Heat Diffusion Based Dissimilarity Analysis for Schizophrenia Classification

Aydın Ulaş[1,*], Umberto Castellani[1], Vittorio Murino[1,2],
Marcella Bellani[3], Michele Tansella[3], and Paolo Brambilla[4]

[1] University of Verona, Department of Computer Science, Verona, Italy
[2] Istituto Italiano di Tecnologia (IIT), Genova, Italy
[3] Department of Public Health and Community Medicine, Verona, Italy
[4] IRCCS "E. Medea" Scientific Institute, Udine, Italy

Abstract. We apply shape analysis by means of heat diffusion and we show that dissimilarity space constructed using the features extracted from heat diffusion present a promising way of discriminating between schizophrenic patients and healthy controls. We use 30 patients and 30 healthy subjects and we show the effect of several dissimilarity measures on the classification accuracy of schizophrenia using features extracted by heat diffusion. As a novel approach, we propose an adaptation of random subspace method to select random subsets of bins from the original histograms; and by combining the dissimilarity matrices computed by this operation, we enrich the dissimilarity space and show that we can achieve higher accuracies.

Keywords: heat diffusion, schizophrenia, dissimilarity space, support vector machines, random subspace.

1 Introduction

Recently, it has been popular to use magnetic resonance imaging (MRI) based image analysis to quantify morphological characteristics of different brains [11,1]. The goal of these studies is to distinguish normal subjects with patients affected by a certain disease.

Towards this aim of classification of healthy controls and patients, advanced computer vision and pattern recognition techniques have become popular in the recent years. This process can be summarized in two steps: i) feature (descriptor, distance, etc.) extraction, and ii) application of classification algorithms. The second part of this process is a well-established part of pattern recognition community and there are various techniques for classification ranging from simple classifiers such as decision trees to more advanced ones such as support vector machines [29]; from single classifier to ensembles of classification algorithms [14]; from feature based pattern recognition to dissimilarity based pattern recognition [17]. The list goes on and the research continues in this domain.

---

* Corresponding author.

Even though there are various classification techniques, there is no single classification algorithm which is the best for all problems and most of the time the algorithm's success depends on the discriminative capability of the features used. Base features change from application to application and one has to find the best representation and the complementary classification algorithm to succeed in the aim towards accurate classification. In this paper, we aim towards the question "Can schizophrenia be detected used MRI images?". Structural and morphological abnormalities have been demonstrated in patients [25,23] and there has been several studies which use different features and pattern recognition techniques [10,27,21,7,30]. Some approaches use the deformation information in the registration of pairwise brains which requires extensive computation time and the solution of complex mathematical problems due to the nonlinearity of the process. The usually preferred approach is to detect volume variations in specific parts of the brain called regions of interest (ROIs) [25,3]. The following works are examples of this methodology which we also used in this study. Gerig *et al.* [10] introduced a ROI-based morphometric analysis by defining spherical harmonics and 3D skeleton as shape descriptors. Timoner *et al.* [27] introduced a new descriptor by encoding the displacement fields and the distance maps for amygdala and hippocampus at the same time. Shape-DNA signature has been proposed in [21]. Liu *et al.* [16] analyzed image features encoding general statistical properties and Law's texture features from the whole brain. In a recently accepted work [7], we used *Heat Kernel Signature* (HKS) [26] using different scales and applied it to classification of schizophrenia. In [30], Yushkevich *et al.* used the Jacobian fields of deformation.

In this work, as the classification technique, we pursue dissimilarity based classification [17] which differs from usual pattern recognition in the sense that objects are represented as pairwise dissimilarities instead of feature vectors per object. This technique has been show to be promising in brain disease research [28]. The dissimilarities are then transferred into a new vector space [17] (the dissimilarity space) where traditional classification algorithms can be applied. This paper is a continuation of our two previous works [7,28] where we use the heat kernel signatures to extract the features and analyse the effect of several dissimilarity measures on the features extracted using HKS and demonstrate the improvement in classification accuracy when we utilize the dissimilarity space.

