



Project acronym	SIMBAD
Project full title	Beyond Features: Similarity-Based Pattern Analysis and Recognition
Deliverable Responsible	Dipartimento di Informatica Università degli studi di Verona Strada le Grazie, 15 – 37134 Verona (Italy) http://www.di.univr.it/
Project web site	http://simbad-fp7.eu
EC project officer	Teresa De Martino
Document title	Mid-term report
Deliverable n.	D7.1
Document type	Report
Dissemination level	Public
Contractual date of delivery	M 24
Project reference number	213250
Status & version	Definitive version
Work package, Deliverable responsible	WP7, UNIVR
Author(s)	A. Ulaş, M. Bicego, U. Castellani, M. Cristani, V. Murino, A. Perina
Additional contributor(s)	

Contents

1	Overview	5
2	Related Work	5
3	Materials	7
3.1	Study population	7
3.2	MRI procedure	8
3.3	Tracing	8
3.3.1	Intracranial Volume	9
3.3.2	Entorhinal Cortex	9
3.3.3	Hippocampus	9
3.3.4	Thalamus	9
3.4	Data set	10
3.5	Pre-processing	11
3.5.1	MRI intensity scale normalization	12
3.6	Dissimilarity calculations	14
3.6.1	Dissimilarities based on histograms	14
3.6.2	Dissimilarities based on registration of MRIs	16
3.7	Delivered data	16
4	Proposed methods	17
4.1	Single ROI	17
4.1.1	Statistical analysis	17
4.1.2	Classification experiments	17
4.1.3	Conclusions	21
4.1.4	Preliminary 1nn results of dissimilarity matrices	22
4.2	Multiple ROIs	22
4.2.1	Combining ROIs and classifiers	22
4.2.2	Classifier combination and ensemble techniques	25
4.2.3	Our contributions	25
4.2.4	Experiment setup	26
4.2.5	Ensemble methods applied	26
4.2.6	Divison of data set	27
4.2.7	Results and Discussions	28
4.2.8	Discussion	32
5	Work in progress	32
5.1	Schizophrenia detection by generative kernels	33
5.1.1	The generative part	33
5.1.2	The discriminative part: the Fisher kernel	34
5.1.3	Preliminary results	35
5.2	Schizophrenia detection by feature-based morphometry approach	35
5.2.1	Proposed method	36
5.2.2	Preliminary results	37

5.3	Schizophrenia detection from multiple modalities	38
6	Cooperation with other partners	38
6.1	Schizophrenia detection by Information Theoretic Kernels	40
6.1.1	Proposed Approach	41
6.2	Schizophrenia detection on the Dissimilarity space	41
6.2.1	Dissimilarity Space	43
6.2.2	Experiments	43
6.3	Work in progress	45
7	Final discussion and future work	46

WP7: Mid-term report on preliminary results

Abstract

This document results from work carried out in the context of the Workpackage 7 of the SIMBAD project (Information and Communication Technologies, FET Open Collaborative Project, 7th Framework Programme). This is the mid-term report on preliminary results with indications on refining individual and collaborative approaches.

This WP should answer in some way to the following question: Is it possible to identify schizophrenia just by analyzing magnetic resonance imaging (MRI) brain images? This is one of the fundamental questions of MRI-based studies of human brains for people affected by mental illnesses, like schizophrenia, traditionally diagnosed by self-reports and behavioral observations. This problem is challenging because current diagnosis of schizophrenia is not carried out using MRI images. The second challenging aspect of the problem is the lack of enough data to automatize or train suitable pattern recognition learners to help the diagnosis. In the state of the art, there have been several studies which use smaller number of subjects, but the data we analyze in this WP is one of the largest in the literature. Actually, the data set consists of MRI images of subjects which are manually divided into subparts, called Regions of Interest (ROIs), by experts using internationally accepted protocols. The appeal of our approach in this context is at least twofold: first, to provide a non-invasive diagnostic tool for mass analyses and early diagnoses, and, second, to characterize mental illnesses with specific and detectable brain abnormalities.

Our preliminary basic classification results are promising; hence we are encouraged to carry out further steps up to the pattern recognition ladder. In our experiments, we apply a multi-classification approach, involving classifiers and ROIs, as well as the methods and tools which were developed in the context of SIMBAD project (IT kernels, generative kernels, dissimilarity combinations). We have established collaborations with partners to share knowledge and experience. Some of these collaborations have already borne fruit; we have submitted a joint paper with the Delft group. The SIMBAD workshop in Zurich last February was a great opportunity to create new collaborations and strengthen the ones already on going, and also it was useful to share knowledge and experience between the partners in the project. We are now pursuing several new directions and establishing collaborations in the search for a pattern recognition solution to schizophrenia detection from MRI images. We have already published three conference papers and submitted one journal and one conference paper in the context of WP7.

In this report, we initially report a description of the data set and an overview of the state of the art techniques for processing this kind of information, followed by our preliminary results and work in progress (both own and collaborative), in order to prepare a roadmap for future analysis.

1 Overview

Computational neuroanatomy using magnetic resonance imaging (MRI) is a growing research field that employs 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.

In this project, we focus on schizophrenia. Schizophrenia is a heterogeneous psychiatric disorder characterized by several symptoms such as hallucinations, delusions, cognitive and thought disorders (Bellani and Brambilla, 2008). Although genetic and environmental factors play a role in the disorder its etiology remains unknown and substantial body of research has demonstrated numerous structural and functional brain abnormalities in patients with both chronic and acute forms of the disorder (Shenton et al., 2001; Rujescu and Collier, 2009).

The main aim in this workpackage is to exploit the capability of advanced pattern recognition techniques in distinguishing between normal subjects and patients affected by schizophrenia (Davatzikos, 2004). We focus our analysis on our SIMBAD data set composed of a wide number of brain scans for which several Regions of Interest (ROIs) have been manually traced by experts. We start our analysis by adopting standard feature-based classification techniques on single ROIs. Then, we show drastic improvements when the several ROIs are combined by employing multi-classification strategies. Moreover, we report the work in progress devoted to the application of the new pattern recognition techniques of the SIMBAD project to our medical application. To this aim, we have computed several dissimilarity matrices which have been delivered to the SIMBAD partners. In particular, some experiments have been already carried out by employing: i) generative kernels, ii) IT-kernels, iii) dissimilarity-based methods. Preliminary results are encouraging, showing the effectiveness of the new methods proposed in the SIMBAD project for the problem of brain classification in Schizophrenia.

Roadmap. The report is organized as follows. In Section 2, we present the state of the art in schizophrenia detection, and in Section 3 we introduce the data set used in this study, together with the pre-processing procedures and the feature (dissimilarity) extraction phase. In Section 4, we describe the methods developed to process this data followed, in Section 5, by the work in progress constituted by the use of generative kernels and the morphometry-based approach. Finally, we highlight the work in cooperation with the other partners in in Section 6, and we draw some conclusions in Section 7.

2 Related Work

Several works have been proposed for human brain classification in the context of schizophrenia research (Shenton et al., 2001). Standard approaches are based on detecting morphological differences on certain brain regions, namely Regions of Interest (ROIs). Usually, the aim is the observation of volume variations (Pruessner et al., 2000; Shenton et al., 2001; Baiano et al., 2008), but recently more advanced morphological region descriptors have been proposed (Gerig et al., 2001; Timoner et al., 2002; Reuter et al., 2009). Gerig et al. (2001) introduced a ROI-based morphometric analysis by defining spherical harmonics and 3D skeleton as shape descriptors. Improvement of such shape-descriptor-based approach with

respect to classical volumetric techniques is shown experimentally. Timoner et al. (2002) introduced a new morphological descriptor by properly encoding both the displacement fields and the distance maps for amygdala and hippocampus. In Reuter et al. (2009), the so called Shape-DNA signature has been introduced by taking the eigenvalues of the Laplace-Beltrami operator as region descriptor for both the external surface and the volume. Moreover, possible morphological anomalies can be localized by the analysis of the eigenfunctions.

In general, ROI-based techniques require the manual tracing of brain subparts. In order to avoid such expensive procedure, Voxel Based Morphometry (VBM) techniques have been introduced (Ashburner and Friston, 2000; Kawasaki et al., 2007) for which the entire brain is transformed onto a template, namely the stereotaxic space. In Kawasaki et al. (2007), a multivariate Voxel-Based Morphometry approach method is proposed to differentiate schizophrenic patients from normal controls. Inferences about the structural relevance of gray matter distribution are carried out on several brain sub-regions. Around 80% of classification accuracy was observed on an experiment with 16 patients and 16 controls. In Voets et al. (2008), cortical changes in adolescent on-set schizophrenic patients are analyzed by combining Voxel-Based with Surface-Based Morphometry (SBM). The experiment has involved 25 patients and 25 controls by observing an average of about 80% of accuracy. A SBM approach was also proposed in Yoon et al. (2007). A support vector machine (SVM) has been introduced to classify cortical thickness which has been measured by calculating the Euclidean distance between linked vertices on the inner and outer cortical surfaces. More than 90% of accuracy is observed for the classification with a sample of 53 schizophrenic patients and 52 healthy subjects. Recently, more complex brain signatures have been introduced by analyzing the deformation (or the tensors) which brings the current brain onto the template (Dubb et al., 2002; Fan et al., 2007). In Dubb et al. (2002), a statistical analysis has been carried out on corpus callosum for a large population (i.e., 100 patients and 190 controls) and a Tensor-Based Morphometry (TBM) approach is introduced as a curve registration algorithm. Statistical plots revealed a substantial area of contraction in the anterior callosum associated with schizophrenia. In Fan et al. (2007), a new morphological descriptor has been defined by combining deformation-based morphometry with SVM. In this fashion, multivariate relationships among various anatomical regions have been captured to characterize more effectively the group differences.

It is worth to note that in most of the mentioned works, the involved classifier was a Support Vector Machine, but more general approaches are also proposed (i.e. Liu et al. (2004)). Here, a set of image features which encode both general statistical properties and Law's texture features from the whole brain are analyzed. Such features are concatenated onto a very high dimensional vector which represents the input for a classic learning-by-example classification approach. Several classifiers are then evaluated such as decision trees or decision graphs, and the observed classification accuracy resulted about 80% for 27 controls and 24 schizophrenic patients. In Browne et al. (2008), the authors proposed a neural network to measure the relevance of thalamic subregions implicated in schizophrenia. The study is based on the metabolite N-acetylaspartate (NAA) using *in vivo* proton magnetic resonance spectroscopic imaging. The experiment has been carried out on 18 subjects (all males), 9 patients and 9 controls. After leave-one-out cross validation the achieved performance was 100%.

3 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.

Roughly speaking there are two main categories of methods: (i) methods based on the analysis of regions of interest (ROI), 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 report, we introduce our data set which consists of ROIs of MRI images and we introduce our results using proposed methodologies.

3.1 Study population

This study involves a 64 patients 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 data set is composed of MRI brain scans of 64 patients recruited from the area of South Verona (i.e., 100000 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 (American Psychiatric Association, 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 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 psycho-pathological item groups completed by the two raters were compared, in order 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 in order 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.

All but two patients were receiving antipsychotic medication at the time of imaging. 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. 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, 60 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 the dizziness. Also, none of these participants was taking medication, including drugs for nausea or vertigo.

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.

3.2 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 minutes 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. T1-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, matrix size = 384 *times* 512, NEX = 2). Proton density (PD)/T2-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, NEX = 2) according to an axial plane running parallel to the anterior-posterior (AC-PC) commissures to exclude focal lesions. Subsequently, a coronal 3-dimensional 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, inversion time = 1100) to obtain 144 images covering the entire brain.

3.3 Tracing

Every ROI has a protocol based on the individuation of particular anatomical landmarks in the brain. Here we detail some of these protocols as reference.

3.3.1 Intracranial Volume

Intracranial volume (ICV) was traced in the coronal plane along the border of the brain and included the cerebrospinal fluid (CSF), dura mater, sinus, optic chiasma, brainstem, cerebral and cerebellar matter. The inferior border did not extend below the base of the cerebellum. The first and last 10 slices including brain matter were traced, after which 1 of every 5 slices was traced. ICV measurements were obtained by a well-trained rater who achieved interrater reliability of 0.97 (intraclass correlation coefficient), established by blindly tracing 10 randomly selected scans.

3.3.2 Entorhinal Cortex

The entorhinal cortex was traced on MRI coronal slices based on prior published methods (Bernasconi et al., 1999; Prasad et al., 2005). The frontotemporal stem delimited the region of interest anteriorly. The intersection of the line along the grey-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 uncal cleft. The intersection of the line along the grey-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 (Bernasconi et al., 1999). It should be noted that the prior methods were slightly modified since the lateral geniculate body was poorly detectable in most of our scan and therefore it was not to used as a posterior limit. Also, the perirhinal cortex was included in our tracing.

3.3.3 Hippocampus

Hippocampus was detected on MRI coronal slices, where the corona radiata is and then the ambient cistern were used as the superior border; the white matter acted as the inferior border, and the inferior horn of the lateral ventricle as the lateral one (Brambilla et al., 2003). Parahippocampal gyrus was not included in the tracing.

3.3.4 Thalamus

After the thalamus was identified in the coronal plane, its limits were manually traced.^{47, 48} All thalamic measurements were done by a well-trained rater blinded to the participant's identity. Interrater reliability of $r = 0.96$ for the right thalamus and $r = 0.95$ for the left thalamus was achieved. Thalamic volumes (mL) were obtained by summing the volumes of all relevant slices and were expressed in cubic centimeters. The tracing of the thalamus was performed on the T1-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.

3.4 Data set

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.

The raw images are acquired using a 1.5 tesla MRI machine and 144 slices are acquired using 384×512 resolution. These images are then transferred to PC workstations in order 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. This latter procedure is the manual annotation of the images, performed by drawing contours enclosing the intended region. It is carried out by a trained expert following a specific protocol for each ROI. We begin introducing our data set with some demographic and clinical characteristics of the study groups Table 1.

Table 1: Some demographic and clinical characteristics of the study groups. The Student’s *t*-test of the age means rejects (at a two-tailed significance level of $p < 0.05$) the hypothesis that the study groups are significantly different in age, and Pearson χ^2 confirms the same for the gender differences.

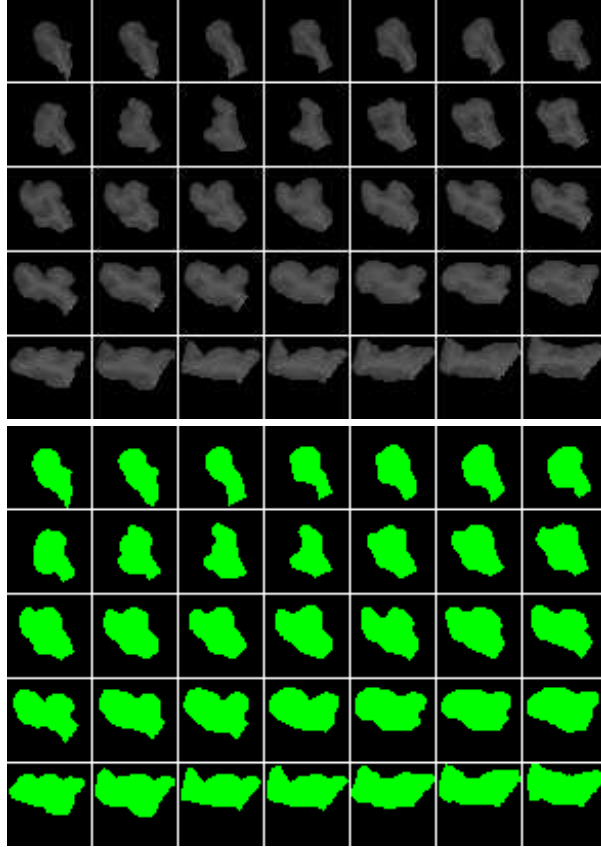
Characteristic	Group mean (and SD)*		Statistics		
	Control <i>n</i> = 60	Schizophrenia <i>n</i> = 64	Test	<i>df</i>	<i>p</i>
Age, yr	39.95 (11.25) [23-60]	38.84 (11.96) [18-62]	$t = 0.53$	122	0.60
Male/female	32/28	43/21	$\chi^2 = 2.49$	1	0.11
Age at onset, yr		26.28 (9.17)			
Duration of illness, yr		13.37 (10.30)			

SD = standard deviation; *df* = degrees of freedom; *p* = value of significance
* Unless otherwise indicated.

