distortions. The standard approach is to follow well-established guidelines, issued by international organizations, such as the World Health Organization (WHO) or codified by respected institutions, such as leading universities.

In the analysis of MR images, two are the main categories: (i) methods based on the analysis of ROIs and (ii) methods based on voxel-based morphometry (VBM) (Ashburner and Friston, 2000). ROI-based methods focus on a limited set of brain subparts, which are manually traced by experts. Methods based on VBM use the whole brain after a normalization procedure, which maps the current brain onto a standard reference, namely the stereotaxic space. In this fashion, a voxel-by-voxel correspondence is available among the analyzed subjects. In this article, we introduce our dataset that consists of ROIs of MRI images. The choice of ROIs is based on earlier investigations into their abnormal activity in case of schizophrenia (Baiano et al., 2008; Potkin et al., 2009; Corradi-Dell'Acqua et al., in press).

A. Study Population. This study involves a 59-patient subset of a larger database cared by the Research Unit on Brain Imaging and Neuropsychology (RUBIN) at the Department of Medicine and Public Health-Section of Psychiatry and Clinical Psychology of the University of Verona. The dataset is composed of MRI brain scans of 59 patients recruited from the area of South Verona (i.e., 100,000 inhabitants) through the South Verona Psychiatric Case Register (Tansella and Burti, 2003). All had received a diagnosis of schizophrenia according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV, 1994) and were being treated by the South Verona Community-based Mental Health Service (for a detailed description please refer to Andreone et al., 2007) and by other clinics reporting to the South Verona Psychiatric Care Register (Amaddeo and Tansella, 2009). Diagnoses for schizophrenia were obtained using the Item Group Checklist of the Schedule for Clinical Assessment in Neuropsychiatry (World Health Organization, 1992), administered by research clinical psychologists who had extensive experience with it. They were required to show inter-rater reliability (IRR) both blindly and independently with those of a senior investigator also trained in the procedure by achieving similar diagnosis for at least 8 of 10 assessments.

Moreover, the psychopathological item groups completed by the two raters were compared, to discuss any major symptom discrepancies. The reliability of the IGC-SCAN diagnoses was also ensured by holding regular consensus meetings with the psychiatrists treating the patients and a senior investigator. The Italian version of the SCAN was edited by the RUBIN group (World Health Organization, 1996), and our investigators attended specific training courses to learn how to administer the IGC-SCAN.

Subsequently, diagnoses for schizophrenia according to the DSM-IV criteria were corroborated by the clinical consensus of two staff psychiatrists. Patients with comorbid psychiatric disorders, alcohol or substance abuse within the 6 months preceding the study, history of traumatic head injury with loss of consciousness, epilepsy, or other neurological diseases were excluded.

At the time of imaging, only two patients were not receiving antipsychotic medication. More specifically, 25 patients were on typical antipsychotic drugs (16 on haloperidol, three on chlorpromazine, two on fluphenazine, two on clotiapine, one on thioridazine, one on zuclopenthixol) and 45 on atypical antipsychotic medication (25 on olanzapine, nine on clozapine, nine on risperidone, two on quetiapine). Patients' clinical information was retrieved from psychiatric interviews, the attending psychiatrist, and medical charts.

Table I. Demographic and clinical characteristics of the study groups

Characteristic	Group mean (and standard deviation)		Statistics		
	Control ($n = 55$)	Patient ($n = 59$)	Test	df	p
Age	40.24 (11.13) [range 23–59]	38.68 (11.78) [range 18–62]	$t = 0.72$	112	0.47
Male/female	30/25	39/20	$\chi^2 = 1.59$	1	0.21
Age at onset		26.02 (9.10)			
Illness duration		13.50 (10.23)			
BPRS		44.65 (17.49)			

df , degrees of freedom; p , significance value; BPRS, brief psychiatric rating scale. Student's t -test of the age means rejects (at a two-tailed significance level of $p < 0.05$) the hypothesis that the groups are significantly different in age, and Pearson's χ^2 confirms the same for the gender differences.

The Brief Psychiatric Rating Scale (24-item version) (Ventura et al., 2000) was used to characterize clinical symptoms. Again, it was administered by trained research clinical psychologists following the same reliability procedure as outlined above for the IGC-SCAN.

Additionally, 55 individuals without schizophrenia (control subjects) were also recruited. They had no DSM-IV axis I disorders, as determined by a modified version of the Structured Clinical Interview for DSM-IV nonpatient version. As well, they had no history of psychiatric disorders among first-degree relatives, no history of alcohol or substance abuse, and no current major medical illness. Typical control subjects were hospital/university staff volunteers or individuals undergoing imaging for dizziness whose MRI showed no evidence of central nervous system abnormalities when reviewed by the neuroradiologist. Any dizziness was due to benign paroxysmal positional vertigo or to nontoxic labyrinthitis. Participants in the control group were scanned only after a full medical history was taken and general neurological, otoscopic, and physical examinations were carried out; they had completely recovered from dizziness. Also, none of these participants was taking medication, including drugs for nausea or vertigo. For details see Tomasino et al. (2010).

This research study was approved by the Biomedical Ethics Committee of the Azienda Ospedaliera of Verona. All participants provided signed informed consent after they understood all aspects of study participation.

Table I shows relevant demographic and clinical characteristics of both groups.

B. MRI Procedure. MRI scans were acquired with a 1.5-T Magnetom Symphony Maestro Class Syngo MR 2002B (Siemens), and in total, it took about 19 min to complete an MR session. A standard head coil was used for radio frequency transmission and reception of the MR signal, and restraining foam pads were used to minimize head motion. T_1 -weighted images were first obtained to verify the participants head position and image quality (TR = 450 ms, TE = 14 ms, flip angle = 90° , FOV = 230×230 , 18 slices, slice thickness = 5 mm, and matrix size = 384×512 , NEX = 2). Proton density (PD)/ T_2 -weighted images were then acquired (TR = 2500 ms, TE = 24/121 ms, flip angle = 180° , FOV = 230×230 , 20 slices, slice thickness = 5 mm, matrix size = 410×512 , and NEX = 2) according to an axial plane running parallel to the anterior-posterior (AC-PC) commissures to exclude focal lesions. Subsequently, a coronal 3D magnetization-prepared rapid gradient echo (MP-RAGE) sequence was acquired (TR = 2060 ms, TE = 3.9 ms, flip angle = 15° , FOV = 176×235 , slice thickness = 1.25 mm, matrix size = 270×512 , and inversion time = 1100) to obtain 144 images covering the entire brain.

C. Diffusion-Weighted Imaging Procedure. MRI performing different types of diagnostic images including high-resolution volumetry and diffusion imaging might provide a new strategy for the

detection of subtle structural alterations that cannot be visualized by conventional volumetric imaging.

Diffusion-weighted imaging (DWI) investigates molecular water mobility within the local tissue environment, providing information on tissue microstructural integrity. The diffusion of water in the brain is characterized by its apparent diffusion coefficient (ADC), which represents the mean diffusivity of water along all directions (Taylor et al., 2004). Thus, ADC gives potential information about the size, orientation, and tortuosity of both intracellular and extracellular spaces, providing evidence of disruption when increased (Rovaris et al., 2002). ADC has also been used to explore regional grey matter microstructure, being higher in the case of potential neuron density alterations or volume deficit (Ray et al., 2006).

Diffusion weighted echoplanar images in the axial plane parallel to the AC-PC line (TR = 3200 ms, TE = 94 ms, FOV = 230×230 , 20 slices, slice thickness = 5 mm with 1.5-mm gap, matrix size = 128×128 , echo-train length = 5; these parameters were the same for $b = 0$, $b = 1000$, and the ADC maps). Specifically, three gradients were acquired in three orthogonal directions. ADC maps (denoted by D_{ADC}) were obtained from the diffusion images with $b = 1000$, according to the following equation:

$$-bD_{ADC} = \ln[A(b)/A(0)]$$

where $A(b)$ is the measured echo magnitude, b is the measure of diffusion weighting, and $A(0)$ is the echo magnitude without diffusion gradient applied.

In Figure 1, we can see two slices of the same subject acquired by using MRI and DWI techniques. We can see that while MRI images are more reliable, DWI resolution is very low and it is hard to segment ROIs from these DWI images. That is why we adopted a registration approach to segment DWI.

III. PREPROCESSING

This database has been investigated several times, for example, to produce large sample studies aimed at confirming previous reports of pathophysiological abnormalities associated with the given mental illnesses (Agarwal et al., 2008; Baiano et al., 2008). Each of these studies focuses on a particular subregion of the brain, a so-called (ROI), whose abnormal activity is known to affect cognitive processes. We use a subset of the data we used in our previous work (Ulaş et al., 2010) where we used a total of 124 subjects but only one sensor and one modality. In this work, we extract four modalities, which are intensity histograms of structural MRI images, ADC histograms of DWI images, and two geometric descriptors, computed from the surfaces of the ROI, which are Shape Index and Mean Curvature (details are in Section 4).