The novel methodology proposed in this work is the adaptation of Ho's random subspace [12] method to the creation of dissimilarities by the use of four dissimilarity measures on histograms. The method differs from prototype selection [18] because instead of selection prototypes from already available dissimilarities, we attempt at the transpose of this problem where we select (randomly) different bins from the original histograms and then compute the dissimilarities and combine them. As a result we do not reduce the dimension but we use more information and enrich the dissimilarity space. In a similar recent work, Carli *et al.* [6] used the dissimilarities between *local parts* for scene categorization.

The paper is organized as follows: in Section 2, we introduce the heat kernel based shape analysis; in Section 3 we introduce the methods and the application

of the methodology. We show our experiments in Section 4 and we conclude in
Section 5.

2   Shape Analysis by Heat Diffusion

When we considering a shape M as a compact Riemannian manifold [5], the
heat diffusion on shape[1] is defined by the *heat* equation:

$$(\Delta_M + \frac{\partial}{\partial t})u(t, \boldsymbol{m}) = 0; \tag{1}$$

where u is the distribution of heat on the surface, $\boldsymbol{m} \in M$, Δ_M is the *Laplace-Beltrami* operator which, has discrete eigendecomposition of the form $\Delta_M = \lambda_i \phi_i$ in compact spaces. The *heat kernel* can then be decomposed as:

$$h_t(\boldsymbol{m}, \boldsymbol{m}') = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(\boldsymbol{m}) \phi_i(\boldsymbol{m}'), \tag{2}$$

where λ_i and ϕ_i represent the ith eigenvalue and the ith eigenfunction of the
Laplace-Beltrami operator, respectively. The heat kernel $h_t(\boldsymbol{m}, \boldsymbol{m}')$ is the solu-
tion of the heat equation with initial point heat source in $\boldsymbol{m}$ at time $t = 0$, and
heat value in ending point $\boldsymbol{m}' \in M$ after time t. The heat kernel has been shown
to be *isometric invariant*, *informative*, and *stable* [26].

For volumetric representations, the volume is sampled by a regular Cartesian
grid composed of voxels, which then allows us to use the standard Laplacian in
R^3 as the Laplace-Beltrami operator. Finite differences are used to evaluate the
second derivative in each direction of the volume. The heat kernel on volumes
is invariant to volume isometries, in which shortest paths between points inside
the shape do not change. Note that in real applications exact volume isometries
are limited to the set of rigid transformations [20], however, also non-rigid de-
formations can be modelled as approximations to volume isometries in practice.
It is also worth noting that, for small t the autodiffusion heat kernel $h_t(\boldsymbol{m}, \boldsymbol{m})$
of a point $\boldsymbol{m}$ with itself is directly related to the *scalar* curvature $s(\boldsymbol{m})$ [26,20]:

$$h_t(\boldsymbol{m}, \boldsymbol{m}) = (4\pi t)^{-3/2}(1 + \frac{1}{6}s(\boldsymbol{m})). \tag{3}$$

In practice, Equation 3 states that the heat tends to diffuse slower at points with
positive curvature, and viceversa. This gives an intuitive explanation about the
geometric properties of $h_t(\boldsymbol{m}, \boldsymbol{m})$, and suggests the idea of using it to build a
shape descriptor [26].

3   Dissimilarities

3.1   Dissimilarity Measures

The computed histograms of the data have been used to calculate dissimilarities
between subjects using dissimilarity measures for histograms. There are various

---

[1] The notation is taken from [26,5].

dissimilarity measures that can be applied to measure the dissimilarities between histograms [8,24]. 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 and the L1 norm (L1).

Given two histograms S and M with n bins, and two discrete probability density functions p and q, we define the number of elements in S and M as $|S|$ and $|M|$ respectively.

Earth Mover's Distance (EMD). Proposed by Rubner *et al.* [22], the basic idea is to use the cost to transform one distribution into another as dissimilarity. It is calculated using linear optimization by defining the problem as a transportation problem. For 1D histograms, it reduces to a simple calculation [8] which was implemented in this study.