The ROIs traced in this data set are 7 pairs (for the left and the right hemisphere respectively) of disconnected image portions which can be described as the following:

- 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*);
- Superior Temporal Gyrus (*lstg* and *rstg*);

Figure 1: 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.



- Thalamus (*lthal* and *rthal*).

In Fig. 1, we show a sample from the data set, specifically the ROI volume of the right superior temporal gyrus 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 bounding box, the ROI itself is irregularly shaped, as can be clearly seen from the corresponding binary masks on the right, artificially colored to highlight the ROI shape.

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.

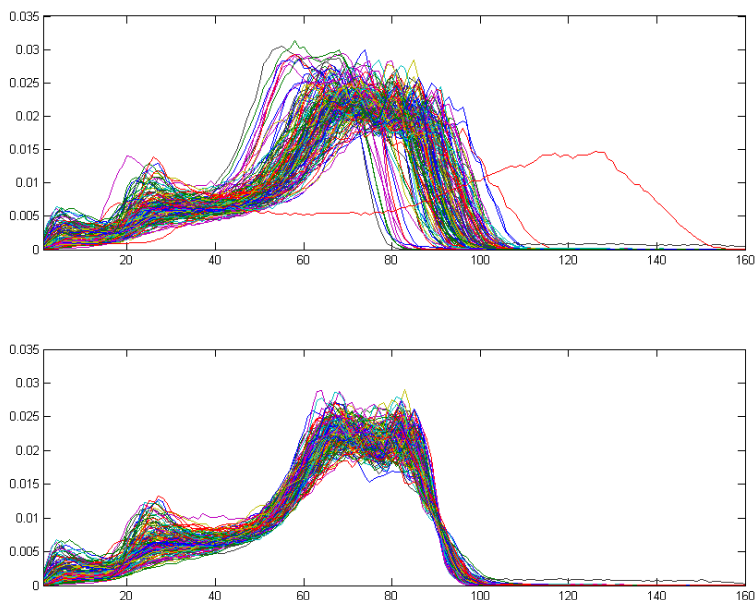
3.5 Pre-processing

From the slices of ROIs, we compute histograms of normalized intensities and use them for our feature based experiments.

3.5.1 MRI intensity scale normalization

A major disadvantage of MRI compared to 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 (Nyúl 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.

Figure 2: ICV intensity histograms (treated like probability density functions), before and after the normalization process.



A successful technique used to calibrate MR signal characteristics at the time of acquisition employs *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; Nyúl et al., 2000).

In this work, we retrieve the rescaling parameters to form intensity histograms from the ICV histograms (Figure 2). 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., 2009a). 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, since 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 feature based pattern recognition experiments.

In addition to using raw histograms, we tried a variety of methods to extract possible representations to calculate the dissimilarity matrices by employing feature selection or processing techniques and dimensionality reduction procedures. The following list shows the methods we applied:

1. Raw histograms, i.e., with frequencies of values in natural bins (no quantization)
2. Histograms normalized as probability density functions (pdf), to emphasize the relative distribution of frequencies
3. Histograms scaled bin by bin to have zero mean and unit variance (bin scaling), to emphasize deviations from average frequencies
4. Pdfs with bin scaling, to emphasize the relative distribution of deviations
5. Histograms with zero variance bins eliminated, to eliminate constant bins
6. Pdfs with zero variance bins eliminated
7. As (5) with bin scaling
8. As (6) with bin scaling
9. Histograms quantized into 4 bins
10. Pdfs quantized into 4 bins
11. As (9) with bin scaling
12. As (10) with bin scaling
13. Dimensionality reduction with principal component analysis (PCA) at 99
14. PCA of (6);
15. PCA of (7);
16. PCA of (8);
17. For each histogram vector in (1), we calculated the cumulative sum.
18. Cdf of (2);
19. Cumulative sum of (5);
20. Cdf of (6)
21. Cumulative sum of (9);
22. Cdf of (10)

3.6 Dissimilarity calculations

The computed histograms (and their pdfs) of intensities have been used to calculate dissimilarities between subjects using dissimilarity measures for histograms and pdfs. Also we have computed some dissimilarities based on the registration of brain MRIs.

3.6.1 Dissimilarities based on histograms

There are various dissimilarity measures that can be applied to measure the dissimilarities between histograms (Cha and Srihari, 2002; Serratosa and Sanfeliu, 2006). Each dissimilarity measure has its own assumptions about the data. Some are cross-bin measures, some are based on the corresponding bins only; some are metrics, some are not. We selected a palette of histogram dissimilarity measures to be used in this study. Also, histograms can be considered as (converted to) pdfs. Various algorithms exist which measure the dissimilarity between two discrete distributions. We also applied some of these to create other dissimilarity matrices.

We used the following measures for calculating the dissimilarity between two vectors:

- *bs*: Bhattacharyya distance
- *chi*: χ^2 distance
- *diffusion*: Diffusion distance
- *emd*: Earth mover's distance
- *euclid*: Euclid distance
- *intersect*: Histogram intersection
- *kl_orig*: Original Kullback-Leibler divergence which creates a nonsymmetric dissimilarity matrix
- *kl*: Symmetrized KL divergence
- *js*: Jensen-Shannon divergence
- *l1*: L1 distance.

Suppose that we have two histograms S and M with n bins S_i and M_i ; and their corresponding pdfs p and q . We define the number of elements in S and M as $|S|$ and $|M|$ and we want to calculate the dissimilarity D between M and S .

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

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

where

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

We applied this method to pdfs.

χ^2 distance This metric is based on the χ^2 test for testing the similarity between two distributions. It is defined as:

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

We applied this method to histograms.

Diffusion distance Ling and Okada (2006) proposed the diffusion distance between two histograms. In this method, the distance between two histograms is defined as a temperature field. It is derived as the sum of dissimilarities over scales. For more information see Ling and Okada (2006)¹.

We applied this method to histograms.

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

$$C_i = \left| \sum_{i=1}^n (S_i - M_i) \right|$$

$$D = \sum_{i=1}^n C_i$$

We applied this method to histograms and pdfs.

Histogram intersection *intersect* is defined (Swain and Ballard, 1991) as a measure for distinguishing between objects using color information. It is defined as:

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

which is the number of intersecting values in each cell, normalized by the total number of elements in the histograms (if the histograms have different number of elements, normalized by their minimum). Since this is a similarity measure, we convert it to dissimilarity using:

$$D = \min(|M|, |S|) \times (1 - Sim(S, M))$$

We applied this method to histograms.

Dissimilarities based on Kullback-Leibler divergence Kullback-Leibler divergence is defined as:

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

¹The code has been taken directly from the author’s homepage: http://www.ist.temple.edu/~hbling/code_data.htm

This measure is not a distance metric but relative entropy since $D(p, q) \neq D(q, p)$. *kl_orig* is this dissimilarity so the dissimilarity matrix is not symmetric. There are various ways to symmetrize this dissimilarity. We used two ways: *kl* is defined as:

$$D = D(p, q) + D(q, p)$$

js is the Jensen-Shannon divergence and is defined as:

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

where r is the average of p and q . We applied these measures to pdfs.

***L1* and *L2* distances** These are the well-known *L1* and Euclid measures, which we applied to all representations.

3.6.2 Dissimilarities based on registration of MRIs

Since calculating histograms of intensities and calculating dissimilarities from these features involves preprocessing of data, we also prepared some dissimilarity measures based on deformable registration of brain MRIs for each subjects. For every pair of subjects we used the CAMP² software to register the subjects and calculated dissimilarities based on the deformation field. We used two measures for calculating the dissimilarity between two subjects based on the deformation field:

- The sum of the deformations on every point (i, j, k) summed over all points.
- The length of the deformations on every point (i, j, k) summed over all points.

The registration was carried out on the whole MRIs and the portion that concerns every ROI has been extracted to calculate the dissimilarity between the ROIs of each pair of subjects.

3.7 Delivered data

For the purpose of using these data in the context of SIMBAD project, several intermediate representations of this data have been computed and put on the SIMBAD web site for our partners to use³. The intermediate representations consist of:

- Histograms and pdfs of normalized intensity values of MRI images. These are the first two representations introduced in Section 3.5.1 and have been provided for partners who would like to work on raw histograms.
- Dissimilarity matrices using these representations have been computed and provided for partners who would like to work on dissimilarity matrices.
- Dissimilarity matrices based on nonrigid registration of brain MRIs have been provided for partners who would like to work on dissimilarities based on unprocessed data.

²The software is available from <http://www.mrf-registration.net>

³The data are available at: http://simbad-fp7.eu/restricted/SIMBAD_Dataset_Dissimilarity

4 Proposed methods

With the MRI classification data we have carried out several experiments using various pre-processing methods and classifiers. We can divide the methods we tried into two categories. The first category is the single ROI category where we apply several classification methods on the ROIs independently. The second category is the multiple ROI category where we apply classifier combination techniques using multiple ROIs. Our experiments prove that it's better to combine multiple ROIs in order to get better classification accuracy.

4.1 Single ROI

In this subsection, we will present our preliminary classification results using ROIs independently.

4.1.1 Statistical analysis

An analysis of covariance (ANCOVA) using age, gender and ICV as covariates is usually performed to compare the volumes of ROIs between patients with schizophrenia and healthy normal controls. The purpose of ANCOVA is to find out whether data from several groups have a common mean, after adjusting for sources of bias in observational studies. That is, to determine whether the groups are actually different in the measured characteristic when some known sources are factored in. When studying volumetric properties of the brain or parts of it, it is well known that the overall size shrinks with age, it is smaller on average for females, and there is considerable variation between subjects. ANCOVA retrieves adjusted means for the population groups, in this case patients and controls that can be tested for significance against the simpler hypothesis that there is no difference between them (the so-called null hypothesis). This test is a straightforward application of the F test, with an accompanying p value of significance. From the results reported in Table 2, we can draw the following evidence:

- two ROIs, *rdlpfc* and *rec*, show significant differences (respectively, $p = 0.004$ and $p = 0.002$);
- three additional ROIs, *lamyg*, *ramyg* and *lthal*, are significant at the $p = 0.05$ level.
- some other ROIs are close to being significant, but, overall, the normal distributions of volumes are too close or too overlapping in spread to be usable for discrimination.

These results are well in accordance with previous studies (Agarwal et al., 2008; Baiano et al., 2008)

4.1.2 Classification experiments

We performed several classification experiments under varying conditions of histogram pre-processing, feature selection and classifiers to seek the most promising settings for further investigations. Experiments were carried out in Matlab using PRTools (Duin, 2005) and accuracy figures for each test run were obtained through leave-one-out (LOO) cross-validation.

Table 2: Analysis of covariance for ROI volumes in normal controls and patients with schizophrenia show significant differences in the group means only for *rec* and *rdlpfc*.

ROI volumes (cm ³)	Group mean (and SD)		Statistics	
	Control <i>n</i> = 60	Schizophrenia <i>N</i> = 64	<i>F</i>	<i>p</i>
<i>lamyg</i>	1.46 (0.27)	1.37 (0.28)	3.07	0.08
<i>ramyg</i>	1.53 (0.27)	1.43 (0.30)	3.74	0.06
<i>ldlpfc</i>	15.08 (7.07)	14.35 (7.25)	0.42	0.52
<i>rdlpfc</i>	15.93 (6.83)	13.32 (6.75)	5.73	0.02
<i>lec</i>	1.05 (0.22)	1.02 (0.22)	0.67	0.41
<i>rec</i>	1.16 (0.23)	1.08 (0.24)	7.27	0.008
<i>lhg</i>	2.22 (0.70)	2.37 (0.67)	2.33	0.13
<i>rhg</i>	2.04 (0.58)	2.16 (0.70)	1.33	0.25
<i>lhippo</i>	1.73 (0.29)	1.75 (0.41)	0.05	0.82
<i>rhippo</i>	1.77 (0.32)	1.76 (0.33)	0.09	0.76
<i>lstg</i>	13.75 (1.95)	13.78 (2.17)	0.01	0.99
<i>rstg</i>	14.55 (1.98)	14.56 (2.36)	0.01	0.94
<i>lthal</i>	4.77 (0.52)	4.66 (0.59)	3.10	0.08
<i>rthal</i>	5.00 (0.63)	5.02 (0.67)	0.17	0.68

SD = standard deviation.

In this preliminary work, each ROI was treated independently of the others, much like in common medical analyses, to assess the individual discriminatory capabilities and to be able to effectively compare results with previous medical studies. In the following, every reference to “raw histograms” are intended to be the histograms scale-normalized as in Subsection 3.5.1. We used the first 16 representations of Section 3.5.1 in these experiments.

We tested the following classifiers (Duda et al., 2000):

- Gaussian radial basis support vector classifier (svm), where the standard deviation is estimated by cross validation;
- Parzen classifier (parzen), where the optimum smoothing parameter is retrieved by LOO using the (Lissack and Fu, 1976) estimate for the classification error;
- Fisher’s least square linear classifier (fisher);
- K-nearest neighbor classifier (knn), with K automatically optimized with respect to the LOO error;
- Nearest neighbor (1nn);
- Linear Bayes normal classifier (ldc);
- Logistic linear classifier (loglc).

Table 3: Best and average accuracies of classification for each ROI. In brackets, the combination of classifier and features (more than one number indicates that multiple settings attained the same accuracy) that achieved the given performance (see Section 3.5.1 for the numbering of feature methods). * In brackets, the average of only the three best classifiers (*svm*, *knn*, *parzen*).

ROI	Best % accuracy (classifier,features)	Average % accuracy over classifiers and features*
<i>lamyg</i>	71.0 (svm,13/15)	60.0 (63.7)
<i>ramyg</i>	68.5 (knn,13)	57.2 (59.9)
<i>ldlpfc</i>	71.0 (svm,1/5)	59.3 (60.9)
<i>rdlpfc</i>	70.2 (svm,16)	55.4 (55.0)
<i>lec</i>	66.1 (svm,15)	58.1 (59.0)
<i>rec</i>	66.9 (parzen,13)	58.6 (61.2)
<i>lhg</i>	66.1 (svm,12)	56.4 (58.5)
<i>rhg</i>	64.5 (knn,16)	53.0 (53.4)
<i>lhippo</i>	73.4 (svm,8)	58.8 (61.0)
<i>rhippo</i>	65.3 (knn,1/5/12)	53.7 (55.5)
<i>lstg</i>	64.5(knn,13)	56.0 (58.1)
<i>rstg</i>	65.3 (knn,10)	55.1 (57.5)
<i>lthal</i>	68.5 (knn,10)	55.3 (56.9)
<i>rthal</i>	66.9 (knn,4/8/12)	57.0 (58.8)

Table 4: Best and average accuracies for each classifier. In brackets, the combination of ROI and features that achieved the given performance.

Classifier	Best % accuracy (ROI,features)	Average % accuracy over ROIs and features
<i>svm</i>	73.4 (<i>lhippo</i> ,8)	59.6
<i>knn</i>	69.4 (<i>ldlpfc</i> ,12)	58.1
<i>parzen</i>	69.4 (<i>lamyg</i> ,16)	57.9
<i>fisher</i>	67.7 (<i>lamyg</i> ,16)	55.8
<i>loglc</i>	66.9 (<i>lhippo</i> ,2/4)	55.6
<i>1nn</i>	69.4 (<i>ldlpfc</i> ,12)	55.0
<i>ldc</i>	67.7 (<i>lamyg</i> ,16)	55.1

Table 5: Best and average accuracies for each feature method. In brackets, the combination of classifiers and ROIs that achieved the given performance.