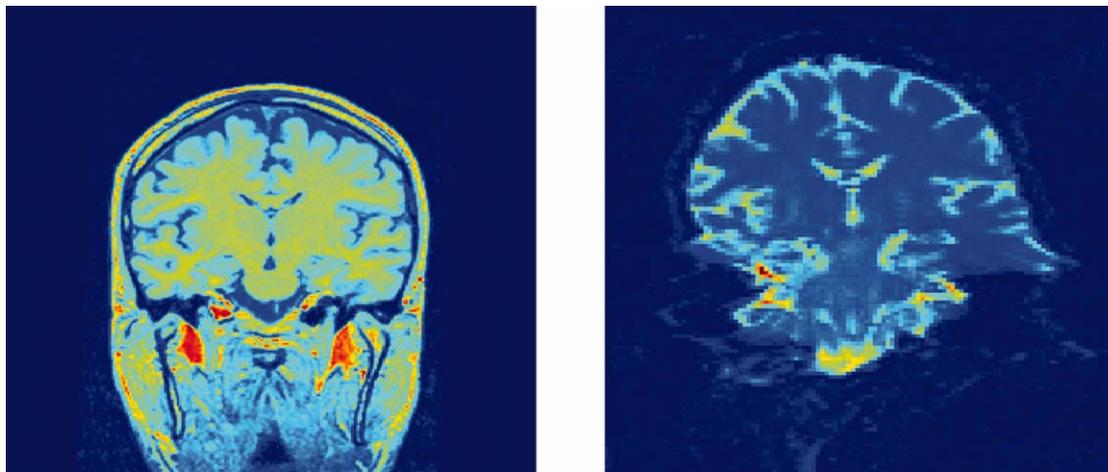


Figure 1. Two slices acquired by 3D morphological (left) and diffusion-weighted imaging (right) techniques. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

A. Segmentation of ROIs. The raw images are transferred to PC workstations to be processed for ROI tracing. Based on manual identification of landmarks, these slices are resampled and realigned by the medical personnel using the Brains2 software. The same software is used to manually trace the ROIs by manually drawing contours enclosing the intended region. This was carried out by a trained expert following a specific protocol for each ROI (Baiano et al., 2008) without knowledge of the class labels. There are methods that automatically segment the ROIs, but their accuracy is lower than the manual methods, so manual segmentation was preferred. In this sense, every ROI has a protocol based on the individuation of particular anatomical landmarks in the brain. The detailed guidelines can be found in APPENDIX: GUIDELINES FOR ROI TRACING.

The ROIs traced are seven pairs (for the left and the right hemisphere respectively) of disconnected image areas:

- Amygdala (*lamyg* and *ramyg* in short);
- Dorso-lateral preFrontal cortex (*ldlpfc* and *rdlpfc*);
- Entorhinal cortex (*lec* and *rec*);
- Heschl's gyrus (*lhg* and *rhg*);
- Hippocampus (*lhippo* and *rhippo*);
- STG (*lstg* and *rstg*);
- Thalamus (*lthal* and *rthal*).

We select these ROIs because they have consistently been found to be impaired in schizophrenia, and in a recent work, some of them have been found to support a specific altered neural network (Corradi-Dell'Acqua et al., in press). The IRR values for each brain hemisphere and ROI can be seen in Table II, which shows us the reliability of the segmentation. Higher value means that the segmentation is more reliable.

Table II. IRR values for ROI segmentation

ROI	Left	Right
<i>amyg</i>	0.91	0.98
<i>dlpfc</i>	0.93	0.98
<i>ec</i>	0.94	0.96
<i>hg</i>	0.96	0.98
<i>hippo</i>	0.96	0.96
<i>stg</i>	0.93	0.99
<i>thal</i>	0.95	0.96

In Figure 2, we show a sample from the dataset, specifically the ROI volume of the right STG for subject 11. This volume is made up of 35 slices of size 41×40 and can be arranged as a montage of images (ordered from left to right, top to bottom). Within this

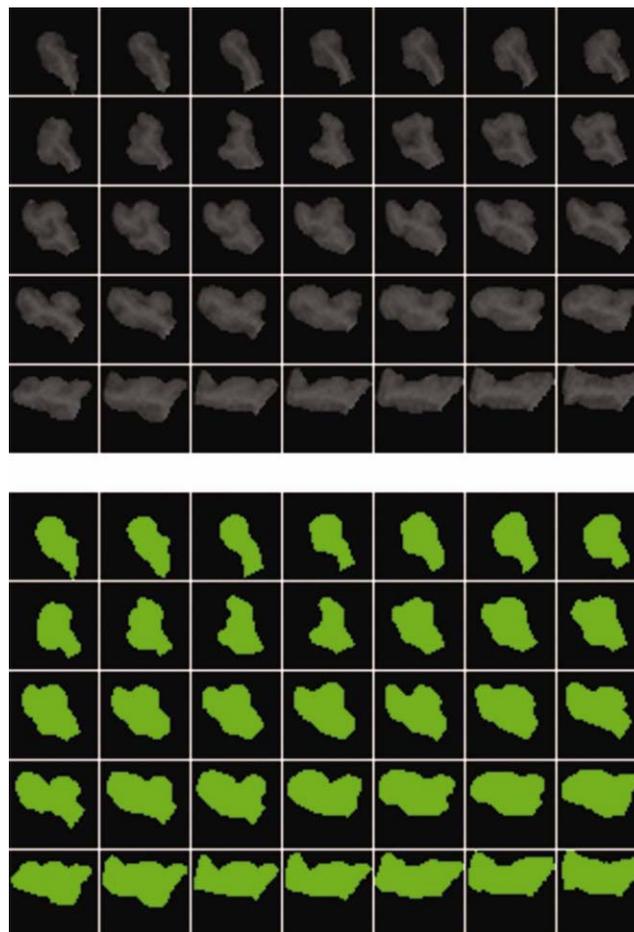


Figure 2. Montage of the slices in the ROI volume ($41 \times 40 \times 35$) of *rstg* for subject 11. At the top, the MRI values; at the bottom, the corresponding binary masks. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

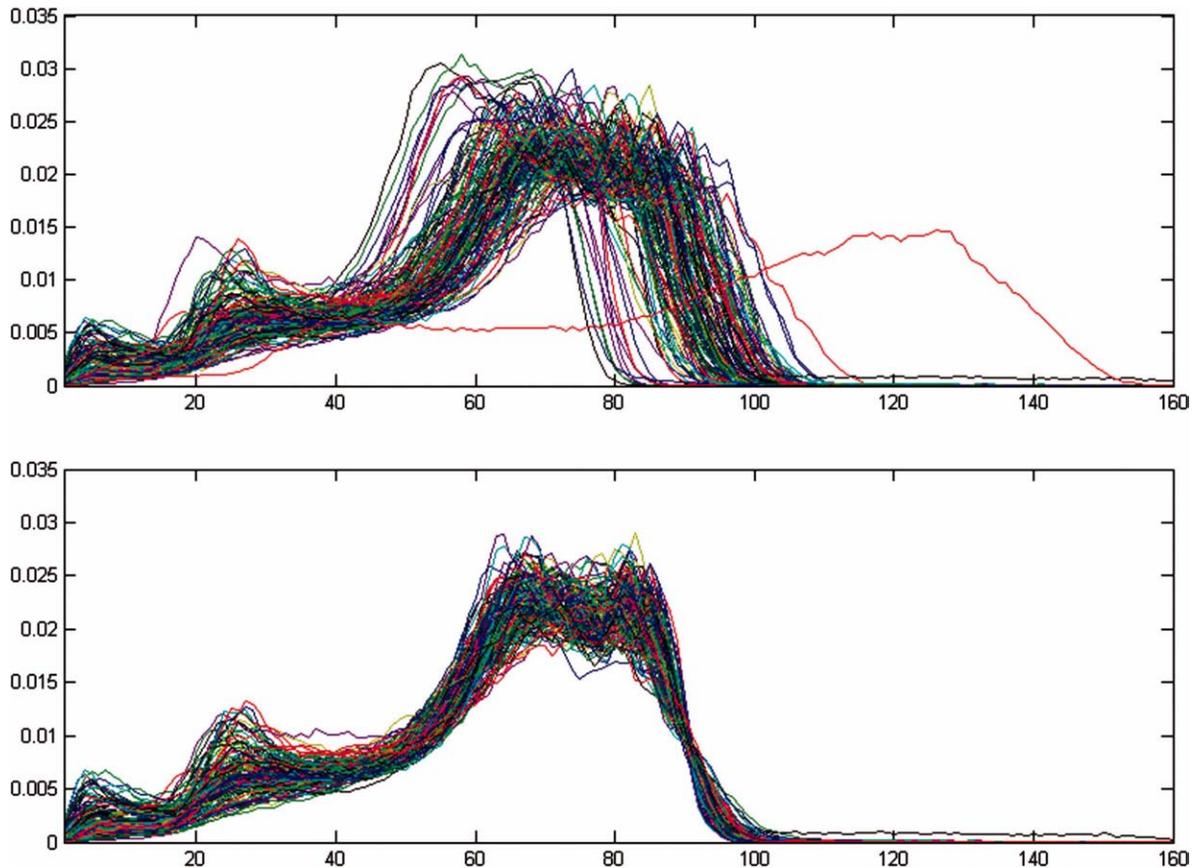


Figure 3. ICV intensity histograms (treated like probability density functions) before and after the normalization process. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

bounding box, the ROI itself is irregularly shaped, as can be clearly seen from the corresponding binary masks (shapes) at the bottom, artificially colored to highlight the ROI shape. These are the images where intensities and shape information is extracted.

Additionally, another important ROI that is traced is the intracranial volume (ICV) that is the volume occupied by the brain in the cranial cavity leaving out the brainstem and the cerebellum. This information is extremely useful for normalizing volume values against differing overall brain sizes.

IV. FEATURE EXTRACTION

From source data, we extract four different feature sets, which we explain in this section.

A. Intensity Histograms of Structural MRI Images. From the slices of ROIs, we compute histograms of normalized intensities (will be called s_{MRI} throughout the text) and use them for our feature-based experiments and calculating the dissimilarity matrices. A major disadvantage of MRI compared with other imaging techniques is the fact that its intensities are not standardized. Even MR images taken for the same patient on the same scanner with the same protocol at different times may differ in content due to a variety of machine-dependent reasons, therefore, image intensities do not have a fixed meaning (Ny'ul et al., 2000). This implies a significant effect on the accuracy and precision of the following image processing, analysis, segmentation, and registration methods relying on intensity similarity.