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

KullbackŰLeibler (KL) Divergence (KL) . KullbackŰLeibler 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 used the so-called Jensen-Shannon divergence (JS): $D = \frac{1}{2} D(p, r) + \frac{1}{2} D(q, r)$, where r is the average of p and q.

In addition to these dissimilarity measures, we propose to combine different "sub-dissimilarities" computed by random subspace method by using averaging and concatenating. We also test the accuracy of these combinations against the support vector machines on the original feature space. Further details of dissimilarity combination are provided in Section 3.4.

3.2 Dissimilarity Space

Suppose that we have n objects and we have a dissimilarity matrix D of size $n \times n$. And suppose that the dissimilarity between two objects o and $\hat{o}$ are denoted by $D(o, \hat{o})$. There are several ways to transform an $n \times n$ dissimilarity matrix D with elements $D(o, \hat{o})$ into a vector space with objects represented by vectors $X = \{x'_1, \ldots, x'_o, \ldots, x'_{\hat{o}}, \ldots, x'_n\}$ [17]. If the dissimilarities are already proper Euclidean, classical scaling can be used. One can utilize pseudo-Euclidean embedding for arbitrary symmetric dissimilarities. These yield vector spaces in which vector dissimilarities can be defined that produce the given dissimilarities D. When the dissimilarities are non-Euclidean, classification on these pseudo-Euclidean spaces are ill-defined since the corresponding kernels are indefinite.

A more general solution is to work directly in the *dissimilarity space*. It creates an Euclidean vector space by using the given dissimilarities as a set of features. In this case, the dissimilarities in this space do not correspond to the dissimilarities on the original space; which is sometimes an advantage if it is not known that they really are dissimilarities. 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 60 in our case. 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.

3.3 Random Subspace Method and Adaptation to Dissimilarity Computation

In classical pattern recognition, there are several classifier combination techniques to get the best of a data set in hand. For the combination to be effective, one has to create diverse classifiers and combine them in a proper way. Most combination methods aim at generating uncorrelated classifiers, and it has been proposed [14] to use different (i) learning algorithms, (ii) hyperparameters, (iii) input features, and (iv) training sets. For example Bagging [4] uses bootstrapping to generate different training sets and takes an average, the random subspace method [12] trains different classifiers with different subsets of a given feature set. In dissimilarity based classification, since most of the time the k-nearest classifier is used, the research has focused on prototype selection [18]. In prototype selection, the aim is to select a subset of the data which explains the whole set of data in a proper way. In this paper, we aim for the transpose of this problem, instead of selecting a set of instances, we aim to select a set of features and combine them as in [12]. This can be done after we have the dissimilarities which was also targeted in [18,19] or before the dissimilarities are actually computed. In this paper, we propose the latter approach and adopt Ho's random subspace method to dissimilarity computation. What we do is we select a random subset of the original set of bins and compute the dissimilarities according to these subsets. Then we combine these subsets to get the final classification. In this way we enrich the dissimilarity space by using more information. This is possible to achieve because we calculate the dissimilarities using histogram bins and is logical because some bins may be more descriptive of the data and computing the dissimilarities using only these bins may increase the classification accuracy. Dividing the feature sets into two and combining Euclidean distances has been investigated recently in [15]. Once these dissimilarities are computed, they are combined using averaging and concatenation. We show the results using both techniques: see below for the combination of dissimilarity matrices.

3.4   Dissimilarity Combination

Combined dissimilarity spaces can be constructed by combining dissimilarity representations. As in normal classifier combination [13], a simple and often effective way is using an (weighted) average of the given dissimilarity measures:

$$D_{combined} = \frac{\sum \alpha_i D_i}{\sum \alpha_i} \,. \tag{4}$$

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. In this work, we used equal weights while combining multiple dissimilarity matrices and all the dissimilarity matrices are scaled such that the average dissimilarity is one, i.e.:

$$\frac{D(i,j)}{\frac{1}{m(m-1)} \sum_{i,j} D(i,j)} = 1 \tag{5}$$

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

$$\tilde{D}(i,j) := \frac{\tilde{D}(i,j) + \tilde{D}(j,i)}{2} \tag{6}$$

After this scaling, we use averaging or concatenation to get the final space for classification.