Feature	Best % accuracy (classifier,ROI)	Average % accuracy over classifiers and ROIs
1	71.0 (svm, <i>ldlpfc</i>)	56.1
2	66.9 (loglc, <i>hippo</i>)	54.7
3	72.6 (svm, <i>hippo</i>)	55.9
4	71.8 (svm, <i>hippo</i>)	55.4
5	71.0 (svm, <i>ldlpfc</i>)	56.2
6	66.1 (svm, <i>ldlpfc</i>)	54.8
7	71.0 (svm, <i>hippo</i>)	55.7
8	73.4 (svm, <i>hippo</i>)	55.3
9	69.4 (svm, <i>lamyg/ldlpfc</i>)	58.7
10	68.5 (knn, <i>lthal</i>)	58.8
11	68.5 (svm, <i>ldlpfc</i>)	58.8
12	69.4 (knn/1nn, <i>ldlpfc</i>)	58.3
13	71.0 (svm, <i>lamyg</i>)	57.5
14	66.1 (svm, <i>lamyg</i>)	57.1
15	71.0 (svm, <i>lamyg</i>)	57.1
16	72.6 (svm, <i>hippo</i>)	56.8

In Tables 3, 4 and 5, we report a summary of the results we obtained. Table 3 shows the best and average classification accuracies for each ROI over all tested classifiers and feature methods. It gives a rough indication of which ROIs are more discriminatory. Table 4 shows the best and average classification accuracies for each classifier over all feature methods and ROIs, giving a ranking of the most performing classifiers. Finally, Table 5 shows the best and average accuracies for each feature method over all classifiers and ROIs, highlighting the most promising data processing techniques: quantization and PCA.

Note that the average accuracies can be used to sort the *relative* discriminatory power between different choices (*e.g.*, *svm* is more powerful than *1nn* among classifiers), but they have no meaning in absolute terms (59.6% for *svm* is a rather dreadful classification rate, barely above chance).

Overall, results are suggestive and encouraging; in fact they seem to support the main scientific claim that it is possible to identify schizophrenic patients from healthy people. However, best and average accuracies are not conclusive, showing a significant rate of false positives and false negatives (see Figure 3). This confirms how tough the problem is.

Nevertheless, we can draw some limited conclusions:

- *svm* was the best performing classifier, both on average and with the best overall accuracy, obtained on *hippo* with 73.4%;
- the three best performing ROIs are *hippo*, *lamyg* and *ldlpfc* with respectively 73.4%, 71.0% and 71.0%;

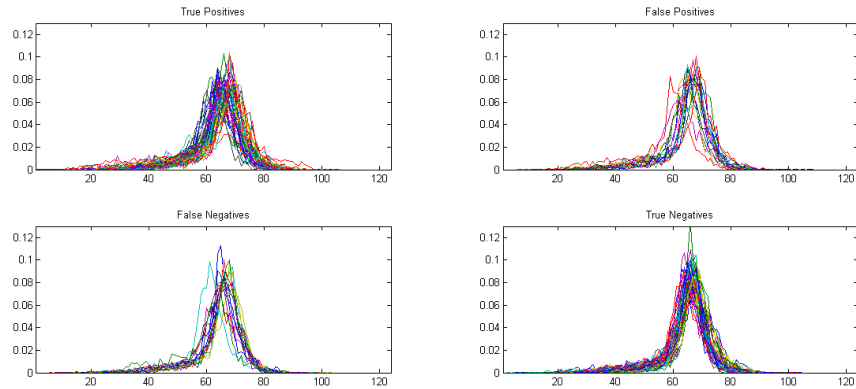


Figure 3: Plots of the subjects’ histograms divided according to the classification experiment that achieved the best accuracy (*svm* with feature method 8 on *hippo*). From top to bottom, left to right, the four plots identify true positives, false positives, false negatives and true negatives (precision = 0.74, recall = 0.75).

- left portions of the ROIs are almost always more discriminative, which is interesting since medical analyses found an overall difference between healthy subjects and schizophrenic patients with higher statistical difference on right ROIs of *amyg*, *hippo* and *dlpfc*;
- while not achieving the best peak performances, quantization seems to enhance discrimination power over all classifiers and ROIs;
- average accuracies are very low, in fact just above chance;
- *1nn*, *ldc* and *loglc* always perform worse or equal to the other classifiers, bringing the averages down.

The abnormalities in the amygdala, hippocampus and dorsolateral prefrontal cortex, in particular in the left side, are among the most consistent findings in MRI studies on schizophrenia (Meisenzahl et al., 2008; Shenton et al., 2001). Suggesting that these structures play a major role for the pathophysiology of the disease (Lopez-Garcia et al., 2006). In particular, the dorsolateral prefrontal cortex, along with the thalamus and the hippocampus, is a critic component of the brain circuitry underlying higher cognitive functions, such as attention, executive function and context processing (Procyk and Goldman-Rakic, 2006).

The amygdala plays a critical role in the neural system that is involved in emotional and in fear-related responses (Swanson and Petrovich, 1998); the hippocampus is involved in long term memory and in regulating stress response (Sala et al., 2004; Tulving and Markowitsch, 1998).

4.1.3 Conclusions

In these first set of experiments, we have provided some supportive evidence that it is possible to discriminate between schizophrenic patients and healthy people based on analyses of brain

MR images. We have built upon previous medical studies that have focused on volumetric measurements of selected portions of the brain, namely ROIs, corresponding to well-known functional units that affect human cognitive behavior.

Classification results achieved with several off-the-shelf pattern recognition techniques under different data processing methods seem to suggest that the content of these ROIs, in the form of histograms, can be used to characterize the population of schizophrenic patients. Although intended to go beyond volumetric statistical analysis, our approach was expected to confirm previous medical findings. Instead, we struggle to find a clear interpretation of these results. In fact, we find that the left parts of the ROIs are almost always more discriminative than their counterparts, which is significant but unconnected to the relevant preexisting findings.

Moreover, we find that SVMs perform best both on peak and average accuracy, but not by a large margin with respect to a simpler classifier such as k nn. At the same time, the feature processing methods we employed have increased accuracies, but not dramatically and decisively. The results of this study has been published (Cheng et al., 2009b).

4.1.4 Preliminary 1nn results of dissimilarity matrices

We also computed the LOO accuracies from the dissimilarity matrices calculated using the dissimilarity measurements to be a baseline for dissimilarity based computations. In Tables 6, 7 and 8 we can see the 1nn results of the dissimilarity matrices using histogram dissimilarities, pdf dissimilarities and registration based dissimilarities respectively⁴.

4.2 Multiple ROIs

In this section techniques coming from the fusion theory (multiclassifier systems) have been applied to the dataset. In a preliminary set of experiments we used ROIs independently and trained/used/combined multiple classifiers for each ROI. In a further experiment setting, we combine multiple ROIs to see the effect of combination on classification accuracy. Medically speaking, it's a good idea to combine different ROIs because the contribution of each ROI to brain functioning is different, what we want to test in these experiments is if these contributions are complementary from the schizophrenia classification point of view.

4.2.1 Combining ROIs and classifiers

Our main aim in these experiments is to show that information contained in each Region of Interest (ROI) is different, and testing the effect of using multiple ROIs for detecting diseases in the brain using MRI images. Advanced pattern recognition techniques may deeply help the understanding of brain characteristics and functionalities and there has been several studies where these techniques have been applied (Davatzikos, 2004; Fan et al., 2007; Cheng et al., 2009b). The results of these experiments have been submitted to the "Multivariate Decoding and Brain reading" special issue of the Neuroimage journal (Ulaş et al., 2009a).

For the classification purpose, different and sophisticated approaches have been proposed and applied in this context (e.g., principal components (Yoon et al., 2007), SVMs (Fan et al.,

⁴These results have small differences between the ones in Table 22 because of differences in the normalization procedure

Table 6: Inn results of dissimilarities calculated using histograms (%).

	L2	L1	Intersect	Diffusion	Chi	EMD	max	mean	std
<i>lamyg</i>	57.26	62.90	54.84	62.90	58.06	51.61	62.90	57.93	4.46
<i>ramyg</i>	58.87	58.87	57.26	58.87	56.45	56.45	58.87	57.80	1.21
<i>ldlpfc</i>	53.23	49.19	54.03	50.00	54.84	53.23	54.84	52.42	2.28
<i>rdlpfc</i>	46.77	50.81	66.13	50.81	47.58	40.32	66.13	50.40	8.61
<i>lec</i>	51.61	50.00	50.81	50.81	50.81	42.74	51.61	49.46	3.33
<i>rec</i>	57.26	61.29	58.87	61.29	62.10	45.16	62.10	57.66	6.39
<i>lhg</i>	55.65	54.03	41.13	54.03	58.87	53.23	58.87	52.82	6.07
<i>rhg</i>	57.26	62.10	45.97	62.10	62.10	42.74	62.10	55.38	8.80
<i>lhippo</i>	54.03	56.45	52.42	56.45	58.06	50.00	58.06	54.57	3.00
<i>rhippo</i>	62.10	59.68	63.71	59.68	60.48	54.03	63.71	59.95	3.29
<i>lstg</i>	43.55	42.74	57.26	42.74	43.55	48.39	57.26	46.37	5.74
<i>rstg</i>	58.87	54.84	54.84	54.84	54.84	51.61	58.87	54.97	2.30
<i>lthal</i>	61.29	60.48	62.10	60.48	57.26	52.42	62.10	59.01	3.62
<i>rthal</i>	55.65	54.03	47.58	54.03	48.39	53.23	55.65	52.15	3.33
max	62.10	62.90	66.13	62.90	62.10	56.45	66.13		
mean	55.24	55.53	54.78	55.65	55.24	49.65		54.35	
std	5.19	5.86	6.96	5.74	5.67	4.99			5.98

Table 7: Inn results of dissimilarities calculated using pdfs (%).

	L2	L1	EMD	BC	KL	KL-asym	JS	max	mean	std
<i>lamyg</i>	59.68	62.90	54.84	54.03	58.87	60.48	57.26	62.90	58.29	3.15
<i>ramyg</i>	60.48	58.06	58.06	58.87	58.87	61.29	59.68	61.29	59.33	1.22
<i>ldlpfc</i>	47.58	50.00	38.71	46.77	45.97	43.55	45.97	50.00	45.51	3.57
<i>rdlpfc</i>	54.84	52.42	57.26	56.45	56.45	54.03	56.45	57.26	55.41	1.72
<i>lec</i>	61.29	63.71	59.68	58.87	59.68	58.87	59.68	63.71	60.25	1.72
<i>rec</i>	50.00	57.26	53.23	55.65	58.06	53.23	56.45	58.06	54.84	2.83
<i>lhg</i>	50.81	54.03	47.58	50.81	50.00	50.81	50.81	54.03	50.69	1.89
<i>rhg</i>	51.61	52.42	52.42	54.03	54.03	56.45	54.84	56.45	53.69	1.67
<i>lhippo</i>	60.48	59.68	64.52	54.84	59.68	55.65	54.03	64.52	58.41	3.75
<i>rhippo</i>	49.19	52.42	44.35	50.00	50.00	47.58	50.00	52.42	49.08	2.53
<i>lstg</i>	52.42	54.03	55.65	57.26	56.45	54.03	57.26	57.26	55.30	1.85
<i>rstg</i>	60.48	63.71	54.84	60.48	59.68	59.68	60.48	63.71	59.91	2.62
<i>lthal</i>	59.68	57.26	57.26	54.84	57.26	54.84	54.03	59.68	56.45	1.98
<i>rthal</i>	61.29	63.71	64.52	60.48	62.10	61.29	60.48	64.52	61.98	1.57
max	61.29	63.71	64.52	60.48	62.10	61.29	60.48	64.52		
mean	55.70	57.26	54.49	55.24	56.22	55.13	55.53		58.99	
std	5.23	4.86	7.14	4.01	4.61	5.24	4.30			4.71

Table 8: 1mm results of dissimilarities based on deformable registration (%).

	sum of deformations	sum of length of deformations
<i>lamyg</i>	49.19	50.00
<i>ramyg</i>	49.19	47.58
<i>ldlpfc</i>	56.45	57.26
<i>rdlpfc</i>	50.81	49.19
<i>lec</i>	48.39	52.42
<i>rec</i>	50.00	49.19
<i>lhg</i>	52.42	47.58
<i>rhg</i>	57.26	55.65
<i>lhippo</i>	53.23	53.23
<i>rhippo</i>	50.00	50.00
<i>lstg</i>	53.23	50.81
<i>rstg</i>	56.45	52.42
<i>lthal</i>	44.35	43.55
<i>rthal</i>	46.77	51.61
<i>icv</i>	55.65	53.23
whole-1	53.23	56.45
whole-2	58.87	58.06

2007; Cheng et al., 2009b), hybrid methods (Cheng et al., 2009c), etc.). In these experiments, we propose to apply one of these techniques in the context of ROI-based MRI analysis for classification of schizophrenic patients. In particular, we apply a quite recent and promising pattern recognition theory, namely the multi-classification theory. Classifier combination can be done at different levels and in different ways: (i) sensor fusion, (ii) representation fusion, (iii) algorithm fusion, (iv) decision fusion and others (Alpaydm, 2004; Kuncheva, 2004). One can also focus on extracting more complex morphological features from the given MRI image to get better classification accuracy; we do not choose this direction, instead, we focus on the classification and the combination of classifiers for achieving better performances using basic brain features. As a second contribution, we apply a correlation analysis to the outputs of classifiers trained on different ROIs to measure the level of overlap of information contained among different ROIs.

Typically, the employed approach is to train different classifiers, and fuse together the outputs of each classifier to come up with the final classification. In this study, we exploit the nature of the problem and investigate the use of a single classifier for each ROI, finally fusing the contributions coming from the different ROIs. This method has several advantages: (i) while exploiting all the potentials of classifier combination, we provide an alternative point of view, (ii) the results may have immediate medical interpretation, because results are presented in terms of ROIs, not in terms of complex classification algorithms' outcomes.

In particular, we use the histograms of properly normalized intensity values (Section 3.5.1) from the raw MRI images for each subject and ROI, then, we use different classifier combination techniques to form ensembles of ROIs, and compare the accuracies of these methods.

Besides classifier combination, we provide a correlation analysis on different ROIs (computed on outputs of classifiers, not on images or representations, as usually carried out), which allows us to realize that ROIs are mostly uncorrelated, and that fusing different ROIs may increase the performance of the selected ensemble. The obtained results show the effectiveness of this approach.

4.2.2 Classifier combination and ensemble techniques

There are several classification algorithms which can be used for any classification problem. These algorithms differ in the way they model the input data: actually, the accuracy or the success of them depend on the problem, data at hand, parameters of the classifier, etc.. This also complicates the selection of a successful classifier for any given task for people who are not experts in machine learning or pattern recognition. Furthermore, there is no universal classifier which is always better than all the other classifiers for all problems (no free lunch theorem (Duda et al., 2000)). Instead of selecting a single classifier, a possible and more accurate (but more complex) approach is to use multiple classifiers and combine them to get a final decision (known as ensemble construction, classifier fusion, classifier combination, etc. in the literature). There are several methods for combining multiple classifiers (Alpaydm, 2004; Kuncheva, 2004). The simplest method is to use voting which corresponds to the sum rule (Kittler et al., 1998) or other fixed rules (minimum, maximum, product and median). Methods based on resampling from a single data set, such as bagging (Breiman, 1996) and AdaBoost (Freund and Schapire, 1996), are not used to combine different learning algorithms. Combination is carried out using a trained, second layer, classifier which estimates the real output from the outputs of base classifiers (Wolpert, 1992) in stacking. In our experiments, we do not apply stacking because we need to keep part of the data to train the combiner and our data set is small. Ting and Witten (1999) propose the MLR (Multiresponse Linear Regression) algorithm which combines linearly the outputs of base learners. In a mixture of experts architecture (MoE), classifiers are local and a separate gating network selects one of the local experts based on the input (Jacobs et al., 1991).