A successful technique used to calibrate MR signal characteristics at the time of acquisition uses phantoms (Edelstein et al., 1984), by placing physical objects with known attributes within the scanning frame. Unfortunately, this technique is not always exploited, which is our present case. Alternatively, it is possible to obtain good results by retrieving deformation mappings for the image intensities, that is, by developing histogram mappings (Jager and Hornegger, 2009; Ny'ul et al., 2000) or using an iterative non-parametric approach (Sled et al., 1998).

In this work, we retrieve the rescaling parameters to form intensity histograms from the ICV histograms (Figure 3). In this way, we focus on the interesting content of the images, which usually contain “noise” in the form of bone and muscle tissue surrounding the brain matter proper (Cheng et al., 2009b). It is also easier to identify landmarks on the histograms that match the canonical subdivision of intracranial tissue into white matter, gray matter, and cerebrospinal fluid. We opt to select a simple rescaling mapping that conserves most of the signal in the gray matter—white matter area, corresponding to the two highest bumps in the range 60–90, because ROIs primarily contain those kinds of tissue. With this technique, we form the histograms of intensity values of images using the whole ROI and use them as bases for our experiments.

B. Histograms of ADC Values. Although we do not have manually segmented ROIs for DWI images, we use deformable registration to segment DWI images into ROIs. For this purpose, every subjects' DWI image is registered into the corresponding structural

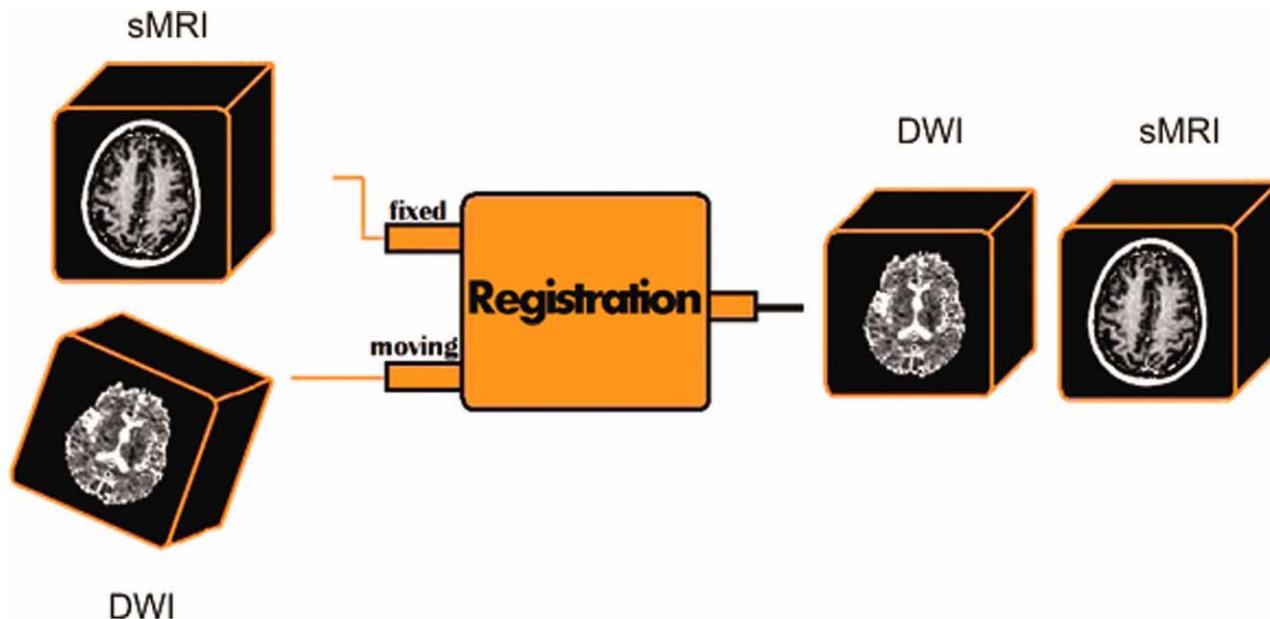


Figure 4. Registration of DWI to sMRI. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

MRI image. Then ADC values are calculated using these images. Figure 4 shows how the registration is performed.

The registration consists of matching high-resolution (also known as T_1 -w) and DWI images defined in different coordinate systems. Open source libraries of National Library of Medicine ‘‘Insight Segmentation and Registration Toolkit’’ are adopted for the coregistration procedure, while Tcl/Tk code and VTK open source libraries are chosen for the graphic interface. Digital Imaging and Communications in Medicine format (DICOM) tag parameters necessary for the coregistration are Image Origin, Image Spacing, Patient Image Orientation, and Frame of Reference.

Assuming the same anatomy topology for different studies, a Mutual Information technique based on Mattes algorithm is applied. An in-house software for multimodal registration has been developed. The program 3D Slicer™, a free open source software for visualization and image computing, is used for the graphic interface. The process has been performed in several steps.

The source DWI study (moving image) is aligned through a roto-translational matrix with the T_1 -w data (fixed image); the two studies are acquired in straight succession with the same MR unit without patient repositioning; the parameters related to algorithm implementation are automatically defined; then, by applying a multi-resolution pyramid, we are able to reach a registration within eight iterations avoiding local minimal solution.

The results of the registration are visually inspected in a check-board, where each block alternately displayed data from each study, verifying alignment of anatomical landmarks (ventricles, etc.) for confirmation.

This procedure is needed because sMRI images have better resolution, and the anatomy can better be seen for manual ROI segmentation. We use this procedure to extract ADC values for each of the ROIs instead of the whole image (Will be referred to as ADC throughout the text).

C. Geometric Shape Descriptors. From the set of 2D ROIs of the shapes (slices), the 3D surface is computed as triangle mesh using marching cubes. A minimal smoothing operation is applied to

remove noise and voxelization effect. We encode geometric properties of the surface using the Shape Index (Koenderink and van Doorn, 1992), which, for every point on the surface, is defined as:

$$sh = -\frac{2}{\pi} \arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right) \quad k_1 > k_2,$$

where k_1 and k_2 are the principal curvatures of a generic surface point. The Shape Index (denoted by sh) varies in $[-1, 1]$ and provides a local categorization of the shape into primitive forms such as spherical cap and cup, rut, ridge, trough, or saddle (Koenderink and van Doorn, 1992). Shape index is pose and scale invariant (Koenderink and van Doorn, 1992), and it has already been successfully used in biomedical domain (Awate et al., 2009). The shape index is computed at each vertex of the extracted mesh. Then, all the values are quantized, and a histogram of occurrences is computed. Such histograms represent the descriptor of a given subject, and it basically encodes the brain local geometry of a subject, disregarding the spatial relationships (will be referred as sh throughout the text).

Figure 5 shows the 3D surface of the left-Amygdala (left), the surface colored according with Shape Index values (right), and the histogram of Shape Index occurrences (bottom). It is worth noting that convex regions (in blue) are clearly distinguished from concave regions (in red) by the Shape Index values. As a further step, we also calculate the mean curvature using the same methodology (will be referred as mCUR):

$$m = \frac{k_1 + k_2}{2}$$

One can also convert histograms into discrete probability distributions and use the probability density functions (pdfs) as features. We use these (histograms which will be referred to as *hist* and their corresponding pdfs which will be referred to as *pdf*) as features for our feature-based classification experiments. From these four modalities, several dissimilarity measures are defined between pairs