4   Experiments

4.1   MRI Data Collection

Data collection and processing in MRI-based research faces several methodological issues to minimize biases and distortions. The standard approach is to follow well-established guidelines by international organizations, such as the World Health Organization (WHO), or by respected institutions. The data set used in this work is composed by MRI brain scans of 30 patients affected by schizophrenia and 30 healthy control subjects. All patients received a diagnosis of schizophrenia according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders [2]. After the data collection phase, we employ a ROI-based approach [11], so only a well defined brain subpart has been considered in this study. More specifically, we focus our analysis on the left-Thalamus whose abnormal activity has been already investigated in schizophrenia [9]. ROIs have been manually traced by experts, according to well defined medical protocols.

4.2   Histogram Computation Using HKS

In order to employ a *learning-by-example* approach, we need a collection of samples for both healthy subjects and patients. Source data are MRI scans where shape information can be provided in terms of volumetric data. According to the shape diffusion analysis described in Section 2, for each subject geometric features are extracted at a certain scale according to a time value t, and the autodiffusion value is computed for each voxel $\boldsymbol{m}$, leading to:

$$H_t(M) = \{h_t(\boldsymbol{m}, \boldsymbol{m}), \forall \boldsymbol{m} \in M\}.$$

Then, such values are accumulated into a histogram $r = hist(H_t(M))$. In this manner, we obtain a shape representation r, encoding the global shape at a certain scale. In our experiments, the number of bins for the histograms have been chosen as 100.

4.3   Experimental Methodology

In our experiments, using the 60 patients, we apply leave-one-out cross-validation to assess the performance of the methods. We first train support vector machines (SVMs) with linear kernels as a base-line for classification accuracies using the original histogram features. The C parameter is chosen using cross validation. We then compute the full dissimilarity space in a transductive way using the dissimilarity measures mentioned in Section 3. We again use the linear kernel on this space to compare the results with the base-line original space accuracy and random subspace accuracy. Random bins are chosen using different number of bins and different number of random subspaces and the dissimilarity space constructed using the combination of these dissimilarities are applied for the final classification. Note that, in this paper we do not aim at combining different modalities or data sets. What we are trying to achieve is to get the best accuracy using the single data set at hand; and with this in mind, we demonstrate the effect of different dissimilarity measures.

4.4   Results

Table 1 reports the accuracies of the SVM with the linear kernel on the original space (`LIN`), linear SVM on the dissimilarity spaces (`DIS`, where `-AVG` shows combining using averaging and `-CONC` shows combination using concatenation) and random subspace accuracies (`RAND-`N, where N is the selected number of bins. For this example, we used 20 random subspaces; the results for other numbers are similar. See also Figures 1 and 2). We can see that the dissimilarity space accuracies are always better than the original space accuracies. This also corresponds with our previous work [28]. We can also see that with a suitable number of random bins, one can always achieve a better accuracy than the full dissimilarity space accuracies. We can see that the average operation usually

Table 1. Accuracies using the dissimilarity space and the random subspace. 20 random subspaces are created

	L1-AVG	L1-CONC	EMD-AVG	EMD-CONC	KL-AVG	KL-CONC	JS-AVG	JS-CONC
LIN				66.67				
DIS	76.67		61.67		75.00		68.33	
RAND-50	76.67	75.00	70.00	58.33	73.33	68.33	70.00	75.00
RAND-25	78.33	70.00	65.00	58.33	71.67	58.33	71.67	60.00
RAND-20	**86.67**	83.33	65.00	55.00	71.67	70.00	68.33	71.67
RAND-15	80.00	73.33	73.33	68.33	75.00	**88.33**	73.33	63.33
RAND-10	71.67	56.67	65.00	60.00	71.67	66.67	70.00	53.33

creates more accurate results (though the best result is using concatenation on `KL`). We believe that this is because of the curse of dimensionality. When one uses the concatenation operator, the number of features increase and the classification accuracy may decrease accordingly.