Additional to its effect on accuracy, each further classifier increases the space and the computational complexity of a multiclassifier system. A new classifier may also be using a costly representation which can be saved if the classifier is considered redundant. Therefore, methods have been proposed to choose a small subset from a large set of candidate classifiers. Since there are 2^L possible subsets of L classifiers one cannot try all possible subsets unless L is small, and various methods have been proposed to get a reasonable subset of size $m < L$ in reasonable time. In this work, we applied the fixed rules (min/max, median and sum) and the incremental methods used by Ulaş et al. (2009b) to select a subset of classifiers from a given set.

4.2.3 Our contributions

Most methods which check for brain diseases using pattern recognition techniques in the literature are based on extracting complex or more informative descriptors from the given MRI images and/or applying a single classification algorithm to these descriptors to have better classification accuracy. In this study, our main contribution is to choose a multi-classification approach. Instead of addressing the extraction of complex descriptors from

these images, we use the histograms of properly normalized intensity values as our feature vectors (Section 3.5.1). We then exploit the nature of the problem and try to combine the information of multiple ROIs to come up with better classification accuracy. In principle, one can also apply multi-classification using different type of classifiers on the same ROI (or the whole MRI) and come up with better ensemble accuracy. This approach has the problem of interpreting the results even for simple classifier combination algorithms. The aim of our study is to show that, even using a single base classifier, if we use (choose) informative ROIs, we have better ensemble accuracy and the results can be medically interpreted. Our second contribution is the analysis of correlation among different ROIs. We look at the problem from a different perspective. Instead of analyzing the correlation of the data, we analyze the correlation of the results of classification of each ROI. We have multiple classifiers trained using different ROIs and we apply a correlation analysis to measure the overlap of information contained in each ROI in terms of their classification outputs.

4.2.4 Experiment setup

In these experiments, we use the same 14 ROIs from Section 3.4 and the first 16 representations from Section 3.5.1. We use the following 13 classifiers⁵ to span a large collection of classification algorithms:

- 1–4) *knn*: k -nearest neighbor with $k = 1, 3, 5$ and 7 ;
- 5–9) *mlp*: Multilayer perceptron where with D inputs and K classes, the number of hidden units is taken as D (*ml1*), K (*ml2*), $(D + K)/2$ (*ml3*), $D + K$ (*ml4*), $2(D + K)$ (*ml5*);
- 10) *lnp*: Linear perceptron with softmax outputs trained by gradient-descent to minimize cross-entropy.
- 11) *c45*: The most widely-used C4.5 decision tree algorithm;
- 12–13) *svm*: Support vector machines (SVM) with a linear kernel (*svl*) and a radial (Gaussian) kernel (*svr*). We use the LIBSVM 2.82 library that implements pairwise SVMs (Chang and Lin, 2001) but we do not optimize the kernel parameters;

4.2.5 Ensemble methods applied

There are many sophisticated ensemble construction methods to apply for any given problem (Alpaydm, 2004; Kuncheva, 2004). In this study, our main aim is to show how the information stored in each ROI affects the outcome of the classification problem. Considering this aim, instead of opting to more complex pattern recognition techniques, we use fixed rules (sum, median, min/max) and an incremental algorithm (Ulaş et al., 2009b) called ICON which selects a subset of classifiers from a given set.

Let $P(C_k|\mathbf{x})$ denote the true posterior probability of class C_k given instance \mathbf{x} , and $P(C_k|\mathbf{x}, M)$ denote the posterior probability of class C_k estimated by classifier M . Estimated class for classifier M is calculated by $\text{argmax}_k P(C_k|\mathbf{x}, M)$. Suppose that the ensemble is denoted by E and we have L base classifiers $M_i, i = 1 \dots L$. The fixed rules used in the literature are given in Table 9 (Kittler et al., 1998):

⁵All classifiers were trained using the ISELL software by Olcay Taner Yildiz

Table 9: Fixed rules in classifier combination.

SUM:	$P(C_k \mathbf{x}, E) = \sum_{i=1}^L P(C_k \mathbf{x}, M_i)$
MAX:	$P(C_k \mathbf{x}, E) = \max_i P(C_k \mathbf{x}, M_i)$
MIN:	$P(C_k \mathbf{x}, E) = \min_i P(C_k \mathbf{x}, M_i)$
MED:	$P(C_k \mathbf{x}, E) = \text{median}\{P(C_k \mathbf{x}, M_1), \dots, P(C_k \mathbf{x}, M_L)\}$
PRO:	$P(C_k \mathbf{x}, E) = \prod_{i=1}^L P(C_k \mathbf{x}, M_i)$

Sum rule is known to work best in practice. Alkoot and Kittler (1999) investigate fixed rules and see that the sum and median rules are more robust to noise. Kuncheva (2002) discusses six fixed rules for classifier combination (minimum, maximum, sum, median, majority vote, and the oracle) on two-class problems with Gaussian and uniform error, under the assumption that the outputs are independent. She concludes that the min/max rule (they are the same for two-class problems) finds the best ensemble when uniform error is the case. For Gaussian distributed errors, the combination rules behave similarly. Tax et al. (2000) show that in multi-class problems product rule may be superior to sum rule when the independent data representation assumption is met. On the other hand, sum rule is more robust to noise. Especially in the cases when one of the classifiers is an outlier, product rule “acts as a veto” and this decreases the combination performance. We did not use product rule in our experiments because if one of the classifiers predict the posterior of one class as 0 (this is the case for *1nn*), the outcome is random. The analysis in these studies is under the assumption that classifiers are independent.

ICON ICON is an incremental algorithm which selects a subset of classifiers given a set. The pseudocode of the algorithm is given in Figure 4. We start with $E^{(0)} = \emptyset$. At iteration t of the algorithm, we have ensemble $E^{(t)}$ containing t models (representation and the classifier using that representation). Given the set of remaining $L - t$ candidate models, $M_k \notin E^{(t)}$, we have new candidate ensembles for iteration $t + 1$ as $S_k^{(t+1)} \equiv E^{(t)} \cup M_k, k = 1, \dots, L - t$. Among these, we choose the one that is preferred to all the other candidates and is also preferred to the current ensemble:

$$E^{(t+1)} \leftarrow S_j^{(t+1)} \quad \text{if} \quad S_j^{(t+1)} \prec S_k^{(t+1)}, \forall k \neq j \text{ and } S_j^{(t+1)} \prec E^{(t)}$$

$E_i \prec E_j$ denotes the binary relation of “preference” comparing two ensembles and holds if E_i is preferred to E_j according to the model selection criterion used (in this work, we use average accuracy). If none of the candidates is preferred to $E^{(t)}$, the algorithm stops and $E^{(t)}$ is taken as the final ensemble.

4.2.6 Divison of data set

In our previous experiments (Cheng et al., 2009b), we show the accuracies using leave-one-out (LOO) method. We define the whole data set as D . For each subject $\mathbf{x} \in D$, we train the classifier with the 123 subjects left-out ($D - \mathbf{x}$) and record the accuracy of \mathbf{x} using the classifier. We have 124 accuracies and we report the average of them. In this work, we use

```

1  function icon( $P$ )
2   $E^0 \leftarrow \emptyset$ 
3  for  $t = 0$  to  $L - 1$ 
4     $S_k^{(t+1)} \leftarrow E^{(t)} \cup M_k, \forall M_k \in P$  where  $M_k \notin E^{(t)}$ 
5    if  $\exists S_j^{(t+1)}$  such that  $S_j^{(t+1)} \prec S_k^{(t+1)}, \forall k \neq j$ 
6      and  $S_j^{(t+1)} \prec E^{(t)}$ 
7      then  $E^{(t+1)} \leftarrow S_j^{(t+1)}, t \leftarrow t + 1$ 
8      else break
9  end for
10 return  $E^{(t)}$ 

```

Figure 4: Pseudocode of the ICON algorithm.

a multiclassifier approach for analyzing classification of schizophrenia patients using MRI scans.

For each LOO fold ($D - \mathbf{x}$), we use 20-fold cross validation to select which classifier to add at each step when we use ICON. We use the 19 folds to train the base classifiers, the 20th fold to assess the classification accuracy to select the final ensemble. We do not need to have 20-fold cross validation when we use the fixed rules. We again record the classification accuracy of \mathbf{x} .

4.2.7 Results and Discussions

In this study, our main aim is to show the importance of ROIs in classifying brain disorders using MRI images. Although we went through exhaustive experiments using all the ROIs, classification algorithms and the representations constructed using different techniques, we provide here results using one representation and we want to stress the importance of using multiple ROIs so we skip combining multiple classifiers per ROI as summary and present results for combining multiple ROIs using a single classifier.

We present our multi-ROI results using the second representation (Histograms normalized as probability density functions) and in our correlation analysis we present the results using *svr* using this representation. Except for a small number of cases, the results are general and apply to all the other representations and classifiers used.

Classification experiments We can see in Table 10 the LOO accuracies of base classifiers, of the single best classifier (BEST), using sum, max (min) and median combination rules (MAX, SUM, MED), and using ICON⁶. In this table, we see that using all ROIs (for each classifier) is better than using the single best (9 out of 13) or using a subset selection algorithm. It has been shown by many researchers (Sharkey et al., 2000; Roli et al., 2001; Zhou et al., 2002; Caruana et al., 2004; Ruta and Gabrys, 2005; Yang et al., 2007; Martínez-Munoz et al., 2009; Ulaş et al., 2009b) that choosing a proper subset of given classifiers (in

⁶The base classifier results of this section and Section 4.1.2 are different because different tools were used to train the base classifiers.

our case ROIs) achieves better accuracy than using all classifiers or the single best classifier. We observe in this table that using a subset selection method does not yield better results than using all the classifiers and the single best classifier. The reason for this is that the validation set accuracies are not a good representative of the test set accuracy, so the subset is better than ALL on the validation set, but not on the test set. Another explanation is that, when the base classifiers are not correlated, combining all the base classifiers yields good combination accuracy, whereas the incremental algorithm may have been stuck in a local optima. Looking at Table 10 we can see that for nine classification algorithms, SUM is better than BEST. We did not include our results using the same ROI with different classifiers in this study because it's not easy to derive a medical interpretation, but our results show that using multiple classifiers but the same ROI has less effect on the overall classification accuracy than using the same classification algorithm and different ROIs. For none of the ROIs, SUM is better than the most accurate base classifier. This shows the importance of the ROIs. In this problem, the selection of the ROIs is more important than the selection of the classifier.

Table 10: Accuracies (%) of base classifiers and combination methods on the pdfs using all classifiers and ROIs.

	c45	1nn	3nn	5nn	7nn	lnp	ml1	ml2	ml3	ml4	ml5	svl	svr
<i>lamyg</i>	51.61	52.42	58.06	58.87	57.26	58.87	58.06	62.90	67.74	62.90	67.74	66.94	60.48
<i>ramyg</i>	50.00	62.10	63.71	63.71	60.48	58.06	53.23	61.29	60.48	55.65	58.06	54.03	67.74
<i>ldlpfc</i>	49.19	58.87	60.48	63.71	58.87	58.87	65.32	61.29	62.10	60.48	60.48	65.32	60.48
<i>rdlpfc</i>	56.45	49.19	50.00	49.19	47.58	50.81	62.10	52.42	56.45	58.06	47.58	56.45	44.35
<i>lec</i>	67.74	54.03	55.65	55.65	58.06	65.32	63.71	59.68	58.87	66.13	62.10	61.29	66.94
<i>rec</i>	47.58	56.45	62.10	63.71	64.52	61.29	57.26	57.26	66.94	67.74	65.32	56.45	65.32
<i>lhg</i>	55.65	49.19	55.65	54.84	55.65	55.65	61.29	60.48	63.71	62.90	58.87	56.45	53.23
<i>rhg</i>	50.81	48.39	52.42	51.61	51.61	54.84	58.06	58.06	54.84	54.03	54.84	42.74	61.29
<i>lhippo</i>	52.42	50.00	55.65	58.87	51.61	58.87	62.10	69.35	60.48	65.32	58.87	67.74	64.52
<i>rhippo</i>	50.81	59.68	61.29	61.29	61.29	54.03	58.87	60.48	62.10	53.23	55.65	58.87	51.61
<i>lstg</i>	44.35	49.19	50.81	49.19	55.65	59.68	58.06	63.71	64.52	67.74	59.68	61.29	59.68
<i>rstg</i>	56.45	52.42	56.45	57.26	60.48	60.48	60.48	65.32	57.26	57.26	62.10	62.90	63.71
<i>lthal</i>	53.23	58.87	58.06	58.87	54.84	50.81	54.84	53.23	55.65	58.06	57.26	55.65	58.87
<i>rthal</i>	51.61	59.68	55.65	50.81	57.26	62.10	54.03	56.45	53.23	58.87	57.26	62.90	62.90
Best	67.74	62.10	63.71	63.71	64.52	65.32	65.32	69.35	67.74	67.74	67.74	67.74	67.74
Icon	57.26	46.77	56.45	55.65	52.42	64.52	56.45	62.90	66.13	62.90	55.65	66.13	62.90
Sum	59.68	63.71	62.10	60.48	63.71	70.97	73.39	73.39	73.39	75.81	68.55	73.39	74.19
Max	62.10	49.19	53.23	64.52	68.55	61.29	66.94	66.13	74.19	70.16	67.74	67.74	65.32
Med	52.42	63.71	62.10	64.52	66.13	70.16	70.16	68.55	72.58	73.39	68.55	70.97	75.81

From the 14 ROIs we used, we chose the seven ROIs which have low discriminative power (*rdlpfc*, *lhg*, *rhg*, *rhippo*, *lstg*, *lthal* and *rthal*) and discard them in Table 11. This time, we can see that for 11 out of the 13 classifiers, SUM is better than BEST. Also, we see that when we select a subset of ROIs carefully, the accuracy of the ensemble also increases. We can see that, for only two out of the 13 classifiers, SUM accuracy of drops compared to the case where we include all ROIs. This shows us two things: first, some ROIs have less discriminative

power and discarding them leads to better ensemble performance, which proves our initial claim that ROIs are more descriptive in this context. Second, one can apply other complex combination techniques by choosing the right ROIs and classifiers to have better ensemble accuracy (which our incremental ICON algorithm fails to discover in this case). Table 11 shows the algorithms using only the selected ROIs and classification algorithms.