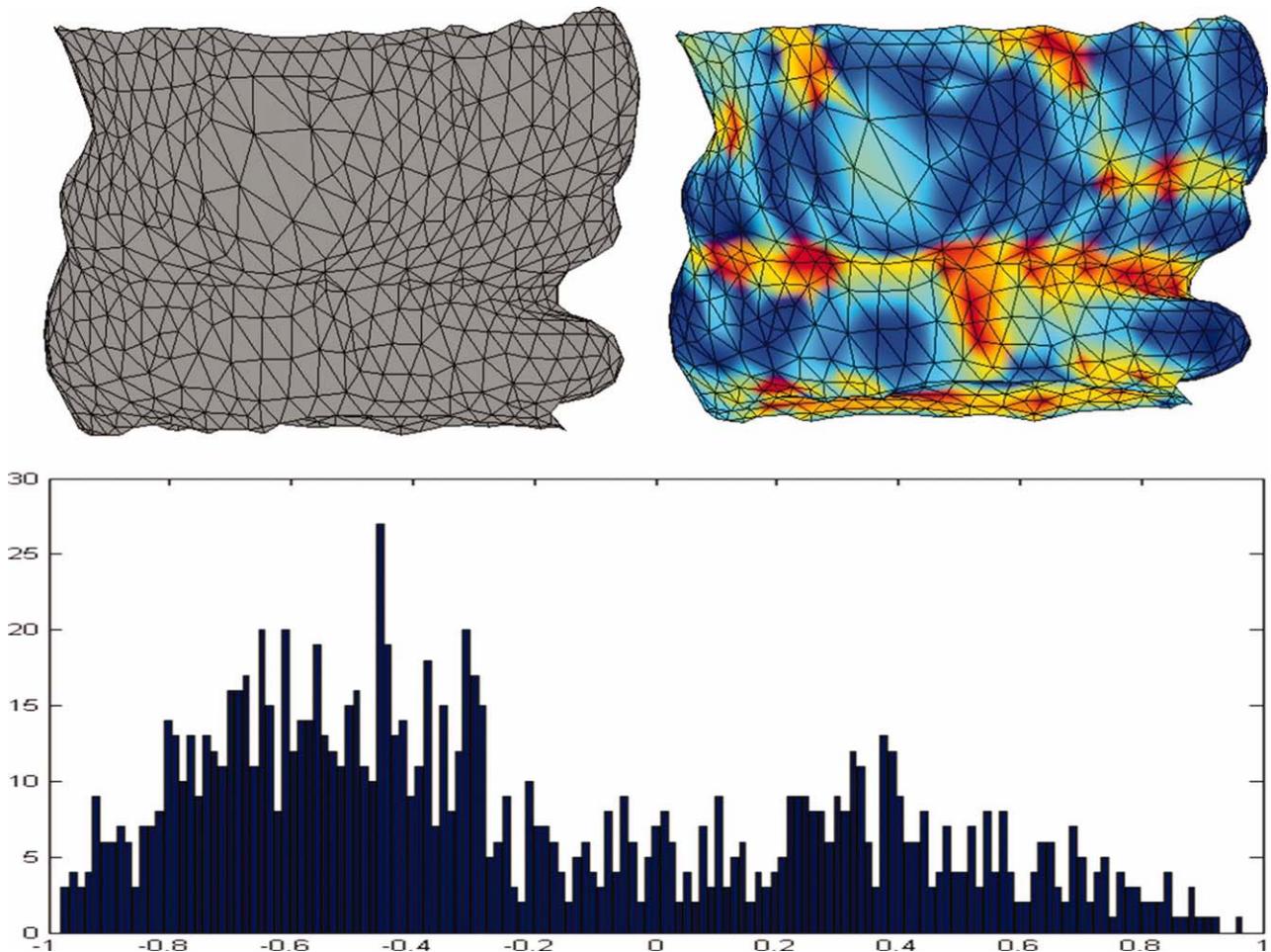


Figure 5. Geometric feature extraction: 3D surface of the left-amygdala (left), the surface colored according with shape index values (right), and the histogram of shape index occurrences (bottom). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

of histograms (and their pdfs) all of which are described in the next section. In the spirit of dissimilarity-based classification, there are various other ways to calculate dissimilarities between brain images based on registration of brains (Klein et al., 2010). In this study, we opted not to choose this direction because of the computational complexity and time required to pairwise register brain images and leave this as feature work to calculate dissimilarities between images without resorting to histograms.

V. DISSIMILARITY MEASURES

There are several ways to calculate dissimilarities between two feature sets or two objects, depending on the origin of the objects. One can use dissimilarities based on graph representations, Euclidean distance and other dissimilarities between two vectors. There are also some problems where the objects are already represented as dissimilarities. In this study, our feature sets consist of histograms of different measures (i.e. intensities and ADC values), so we use different dissimilarity measures specifically designed for computing the dissimilarity between histograms (and pdfs). There are various dissimilarity measures that can be applied for this purpose (Cha and Srihari, 2002; Serratosa and Sanfeliu, 2006). All measures have their own assumptions, which we believe covers different relations

between the subjects at hand. The relation between these measures and DSM-IV criteria can be a new line of research, which we wish to exploit in the future.* Moreover, histograms are converted to pdfs, and dissimilarity measures between two discrete distributions are used as well. The measures we investigate in this article are the most commonly used and accepted measures in the literature, which are described in the following.

Given two histograms S and M with n bins, we define the number of elements in S and M as $|S|$ and $|M|$, respectively.

A. Histogram Intersection. It measures the number of intersecting values in each bin (Swain and Ballard, 1991):

$$\text{sim}(S, M) = \frac{\sum_{i=1}^n \min(S_i, M_i)}{\min(|S|, |M|)}.$$

As this is a similarity measure, we convert it to a dissimilarity using

$$D = \min(|M|, |S|) \times (1 - \text{sim}(S, M)).$$

*We thank the anonymous reviewer for this suggestion.

B. Diffusion Distance. In diffusion distance (Ling and Okada, 2006),[†] the distance between two histograms is defined as a temperature field $T(x, t)$ with $T(x, 0) = S(x) - M(x)$. Using the heat diffusion equation

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2}$$

which has a unique solution

$$T(x, t) = T(x, 0) \times \phi(x, t)$$

where

$$\phi(x, t) = \frac{1}{(2\phi)^{1/2} t} \exp -\frac{x^2}{2t^2},$$

we can compute D as:

$$D = \int_0^T k(|T(x, t)|) dt$$

C. χ^2 Distance. This metric is based on the χ^2 test for testing the similarity between histograms. It is defined as

$$D = \sum_{i=1}^n \frac{(S_i - M_i)^2}{S_i + M_i}.$$

It is a standard measure for histograms.

D. Earth Mover's Distance. This distance was originally proposed by Rubner et al. (2000). It is basically defined as the cost to transform one distribution into another. It is calculated using linear optimization by defining the problem as a transportation problem. For 1D histograms, it reduces to a simple calculation (Cha and Srihari, 2002), which is implemented in this work.

$$C_i = \left| \sum_{j=1}^i (S_j - M_j) \right| D = \sum_{i=1}^n C_i.$$

Similarly, we consider the following dissimilarities between pdfs:

E. Bhattacharyya. It is used to measure the similarity of discrete probability distributions p and q . It is defined as:

$$D(p, q) = -\log BC(p, q),$$

where

$$BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)}.$$

F. Kullback–Leibler (KL) Divergence. Kullback–Leibler divergence is defined as

$$D(p, q) = \sum_{i=1}^n q_i \log \frac{q_i}{p_i}.$$

This measure is not a distance metric but a relative entropy because $D(p, q) \neq D(q, p)$, that is, the dissimilarity matrix is not symmetric. There are various ways to symmetrize this matrix. We simply used $D = D(p, q) + D(q, p)$ and the so-called Jensen–Shannon divergence:

$$D = \frac{1}{2} D(p, r) + \frac{1}{2} D(q, r),$$

where r is the average of p and q .

In summary, we used the following 13 measures:

- *hist-euclid*: Euclidean distance between histograms.
- *hist-l1*: $L1$ distance between histograms.
- *hist-intersect*: Intersection between histograms.
- *hist-diffusion*: Diffusion distance between histograms.
- *hist-chi*: χ^2 distance between histograms.
- *hist-emd*: Earth mover's distance between histograms.
- *pdf-euclid*: Euclidean distance between pdfs.
- *pdf-l1*: $L1$ distance between pdfs.
- *pdf-emd*: Earth mover's distance between pdfs.
- *pdf-bs*: Bhattacharyya distance between pdfs.
- *pdf-kl*: Symmetrized KL divergence between pdfs.
- *pdf-kl-orig*: Original, asymmetric KL divergence.
- *pdf-js*: Jensen–Shannon divergence between pdfs.

Summarizing, there are 14 ROIs and 13 different dissimilarity measures per modality, which yield a total of 182×4 dissimilarity matrices. In addition to these, we propose to merge the different dissimilarity matrices into one overall dissimilarity matrix per modality potentially exploiting complementary information useful to improve the classification accuracy. We also test the accuracy of these combinations against combining classifiers in the original feature space (histograms and pdfs for each of the four modalities). Further details of this combination are provided below.

VI. DISSIMILARITY SPACE

Suppose that we have n objects and we have a dissimilarity matrix D of size $n \times n$, which represents the dissimilarities between all pairs of objects. And suppose that the dissimilarity between two objects o and \hat{o} is denoted by $D(o, \hat{o})$. There are several ways to transform an $n \times n$ dissimilarity matrix D with elements $D(o, \hat{o})$ into a vector space with objects represented by vectors $X = \{x_1, \dots, x_o, \dots, x_{\hat{o}}, \dots, x_n\}$ (Pekalska and Duin, 2005). Classical scaling (for proper Euclidean dissimilarities) and pseudo-Euclidean embedding (for arbitrary symmetric dissimilarities) yield vector spaces in which vector dissimilarities can be defined that produce the given dissimilarities D . As almost all dissimilarity measures studied here are non-Euclidean, classification procedures for these pseudo-Euclidean spaces are ill-defined, as for instance the corresponding kernels are indefinite.

A more general solution is to work directly in the dissimilarity space. It postulates a Euclidean vector space using the given dissimilarities to a representation set as features. As opposed to the previously mentioned techniques, it is not true anymore that dissimilarities in this space are identical to the given dissimilarities, but this is an advantage in case it is doubtful whether they really represent dissimilarities between the physical objects. As this holds in our case, we construct such a dissimilarity space using all available objects by taking X equal to D . In the dissimilarity space basically any traditional classifier can be used. The number of dimensions, however, equals

[†]The code has been taken from the author's home page: http://www.ist.temple.edu/~hbling/code__data.htm.

the number of objects, which is 114 in our case. Many classifiers will need dimension reduction techniques or regularization to work properly in this space. Here, we used the linear SVM to avoid this.

It is a well-known fact that there is no single most accurate classification algorithm, so methods have been proposed to combine classifiers based on different assumptions (Kuncheva, 2004; Alpaydm, 2004). Every algorithm makes a different assumption about the data and makes errors on different instances; therefore, it seems reasonable that by combination; the overall error can be decreased. Keeping this in mind, we use the same idea to combine dissimilarity measures to achieve higher accuracy.

