Dissimilarity-based Detection of Schizophrenia

A. Ulaş[†], R.P.W. Duin[*], U. Castellani[†], M. Loog[*], M. Bicego[†‡],
V. Murino[†‡], M. Bellani[†], S. Cerruti[†], M. Tansella[†], P. Brambilla[§]
[†]*University of Verona*
[*]*Pattern Recognition Laboratory, Delft University of Technology, The Netherlands*
[‡]*Istituto Italiano di Tecnologie (IIT)*
[§] *University of Udine*

***Abstract*—We propose to approach the detection of patients affected by schizophrenia by means of dissimilarity-based classification techniques applied to brain magnetic resonance images. Instead of working with features directly, pairwise distances between expert delineated regions of interest (ROIs) are considered as representations based on which learning and classification can be performed. Experiments were carried out on a set of 64 patients and 60 controls and several pairwise dissimilarity measurements have been analyzed. We demonstrate that good results are possible and especially significant improvements can be obtained when combining over different ROIs and different distance measures. The lowest error rate obtained is 0.210.**

I. Introduction

This work exploits advanced pattern recognition techniques in order to discriminate between subjects affected by schizophrenia and healthy controls on the basis of magnetic resonance brain imaging. We adopt a dissimilarity approach that exploits a particular selection of ROIs in the brain. The choice of ROIs is based on earlier investigations into their abnormal activity in case of schizophrenia [1], [2].

Several works have been proposed for human brain classification in the context of schizophrenia research [3], [4], [5]. Besides standard volumetric methods [6], [7], the most promising approaches focus on: (i) shape characterization [4], (ii) surface computation [5], and (iii) high dimension pattern classification [3]. In [4] a ROI-based morphometric analysis is introduced by defining spherical harmonics and a 3D skeleton as shape descriptors. In [5] a support vector machine (SVM) has been proposed to classify cortical thickness using the Euclidean distance between linked vertices on the inner and outer cortical surfaces. In [3] a new morphological signature has been defined by combining deformation-based morphometry with SVM.

The dissimilarity-based paradigm pursued in this work differs from typical pattern recognition approaches where objects to be classified are represented by feature vectors. In the dissimilarity approach, objects are described using pairwise (dis)similarities to a representation set of objects [8]. This offers the analyst 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.

Encouraged by our previous studies [9], [10], this work extends our earlier work and goes beyond volumetric measurements by classifying intensity histograms of the given ROIs. Our main contribution is the application of the dissimilarity-based classification approach to the detection of schizophrenia in MR images and the demonstration of its accuracy.

II. Data and Feature Extraction

The dataset used in this work is composed of MRI brain scans of 64 patients affected by schizophrenia and 60 healthy control subjects. Table I shows relevant demographic and clinical characteristics of both groups. The ROIs in this study were obtained by manually drawing contours enclosing the intended region. This was carried out by a trained expert following a specific protocol for each ROI [1]. The ROIs traced are 7 pairs (for the left and the right hemisphere respectively) of disconnected image areas:

- Amygdala (*l_amyg* and *r_amyg* in short);
- Dorso-lateral PreFrontal Cortex (*l_dlpfc* and *r_dlpfc*);
- Entorhinal Cortex (*l_ec* and *r_ec*);
- Heschl's Gyrus (*l_hg* and *r_hg*);
- Hippocampus (*l_hippo* and *r_hippo*);
- Superior Temporal Gyrus (*l_stg* and *r_stg*);
- Thalamus (*l_thal* and *r_thal*).

From the various ROIs, gray value histograms are determined and in order to reduce the effect of inter-subjects intensity variations, the extracted histograms are properly normalized[1]. Subsequently, several dissimilarity measures are defined between pairs of histograms all of which are described in the next section. In the spirit of dissimilarity based classification, there are various other ways to calculate dissimilarities between brain images based on registration of brains [11]. In this study we opted not to choose this

[1]See [9] for more details on the histogram normalization procedure.

Table I
DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF THE STUDY GROUPS. STUDENT'S t-TEST OF THE AGE MEANS REJECTS (AT A TWO-TAILED SIGNIFICANCE LEVEL OF $p < 0.05$) THE HYPOTHESIS THAT THE GROUPS ARE SIGNIFICANTLY DIFFERENT IN AGE, AND PEARSON'S χ^2 CONFIRMS THE SAME FOR THE GENDER DIFFERENCES.

	Group mean (and SD)*		Statistics		
Characteristic	Control (n = 60)	Schizophrenia (n = 64)	Test	*df*	*p*
Age, yr	39.95 (11.25) [range 23-60]	38.84 (11.96) [range 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* = significance value.
* Unless otherwise indicated.

direction because of the computational complexity and time required to pairwise register brain images.

III. Dissimilarity Measures

The computed histograms (and their pdfs) of intensities have been used to calculate dissimilarities between subjects using dissimilarity measures for histograms and pdfs. There are various dissimilarity measures that can be applied to measure the dissimilarities between histograms [12], [13]. Moreover, histograms can be converted to pdfs and dissimilarity measures between two discrete distributions can be used as well. All in all, we decided to study measures below.