Table 11: Accuracies (%) of base classifiers and combination methods on the pdfs using selected ROIs.

	c45	1nn	3nn	5nn	7nn	lnp	ml1	ml2	ml3	ml4	ml5	svl	svr
<i>lamyg</i>	51.61	52.42	58.06	58.87	57.26	58.87	58.06	62.90	67.74	62.90	67.74	66.94	60.48
<i>ramyg</i>	50.00	62.10	63.71	63.71	60.48	58.06	53.23	61.29	60.48	55.65	58.06	54.03	67.74
<i>ldlpfc</i>	49.19	58.87	60.48	63.71	58.87	58.87	65.32	61.29	62.10	60.48	60.48	65.32	60.48
<i>lec</i>	67.74	54.03	55.65	55.65	58.06	65.32	63.71	59.68	58.87	66.13	62.10	61.29	66.94
<i>rec</i>	47.58	56.45	62.10	63.71	64.52	61.29	57.26	57.26	66.94	67.74	65.32	56.45	65.32
<i>lhippo</i>	52.42	50.00	55.65	58.87	51.61	58.87	62.10	69.35	60.48	65.32	58.87	67.74	64.52
<i>rstg</i>	56.45	52.42	56.45	57.26	60.48	60.48	60.48	65.32	57.26	57.26	62.10	62.90	63.71
Best	67.74	62.10	63.71	63.71	64.52	65.32	65.32	69.35	67.74	67.74	67.74	67.74	67.74
Icon	56.45	50.81	57.26	56.45	58.87	66.94	64.52	66.13	67.74	58.06	60.48	68.55	65.32
Sum	61.29	58.87	64.52	64.52	65.32	74.19	69.35	73.39	75.00	75.81	71.77	74.19	77.42
Max	62.10	50.00	60.48	66.13	66.13	65.32	67.74	67.74	74.19	66.94	69.35	70.97	67.74
Med	54.84	58.87	66.94	69.35	63.71	75.00	66.13	73.39	73.39	74.19	73.39	71.77	74.19

Correlation analysis Multiclassifier techniques are commonly used in machine learning and pattern recognition areas. We apply fixed rules and an incremental algorithm to form ensembles of classifiers in this study. For ensembles to have better accuracy, we need to have diverse base classifiers. Various measures of diversity exist (Kuncheva and Whitaker, 2003; Banfield et al., 2005; Ko et al., 2006), but in general, diversity can be defined as the classifiers responding differently to the same input. Clearly, being diverse is not enough for achieving better ensemble accuracy; one should also have accurate base classifiers. A special journal issue edited by Kuncheva (2005) contains papers on diversity measures and their role in ensemble construction. It is known and investigated by many researchers that if the experts in an ensemble are correlated, the ensemble accuracy decreases. Better than having no correlation is the case where experts are negatively correlated (Kuncheva et al., 2000), which in practice is not possible when number of base classifiers is large. For an ensemble to have better accuracy, we need uncorrelated and accurate base classifiers. Analysis of correlations is not novel and has been studied empirically (Petraikos et al., 2000; Goebel et al., 2002; Goebel and Yan, 2004) in the context of decision fusion, and theoretically by assuming different distributions of the outputs of base classifiers (Tumer and Ghosh, 1999; Kuncheva, 2002; Tumer and Ghosh, 2002; Fumera and Roli, 2005). All studies point out that in an ensemble, base classifiers are correlated and for better ensemble accuracy, we need to get rid of correlation. In this work, we apply the correlation computation proposed by Petraikos et al. (2000) to see how the classification algorithms and the ROIs are correlated. We do not present the results of correlations among different type of classifiers on the same

ROI because it is more difficult to interpret the medical meaning with such approach. We present the correlations among ROIs. In principle, one would use a validation set to compute the correlation of the classification algorithms and then form the ensemble according to this information (Goebel et al., 2002; Kuncheva and Whitaker, 2003; Goebel and Yan, 2004). Instead, what we do in this study is to calculate the correlation of pairwise ROIs on the test set to observe the information contained in different ROIs. Analysis of correlation among ROIs using the outputs of classifiers and correlation is novel as far as we know. Most studies in the literature which analyze the correlation between the ROIs are based on the correlation analysis of images.

The correlation between two classifiers is calculated as in Petrakos et al. (2000):

$$\rho_2 = \frac{2 \times N_{00}}{N_{01} + N_{10} + 2 \times N_{00}}$$

where N_{ij} are defined as in Table 12 and M_i and M_j are the two classifiers.

Table 12: Confusion matrix used to compute correlation.

	M_j correct	M_j wrong
M_i correct	N_{11}	N_{10}
M_i wrong	N_{01}	N_{00}

We can see an example correlation matrix in Table 13. This table shows important results considering the medical aspects of this study. Correlation analysis of ROIs using classification algorithms show us how the information contained in each ROI is related to one another. If the ROIs were highly correlated, this would have practical implications. For example, if two of the ROIs have highly correlated classifier outputs, this means that one does not have to trace one of the ROIs which allows us to save time. Analyzing these tables we see that this is not true. The largest correlation in Table 13 is 0.77 between *rhg* and *rstg*, and the second largest is between *lhg* and *lstg* and it's 0.70. With some small number of exceptions, this is what we observe from the 16×13 tables, the highest correlations are between *rstg* and *lstg*, *lhg* and *rhg*, and *rhg* and *rstg* and the other correlations are mostly below 0.6. These results suggest that the information contained in each ROI is different from the other. The information may be useful in discriminating patients with schizophrenia or not, but these results show us clearly that the information contained in each ROI has its own importance. Not all the ROIs may be needed for schizophrenia classification, but we may need to use all the ROIs for other studies of the brain. These results also suggest us that since the correlation between ROIs is smaller with regard to the correlations between classifiers, it is better to combine different ROIs, rather than different classifiers to get better ensemble accuracy, which also supports our results in the previous section. What we need is to analyze the contribution of each ROI, and select a careful subset of ROIs for better ensemble accuracy. In this study we use a very basic approach to do this but for further studies, different combination techniques may be used.

Table 13: Correlation analysis of ROIs on the pdfs using *svr*. High correlations ($\geq .70$) are shown in bold face.

<i>svr</i>	<i>lamyg</i>	<i>ramyg</i>	<i>ldlpfc</i>	<i>rdlpfc</i>	<i>lec</i>	<i>rec</i>	<i>lhg</i>	<i>rhg</i>	<i>lhippo</i>	<i>rhippo</i>	<i>lstg</i>	<i>rstg</i>	<i>lthal</i>	<i>rthal</i>
<i>lamyg</i>	1.00	0.49	0.45	0.54	0.42	0.57	0.56	0.45	0.45	0.48	0.46	0.38	0.38	0.34
<i>ramyg</i>	0.49	1.00	0.47	0.42	0.42	0.41	0.45	0.32	0.40	0.44	0.38	0.33	0.42	0.40
<i>ldlpfc</i>	0.45	0.47	1.00	0.59	0.44	0.33	0.54	0.52	0.34	0.53	0.48	0.51	0.50	0.48
<i>rdlpfc</i>	0.54	0.42	0.59	1.00	0.42	0.45	0.60	0.51	0.44	0.57	0.50	0.47	0.48	0.50
<i>lec</i>	0.42	0.42	0.44	0.42	1.00	0.36	0.36	0.36	0.28	0.32	0.29	0.33	0.33	0.37
<i>rec</i>	0.57	0.41	0.33	0.45	0.36	1.00	0.50	0.42	0.34	0.43	0.43	0.36	0.32	0.36
<i>lhg</i>	0.56	0.45	0.54	0.60	0.36	0.50	1.00	0.57	0.37	0.64	0.70	0.60	0.46	0.52
<i>rhg</i>	0.45	0.32	0.52	0.51	0.36	0.42	0.57	1.00	0.37	0.54	0.57	0.77	0.34	0.36
<i>lhippo</i>	0.45	0.40	0.34	0.44	0.28	0.34	0.37	0.37	1.00	0.46	0.47	0.47	0.42	0.40
<i>rhippo</i>	0.48	0.44	0.53	0.57	0.32	0.43	0.64	0.54	0.46	1.00	0.53	0.55	0.50	0.51
<i>lstg</i>	0.46	0.38	0.48	0.50	0.29	0.43	0.70	0.57	0.47	0.53	1.00	0.65	0.46	0.42
<i>rstg</i>	0.38	0.33	0.51	0.47	0.33	0.36	0.60	0.77	0.47	0.55	0.65	1.00	0.33	0.37
<i>lthal</i>	0.38	0.42	0.50	0.48	0.33	0.32	0.46	0.34	0.42	0.50	0.46	0.33	1.00	0.54
<i>rthal</i>	0.34	0.40	0.48	0.50	0.37	0.36	0.52	0.36	0.40	0.51	0.42	0.37	0.54	1.00

4.2.8 Discussion

In this work, we apply an ensemble-based approach to classify schizophrenic patients using MRI images. We use fixed rules and an incremental algorithm for creating ensembles of classifiers applied to ROIs, and compare their performances with the single best classifier, also providing a medical interpretation of the obtained results. Moreover, we apply correlation analysis on the classifiers' outputs to analyze the behavior of each ROI. Indeed, we observe that in brain schizophrenia classification, the most relevant element is not the choice of the classification algorithm, but the choice of the ROI. This choice is more important since choosing a proper subset of ROIs leads to better ensemble accuracy. The novel approach presented in this study consists of calculating the correlations of the ROIs using the classifiers. For this purpose, we calculate the output of the classifier on each test subject and compute the correlations between each pair of ROIs. Hence, we also observe that the correlations of the ROIs are (most of the time) below 0.6, that is, they are nearly uncorrelated. From the medical point of view, this confirms that there is different information encoded in each different part of the brain. Detecting and combining single ROI featuring schizophrenia in the context of multiple ROI classification would provide an interesting opportunity to characterize structural markers of the disease in the context of impaired structural network. This also proves that our multi-ROI approach is to be more successful than using single base classifiers. By selecting a suitable subset of ROIs (in our case we use a simple approach to discard ROIs), one can achieve better performances. This work has a notable potential for improvement in many directions. One can try various classifier combination schemes to increase accuracy, as well as applying other diversity-based techniques for classifier combination or information measures stored in each ROI. One may also apply distance-based classification techniques once we compute distances between pairs of subjects.

5 Work in progress

In this section we report the work in progress on the application of advanced pattern recognition techniques developed under the SIMBAD project on our medical problem. In particular

we focus on: i) schizophrenia detection by generative kernels, ii) schizophrenia detection by feature-based morphometry approach, and iii) schizophrenia detection from multiple modalities.

5.1 Schizophrenia detection by generative kernels

The main goal of this study is to verify whether a significant improvement in brains classification can be obtained by exploiting more sophisticated pattern recognition techniques, instead of investigating more complex morphological features from MRI data. Inspired by recent trends in machine learning and pattern recognition research, we explore a hybrid generative/discriminative approach using the Fisher Score Space (Jaakkola and Haussler, 1998) to represent our data and employing support vector machines (SVM) as classifiers (Shawe-Taylor and Cristianini, 2004). In particular, we based our framework on the so called *constellation* generative model which has been recently successfully applied for object recognition (Holub et al., 2008).

Generative and discriminative approaches are the two broad categories within which learning and classification methods fall: a generative approach will estimate the joint probability density function (pdf) of the data and class labels and will classify using the posterior probabilities obtained by Bayes' rule, while a discriminative approach will estimate a classification function directly. Generally, methods falling in the latter category obtain lower asymptotic errors. However, generative models remain popular for their ability to capture explicit data attributes and to incorporate missing features. Fisher kernels are designed to get the best of both worlds. Jaakkola and Haussler (1998), showed that it is possible to extract Fisher scores from a generative model and convert them into a Fisher kernel, which may be used for classification by a kernel method, such as the SVM.

In this work, we propose to classify between healthy and diseased subjects by using a hybrid generative-discriminative framework. The most known and applied class of hybrid methods relies on the so called generative kernels (Jaakkola and Haussler, 1998), to be employed with a Support Vector Machine: the basic idea is to employ a generative model to define feature vectors and project objects to the resulting feature space. Therefore, a meaningful similarity/distance measure is defined, leading to a kernel. In the following all the parts of the proposed approach are detailed, namely the generative and discriminative parts.

5.1.1 The generative part

For this part of the approach, the general idea is to choose a generative model capable of considering all the ROIs at the same time, together with the relations between them. To this end, we based our framework on the same concepts behind the constellation probabilistic model (Holub et al., 2008), which foresees the encoding of one object in terms of a fixed number N of object subparts M_j , and relative spatial relationships. In general, object subparts are represented by their appearance A_i , while the spatial relations are encoded by the *shape* X_i (i.e., the relative subparts positions) of the overall configuration.

Here, we apply this intuition to MRI brain scans, by looking at the ROIs as subparts with a definite spatial configuration within the cerebral volume. Ideally, we expect the content and configuration of the ROIs in each subject to be informative enough to recognize patients

from controls. Note that, while in Holub et al. (2008) the correspondences between subparts in different objects is considered missing data, this is not our case, since ROIs identities (e.g., amygdala or thalamus) and matching between subjects are pre-determined. Consequently, instead of having to evaluate multiple combinatorial hypotheses, our model has to evaluate a single combination, incorporating this prior information by design.

Ultimately, we define the following expression for the log-likelihood of a particular class of L subjects $\{O_i\}_{i=1}^L$ (assumed to be independent and identically distributed):

$$\log p(\{O_i\}) = \log \prod_{i=1}^L p(O_i) = \sum_{i=1}^L \log p(O_i) = \sum_{i=1}^L \log [p(A_i | \theta_a)p(X_i | \theta_s)], \quad (1)$$

where $\theta = \{\theta_a, \theta_s\}$ is the set of appearance and shape model parameters, respectively. Training is performed by estimating the Maximum Likelihood solution θ^{MLE} .

Based on preliminary experiments, we observed that the shape information was not significant and therefore we further simplified the model by using only the appearance model part. Note that, similar to Holub et al. (2008), the appearances A_i are represented by PCA components obtained from the intensity histograms of the ROIs, in order to reduce data dimensionality and highlight consistent variations in the distribution of MRI values. As in Holub et al. (2008), the appearance model $p(A_i | \theta_a)$ is then a Gaussian over the PCA components.

5.1.2 The discriminative part: the Fisher kernel

Fisher kernels (Jaakkola and Haussler, 1998) allow an effective general way of mixing generative and discriminative models for classification. In particular, the Fisher kernel approach measures the similarity between the objects by comparing them in the tangent space induced by the trained generative model, which is considered as a point in the Riemannian manifold defined by the chosen family of generative models. In practice, each object is represented by a feature vector, whose components are called Fisher scores, defined by the evaluation of the gradient of the model log-likelihood on the MLE solution. The dimensionality of this space equals the number of parameters. More in detail, given a probabilistic model, the Fisher scores $\phi(O_i)$ are defined through the following derivatives:

$$\phi(O_i) = \frac{\partial}{\partial \theta} \log p(O_i | \theta). \quad (2)$$

In particular, starting from Equation 1 and discarding the shape contribution, the Fisher score we obtain is

$$\phi(O_i) = \frac{\partial}{\partial \theta_a} \log p(A_i | \theta_a), \quad (3)$$

where θ_a represent the mean and variance parameters of the Gaussian appearance model. Following Jaakkola and Haussler (1998), we employ and train one generative model for both classes.

A kernel can be defined in various ways in the resulting space: the inner product was used in Jaakkola and Haussler (1998), while *RBF* and polynomial kernel have been proposed in Holub et al. (2008).

5.1.3 Preliminary results

In this part, we will show the effectiveness of the proposed approach using the above described data set. The goal of the experimental evaluation is twofold: on one hand, we want to provide evidence that using all ROIs at the same time is advantageous with respect to using individual ROIs. On the other hand, we want to show that a hybrid generative/discriminative model may outperform a simply discriminative approach. To this end, the proposed approach has been compared with two different techniques:

- Single ROI - SVM (*RBF*kernel): in this case the classification has been carried out using a single ROI. We used the same descriptor employed in the proposed approach, namely histograms whose dimensionality has been reduced with the PCA analysis.
- Multiple ROI - SVM (*RBF* kernel): in this case, the information coming from all the ROIs is merged. There are many methods for fusing information from different sources (see the huge Multi Classifier System Theory). In this case, we performed a feature level fusion, obtained by simply concatenating the vectors coming from different ROIs. Subsequently, the concatenated PCA-reduced vectors have been classified using again a SVM (*RBF* kernel).

Experiments were carried out in MATLAB and C, whereas accuracy figures for each test run were obtained through leave-one-out (LOO) cross-validation.

Results are proposed in Table 14. From the table, it is evident that most ROIs do not possess significant discriminative powers, but that using all of them at the same time achieves higher accuracy than the individual best ROI. Moreover it is evident that the hybrid generative discriminative approach outperforms the purely discriminative approach, confirming the findings obtained in other fields. Overall, results are suggestive, encouraging in a way, in fact they seem to support the main scientific claim that it is possible to identify schizophrenic patients from healthy people.