Combined dissimilarity spaces can be constructed by combining dissimilarity representations. As in normal classifier combination (Kittler et al., 1998), a simple and often effective way is using an (weighted) average of the given dissimilarity measures:

$$D_{combined} = \sum_i \alpha_i D_i \quad (1)$$

It is related to the sum-rule in the area of combining classifiers. The weights can be optimized for some overall performance criterion, or determined from the properties of the dissimilarity matrix D_i itself, for example, its maximum or average dissimilarity. In this work, we used equal weights while combining multiple dissimilarity matrices, because it is computationally fast and we do not need to keep part of the data to train and compute the weights. All the dissimilarity matrices are scaled such that the average dissimilarity is one, that is:

$$\frac{D(o, \hat{o})}{\frac{1}{n(n-1)} \sum_{o, \hat{o}} D(o, \hat{o})} = 1$$

This is done to assure that the results are comparable over the dissimilarities as we deal with dissimilarity data in various ranges and scales. Such scaled dissimilarities are denoted as \tilde{D} . In addition, we assume here that the dissimilarities are symmetric. So, every dissimilarity $\tilde{D}(o, \hat{o})$ is transformed by

$$\tilde{D}(o, \hat{o}) := \frac{\tilde{D}(o, \hat{o}) + \tilde{D}(\hat{o}, o)}{2}$$

VII. EXPERIMENTS

We considered 182 dissimilarity matrices using four different modalities. For assessing the performance of each technique, we report the leave-one-out accuracy using a two-class classification task to discriminate between healthy and schizophrenic subjects. For every subject x_i , we train the model (classifier) using the remaining set of subjects, and we use this model to check if x_i is classified correctly. The accuracy is then the number of correctly classified subjects divided by the number of all subjects. For comparing the performances of the provided techniques, we used one-sided paired t -test. All accuracy differences reported in this article are significant at $p = 0.05$. The dissimilarity spaces have been built in a transductive way by using all available subjects for dissimilarity (of course labels are ignored in this phase). Three classifiers are considered to compare the performances:

- Linear SVM classifier on the original feature space (called *svm*)
- The 1-nearest neighbor (NN) rule on the dissimilarity matrices (called *1nn*)
- Linear SVM classifier on the dissimilarity space (called *sv0*)

The linear SVM in dissimilarity space avoids complications that could arise from the dissimilarity measures being non-Euclidean

because we treat the dissimilarities as features in this new space. While combining dissimilarities, we use for α_i in [Eq. (1)] the reciprocal of the number of dissimilarity matrices to be combined (Lee et al., 2010). On the original feature space, the SVM classifiers produce posterior probability outputs, and these outputs are combined using the SUM rule (Kittler et al., 1998). So, on the original feature space, we combine after training the classifiers, whereas on the dissimilarity space, we combine before we do classification. The experiments are carried out using the Matlab package PRTTools (Duin, 2005). We designed three experiments to show the improvements of dissimilarity-based pattern recognition techniques and combination of dissimilarities using multiple ROIs and modalities:

1. ROI-based classification: for each modality, we report the highest accuracy that a classifier reaches without combination (on the original feature space and on the dissimilarity space). We use these results as baseline for comparison.
2. Multi-ROI classification: in these set of experiments, for each modality, we fix the dissimilarity measure and combine all ROIs using this dissimilarity measure.
3. Multimodal classification: in this experiment, we go one further step to combine information coming from different sources by combining different dissimilarity matrices from different modalities.

We note that, through this section, we will use the following notation: every dissimilarity representation will be referred to as modality-ROI-dissimilarity-measure (i.e., *SMRI-ldlpfc-pdf-js* shows the dissimilarity matrix for the structural MRI of ROI *ldlpfc* using the dissimilarity measure of Jensen–Shannon divergence). The modality, ROI, or the dissimilarity measure will be omitted when it is clear from the context.

A. ROI-Based Classification. We evaluate the classification accuracies for each of the original dissimilarity matrices. Table III summarizes the results for structural MRI. For each ROI, the best performance is reported with respect to various dissimilarity measures. First column reports the accuracy estimates for *svm* using the original feature space (histograms and pdfs). Second column reports the maximum accuracy of *1nn* on different dissimilarity measures. Third column reports the leave-one-out accuracy estimates of the linear SVM in dissimilarity space. It shows clearly the improvements of our dissimilarity-based approach. Except for two ROIs (*rhg* and *rthal*) SVM classifier in the dissimilarity space is always better than classifiers in the standard space. Although the best accuracy of standard approaches is 68.42%, we can reach 78.07% accuracy on dissimilarity space, and dissimilarity space accuracies are more stable.

Table IV shows the same results for the ADC values extracted from DWI images. We can again note that when we switch to dissimilarity-based classification, we get better accuracies (either *1nn* or *sv0*) except for two ROIs (*lamyg* and *rec*). We can again see that with a single ROI and dissimilarity measure, we can reach 70.18%, whereas the highest accuracy we can obtain in the original space is 64.04%. The same trend can be observed when we look at Table V and VI. Also in these modalities, the best accuracy can be achieved using dissimilarities. We can see that on *SH*, we reach 68.42% using *1nn* and 65.79% using *sv0*. The best accuracy using the features on the original space is 55.26%. Also on *MCUR*, best accuracy is reached using *sv0*.

Table III. Best single ROI accuracies for each ROI on histograms of intensities

ROI	<i>svm</i>	<i>1nn</i>	<i>sv0</i>
<i>lamyg</i>	68.42 (pdf)	64.04 (hist-11)	78.07 (pdf-bc)
<i>ramyg</i>	54.39 (hist)	65.79 (pdf-11)	66.67 (hist-chi)
<i>ldlpfc</i>	60.53 (hist)	62.28 (pdf-kl-orig)	76.32 (pdf-js)
<i>rdlpfc</i>	64.04 (hist)	57.89 (hist-intersect)	68.42 (pdf-kl-orig)
<i>lec</i>	64.04 (pdf)	56.14 (pdf-emd)	64.91 (hist-11)
<i>rec</i>	64.91 (pdf)	65.79 (pdf-11)	71.05 (hist-intersect)
<i>lhg</i>	51.75 (pdf)	60.53 (pdf-11)	63.16 (pdf-bc)
<i>rhg</i>	50.00 (hist)	<i>63.16 (hist-11)</i>	59.65 (hist-emd)
<i>hippo</i>	63.16 (pdf)	60.53 (hist-intersect)	72.81 (pdf-kl-orig)
<i>rhippo</i>	60.53 (pdf)	64.04 (pdf-emd)	66.67 (pdf-bc)
<i>lstg</i>	55.26 (pdf)	59.65 (hist-intersect)	69.30 (hist-chi)
<i>rstg</i>	63.16 (hist)	57.02 (pdf-emd)	64.91 (pdf-kl)
<i>lthal</i>	58.77 (pdf)	64.91 (pdf-11)	67.54 (hist-11)
<i>rthal</i>	<i>64.91 (pdf)</i>	59.65 (pdf-12)	64.04 (pdf-emd)

B. Multi-ROI Classification. In this subsection, we will show our experiments where we combine multiple ROIs, fixing the modality, and the dissimilarity measure. We also conducted experiments by fixing the ROIs and combining multiple dissimilarity matrices using the same ROI. We see that the accuracy does not increase when compared with combining ROIs with fixed dissimilarity measure. This conforms to our previous studies, therefore in this article, we do not report combination of dissimilarity measures with fixed ROI.

In this experiment, a multi-ROI approach is adopted to use all ROIs at the same time. All the dissimilarity matrices for each ROI are combined by averaging the normalized dissimilarity matrices. Second and third columns of Table VII report the results on intensity histograms, using 1-NN rule on the dissimilarity matrices and the support vector classifiers in the dissimilarity spaces. Also in this case, the classification on the dissimilarity space clearly outperforms the standard approach. Moreover, the multi-ROI approach brings an improvement by confirming the complementary information enclosed onto the different brain subparts when we use *sv0* on the dissimilarity space. In most of the cases, the results from the averaged similarity matrices are better than the respective best single-ROI results. The row average in Table VII reports the error estimates computed on the overall dissimilarity matrix (combining all the measures and ROIs), which has the highest accuracy 76.32% (same as combining all ROIs for *pdf-11*) for both standard approach

Table IV. Best single ROI accuracies for each ROI on histograms of ADC values

ROI	<i>svm</i>	<i>1nn</i>	<i>sv0</i>
<i>lamyg</i>	<i>64.04 (hist)</i>	57.89 (hist-emd)	62.28 (hist-emd)
<i>ramyg</i>	54.39 (pdf)	56.14 (pdf-12)	59.65 (pdf-bc)
<i>ldlpfc</i>	56.14 (pdf)	51.75 (pdf-12)	61.40 (pdf-kl)
<i>rdlpfc</i>	54.39 (hist)	56.14 (hist-emd)	65.79 (pdf-emd)
<i>lec</i>	53.51 (pdf)	62.28 (pdf-emd)	61.40 (hist-emd)
<i>rec</i>	<i>64.04 (pdf)</i>	58.77 (pdf-12)	60.53 (pdf-kl-orig)
<i>lhg</i>	55.26 (hist)	61.40 (pdf-11)	54.39 (pdf-12)
<i>rhg</i>	50.88 (hist)	58.77 (hist-emd)	58.77 (hist-emd)
<i>hippo</i>	52.63 (hist)	57.02 (hist-emd)	59.65 (pdf-11)
<i>rhippo</i>	48.25 (hist)	55.26 (hist-emd)	52.63 (pdf-kl)
<i>lstg</i>	54.39 (pdf)	56.14 (hist-12)	60.53 (hist-intersect)
<i>rstg</i>	64.04 (hist)	60.53 (hist-emd)	70.18 (pdf-bc)
<i>lthal</i>	57.89 (hist)	57.89 (pdf-11)	58.77 (pdf-kl-orig)
<i>rthal</i>	53.51 (pdf)	60.53 (pdf-bc)	59.65 (pdf-js)