We can see that the random subspace method applied to the creation of the dissimilarity space achieves better accuracies than the full dissimilarity space and the original space. One disadvantage of this method is to choose the proper number of bins and proper number of random subspaces. One can use cross validation to decide these values but we leave the exploration of this topic as a future work. Nevertheless, we want to present the performance of these methods when we apply different number of bins and different number of random subspaces. In Figure 1, we can see the change in accuracy versus the number of subspaces created. The numbers on the plots show the number of bins chosen and the peak accuracy using that number of bins. Although there is no clear correlation between accuracy and the number of subspaces, we can still observe that with a fairly small number of subspaces, one can achieve relatively high accuracies. We can also observe that (as expected), with the increasing number of subspaces, the diversity decreases and the accuracy converges. We can observe that the dissimilarity space is always superior to the original space and (except for `KL` where the choice of the number of subspaces is critical), the random subspace accuracy is higher than full dissimilarity space accuracy.

In the second set of experiments, we fix the number of random subspaces and see the change in accuracy when we change the number of selected bins. These results are presented in Figure 2. Again, the numbers on the figure show the number of random subspaces and peak accuracy obtained by this selection. We can see a clearer picture in this setup because when we increase the number of bins, we get to a peak point and then the accuracy decreases and in the end levels off as expected. The peak point is usually between 20 and 30 which we suggest to use for a 100 bin histogram representation. This shows us that the selection of number of bins is the critical point in this methodological setup and the proper selection of this parameter yields the best accuracy.
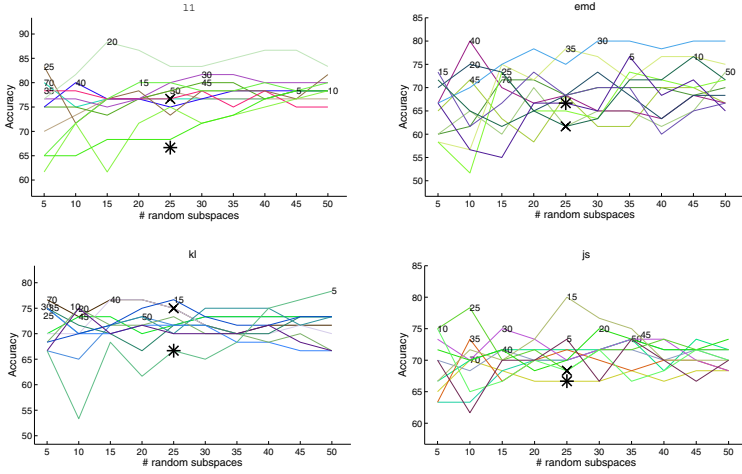
Fig. 1. # of subspaces vs accuracy. The numbers in the plot show the number of histogram bins used. X is the accuracy on the full dissimilarity matrix of the corresponding dissimilarity measure and * is the base-line original space accuracy.
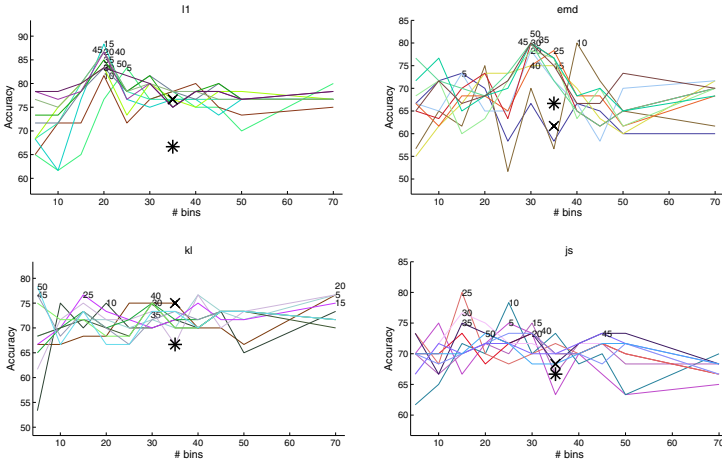
Fig. 2. # of bins vs accuracy. The numbers in the plot show the number of random subspaces used. X is the accuracy on the full dissimilarity matrix of the corresponding dissimilarity measure and * is the base-line original space accuracy.