5.2 Schizophrenia detection by feature-based morphometry approach

In this work we exploit the use of sparse point-based local features for brain classification. Few and significant landmarks are detected and characterized by local region descriptors, estimated by the Scale Invariant Feature Transform (SIFT) operator (Lowe, 2004). In this fashion the brain abnormalities can be characterized in terms of intra-ROI local pattern which are not necessarily spatially coherent. The underlying hypotheses consist of relaxing the common constraint that morphological anomalies appear at the same voxel location for all the population. In order to deal with brains represented by an unordered set of features a local *kernel* of a SVM is employed. The proposed method is inspired by the *Bag-of-Words* (Cruska et al., 2004) paradigm which implicitly implements the feature matching within the SVM framework (Grauman and Darrell, 2007). Preliminary results have been carried out on the *dlpfc* region.

Table 14: Leave-one-out cross-validation accuracies (%). Our hybrid approach performs best. Taking all the ROIs performs better than considering them individually.

Method	ROI	% Accuracy
SVM Single ROIs	<i>lamyg</i>	70.97
	<i>ramyg</i>	58.87
	<i>ldlpfc</i>	68.55
	<i>rdlpfc</i>	47.58
	<i>lec</i>	59.68
	<i>rec</i>	58.06
	<i>lhg</i>	58.06
	<i>rhg</i>	60.48
	<i>lhippo</i>	62.10
	<i>rhippo</i>	50.00
	<i>lstg</i>	59.68
	<i>rstg</i>	56.45
	<i>lthal</i>	61.29
<i>rthal</i>	59.68	
SVM Multiple ROIs	<i>all</i>	77.42
Hybrid Approach Multiple ROIs	<i>all</i>	80.65

5.2.1 Proposed method

After landmarks extraction⁷, feature points are properly clusterized in order to obtain the *visual words*. After this step, each landmark can be easily assigned to the visual word associated to the closest cluster centroid (Cruska et al., 2004). Then, given the set of feature vectors coming from one brain, the Bag of Words representation is obtained by computing the histogram of visual word occurrences (Cruska et al., 2004). Moreover, for each cluster of feature we measure its discriminative *relevance* by counting the occurrences of patients and controls respectively. In particular the following weighting function is defined for each *visual world*:

$$w_i(n_p^{c_i}, n_c^{c_i}) = \begin{cases} 1.5 & \text{if } ||n_p^{c_i} - n_c^{c_i}|| \geq \Delta \\ 0.5 & \text{otherwise} \end{cases} \quad (4)$$

where c_i is the i^{th} centroid (i.e., the i^{th} visual world, $i = 1 \dots K$), $n_p^{c_i}$ and $n_c^{c_i}$ are the percentage of patients and controls in c_i , and Δ is a constant heuristically chosen. In this fashion clusters composed of a clear majority of population (i.e., patients or controls) are considered as more discriminative for the classification purpose.

Indeed, given two brains A and B , the kernel function is defined as:

$$k(h^A, h^B) = \sum_{i=1}^K w_i \cdot \min(h_i^A, h_i^B), \quad (5)$$

⁷We use the SIFT implementation available from <http://vision.ucla.edu/vedaldi>

where h_i^Θ denotes the count of the i^{th} bin of the histogram h^Θ ($\Theta \in \{A, B\}$), with K bins, and w_i are computed from Eq. (4). Such kernel is called weighted *histogram intersection* function and it is shown to be a valid kernel (Grauman and Darrell, 2007). Histograms are assumed to be normalized such that $\sum_{i=1}^n h_i = 1$.

Note that, as observed by Grauman and Darrell (2007), the proposed kernel implicitly encodes the point-to-point matching since corresponding features are likely to belong to the same histogram bin. In other words, when points share a bin, they are counted as matched, since the size of that bin indicates the farthest distance any two points in it could be from one another. Indeed, the histogram intersection function counts the number of feature matchings being intermediated by the visual vocabulary.

5.2.2 Preliminary results

Table 15 shows the performance of classification. Scores are obtained by LOO cross validation. The performances have been evaluated also without the weighting strategy. Moreover, in order to take into account the intra-class variability the whole data set has been stratified by sex and age. The Matlab⁸ version of the K -means algorithm has been used to clusterize the feature points. Here the number of clusters (i.e., visual words) has been fixed experimentally $K = 30$.

In general, a drastic improvement was observed when weights were applied. Satisfactory results are obtained for the classification of the whole data set. Moreover, performance increased when either only women or only senior people were considered. It is worth nothing that, a part of the case of men, better performances are observed from the left hemisphere. This conforms with previous studies in the field (Andreone et al., 2007).

Table 15: Classification rate (%). For each experiment the number of involved controls and patients are reported. The scores are reported with weights (w-score) and without weights (score) respectively. Results for both the hemispheres are reported (i.e., l. and r.).

Exp.	<i>n.Cont.</i>	<i>n.Pat.</i>	w-score	score
All l.	54	54	75.00	62.93
All r.	54	54	66.38	59.48
Wom l.	25	19	84.09	77.27
Wom r.	25	19	77.27	72.73
Men l.	29	35	60.00	44.62
Men r.	29	35	67.69	50.77
Senior l.	23	25	81.25	73.52
Senior r.	23	25	70.83	64.12
Junior l.	31	29	71.67	55.27
Junior r.	31	29	63.33	51.18

⁸<http://www.mathworks.com>

5.3 Schizophrenia detection from multiple modalities

So far, we have worked on single classifiers using each ROI, combining multiple classifiers using each ROI and combining multiple ROIs to increase accuracy of classification and also add interpretability to the results of these experiments which is crucial in medical point of view. In a further experiment setup, we will try to use other modalities of the same data, i.e., Diffusion Weighted Imaging (DWI) data. The DWI data has been acquired from the same set of subjects but the data consists of the whole brain. Initially, we have done a registration of the DWI data with the morphological data and with this method, we have the ROIs for the DWI data. Unfortunately, the registration was not successful for all the 124 subjects, so in these experiments in the future, we plan to use the 114 subjects which could be registered properly. In Tables 16, 17 and 18 we see the initial single base classifier and mean combiner accuracies for the morphological data (MPH), DWI data (DWI) and the concatenation of both (CON) respectively.

Table 16: Single base classifier results and mean combiner accuracy (%) for MPH.

MPH	<i>inn</i>	<i>svl</i>	<i>svr</i>	meanc
<i>lamyg</i>	57.02	56.14	64.91	63.16
<i>ramyg</i>	60.53	52.63	64.91	63.16
<i>ldlpfc</i>	57.02	59.65	71.93	64.91
<i>rdlpfc</i>	46.49	61.40	57.02	63.16
<i>lec</i>	48.25	48.25	57.02	50.00
<i>rec</i>	61.40	63.16	61.40	65.79
<i>lhg</i>	53.51	47.37	55.26	52.63
<i>rhg</i>	57.02	50.00	53.51	51.75
<i>lhippo</i>	58.77	56.14	70.18	64.04
<i>rhippo</i>	57.89	50.88	50.88	55.26
<i>lstg</i>	42.98	57.02	62.28	59.65
<i>rstg</i>	57.02	62.28	57.02	58.77
<i>lthal</i>	63.16	62.28	55.26	60.53
<i>rthal</i>	53.51	61.40	53.51	58.77

6 Cooperation with other partners

In the context of SIMBAD project, we have started some collaborations with other SIMBAD partners. In this section, we highlight some of the collaborative work in progress. From some of the collaborations we already have obtained some preliminary results, some others are still under investigation. In particular, we report some experiments obtained in collaboration with Lisbon group (Section 6.1) and Delft group (Section 6.2). Finally, we report the work in progress with other partners in Section 6.3.

Table 17: Single base classifier results and mean combiner accuracy (%) for DWI.

DWI	<i>1nn</i>	<i>svl</i>	<i>svr</i>	meanc
<i>lamyg</i>	54.39	64.04	48.25	61.40
<i>ramyg</i>	49.12	50.88	42.98	50.00
<i>ldlpfc</i>	50.00	49.12	53.51	45.61
<i>rdlpfc</i>	47.37	52.63	52.63	55.26
<i>lec</i>	47.37	53.51	60.53	56.14
<i>rec</i>	42.11	48.25	43.86	48.25
<i>lhg</i>	42.98	54.39	48.25	50.00
<i>rhg</i>	56.14	47.37	57.89	54.39
<i>lhippo</i>	50.88	46.49	50.00	48.25
<i>rhippo</i>	54.39	64.04	62.28	64.91
<i>lstg</i>	54.39	54.39	53.51	54.39
<i>rstg</i>	50.00	57.89	49.12	54.39
<i>lthal</i>	50.88	55.26	49.12	49.12
<i>rthal</i>	45.61	48.25	43.86	43.86

Table 18: Single base classifier results and mean combiner accuracy (%) for the concatenation of MPH and DWI.

CON	<i>1nn</i>	<i>svl</i>	<i>svr</i>	meanc
<i>lamyg</i>	60.53	57.89	63.16	62.28
<i>ramyg</i>	53.51	52.63	55.26	55.26
<i>ldlpfc</i>	54.39	57.89	65.79	60.53
<i>rdlpfc</i>	46.49	56.14	48.25	54.39
<i>lec</i>	56.14	57.02	61.40	59.65
<i>rec</i>	54.39	62.28	66.67	60.53
<i>lhg</i>	40.35	48.25	47.37	47.37
<i>rhg</i>	55.26	55.26	57.89	58.77
<i>lhippo</i>	54.39	60.53	65.79	63.16
<i>rhippo</i>	60.53	57.02	71.93	66.67
<i>lstg</i>	50.88	62.28	59.65	62.28
<i>rstg</i>	56.14	55.26	56.14	55.26
<i>lthal</i>	48.25	56.14	59.65	61.40
<i>rthal</i>	51.75	64.91	55.26	62.28

6.1 Schizophrenia detection by Information Theoretic Kernels

The cooperation with Lisbon unit has been carried out by applying the Information Theoretic kernels, a family of kernels defined on probability measures recently introduced by the Lisbon group, to the ROI-based histograms. Before showing the results, let us briefly summarize such kernels.

Kernels on probability measures have been shown very effective in classification problems involving text, images, and other types of data (Cuturi et al., 2005; Hein and Bousquet, 2005; Jebara et al., 2004). Given two probability measures p_1 and p_2 , representing two objects, several information theoretic kernels (ITKs) can be defined as Martins et al. (2009). The Jensen-Shannon (JS) kernel is defined as

$$k^{\text{JS}}(p_1, p_2) = \ln(2) - JS(p_1, p_2), \quad (6)$$

with $JS(p_1, p_2)$ being the Jensen-Shannon divergence

$$JS(p_1, p_2) = H\left(\frac{p_1 + p_2}{2}\right) - \frac{H(p_1) + H(p_2)}{2}, \quad (7)$$

where $H(p)$ is the usual Shannon entropy.

The Jensen-Tsallis (JT) kernel is given by

$$k_q^{\text{JT}}(p_1, p_2) = \ln_q(2) - T_q(p_1, p_2), \quad (8)$$

where $\ln_q(x) = (x^{1-q} - 1)/(1 - q)$ is the q -logarithm,

$$T_q(p_1, p_2) = S_q\left(\frac{p_1 + p_2}{2}\right) - \frac{S_q(p_1) + S_q(p_2)}{2^q} \quad (9)$$

is the Jensen-Tsallis q -difference, and $S_q(r)$ is the Jensen-Tsallis entropy, defined, for a multinomial $r = (r_1, \dots, r_L)$, with $r_i \geq 0$ and $\sum_i r_i = 1$, as

$$S_q(r_1, \dots, r_L) = \frac{1}{q-1} \left(1 - \sum_{i=1}^L r_i^q \right).$$

Martins et al. (2009) also defined versions of these kernels applicable to unnormalized measures. Let $\mu_1 = \omega_1 p_1$ and $\mu_2 = \omega_2 p_2$ be two unnormalized measures, where p_1 and p_2 are the normalized counterparts (probability measures), and ω_1 and ω_2 arbitrary positive real numbers (weights). The weighted versions of the JT kernels are defined as follows:

- The weighted JT kernel (WJT-v1) is given by

$$k_q^A(\mu_1, \mu_2) = S_q(\pi) - T_q^\pi(p_1, p_2), \quad (10)$$

where $\pi = (\pi_1, \pi_2) = \left(\frac{\omega_1}{\omega_1 + \omega_2}, \frac{\omega_2}{\omega_1 + \omega_2} \right)$ and

$$T_q^\pi(p_1, p_2) = S_q(\pi_1 p_1 + \pi_2 p_2) - (\pi_1^q S_q(p_1) + \pi_2^q S_q(p_2)).$$

- The weighted JT kernel (WJT-v1) is defined as

$$k_q^B(\mu_1, \mu_2) = (S_q(\pi) - T_q^\pi(p_1, p_2)) (\omega_1 + \omega_2)^q. \quad (11)$$

6.1.1 Proposed Approach

The approach proposed in this collaborative work consists of defining a kernel between two histograms as one of the ITKs presented above. Currently, we have the results of using *knn* as the classifier (using the kernels as similarities) but as a further study we plan to apply generative embeddings and support vectors for better classification accuracy.

We used two histogram representations of the ROIs: the histogram of the intensity values (non parametric estimate of the intensity values), and the histogram of some geometric features computed on the surface (in the spirit of Bag of words approach). In this latter case we used Shape Index, curvature max and curvature min. The IT kernels were computed varying q in the range [0:0.05:1]. We used Jensen Shannon, Jensen Tsallis, WJS v1 and WJS v2. We computed the LOO accuracies of such kernels employing the nearest neighbor rule and the *knn* rule (k chosen with cross validation). We computed the accuracies for all possible values of q , picking up the best result. We also have the results of linear and rbf kernels as base results. For rbf, σ was varied using [0.001, 0.005, 0.01, 0.05, 0.1, 0.25, 0.5, 0.6, 1, 1.5, 2, 5, 10, 15, 20, 50, 100], computing the accuracies for all of them and again picking up the best). We can see the results in Tables 19, 20 and 21. The results in italic are the accuracies greater than 60 per cent. The results in bold show the best results.

Table 19: *knn* (%) accuracies (k cross validated) of IT kernels and accuracies of RBF and linear kernels computed using intensity histograms.

	Linear	JS	JT	WJT-v1	WJT-v2	RBF
<i>lamyg</i>	53.23	55.65	58.87	56.45	58.06	57.26
<i>ramyg</i>	52.42	51.61	54.03	54.84	51.61	52.42
<i>ldlpfc</i>	52.42	51.61	54.03	58.87	58.87	58.06
<i>rdlpfc</i>	51.61	54.84	54.84	51.61	51.61	54.03
<i>lec</i>	52.42	51.61	52.42	59.68	58.06	51.61
<i>rec</i>	52.42	55.65	57.26	56.45	59.68	54.03
<i>lhg</i>	54.84	51.61	<i>62.10</i>	51.61	55.65	55.65
<i>rhg</i>	51.61	51.61	55.65	54.03	58.06	58.06
<i>lhippo</i>	55.65	51.61	53.23	54.03	54.03	54.84
<i>rhippo</i>	52.42	51.61	<i>60.48</i>	58.87	59.68	51.61
<i>lstg</i>	52.42	51.61	54.03	65.32	54.84	53.23
<i>rstg</i>	51.61	51.61	54.84	56.45	54.03	56.45
<i>lthal</i>	54.84	51.61	52.42	55.65	51.61	53.23
<i>rthal</i>	52.42	56.45	57.26	52.42	58.87	56.45

6.2 Schizophrenia detection on the Dissimilarity space

In this part, we report some results obtained by approaching the detection of schizophrenia by means of the dissimilarity-based classification techniques developed by the TUDelft unit. Instead of working with features directly, we consider the dissimilarity matrices computed

Table 20: k nn (%) accuracies (k cross validated) of IT kernels and accuracies of RBF and linear kernels computed using shape index histograms.