Table V. Best single ROI accuracies for each ROI on shape index histograms

ROI	<i>svm</i>	<i>1nn</i>	<i>sv0</i>
<i>lamyg</i>	45.61 (hist)	68.42 (hist-emd)	64.91 (hist-emd)
<i>ramyg</i>	49.12 (hist)	53.51 (hist-12)	57.89 (pdf-kl-orig)
<i>ldlpfc</i>	52.63 (hist)	62.28 (hist-12)	57.89 (hist-chi)
<i>rdlpfc</i>	54.39 (hist)	59.65 (hist-intersect)	60.53 (hist-chi)
<i>lec</i>	46.49 (hist)	51.75 (pdf-emd)	54.39 (pdf-emd)
<i>rec</i>	52.63 (hist)	60.53 (hist-emd)	57.02 (hist-chi)
<i>lhg</i>	47.37 (hist)	54.39 (pdf-js)	65.79 (hist-intersect)
<i>rhg</i>	43.86 (hist)	55.26 (pdf-emd)	55.26 (hist-intersect)
<i>hippo</i>	55.26 (hist)	50.00 (hist-diffusion)	57.02 (hist-chi)
<i>rhippo</i>	47.37 (hist)	59.65 (pdf-emd)	57.02 (pdf-kl-orig)
<i>lstg</i>	40.35 (hist)	58.77 (pdf-js)	52.63 (pdf-emd)
<i>rstg</i>	53.51 (hist)	55.26 (pdf-bc)	57.89 (pdf-emd)
<i>lthal</i>	46.49 (hist)	53.51 (pdf-12)	57.89 (hist-12)
<i>rthal</i>	54.39 (hist)	57.89 (hist-emd)	59.65 (hist-11)

and dissimilarity-based approach, respectively. The last row reports the accuracy of combining all SVM classifiers in the original feature space. When we combine all the SVM classifiers on the original space, we get 71.93% accuracy. This shows us that, the dissimilarity space produces better results also when we consider classifier combination. We repeated the same experiments also with the other modalities. In Tables VIII–X, we see the results using the other modalities. We observe that again we get the most accurate results when we combine ROIs in the dissimilarity space using *sv0* except mean curvature histograms where the best results are obtained using *1nn* (using dissimilarities again).

C. Combining Different Modalities. As a further step to understand how information from multiple sources can be combined to reach better classification accuracy, we develop another experiment where we combine information from multiple modalities. We have 182 dissimilarity matrices from each of the four modalities. It is not possible to exhaustively search the whole solution space to find the best solution (optimum subset for combination), so instead, we choose the most accurate four ROI-dissimilarity pairs from each modality and do an exhaustive search on the combination of these matrices to get the best result. We can see the selected dissimilarity matrices and their base accuracies in Table XI. With a total of 16 dissimilarity matrices (modality-ROI-dissimilarity triples), we can get the best accuracy 86.84% (last row in Table XI), which contains

Table VI. Best single ROI accuracies for each ROI on mean curvature histograms

ROI	<i>svm</i>	<i>1nn</i>	<i>sv0</i>
<i>lamyg</i>	43.86 (pdf)	61.40 (hist-intersect)	57.02 (hist-11)
<i>ramyg</i>	46.49 (hist)	58.77 (hist-chi)	57.89 (pdf-kl-orig)
<i>ldlpfc</i>	49.12 (pdf)	53.51 (hist-chi)	53.51 (hist-emd)
<i>rdlpfc</i>	56.14 (pdf)	57.02 (hist-11)	60.53 (hist-chi)
<i>lec</i>	47.37 (pdf)	53.51 (pdf-kl)	53.51 (pdf-emd)
<i>rec</i>	52.63 (hist)	65.79 (hist-emd)	63.16 (pdf-12)
<i>lhg</i>	61.40 (hist)	56.14 (pdf-bc)	62.28 (pdf-kl-orig)
<i>rhg</i>	57.89 (pdf)	64.04 (pdf-emd)	61.40 (pdf-js)
<i>hippo</i>	53.51 (hist)	58.77 (hist-intersect)	62.28 (pdf-emd)
<i>rhippo</i>	54.39 (pdf)	55.26 (pdf-emd)	58.77 (pdf-js)
<i>lstg</i>	50.00 (hist)	50.00 (pdf-bc)	51.75 (pdf-11)
<i>rstg</i>	41.23 (pdf)	63.16 (hist-emd)	57.89 (hist-intersect)
<i>lthal</i>	53.51 (pdf)	56.14 (pdf-emd)	58.77 (pdf-kl-orig)
<i>rthal</i>	48.25 (pdf)	54.39 (hist-chi)	67.54 (pdf-bc)

Table VII. Best accuracies for each dissimilarity measure combining all ROIs on histograms of intensities

Measure	<i>Inn</i>	<i>sv0</i>
<i>hist-l2</i>	62.28	71.05
<i>hist-l1</i>	62.28	74.56
<i>hist-intersect</i>	66.67	68.42
<i>hist-diffusion</i>	62.28	74.56
<i>hist-chi</i>	57.02	71.05
<i>hist-emd</i>	52.63	58.77
<i>pdf-l2</i>	57.02	74.56
<i>pdf-l1</i>	60.53	76.32
<i>pdf-emd</i>	59.65	75.44
<i>pdf-bc</i>	65.79	69.30
<i>pdf-kl</i>	66.67	70.18
<i>pdf-kl-orig</i>	64.04	64.91
<i>pdf-js</i>	65.79	71.93
<i>average</i>	60.53	76.32
<i>svm</i>		71.93

two dissimilarity matrices from intensities (*ldlpfc-pdf-kl-orig* and *ldlpfc-pdf-bc* both having 75.44% accuracy) and one dissimilarity matrix from shape index (*rdlpfc-hist-chi* with 60.53% accuracy). This accuracy is the best accuracy, which has been reached using dissimilarity combination and cannot be reached using only one modality. Applying the same methodology, we can reach only 76.32% accuracy with *Inn* and 83.33% accuracy with *svm* on the original feature space. This also shows us why it is important to combine useful information from different sources to come up with better accuracy. We see that the accuracy can be increased when complementary information using different modalities are combined.

In a medical application, besides increasing accuracy, the interpretability of the results is also important. We use this experiment to deduce information on the use of ROIs, their complementary information, and how each modality relates to the detection of schizophrenia. For this purpose, we select all the combinations of distance matrices with accuracies above 82% (we have 69 different combinations) and count the occurrences of dissimilarity matrices for every combination. From Table XI, we can see that most of the combinations include *ldlpfc* of SMRI and the shape index of *rthal*. This shows us that these two modality-ROI pairs contribute and complement other dissimilarity matrices and by using these two in combination, we increase accuracy. After these two dissimilarity

Table VIII. Best accuracies for each dissimilarity measure combining all ROIs on histograms of ADC values

Measure	<i>Inn</i>	<i>sv0</i>
<i>hist-l2</i>	50.00	60.53
<i>hist-l1</i>	46.49	64.91
<i>hist-intersect</i>	43.86	61.40
<i>hist-diffusion</i>	46.49	64.91
<i>hist-chi</i>	50.88	55.26
<i>hist-emd</i>	58.77	51.75
<i>pdf-l2</i>	57.02	60.53
<i>pdf-l1</i>	54.39	61.40
<i>pdf-emd</i>	57.89	53.51
<i>pdf-bc</i>	53.51	53.51
<i>pdf-kl</i>	55.26	48.25
<i>pdf-kl-orig</i>	49.12	51.75
<i>pdf-js</i>	52.63	54.39
<i>average</i>	51.75	60.53
<i>svm</i>		63.16

Table IX. Best accuracies for each dissimilarity measure combining all ROIs on shape index histograms

Measure	<i>Inn</i>	<i>sv0</i>
<i>hist-l2</i>	57.89	57.89
<i>hist-l1</i>	58.77	60.53
<i>hist-intersect</i>	40.35	53.51
<i>hist-diffusion</i>	58.77	60.53
<i>hist-chi</i>	59.65	57.02
<i>hist-emd</i>	55.26	56.14
<i>pdf-l2</i>	50.88	55.26
<i>pdf-l1</i>	50.88	58.77
<i>pdf-emd</i>	60.53	60.53
<i>pdf-bc</i>	48.25	57.89
<i>pdf-kl</i>	52.63	59.65
<i>pdf-kl-orig</i>	57.02	59.65
<i>pdf-js</i>	48.25	60.53
<i>average</i>	54.39	60.53
<i>svm</i>		51.75

matrices, we see that mean curvature of *rthal* and shape index of *rdlpfc* are used in combination the most. These are followed by *ldlpfc* of histogram intensities and the mean curvature of *rec*. With ADC, we see that most used ROI is *rstg*, which has been selected 38 times. This also shows us that the DWI information is the least complementary modality in this scenario and one can design experiments without this modality, focusing on the other modalities. We can use this information to decrease the costs of the operation, that is, not performing DWI analysis. Also we see that the most accurate dissimilarity matrix (*SMRI-lamyg-pdf-bc*) is the eighth most used dissimilarity when we consider combination. This interesting fact shows us that when doing combination, the complementary information is more important than individual accuracies.

Another interesting fact is that some ROIs are more discriminative when the structural information is considered, and some are more discriminative when we consider DWI. The ROIs selected from the structural analysis in this experiment are those, considered crucial for the impaired neural network in schizophrenia and comply with current studies in the literature (Corradi-Dell'Acqua et al., in press), in contrast DWI is particularly keen in exploring the microstructural organization of white matter therefore providing intriguing information on brain connectivity (Brambilla and Tansella, 2007) but does not have complementary contribution in this context.