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

Histogram intersection: It measures the number of intersecting values in each bin [14]:

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

Since this is a similarity measure, we convert it to a dissimilarity using $D = \min(|M|,|S|) \times (1 - Sim(S,M))$.

Diffusion distance: The distance between two histograms is defined as a temperature field. It is derived as the sum of dissimilarities over scales, see [15][2].

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

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

It is a standard measure for histograms.

Earth mover's distance: This distance was originally proposed by Rubner et. al [16]. It's basically defined as the cost to transform one distribution into another. It is calculated using linear optimization by defining the problem as a transportation problem. For 1D histograms, it reduces to a simple linear calculation [12] which was implemented in this study.

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

Similarly, we have considered the following dissimilarities between pdfs:

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

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

where

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

KullbackLeibler (KL) divergence: KullbackLeibler divergence is defined as

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

This measure is not a distance metric but a relative entropy since $D(p,q) \neq D(q,p)$, i.e., the dissimilarity matrix is not symmetric. There are various ways to symmetrize this dissimilarity. We simply used $D = D(p,q) + D(q,p)$ and the so-called Jensen-Shannon divergence: $D = \frac{1}{2}D(p,r) + \frac{1}{2}D(q,r)$, where r is the average of p and q.

In summary, we used the following 13 measures:

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

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. Further details of this combining are provided below.

IV. Dissimilarity Space

There are several ways to transform an $n \times n$ dissimilarity matrix D with elements $D(o,\hat{o})$ (the dissimilarity between

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

objects o and $\hat{o}$) into a vector space with objects represented by vectors $X = \{x'_1, \ldots, x'_o, \ldots, x'_{\hat{o}}, \ldots, x'_n\}$ [8]. 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 space 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 an (weighted) average of the given dissimilarity measures:

$$D_{combined} = \sum \alpha_i D_i \,. \tag{1}$$

It is related to the sum-rule in the area of combining classifiers. The weights can be optimized for some overall performance criterion, or determined from the properties of the dissimilarity matrix D_i itself, e.g. its maximum or average dissimilarity.

V. 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 considered, the 1-Nearest Neighbour (NN) rule on the original dissimilarities (called the *Standard Classifier*) and the linear SVM in dissimilarity space, called the *Dissimilarity based* classifier. The experiments are carried out using the Matlab package PRTools [17], including LIBSVM for the SVM 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 II summarizes the results. For each ROI the best performance is reported with respect to various dissimilarity measures. First column reports the error estimates for NN using the original dissimilarities (standard approach). Second column reports the leave-one-out error estimates of the linear SVM in dissimilarity space. It shows clearly the improvements of our dissimilarity-based approach. For ROIs like *l_dlpfc* and *l_hippo* the error is less than 0.3, while in the standard approach results are less stable and the error is always higher than 0.331.

Table II
ROI-BASED CLASSIFICATION

ROI	Standard Classifier	Dissimilarity-based
l_amyg	0.355 (his-l1)	0.315 (hist-euclid)
r_amyg	0.379 (pdf-euclid)	0.323 (pdf-euclid)
l_dlpfc	0.355 (pdf-kl)	0.234 (pdf-kl)
r_dlpfc	0.331 (his-intersect)	0.315 (pdf-js)
l_ec	0.403 (pdf-emd)	0.331 (pdf-js)
r_ec	0.355 (his-chi)	0.339 (his-intersec)
l_hg	0.403 (his-chi)	0.309 (pdf-kl-orig)
r_hg	0.387 (his-chi)	0.339 (pdf-kl)
l_hippo	0.444 (his-diffusion)	0.282 (pdf-js)
r_hippo	0.363 (his-euclid)	0.395 (his-intersec)
l_stg	0.427 (his-intersect)	0.355 (pdf-js)
r_stg	0.411 (his-chi)	0.355 (pdf-l1)
l_thal	0.355 (pdf-l1)	0.363 (his-diffusion)
r_thal	0.403 (pdf-euclid)	0.323 (his-emd)

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 (1) the reciprocal of the average dissimilarity value in D_i (cf. [18]). Table III 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 error 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., 0.210).

VI. Conclusion

In this paper a dissimilarity-based approach is proposed for the detection of schizophrenic brains. Several dissimilarity measures are proposed to deal with histograms of MRI intensity for different ROIs. ROI-based classification onto the dissimilarity space shows improvements of the standard NN rule. Moreover, a Multi-ROI classification strategy is obtained by simply averaging the similarity matrices observed in each ROI. Such approach drastically improves the single-ROI one, by highlighting the complementary information enclosed in the several ROIs. This confirms the

Table III
MULTI-ROI CLASSIFICATION.