	Linear	JS	JT	WJT-v1	WJT-v2	RBF
<i>lamyg</i>	58.06	54.03	55.65	58.87	<i>60.48</i>	<i>61.29</i>
<i>ramyg</i>	52.42	55.65	<i>61.29</i>	58.87	57.26	<i>62.10</i>
<i>ldlpfc</i>	51.61	52.42	54.03	51.61	55.65	<i>61.29</i>
<i>rdlpfc</i>	51.61	52.42	53.23	54.03	53.23	58.06
<i>lec</i>	51.61	56.45	59.68	<i>62.10</i>	<i>61.29</i>	56.45
<i>rec</i>	51.61	53.23	56.45	56.45	58.06	53.23
<i>lhg</i>	51.61	57.26	<i>60.48</i>	56.45	56.45	58.87
<i>rhg</i>	51.61	58.06	<i>60.48</i>	54.84	<i>61.29</i>	56.45
<i>lhippo</i>	52.42	58.87	66.13	57.26	<i>63.71</i>	<i>61.29</i>
<i>rhippo</i>	54.84	53.23	56.45	58.87	54.03	54.03
<i>lstg</i>	51.61	55.65	<i>61.29</i>	54.03	59.68	58.87
<i>rstg</i>	51.61	56.45	58.06	56.45	<i>63.71</i>	55.65
<i>lthal</i>	51.61	51.61	56.45	51.61	51.61	51.61
<i>rthal</i>	51.61	51.61	<i>62.10</i>	<i>62.10</i>	<i>62.90</i>	58.06

Table 21: k nn (%) accuracies (k cross validated) of IT kernels and accuracies of RBF and linear kernels computed using min-max curvature histograms.

	Linear	JS	JT	WJT-v1	WJT-v2	RBF
<i>lamyg</i>	55.65	55.65	59.68	58.87	58.06	<i>61.29</i>
<i>ramyg</i>	56.45	55.65	59.68	<i>60.48</i>	<i>62.10</i>	58.06
<i>ldlpfc</i>	54.84	54.03	57.26	59.68	58.87	54.84
<i>rdlpfc</i>	54.84	57.26	59.68	<i>63.71</i>	<i>60.48</i>	54.03
<i>lec</i>	55.65	57.26	64.52	64.52	64.52	53.23
<i>rec</i>	56.45	54.03	56.45	55.65	55.65	54.03
<i>lhg</i>	53.23	51.61	57.26	56.45	58.06	60.48
<i>rhg</i>	57.26	52.42	54.03	55.65	54.84	55.65
<i>lhippo</i>	56.45	54.03	57.26	55.65	57.26	56.45
<i>rhippo</i>	60.48	58.06	<i>61.29</i>	<i>61.29</i>	59.68	55.65
<i>lstg</i>	58.06	55.65	57.26	58.06	58.06	51.61
<i>rstg</i>	56.45	58.06	<i>62.10</i>	<i>62.90</i>	<i>62.90</i>	51.61
<i>lthal</i>	54.03	51.61	58.06	55.65	58.06	51.61
<i>rthal</i>	59.68	<i>60.48</i>	<i>62.10</i>	<i>61.29</i>	<i>61.29</i>	54.84

as described in Section 3.6. This offers the analysis of a different way to express application background knowledge as compared to features. In a second step the dissimilarity representation is transformed into a vector space in which traditional statistical classifiers can be employed. Unlike the related kernel approach, whose application is often restrained by technicalities like fulfilling Mercer’s condition, basically any dissimilarity measure can be used.

As mentioned above, all in all there are 14 ROIs and 13 different histogram dissimilarity measures, which yields a total of 182 dissimilarity matrices. In addition to these, we propose to merge the different dissimilarity matrices into one overall dissimilarity matrix potentially exploiting complementary information useful to improve the classification accuracy (Ulař et al., 2010).

6.2.1 Dissimilarity Space

There are several ways to transform a $n \times n$ dissimilarity matrix D with elements $D(p, q)$ (the dissimilarity between objects p and q) into a vector space with objects represented by vectors $X = \{x'_1, \dots, x'_p, \dots, x'_q, \dots, x'_n\}$ (Pekalska and Duin, 2005). Classical scaling (for proper Euclidean dissimilarities) and pseudo-Euclidean embedding (for arbitrary symmetric dissimilarities) yield vector spaces in which vector distances 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 an 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 distances 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 constructed 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 number of objects, which is 124 in our case. So many classifiers will need dimension reduction techniques or regularization to work properly in this space. Here, we used the linear support vector machine to avoid this.

Combined dissimilarity spaces can be constructed by combining dissimilarity representation. A simple and often effective way is using a (weighted) average of the given dissimilarity measures:

$$D_{combined} = \sum \alpha_i D_i. \quad (12)$$

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, e.g. its maximum or average dissimilarity.

6.2.2 Experiments

We considered all 182 dissimilarity matrices. For each test we evaluated the leave-one-out error. The dissimilarity spaces have been built in a transductive way by using all available subjects for representation (of course labels are ignored in this phase). Two classifiers are

Table 22: ROI-based classification

ROI	Standard Classifier	Dissimilarity-based
<i>lamyg</i>	64.50 (his-l1)	68.50 (hist-euclid)
<i>ramyg</i>	62.10 (pdf-euclid)	67.70 (pdf-euclid)
<i>ldlpfc</i>	64.50 (pdf-kl)	76.60 (pdf-kl)
<i>rdlpfc</i>	66.90 (his-intersect)	68.50 (pdf-js)
<i>lec</i>	59.70 (pdf-emd)	66.90 (pdf-js)
<i>rec</i>	64.50 (his-chi)	66.10 (his-intersec)
<i>lhg</i>	59.70 (his-chi)	69.10 (pdf-kl-orig)
<i>rhg</i>	61.30 (his-chi)	66.10 (pdf-kl)
<i>hippo</i>	55.60 (his-diffusion)	71.80 (pdf-js)
<i>rhippo</i>	63.70 (his-euclid)	60.50 (his-intersec)
<i>lstg</i>	57.30 (his-intersect)	64.50 (pdf-js)
<i>rstg</i>	58.90 (his-chi)	64.50 (pdf-l1)
<i>lthal</i>	64.50 (pdf-l1)	63.70 (his-diffusion)
<i>rthal</i>	59.70 (pdf-euclid)	67.70 (his-emd)

considered, the 1-Nearest Neighbor (NN) rule on the original dissimilarities (called the *Standard Classifier*) and the linear SVM in dissimilarity space, called the *Dissimilarity based classifier*.

We designed two experiments: i) ROI-based classification, and ii) Multi-ROI classification.

ROI-based classification We evaluate the classification errors for each of the original dissimilarity matrices. Table 22 summarizes the results. For each ROI the best performance is reported with respect to various dissimilarity measures. First column reports the accuracy estimates for NN using the original dissimilarities (standard approach). Second 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. For ROIs like *ldlpfc* and *hippo* the accuracy is greater than 70%, while in the standard approach results are less stable the accuracy is always less than 66.90%.

Multi-ROI classification In this experiment a Multi-ROI approach is adopted in order to use all ROIs at the same time. All the dissimilarity matrices for each ROI are combined using for α_i in (12) the reciprocal of the average dissimilarity value in D_i (cf. Lee et al. (2010)). Table 23 reports the results. Also in this case the classification on the dissimilarity space clearly outperforms the standard approach. Moreover, the Multi-ROI approach brings a drastic improvement by confirming the complementary information enclosed onto the different brain subparts. In most of the cases, the results from the averaged similarity matrices are better than the respective best single-ROI results. The last row reports the accuracy estimates computed on the overall dissimilarity matrix (for all the measures and ROIs) for both standard approach and dissimilarity-based approach respectively. This yields the best results so far (i.e., 79 %).

Table 23: Multi-ROI classification.

Measure	Standard Classifier	Dissimilarity-based
<i>his-euclid</i>	62.10	74.20
<i>his-l1</i>	62.10	77.40
<i>his-intersect</i>	71.00	63.70
<i>his-diffusion</i>	62.10	77.40
<i>his-chi</i>	60.50	77.40
<i>his-emd</i>	46.80	58.90
<i>pdf-euclid</i>	58.90	70.20
<i>pdf-l1</i>	58.90	72.60
<i>pdf-emd</i>	60.50	71.00
<i>pdf-bc</i>	63.70	65.30
<i>pdf-kl</i>	64.50	66.10
<i>pdf-kl-orig</i>	63.70	68.50
<i>pdf-js</i>	63.70	67.70
<i>average</i>	61.30	79.00

6.3 Work in progress

In order to improve the collaboration with other SIMBAD partners, we have organized a workshop in Zurich in cooperation with our Zurich partner. The aim of the workshop was to introduce more in detail the work packages WP6 and WP7 and to coordinate the efforts among the partners in exploiting possible new ideas and directions. The medical doctors of both workpackages contributed to the discussion by highlighting the medical aspects of the involved applications. During the workshop we standardized the data set format so that each partner could use and exchange the data from WP6 and WP7 easily. We shared our implementations of histograms dissimilarities so that histograms extracted from both WPs could be used to calculate histogram dissimilarities. Moreover, we discussed possible other similarity measures. The workshop was the opportunity to finalize the collaboration already started with Delft and Lisbon (already reported in Section 6.2 and 6.1, respectively) and to envisage new collaborations with other partners. More in details, together with Venice group we have identified two lines of research: i) graph-based classification and ii) feature-based morphometry classification. Graph-based classification implies the extraction of graph structure from the skeletons of the brain ROIs. This will exploit new morphological representations that will be open to the possibility of applying the graph-based classification methods developed in SIMBAD to our medical application. Also partners from Delft and York are interested in this approach. The feature-based morphometry approach aims at improving the use of local kernels already exploited by our unit in Section 5.2. The overall idea is to propose new local descriptors by possibly taking into account also the relationship between points (i.e. shape context (Belongie et al., 2002)). The York group is utilizing the dissimilarity matrices we extracted from pairwise non-rigid brain registrations. This provides dissimilarity measures without the intermediation of feature representations such as the histograms. We have also identified registration-based distance measures (Klein et al.,

2010) and provided such data in SIMBAD web site. This representation was also useful for the Delft group.

7 Final discussion and future work

In this report, we have created a sort of protocol (actually, a set of techniques) for the detection of schizophrenia from MRI images using ROIs. The problem is challenging in that currently such type of diagnosis is not carried out using MRI images. As a consequence, the first problem we face was to decide where to look in the brain volume. Both our feature-based and dissimilarity-based experiments have shown that the information encoded in different parts (ROIs) of the brain is complementary, and by combining these in proper ways, we achieve better accuracies.

In our preliminary experiments, we used the scaled histograms of intensity values and used these as features for feature-based classification experiments. We used single base classifiers as well as combinations of classifiers for each ROI, and we also combine ROIs. The preliminary results are promising in the sense that, with suitable combination mechanisms, we could reach higher accuracies. In this research direction, we applied ensemble techniques to ROIs and classifiers. We also applied a hybrid generative-discriminative approach which got promising results, as well as a feature-based morphometry approach. We computed several dissimilarity matrices and provided them to the SIMBAD partners for their work. In particular, we used several histogram similarity measures and distances between pdfs, and we have also used brain registration techniques to come up with such similarity matrices.

We have established several collaborations with the partners, some of which have already borne fruit and others which are in progress. We have used the dissimilarity combination techniques developed by the Delft group, which lead to the submission of a joint paper. We have applied the IT kernels developed by the Lisbon group to exploit the capability of these kernels. With the Venice group, we have started to work on graph representations to apply graph similarity measures, and also we are still collaborating in the context of the feature-based morphometry. Together with the York group, we are working on dissimilarities based on deformable registration of brains. The workshop in Zurich was a success on sharing knowledge and experience in the field, and opened new directions for this issue. We have also standardized intermediate representations to easily communicate data between the groups.

As future work, we plan to use multiple modalities to come up with better accuracies. More specifically, our initial results using morphological and diffusion weighted imaging techniques showed that there is much to improve in this direction. Moreover, we will try to use possible 3D shape descriptors, in particular, a surface-based morphometry approach.

Finally, we have applied several methods and established several collaborations, also drawing a roadmap to attack this problem. Initial results are promising and with collaborations and new lines of research mentioned in this report, we believe we will get further improvements and insights in the problem of the classification of schizophrenia by MRI.

References

- Agarwal, N., Rambaldelli, G., Perlino, C., Dusi, N., Kitis, O., Bellani, M., Cerini, R., Isola, M., Versace, A., Balestrieri, M., Gasparini, A., Mucelli, R. P., Tansella, M., Brambilla, P., 2008. Microstructural thalamic changes in schizophrenia: a combined anatomic and diffusion weighted magnetic resonance imaging study. *Journal of Psychiatry & Neuroscience* 33 (5), 440–448.

- Alkoot, F. M., Kittler, J., 1999. Experimental evaluation of expert fusion strategies. *Pattern Recognition Letters* 20 (11-13), 1361–1369.
- Alpaydm, E., 2004. Introduction to machine learning. The MIT Press.
- Amaddeo, F., Tansella, M., 2009. Information systems for mental health. *Epidemiologia e Psichiatria Sociale* 18 (1), 1–4.
- American Psychiatric Association, 1994. Diagnostic and statistical manual of mental disorders, DSM-IV. Washington DC, 4th Edition.
- Andreone, N., Tansella, M., Cerini, R., Versace, A., Rambaldelli, G., Perlini, C., Dusi, N., Pelizza, L., Balestrieri, M., Barbui, C., Nose, M., Gasparini, A., Brambilla, P., 2007. Cortical white-matter microstructure in schizophrenia. diffusion imaging study. *British Journal of Psychiatry* 191, 113–119.
- Ashburner, J., Friston, K. J., 2000. Voxel-based morphometrythe methods. *NeuroImage* 11 (6), 805–821.
- Baiano, M., Perlini, C., Rambaldelli, G., Cerini, R., Dusi, N., Bellani, M., Spezzapria, G., Versace, A., Balestrieri, M., Mucelli, R. P., Tansella, M., Brambilla, P., 2008. Decreased entorhinal cortex volumes in schizophrenia. *Schizophrenia Research* 102 (1–3), 171–180.
- Banfield, R. E., Hall, L. O., Bowyer, K. W., Kegelmeyer, W. P., 2005. Ensemble diversity measures and their application to thinning. *Information Fusion* 6 (1), 49–62.
- Bellani, M., Brambilla, P., 2008. The use and meaning of the continuous performance test in schizophrenia. *Epidemiologia e Psichiatria Sociale* 17 (3), 188–191.
- Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (4), 509–522.
- Bernasconi, N., Bernasconi, A., Andermann, F., Dubeau, F., Feindel, W., Reutens, D., 1999. Entorhinal cortex in temporal lobe epilepsy: a quantitative mri study. *Neurology* 52 (9), 1870–1876.
- Brambilla, P., Harenski, K., Nicoletti, M., Sassi, R. B., Mallinger, A. G., Frank, E., Kupfer, D. J., Keshavan, M. S., Soares, J. C., 2003. Mri investigation of temporal lobe structures in bipolar patients. *Journal of Psychiatric Research* 37 (4), 287–295.
- Breiman, L., 1996. Bagging predictors. *Machine Learning* 24 (2), 123–140.
- Browne, A., Jakary, A., Vinogradov, S., Fu, Y., Deicken, R., 2008. Automatic relevance determination for identifying thalamic regions implicated in schizophrenia. *IEEE Transactions on Neural Networks* 19 (6), 1101–1107.
- Caruana, R., Niculescu-Mizil, A., Crew, G., Ksikes, A., 2004. Ensemble selection from libraries of models. In: *Proceedings of the International Conference on Machine Learning, ICML '04*. pp. 137–144.
- Cha, S.-H., Srihari, S. N., 2002. On measuring the distance between histograms. *Pattern Recognition* 35 (6), 1355–1370.
- Chang, C. C., Lin, C. J., 2001. LIBSVM: a library for support vector machines.
URL <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Cheng, D. S., Bicego, M., Castellani, U., Cerruti, S., Bellani, M., Rambaldelli, G., Atzori, M., Brambilla, P., Murino, V., 2009a. Schizophrenia classification using regions of interest in brain MRI. Tech. rep., Dipartimento di Informatica, University of Verona, Italy.
- Cheng, D. S., Bicego, M., Castellani, U., Cerruti, S., Bellani, M., Rambaldelli, G., Atzori, M., Brambilla, P., Murino, V., 2009b. Schizophrenia classification using regions of interest in brain mri. In: *Proceedings of Intelligent Data Analysis in Biomedicine and Pharmacology, IDAMAP '09*. pp. 47–52.
- Cheng, D. S., Bicego, M., Castellani, U., Cristani, M., Cerruti, S., Bellani, M., Rambaldelli, G., Atzori, M., Brambilla, P., Murino, V., 2009c. A hybrid generative/discriminative method for classification of regions of interest in schizophrenia brain mri. In: *Proceedings of workshop on Probabilistic Models for Medical Image Analysis, MICCAI '09*. pp. 174–184.
- Cruska, G., Dance, C. R., Fan, L., Willamowski, J., Bray, C., 2004. Visual categorization with bags of keypoints. In: *ECCV Workshop on Statistical Learning in Computer Vision*. pp. 1–22.
- Cuturi, M., Fukumizu, K., Vert, J.-P., 2005. Nonextensive information theoretic kernels on measures. *Journal of Machine Learning Research* 6, 1169–1198.
- Davatzikos, C., 2004. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage* 23 (1), 17–20.
- Dubb, A., Avants, B. B., Gur, R., Gee, J. C., 2002. Shape characterization of the corpus callosum in schizophrenia using template deformation. In: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II, MICCAI '02*. pp. 381–388.
- Duda, R. O., Hart, P. E., Stork, D. G., 2000. Pattern classification, 2nd Edition. Wiley-Intersciences.