Table X. Best accuracies for each dissimilarity measure combining all ROIs on mean curvature histograms

Measure	<i>Inn</i>	<i>sv0</i>
<i>hist-l2</i>	49.12	50.88
<i>hist-l1</i>	50.00	51.75
<i>hist-intersect</i>	53.51	50.88
<i>hist-diffusion</i>	50.00	51.75
<i>hist-chi</i>	55.26	48.25
<i>hist-emd</i>	43.86	53.51
<i>pdf-l2</i>	57.02	51.75
<i>pdf-l1</i>	54.39	46.49
<i>pdf-emd</i>	50.00	52.63
<i>pdf-bc</i>	44.74	52.63
<i>pdf-kl</i>	48.25	49.12
<i>pdf-kl-orig</i>	55.26	46.49
<i>pdf-js</i>	53.51	48.25
<i>average</i>	54.39	49.12
<i>svm</i>		47.37

Table XI. Most accurate four dissimilarity matrices from each modality, their single performances, and number of occurrences in the combination of most accurate results

Selected dissimilarity	Accuracy	Occurrences
SMRI-ldlpfc-pdf-js	76.32	60
SH-rthal-hist-l1	59.65	57
MCUR-rthal-pdf-bc	67.54	52
SH-rdlpfc-hist-chi*	60.53	50
SMRI-ldlpfc-pdf-bc*	75.44	48
SMRI-ldlpfc-pdf-kl-orig*	75.44	48
MCUR-rec-pdf-l1	63.16	47
SMRI-lamyg-pdf-bc	78.07	42
ADC-rstg-hist-l2	65.79	38
SH-lamyg-hist-emd	64.91	38
ADC-rstg-pdf-bc	70.18	20
MCUR-rec-pdf-l2	63.16	17
ADC-rdlpfc-pdf-emd	65.79	14
ADC-rstg-pdf-js	65.79	9
SH-lhg-hist-intersect	65.79	8
MCUR-lhippo-pdf-emd	62.28	1
Dissimilarities with * are in the optimum combination	86.84	

With this analysis, we can open a new perspective of how to use each of these modalities to get better accuracies. One can use this information to setup new experiments considering the contributions of these ROIs on these modalities.

D. Discussion on Computational Complexity. In a multi-classification scenario, the computational complexity of the system increases with the number of features, modalities, combinations, and so forth. In this work, main increase for the computational complexity is acquisition and extraction of the features. Concatenation of features is not used, so the classifiers are trained using the original features without increasing complexity. The combination is achieved by averaging the outputs of classifiers after each one is trained. The same is also valid on the dissimilarity space. This time, the distance matrices are combined by averaging before training the classifier. In both cases, the combination is achieved through averaging which is a simple operation. Of course, as the number of modalities increase, number of feature sets to be trained also increase, but considering the modern-day computers and the inherent parallelizability of training these classifiers, this does not increase the complexity that much. In a different scenario, where one has to optimize the weights of combination [α_i in Eq. (1)], one has to leave part of the data to optimize the combiner weights, which is not our case. So, the complexity in the system arises from acquisition and extraction of features from each modality. This is indeed a problem especially when the extraction of a feature requires acquisition of data from a costly sensor, which we also tried to tackle using our exhaustive analysis. One may discard modalities, which contribute less to the combination accuracy by saving from cost and complexity.

VIII. DISCUSSION AND CONCLUSIONS

In this article, a novel approach based on dissimilarity-based pattern recognition is proposed for the detection of schizophrenic brains. Several dissimilarity measures are proposed to deal with histograms of different types for different ROIs. ROI-based classification onto the dissimilarity space shows improvements of the standard NN rule and the support vector classifier on the original space. More-

over, a Multi-ROI classification strategy is obtained by simply averaging the similarity matrices observed in each ROI. Such approach improves the single-ROI one, by highlighting the complementary information enclosed in the several ROIs. This confirms the benefit of combining dissimilarity information and fusing information from various regions in the brain.

We investigate further to combine information from multiple modalities such as intensities, ADC values, and geometric information. We can see that, some ROIs are discriminative when we use intensities; some are useful when DWI data is considered. Geometric properties of some ROIs play a part in schizophrenia detection. We show that we get the best accuracy when we combine multiple modalities.

We can interpret the results of combining multiple modalities to set up further experiments in this context. Our results show that the least contributing modality is the DWI. With this information, one can skip using this modality and focus more on histograms of intensities and geometric information. Also, one can use this result to reduce the costs of this operation, by not performing DWI measurements and without the patient undergoing further medical operations.

We would like to emphasize that in building the (combined) dissimilarities no parameters are optimized w.r.t. performance. The proposed approach of combining dissimilarities on the dissimilarity space opens new perspectives in neuroanatomy classification by allowing the possibility to exploit dissimilarity measures where one does not have to deal with technical difficulties such as the metric requirements of distance-based classification and kernel restrictions of SVMs.

As part of our future work, we aim to develop novel dissimilarity measures (independent or dependent on ROIs), which should reflect the underlying brain structure in a better way and, in turn, further improve the detection of schizophrenia from MR. One possible way to do this is to pairwise register subjects and calculate the distances using registration-based dissimilarity measures. On the more methodological side, we would like to study the use of weighted combinations of dissimilarity matrices to quantify the importance of each ROI-dissimilarity. Introducing a weighting, however, means an increase in the number of parameters to estimate in the classification scheme and as such poses a problem, considering the limited number of subjects a medical application like this typically involves. The careful design of a proper regularizer, however, should make this approach feasible.

APPENDIX: GUIDELINES FOR ROI TRACING

Hippocampus. The first slice to be traced was the one where the superior colliculus completely connected with the thalamus bilaterally. Moving anteriorly through the brain, we traced around the hippocampi using first the corona radiata, and then the ambient cistern as the superior border. The white matter acted as the inferior border and the inferior horn of the lateral ventricle as the lateral one. The anterior limit was one slice posterior to the slice where the mammillary body became visible. On average, 16 slices were traced.

Amygdala. The first slice to be traced was the one where the mammillary body becomes visible. The superior and lateral borders were defined by the temporal lobe white matter and the inferior one by the parahippocampal gyrus white matter. Moving forward, the anterior limit, either right or left, was defined by the point when the

amygdale became too diffuse to be resolved from the temporal lobe gray matter. In average, seven slices were traced.

Entorhinal Cortex. The entorhinal cortex was traced on MRI coronal slices. The frontotemporal stem delimited the region of interest anteriorly. The intersection of the line along the gray-white junction with the medial bank of the collateral sulcus defined the inferolateral border. The superomedial border was defined rostrally by the sulcus semiannularis and caudally by the uncus cleft. The intersection of the line along the gray-white junction with the cortical surface was used to improve the definition of these structures. The most anterior slice in which the body of the hippocampus first became clearly visible was chosen as the posterior limit. It should be noted that the prior methods were slightly modified because the lateral geniculate body was poorly detectable in most of our scan, and therefore, it was not used as a posterior limit. Also, the perirhinal cortex was included in our tracing.

Dorsolateral Prefrontal Cortex. The DLPFC was defined as slices anterior to the posterior border of the genu till the anterior border of the the Sylvian horizontal ramus; the superior border was the superior frontal sulcus, the inferior border was the upper border of the Sylvian fissure posteriorly and the horizontal ramus of the Sylvian fissure anteriorly, the lateral boundary was the edge of the brain, and the medial boundary was the line connecting the most medial point of the superior frontal sulcus and the Sylvian fissure/horizontal ramus.

Thalamus. The tracing of the thalamus was performed on the T_1 -weighted MP-RAGE sequence, beginning at the coronal slice where the anterior pillars of the fornix merge into the mammillary bodies and continuing to the slice in which it was no longer possible to distinguish the thalamus from the surrounding brain matter. The lateral ventricles at the superior border, the red nucleus and the substantia nigra at the inferior border, the posterior limb of the internal capsule at the lateral border separating the thalamus from the adjacent lentiform nucleus, and the third ventricle at the medial border demarcated the limits of the thalamus. The presence of the adhesio interthalamica was also detected.

Superior Temporal Gyrus. STG was traced bilaterally in the coronal plane. The anterior border was defined by the first slice where the temporal stem appeared. Posteriorly, it was traced to the end of the Sylvian fissure. The superior border was the Sylvian fissure, and the inferior one was the superior temporal sulcus.

Heschl's Gyrus. The HG was anatomically identified as an omega or heart-shaped protrusion in the supratemporal plane. It is defined medially by the first transverse sulcus of temporal lobe and laterally by Heschl's sulcus. If there are two complete Heschl's sulci defining two gyri, then the anterior gyrus was used.

REFERENCES

N. Agarwal, J. D. Port, M. Bazzocchi, and P. F. Renshaw, Update on the use of MR for assessment and diagnosis of psychiatric diseases, *Radiology* 255 (2010), 23–41.

N. Agarwal, G. Rambaldelli, C. Perlini, N. Dusi, O. Kitis, M. Bellani, R. Cerini, M. Isola, A. Versace, M. Balestrieri, A. Gasparini, R. P. Mucelli, M. Tansella, and P. Brambilla, Microstructural thalamic changes in schizophrenia: A combined anatomic and diffusion weighted magnetic resonance imaging study, *J Psychiatry Neurosci* 33 (2008), 440–448.

E. Alpaydm, Introduction to Machine Learning, The MIT Press, London, England, 2004.

F. Amaddeo and M. Tansella, Information systems for mental health, *Epidemiol Psichiatr Soc* 18 (2009), 1–4.

DSM-IV, Diagnostic and statistical manual of mental disorders, 4th Edition, American Psychiatric Association, Washington DC, 1994.

N. Andreone, M. Tansella, R. Cerini, A. Versace, G. Rambaldelli, C. Perlini, N. Dusi, L. Pelizza, M. Balestrieri, C. Barbui, M. Nose, A. Gasparini, and P. Brambilla, Cortical white-matter microstructure in schizophrenia. Diffusion imaging study, *Br J Psychiatry* 191 (2007), 113–119.

J. Ashburner and K. J. Friston, Voxel-based morphometry—The methods, *Neuroimage* 11 (2000), 805–821.

S. P. Awate, P. Yushkevich, Z. Song, D. Licht, and J. C. Gee, Multivariate high-dimensional cortical folding analysis, combining complexity and shape, in neonates with congenital heart disease, Proceedings of the 21st International Conference on Information Processing in Medical Imaging, IPMI '09, 2009, pp. 552–563.