Measure	Standard Classifier	Dissimilarity-based
his-euclid	0.379	0.258
his-l1	0.379	0.226
his-intersect	0.290	0.363
his-diffusion	0.379	0.226
his-chi	0.395	0.226
his-emd	0.532	0.411
pdf-euclid	0.411	0.298
pdf-l1	0.411	0.274
pdf-emd	0.395	0.290
pdf-bc	0.363	0.347
pdf-kl	0.355	0.339
pdf-kl-orig	0.363	0.315
pdf-js	0.363	0.323
average	0.387	0.210

benefit of combining dissimilarity representations and fusing information from various regions in the brain.

We like to emphasize that in building the (combined) representations no parameters are optimized w.r.t. performance. The proposed approach generally opens new perspectives in neuroanatomy classification by allowing the possibility to exploit dissimilarity measures which have some medical meaning or are expert-designed without having to worry about metric requirements or other mere technical complications.

ACKNOWLEDGEMENTS

We acknowledge financial support from the FET programme within the EU FP7, under the SIMBAD project (contract 213250).

REFERENCES

[1] M. Baiano, C. Perlini, G. Rambaldelli, R. Cerini, N. Dusi, M. Bellani, G. Spezzapria, A. Versace, M. Balestrieri, R. P. Mucelli, M. Tansella, and P. Brambilla, "Decreased entorhinal cortex volumes in schizophrenia," *Schizophrenia Research*, vol. 102, no. 1–3, pp. 171–180, 2008.

[2] S. G. Potkin, J. A. Turner, G. G. Brown, G. McCarthy, D. N. Greve, G. H. Glover, D. S. Manoach, A. Belger, M. Diaz, C. G. Wible, J. M. Ford, D. H. Mathalon, R. Gollub, J. Lauriello, D. O'Leary, T. G. V. Erp, A. W. Toga, A. Preda, and K. O. Lim, "Working memory and DLPFC inefficiency in schizophrenia: The FBIRN study," *Schizophr. Bulletin*, vol. 35, no. 1, pp. 19–31, 2009.

[3] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, "COMPARE: classification of morphological patterns using adaptive regional elements," *IEEE Trans. on Medical Imaging*, vol. 26, no. 1, pp. 93–105, 2007.

[4] G. Gering, M. Styner, and J. Lieberman, "Shape versus size: Improved understanding of the morphology of brain structures," in *Medical Image Comp. Computer-Assisted Intervention (MICCAI)*, 2001.

[5] U. Yoon, J. Lee, K. Im, W. Shin, B. H. Cho, I. Y. Kim, J. S. Kwon, and S. I. Kim, "Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia," *Neuroimage*, vol. 34, pp. 1405–1415, 2007.

[6] J. Ashburner and K. J. Friston, "Voxel-based morphometry - the methods," *Neuroimage*, vol. 11, pp. 805–821, 2000.

[7] M. Baiano, C. Perlini, G. Rambaldelli, R. Cerini, N. Dusi, M. Bellani, G. Spezzapria, A. Versace, M. Balestrieri, R. Mucelli, M. Tansella, and P. Brambilla, "Decreased entorhinal cortex volumes in schizophrenia," *Schizophr. Res.*, vol. 102, pp. 171–180, 2008.

[8] E. Pekalska and R. P. W. Duin, *The Dissimilarity Representation for Pattern Recognition. Foundations and Applications*. Singapore: World Scientific, 2005.

[9] D. S. Cheng, M. Bicego, U. Castellani, S. Cerruti, M. Bellani, G. Rambaldelli, M. Atzori, P. Brambilla, and V. Murino, "Schizophrenia classification using regions of interest in brain mri," in *Proc. of Intelligent Data Analysis in Biomedicine and Pharmacology, IDAMAP '09*, 2009, pp. 47–52.

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

[11] S. Klein, M. Loog, F. van der Lijn, T. den Heijer, A. Hammers, M. de Bruijne, A. van der Lugt, R. P. Duin, M. M. B. Breteler, and W. J. Niessen, "Early diagnosis of dementia based on intersubject whole-brain dissimilarities," in *Proceedings of IEEE International Symposium on Biomedical Imaging: Macro to Nano*, 2010, p. in press.

[12] S.-H. Cha and S. N. Srihari, "On measuring the distance between histograms," *Pattern Recognition*, vol. 35, no. 6, pp. 1355–1370, 2002.

[13] F. Serratosa and A. Sanfeliu, "Signatures versus histograms: Definitions, distances and algorithms," *Pattern Recognition*, vol. 39, no. 5, pp. 921–934, 2006.

[14] M. J. Swain and D. H. Ballard, "Color indexing," *Int. Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[15] H. Ling and K. Okada, "Diffusion distance for histogram comparison," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 1, 2006, pp. 246–253.

[16] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[17] R. P. W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. De Ridder, and D. M. J. Tax, *PRTools4, A Matlab toolbox for pattern recognition*. Delft University of Technology, 2004.

[18] W.-J. Lee, R. Duin, A. Ibba, and M. Loog, "An experimental study on combining euclidean distances," in *The 2nd International Workshop on Cognitive Information Processing*, 2010, p. in press.