- Duin, R. P., 2005. Prtools, a matlab toolbox for pattern recognition version 4.0.14. available at <http://www.prttools.org/ls.org/>.
URL <http://www.prttools.org/>
- Edelstein, W. A., Bottomley, P. A., Pfeifer, L. M., 1984. A signal-to-noise calibration procedure for NMR imaging systems. *Medical Physics* 11, 180–185.
- Fan, Y., Shen, D., Gur, R. C., Gur, R. E., Davatzikos, C., 2007. Compare: Classification of morphological patterns using adaptive regional elements. *IEEE Transactions on Medical Imaging* 26 (1), 93–105.
- Freund, Y., Schapire, R. E., 1996. Experiments with a new boosting algorithm. In: *Proceedings of the International Conference on Machine Learning, ICML '96*. pp. 148–156.
- Fumera, G., Roli, F., 2005. A theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE Transactions on Pattern Analysis Machine Intelligence* 27 (6), 942–956.
- Gerig, G., Styner, M., Shenton, M. E., Lieberman, J. A., 2001. Shape versus size: Improved understanding of the morphology of brain structures. In: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI '01*. pp. 24–32.
- Giuliani, N. R., Calhouna, V. D., Pearlson, G. D., Francis, A., Buchanan, R. W., 2005. Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. *Schizophrenia Research* 74 (2–3), 135–147.
- Goebel, K. F., Yan, W., 2004. Choosing classifiers for decision fusion. In: Svensson, P., Schubert, J. (Eds.), *Proceedings of the International Conference on Information Fusion, Fusion '04*. Vol. I. pp. 563–568.
- Goebel, K. F., Yan, W., Cheetham, W., 2002. A method to calculate classifier correlation for decision fusion. In: *Proceedings of Information, Decision and Control, IDC '02*. pp. 135–140.
- Grauman, K., Darrell, T., 2007. The pyramid match kernel: Efficient learning with sets of features. *Journal of Machine Learning Research* 8 (2), 725–760.
- Hein, M., Bousquet, O., 2005. Hilbertian metrics and positive definite kernels on probability measures. In: *Proceedings of AISTATS 2005*. pp. 136–143.
- Holub, A. D., Welling, M., Perona, P., 2008. Hybrid generative-discriminative visual categorization. *Int. J. Comput. Vision* 77 (1–3), 239–258.
- Jaakkola, T., Haussler, D., 1998. Exploiting generative models in discriminative classifiers. In: *Advances in Neural Information Processing Systems 11*. MIT Press, pp. 487–493.
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J., Hinton, G. E., 1991. Adaptive mixtures of local experts. *Neural Computation* 3, 79–87.
- Jager, F., Hornegger, J., January 2009. Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging. *IEEE Transactions on Medical Imaging* 28 (1), 137–150.
- Jebara, T., Kondor, R., Howard, A., 2004. Probability product kernels. *Journal of Machine Learning Research* 5, 819–844.
- Kawasaki, Y., Suzuki, M., Kherif, F., Takahashi, T., Zhou, S.-Y., Nakamura, K., Matsui, M., Sumiyoshi, T., Seto, H., Kurachi, M., 2007. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *NeuroImage* 34 (1), 235–242.
- Kittler, J., Hatef, M., Duin, R. P., Matas, J., 1998. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (3), 226–239.
- Klein, S., Loog, M., van der Lijn, F., den Heijer, T., Hammers, A., de Bruijne, M., van der Lugt, A., Duin, R. P., Breteler, M. M. B., Niessen, W. J., 2010. Early diagnosis of dementia based on intersubject whole-brain dissimilarities. In: *Proceedings of IEEE International Symposium on Biomedical Imaging: Macro to Nano, 2010*. p. in press.
- Ko, A. H.-R., Sabourin, R., de Souza Britto Jr., A., 2006. A new objective function for ensemble selection in random subspaces. In: *Proceedings of the International Conference on Pattern Recognition, ICPR '06*. pp. 185–188.
- Kuncheva, L. I., 2002. A theoretical study on six classifier fusion strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2), 281–286.
- Kuncheva, L. I., 2004. *Combining pattern classifiers: methods and algorithms*. Wiley-Interscience.
- Kuncheva, L. I., 2005. Special issue on diversity in multiple classifier systems. *Information Fusion* 6 (1), 1–115.
- Kuncheva, L. I., Whitaker, C. J., 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning* 51 (2), 181–207.

- Kuncheva, L. I., Whitaker, C. J., Ship, C. A., Duin, R. P., 2000. Is independence good for combining classifiers? In: Proceedings of the 15th International Conference on Pattern Recognition, ICPR '00. pp. 168–171.
- Lee, W.-J., Duin, R., Ibba, A., Loog, M., 2010. An experimental study on combining euclidean distances. In: The 2nd International Workshop on Cognitive Information Processing. p. in press.
- Ling, H., Okada, K., 2006. Diffusion distance for histogram comparison. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '06). Vol. 1. pp. 246–253.
- Lissack, T., Fu, K., 1976. Error estimation in pattern recognition via l-distance between posterior density functions. *IEEE Transactions on Information Theory* 22, 34–45.
- Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M., Carter, C. S., Stenger, V. A., Davis, S., Aizenstein, H., Becker, J. T., Lopez, O. L., Meltzer, C. C., 2004. Discriminative mr image feature analysis for automatic schizophrenia and alzheimer's disease classification. In: Proceedings of the Medical Image Computing and Computer-Assisted Intervention, MICCAI '04. pp. 393–401.
- Lopez-Garcia, P., Aizenstein, H. J., Snitz, B. E., Walter, R. P., Carter, C. S., Oct 2006. Automated roi-based brain parcellation analysis of frontal and temporal brainvolumes in schizophrenia. *Psychiatry Res* 147 (2-3), 153–161.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60 (2), 91–110.
- Martínez-Munoz, G., Hernández-Lobato, D., Suárez, A., 2009. An analysis of ensemble pruning techniques based on ordered aggregation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (2), 245–259.
- Martins, A. F. T., Smith, N. A., Xing, E. P., Aguiar, P. M. Q., Figueiredo, M. A. T., 2009. Nonextensive information theoretic kernels on measures. *Journal of Machine Learning Research* 10, 935–975.
- Meisenzahl, E. M., Koutsouleris, N., Bottlender, R., Scheuerecker, J., Jäger, M., Teipel, S. J., Holzinger, S., Frodl, T., Preuss, U., Schmitt, G., Burgermeister, B., Reiser, M., Born, C., Möller, H. J., Sep 2008. Structural brain alterations at different stages of schizophrenia: A voxel-based morphometric study. *Schizophrenia Res.* 104 (1-3), 44–60.
- Nyúl, L. G., Udupa, J. K., Zhang, X., 2000. New variants of a method of mri scale standardization. *IEEE Transactions on Medical Imaging* 19 (2), 143–150.
- Pekalska, E., Duin, R. P. W., 2005. *The Dissimilarity Representation for Pattern Recognition. Foundations and Applications.* World Scientific, Singapore.
- Petrakos, M., Kannelopoulos, I., Benediktsson, J. A., Pesaresi, M., 2000. The effect of correlation on the accuracy of the combined classifier in decision level fusion. In: Proceedings of the IEEE International Geo-science and Remote Sensing Symposium, IGARSS '00. Vol. 6. pp. 2623–2625.
- Prasad, K. M., Sahni, S. D., Rohm, B. R., Keshavan, M. S., 2005. Dorsolateral prefrontal cortex morphology and short-term outcome in first-episode schizophrenia. *Psychiatry Research: Neuroimaging* 140 (2), 147–155.
- Procyk, E., Goldman-Rakic, P. S., Nov 2006. Modulation of dorsolateral prefrontal delay activity during self-organized behavior. *J Neurosci* 26 (44), 11313–23.
- Pruessner, J., Li, L., Serles, W., Pruessner, M., Collins, D., Kabani, N., Lupien, S., Evans, A., 2000. Volumetry of hippocampus and amygdala with high-resolution mri and three-dimensional analysis software: Minimizing the discrepancies between laboratories. *Cerebral Cortex* 10 (4), 433–442.
- Reuter, M., Wolter, F.-E., Shenton, M., Niethammer, M., 2009. Laplace-Beltrami eigenvalues and topological features on eigenfunctions for statistical shape analysis. *Computed-Aided Design* 41 (10), 739–755.
- Roli, F., Giacinto, G., Vernazza, G., 2001. Methods for designing multiple classifier systems. In: Proceedings of the International Workshop on Multiple Classifier Systems, MCS '01. pp. 78–87.
- Rubner, Y., Tomasi, C., Guibas, L. J., 2000. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision* 40 (2), 99–121.
- Rujescu, D., Collier, D. A., 2009. Dissecting the many genetic faces of schizophrenia. *Epidemiologia e Psichiatria Sociale* 18 (2), 91–95.
- Ruta, D., Gabrys, B., 2005. Classifier selection for majority voting. *Information Fusion* 6 (1), 63–81.
- Sala, M., Perez, J., Soloff, P., di Nemi, S. U., Caverzasi, E., Soares, J. C., Brambilla, P., Oct 2004. Stress and hippocampal abnormalities in psychiatric disorders. *Eur. Neuropsychopharmacology* 14 (5), 393–405.
- Serratos, F., Sanfeliu, A., 2006. Signatures versus histograms: Definitions, distances and algorithms. *Pattern Recognition* 39 (5), 921–934.

- Sharkey, A. J. C., Sharkey, N. E., Gerecke, U., Chandroth, G. O., 2000. The “test and select” approach to ensemble combination. In: Proceedings of the International Workshop on Multiple Classifier Systems, MCS '00. Vol. 1857. pp. 30–44.
- Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press.
- Shenton, M. E., Dickey, C. C., Frumin, M., McCarley, R. W., Apr 2001. A review of mri findings in schizophrenia. *Schizophrenia Research* 49 (1–2), 1–52.
- Swain, M. J., Ballard, D. H., 1991. Color indexing. *International Journal of Computer Vision* 7 (1), 11–32.
- Swanson, L. W., Petrovich, G. D., 1998. What is the amygdala? *Trends in Neurosciences* 21 (8), 323–331.
- Tansella, M., Burti, L., 2003. Integrating evaluative research and community based mental health care in verona-italy. *British J. of Psychiatry* 183, 167–169.
- Tax, D. M. J., van Breukelen, M., Duin, R. P., Kittler, J., 2000. Combining multiple classifiers by averaging or multiplying? *Pattern Recognition* 33 (9), 1475–1485.
- Timoner, S. J., Golland, P., Kikinis, R., Shenton, M. E., Grimson, W. E. L., Wells III, W. M., 2002. Performance issues in shape classification. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI '02. pp. 355–362.
- Ting, K. M., Witten, I. H., 1999. Issues in stacked generalization. *Journal of Artificial Intelligence Research* 10, 271–289.
- Tulving, E., Markowitsch, H. J., 1998. Episodic and declarative memory: role of the hippocampus. *Hippocampus* 8 (3), 198–204.
- Tumer, K., Ghosh, J., 1999. Combining Artificial Neural Nets, A. J. C. Sharkey, ed. Springer, Ch. Linear and Order Statistics Combiners for Pattern Classification, pp. 127–155.
- Tumer, K., Ghosh, J., 2002. Robust combining of disparate classifiers through order statistics. *Pattern Analysis and Applications* 5 (2), 189–200.
- Ulaş, A., Castellani, U., Bicego, M., Cheng, D. S., Rambaldelli, G., Bellani, M., Brambilla, P., Murino, V., 2009a. A multiple region of interest analysis for identifying schizophrenia using MRI, submitted to Neuroimage, special issue on “Multivariate Decoding and Brain reading”.
- Ulaş, A., Duin, R. P. W., Castellani, U., Loog, M., Bicego, M., Murino, V., Bellani, M., Cerruti, S., Tansella, M., Brambilla, P., 2010. Dissimilarity-based detection of schizophrenia, submitted to ICPR workshop on “Brain Decoding: Pattern Recognition Challenges in FMRI Neuroimaging”.
- Ulaş, A., Semerci, M., Yıldız, O. T., Alpaydm, E., April 2009b. Incremental construction of classifier and discriminant ensembles. *Information Sciences* 179 (9), 1298–1318.
- Ventura, J., Nuechterlein, K. H., Subotnik, K. L., Gutkind, D., Gilbert, E. A., 2000. Symptom dimensions in recent-onset schizophrenia and mania: a principal component analysis of the 24-item brief psychiatric rating scale. *Schizophrenia Research* 97, 129–135.
- Voets, N. L., Hough, M. G., Douaud, G., Matthews, P. M., James, A., Winmill, L., Webster, P., Smith, S., 2008. Evidence for abnormalities of cortical development in adolescent-onset schizophrenia. *NeuroImage* 43 (4), 665–675.
- Wolpert, D. H., 1992. Stacked generalization. *Neural Networks* 5, 241–259.
- World Health Organization, 1992. Schedules for Clinical Assessment in Neuropsychiatry. WHO, Geneva.
- World Health Organization, 1996. SCAN 2.1.: Schede di valutazione clinica in neuropsichiatria. Il Pensiero Scientifico Editore, Roma.
- Yang, Y., Webb, G. I., Cerquides, J., Korb, K. B., Boughton, J., Ting, K. M., 2007. To select or to weigh: A comparative study of linear combination schemes for superparent-one-dependence estimators. *IEEE Transactions on Knowledge and Data Engineering* 19 (12), 1652–1665.
- Yoon, U., Lee, J., Im, K., Shin, W., Cho, B. H., Kim, I., Kwon, J., Kim, S., 2007. Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *NeuroImage* 34, 1405–1415.
- Zhou, Z.-H., Wu, J., Tang, W., 2002. Ensembling neural networks: many could be better than all. *Artificial Intelligence* 137, 239–263.