M. Baiano, C. Perlini, G. Rambaldelli, R. Cerini, N. Dusi, M. Bellani, G. Spezzapria, A. Versace, M. Balestrieri, R. P. Mucelli, M. Tansella, and P. Brambilla, Decreased entorhinal cortex volumes in schizophrenia, *Schizophrenia Res* 102 (2008), 171–180.

M. Bellani, N. Dusi, P. Brambilla. Longitudinal imaging studies in schizophrenia: the relationship between brain morphology and outcome measures, *Epidemiol e Psichiatr Soc* 19 (2010), 207–210.

P. Brambilla and M. Tansella, Can neuroimaging studies help us in understanding the biological causes of schizophrenia? *Int Rev Psychiatry* 19 (2007), 313–314.

S.-H. Cha and S. N. Srihari, On measuring the distance between histograms, *Pattern Recognit* 35 (2002), 1355–1370.

D. S. Cheng, M. Bicego, U. Castellani, S. Cerruti, M. Bellani, G. Rambaldelli, M. Atzori, P. Brambilla, and V. Murino, Schizophrenia classification using regions of interest in brain mri, *Proc Intell Data Anal Biomed Pharmacol, IDAMAP '09, 2009a*, pp. 47–52.

D. S. Cheng, M. Bicego, U. Castellani, S. Cerruti, M. Bellani, G. Rambaldelli, M. Atzori, P. Brambilla, and V. Murino, Schizophrenia classification using regions of interest in brain MRI. Technical Report, Dipartimento di Informatica, University of Verona, Italy, 2009b.

D. S. Cheng, M. Bicego, U. Castellani, M. Cristani, S. Cerruti, M. Bellani, G. Rambaldelli, M. Atzori, P. Brambilla, and V. Murino, A hybrid generative/discriminative method for classification of regions of interest in schizophrenia brain MRI, Proceedings of workshop on Probabilistic Models for Medical Image Analysis, MICCAI '09, 2009c, pp. 174–184.

C. Corradi-Dell'Acqua, L. Tomelleri, M. Bellani, G. Rambaldelli, R. Cerini, R. Pozzi-Mucelli, M. Balestrieri, M. Tansella, and P. Brambilla, Thalamic-insular dysconnectivity in schizophrenia: Evidence from structural equation modeling, *Hum Brain Mapp* (2011), in press.

C. Davatzikos, Why voxel-based morphometric analysis should be used with great caution when characterizing group differences, *NeuroImage* 23 (2004), 17–20.

R. P. W. Duin, Prtools, a matlab toolbox for pattern recognition version 4.0.14, Available at <http://www.prtools.org/>.

W. A. Edelstein, P. A. Bottomley, and L. M. Pfeifer, A signal-to-noise calibration procedure for NMR imaging systems, *Med Phys* 11 (1984), 180–185.

Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, COMPARE: classification of morphological patterns using adaptive regional elements, *IEEE Trans Med Imaging* 26 (2007), 93–105.

G. Gerig, M. Styner, M. E. Shenton, and J. A. Lieberman, Shape versus size: Improved understanding of the morphology of brain structures. *Proc Int Conf Med Image Comput, MICCAI '01, 2001*, pp. 24–32.

N. R. Giuliani, V. D. Calhoun, G. D. Pearlson, A. Francis, and R. W. Buchanan, Voxel-based morphometry versus region of interest: A comparison

- of two methods for analyzing gray matter differences in schizophrenia, *Schizophrenia Res* 74 (2005), 135–147.
- F. Jager and J. Hornegger, Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging, *IEEE Trans Med Imaging* 28 (2009), 137–150.
- J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, On combining classifiers, *IEEE Trans Pattern Anal Mach Intell* 20 (1998), 226–239.
- S. Klein, M. Loog, F. van der Lijn, T. den Heijer, A. Hammers, M. de Bruijne, A. van der Lugt, R. P. Duin, M. M. B. Breteler, and W. J. Niessen, Early diagnosis of dementia based on intersubject whole-brain dissimilarities. Proceedings of the 2010 IEEE international conference on Biomedical imaging: from nano to Macro, ISBI'10, 2010, pp. 249–252.
- J. J. Koenderink and A. J. van Doorn, Surface shape and curvature scales, *Image Vis Comput* 10 (1992), 557–565.
- L. I. Kuncheva, *Combining pattern classifiers: Methods and algorithms*, Wiley-Interscience, Hoboken, New Jersey, USA, 2004.
- W.-J. Lee, R. P. Duin, A. Ibba, and M. Loog, An experimental study on combining euclidean distances, *The 2nd International Workshop on Cognitive Information Processing, CIP'10*, 2010, pp. 304–309.
- H. Ling and K. Okada, Diffusion distance for histogram comparison. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, CVPR '06*, 2006, Vol. 1, pp. 246–253.
- L. G. Nyul, J. K. Udupa, and X. Zhang, New variants of a method of mri scale standardization, *IEEE Trans Med Imaging* 19 (2000), 143–150.
- E. Pekalska and R. P. W. Duin, *The dissimilarity representation for pattern recognition. foundations and applications*, World Scientific, Singapore, 2005.
- K. Pohl and M. Sabuncu, “A unified framework for MR based disease classification,” *Information Processing in Medical Imaging, Lecture Notes in Computer Science*, J. Prince, D. Pham, and K. Myers (Editors), Springer, Berlin, Heidelberg, 2009, Vol. 5636, pp. 300–313.
- S. G. Potkin, J. A. Turner, G. G. Brown, G. McCarthy, D. N. Greve, G. H. Glover, D. S. Manoach, A. Belger, M. Diaz, C. G. Wible, J. M. Ford, D. H. Mathalon, R. Gollub, J. Lauriello, D. O’Leary, T. G. V. Erp, A. W. Toga, A. Preda, and K. O. Lim, Working memory and DLPFC inefficiency in schizophrenia: The FBIRN study, *Schizophr Bull* 35 (2009), 19–31.
- K. M. Ray, H. Wang, Y. Chu, Y. F. Chen, A. Bert, A. N. Hasso, and M. Y. Su, Mild cognitive impairment: apparent diffusion coefficient in regional gray matter and white matter structures, *Radiology* 24 (2006), 197–205.
- M. Rovaris, M. Bozzali, G. Iannucci, A. Ghezzi, D. Caputo, E. Montanari, A. Bertolotto, R. Bergamaschi, R. Capra, G. L. Mancardi, V. Martinelli, G. Comi, and M. Filippi, Assessment of normal-appearing white and gray matter in patients with primary progressive multiple sclerosis—A diffusion-tensor magnetic resonance imaging study, *Arch Neurol* 59 (2002), 1406–1412.
- Y. Rubner, C. Tomasi, and L. J. Guibas, The earth mover’s distance as a metric for image retrieval, *Int J Comput Vis* 40 (2000), 99–121.
- D. Rujescu and D. A. Collier, Dissecting the many genetic faces of schizophrenia, *Epidemiol Psychiatr Soc* 18 (2009), 91–95.
- F. Serratos and A. Sanfeliu, Signatures versus histograms: Definitions, distances and algorithms, *Pattern Recognit* 39 (2006), 921–934.
- M. E. Shenton, C. C. Dickey, M. Frumin, and R. W. McCarley, A review of mri findings in schizophrenia, *Schizophrenia Res* 49 (2001), 1–52.
- J. G. Sled, A. P. Zijdenbos, and A. C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in mri data, *IEEE Trans Med Imaging* 17 (1998), 87–97.
- M. J. Swain and D. H. Ballard, Color indexing, *Int J Comput Vis* 7 (1991), 11–32.
- M. Tansella and L. Burti, Integrating evaluative research and community based mental health care in Verona-Italy, *Br J Psychiatr* 183 (2003), 167–169.
- W. D. Taylor, E. Hsu, K. R. R. Krishnan, and J. R. MacFall, Diffusion tensor imaging: background, potential, and utility in psychiatric research, *Biol Psychiatr* 55 (2004), 201–207.
- B. Tomasino, M. Bellani, C. Perlini, G. Rambaldelli, R. Cerini, M. Isola, M. Balestrieri, S. Caligrave, A. Versace, R. P. Mucelli, A. Gasparini, M. Tansella, and P. Brambilla, Altered microstructure integrity of the amygdala in schizophrenia: A bimodal MRI and DWI study, *Psychol Med* 41 (2010), 301–311.
- A. Ulaş, R. P. Duin, U. Castellani, M. Loog, M. Bicego, V. Murino, M. Bellani, S. Cerruti, M. Tansella, and P. Brambilla, Dissimilarity based detection of schizophrenia, *ICPR workshop on Brain Decoding: Pattern Recognition Challenges in fMRI Neuroimaging, WBD'10*, 2010, pp. 32–35.
- J. Ventura, K. H. Nuechterlein, K. L. Subotnik, D. Gutkind, and E. A. Gilbert, Symptom dimensions in recent-onset schizophrenia and mania: A principal component analysis of the 24-item brief psychiatric rating scale, *Schizophrenia Res* 97 (2000), 129–135.
- World Health Organization, *Schedules for clinical assessment in neuropsychiatry*, WHO, Geneva, 1992.
- World Health Organization, *SCAN 2.1.: Schede di valutazione clinica in neuropsichiatria*, Il Pensiero Scientifico Editore, Roma, 1996.
- U. Yoon, J.-M. Lee, K. Im, Y.-W. Shin, B. H. Cho, I. Y. Kim, J. S. Kwon, and S. I. Kim, Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia, *Schizophrenia Res* 34 (2007), 1405–1415.