5   Conclusion

In this paper, we use *Heat Kernel Signatures* to extract histogram based features (see also [7]) from MRI scans of 30 schizophrenic patients and 30 healthy controls to detect schizophrenia. We first create several dissimilarity matrices using histogram based dissimilarity measures in the literature and compare our results with the base-line original space accuracies using the support vector machines with the linear kernel. As the novel part of this study, we propose to adapt the random subspace method [12] to the creation of dissimilarity matrices and show that with the combination of these matrices, we achieve higher detection accuracies. Random subspaces have been used also in the dissimilarity based pattern recognition paradigm [19,18], but the method is applied after the dissimilarity matrices are constructed. The novelty of our method is that we use the random subspaces technique before creating the dissimilarity matrices to combine the useful information of the data.

In our previous work, we used the dissimilarity space and dissimilarity combination [28] using intensity and apparent diffusion coefficient (ADC) based histogram representations. The main aim of that study was to show the effect of dissimilarity space using dissimilarity combination and multiple modalities. In this work, we use only one modality and one ROI; our aim is not to combine data from different sources but to get the best from the data as hand as was also targeted in [18]. For this purpose, we use four different dissimilarity measures and analyze the effect of using the dissimilarity space and the dissimilarity space constructed using random subspaces.

We have seen that the dissimilarity space constructed using the dissimilarity measures mentioned in this work always outperform the base-line original space accuracies. Our novel contribution is the adaptation of Ho's random subspace methodology for selecting a subset of bins used in the computation of the dissimilarity matrices. With this method, we see that we can achieve significantly higher results when proper number of bins and random subspaces are selected. For the combination purpose, we compare combining using averaging and concatenation and we see that averaging usually outperforms concatenation (though the best result is achieved using concatenation) and we believe that this is due to the curse of dimensionality. As a future work, we plan to combine the dissimilarities over random subspaces using *Multiple Kernel Learning* (MKL) paradigm.

Our analysis of number of bins and number of random subspaces show that the important parameter is the number of bins. With relatively small number of random subspaces, one can achieve good results. As expected, when the number of subspaces increase, the individual dissimilarity matrices become less diverse and the combination accuracy converges. Same occurs with the number of bins, when we increase the number of bins, the diversity decreases and the accuracy converges. In this setup, the number of bins become more important because of the peak points. In this study we used a source histogram of 100 bins and we have observed that we get the peak accuracies in the 20-30 bin range. As a future study, we would like to apply this methodology to other data sets and

artificial data sets with different number of source bins and draw a theoretical/experimental formula for the selection of the number of bins.

References

1. Agarwal, N., Port, J.D., Bazzocchi, M., Renshaw, P.F.: Update on the use of MR for assessment and diagnosis of psychiatric diseases. Radiology 255(1), 23–41 (2010)
2. American Psychiatric Association: Diagnostic and statistical manual of mental disorders. DSM-IV. Washington DC, 4th edn. (1994)
3. 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.: Decreased entorhinal cortex volumes in schizophrenia. Schizophrenia Research 102(1-3), 171–180 (2008)
4. Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)
5. Bronstein, A.M., Bronstein, M.M., Ovsjanikov, M., Guibas, L.J.: Shape recognition with spectral distances. IEEE Trans. Pattern Analysis and Machine Intelligence 33(5), 1065–1071 (2011)
6. Carli, A., Castellani, U., Bicego, M., Murino, V.: Dissimilarity-based representation for local parts. In: 2010 2nd International Workshop on Cognitive Information Processing (CIP), pp. 299–303 (June 2010)
7. Castellani, U., Mirtuono, P., Murino, V., Bellani, M., Rambaldelli, G., Tansella, M., Brambilla, P.: A new shape diffusion descriptor for brain classification. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 426–433. Springer, Heidelberg (2011)
8. Cha, S.-H., Srihari, S.N.: On measuring the distance between histograms. Pattern Recognition 35(6), 1355–1370 (2002)
9. Corradi-DellAcqua, C., Tomelleri, L., Bellani, M., Rambaldelli, G., Cerini, R., Pozzi-Mucelli, R., Balestrieri, M., Tansella, M., Brambilla, P.: Thalamic-insular dysconnectivity in schizophrenia: Evidence from structural equation modeling. Human Brain Mapping (in press, 2011)
10. Gerig, G., Styner, M.A., Shenton, M.E., Lieberman, J.A.: Shape Versus Size: Improved Understanding of the Morphology of Brain Structures. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 24–32. Springer, Heidelberg (2001)
11. Giuliani, N.R., Calhouna, V.D., Pearlson, G.D., Francis, A., Buchanan, R.W.: 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 (2005)
12. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(8), 832–844 (1998)
13. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(3), 226–239 (1998)
14. Kuncheva, L.I.: Combining pattern classifiers: methods and algorithms. Wiley Interscience (2004)
15. Lee, W.-J., Duin, R.P.W., Loog, M., Ibba, A.: An experimental study on combining euclidean distances. In: 2010 2nd International Workshop on Cognitive Information Processing (CIP), pp. 304–309 (June 2010)

16. Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M.E., Carter, C.S., Stenger, V.A., Davis, S., Aizenstein, H.J., Becker, J.T., Lopez, O.L., Meltzer, C.C.: Discriminative MR Image Feature Analysis for Automatic Schizophrenia and Alzheimer's Disease Classification. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 393–401. Springer, Heidelberg (2004)
17. Pekalska, E., Duin, R.P.W.: The Dissimilarity Representation for Pattern Recognition. Foundations and Applications. World Scientific, Singapore (2005)
18. Pekalska, E., Duin, R.P.W., Paclík, P.: Prototype selection for dissimilarity-based classifiers. Pattern Recognition 39(2), 189–208 (2006)
19. Pękalska, E.z., Skurichina, M., Duin, R.P.W.: Combining fisher linear discriminants for dissimilarity representations. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 117–126. Springer, Heidelberg (2000)
20. Raviv, D., Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Volumetric heat kernel signatures. In: Workshop on 3D Object Retrieval (2010)
21. Reuter, M., Wolter, F.E., Shenton, M., Niethammer, M.: Laplace-Beltrami eigenvalues and topological features on eigenfuntions for statistical shape analysis. Computed-Aided Design 41(10), 739–755 (2009)
22. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision 40(2), 99–121 (2000)
23. Rujescu, D., Collier, D.A.: Dissecting the many genetic faces of schizophrenia. Epidemiologia e Psichiatria Sociale 18(2), 91–95 (2009)
24. Serratosa, F., Sanfeliu, A.: Signatures versus histograms: Definitions, distances and algorithms. Pattern Recognition 39(5), 921–934 (2006)
25. Shenton, M.E., Dickey, C.C., Frumin, M., McCarley, R.W.: A review of mri findings in schizophrenia. Schizophrenia Research 49(1-2), 1–52 (2001)
26. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. In: Proceedings of the Symposium on Geometry Processing, pp. 1383–1392 (2009)
27. Timoner, S.J., Golland, P., Kikinis, R., Shenton, M.E., Grimson, W.E.L., Wells III, W.M.: Performance issues in shape classification. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 355–362. Springer, Heidelberg (2002)
28. Ulaş, A., Duin, R.P.W., Castellani, U., Loog, M., Mirtuono, P., Bicego, M., Murino, V., Bellani, M., Cerruti, S., Tansella, M., Brambilla, P.: Dissimilarity-based detection of schizophrenia. International Journal of Imaging Systems and Technology 21(2), 179–192 (2011)
29. Vapnik, V.N.: Statistical learning theory. John Wiley and Sons (1998)
30. Yushkevich, P., Dubb, A., Xie, Z., Gur, R., Gur, R., Gee, J.: Regional structural characterization of the brain of schizophrenia patients. Academic Radiology 12(10), 1250–1261 (2005)