

COMPARE: Classification of Morphological Patterns Using Adaptive Regional Elements

Yong Fan*, *Member, IEEE*, Dinggang Shen, *Member, IEEE*, Ruben C. Gur, Raquel E. Gur, and Christos Davatzikos, *Senior Member, IEEE*

Abstract—This paper presents a method for classification of structural brain magnetic resonance (MR) images, by using a combination of deformation-based morphometry and machine learning methods. A morphological representation of the anatomy of interest is first obtained using a high-dimensional mass-preserving template warping method, which results in tissue density maps that constitute local tissue volumetric measurements. Regions that display strong correlations between tissue volume and classification (clinical) variables are extracted using a watershed segmentation algorithm, taking into account the regional smoothness of the correlation map which is estimated by a cross-validation strategy to achieve robustness to outliers. A volume increment algorithm is then applied to these regions to extract regional volumetric features, from which a feature selection technique using support vector machine (SVM)-based criteria is used to select the most discriminative features, according to their effect on the upper bound of the leave-one-out generalization error. Finally, SVM-based classification is applied using the best set of features, and it is tested using a leave-one-out cross-validation strategy. The results on MR brain images of healthy controls and schizophrenia patients demonstrate not only high classification accuracy (91.8% for female subjects and 90.8% for male subjects), but also good stability with respect to the number of features selected and the size of SVM kernel used.

Index Terms—Feature selection, morphological pattern analysis, pattern classification, structural MRI, regional feature extraction, schizophrenia, support vector machines (SVM).

I. INTRODUCTION

MORPHOLOGICAL analysis of medical images is used in a variety of research and clinical studies that investigate the effect of diseases and treatments on anatomical structure. Region of interest (ROI) volumetry has been traditionally used to obtain regional measurement of anatomical volumes and to investigate abnormal tissue structures with disease [1]. However, in practice, *a priori* knowledge about abnormal regions is not always available. Even when *a priori* hypotheses can be

made about specific ROIs, a region of abnormality might be part of an ROI, or span over multiple ROIs, thereby potentially reducing statistical power of the underlying morphological analysis. These limitations can be effectively overcome by methods generally referred to as high-dimensional morphological Analysis (HDMA), such as voxel-based and deformation-based morphometrical analysis methods [2]–[5]. However, a voxel-wise analysis is limited by noise, registration error, and excessive interindividual variability of measurements that are too localized, such as voxel-wise displacement fields, Jacobian determinants, or tissue density maps. Although the statistical power of voxel-by-voxel analysis methods can be improved by simply smoothing the morphological measures via a Gaussian filter prior to statistical analysis, the smoothing is typically applied uniformly to all brain regions and all individuals, i.e., it is not adaptive to anatomical structures, shapes, abnormal regions, or specific anatomies.

In neuroimaging studies using HDMA, voxel-wise mass-univariate analysis methods have been widely utilized [3], [5]–[9]. However, these have limited ability to detect complex population differences, because they do not take into account the multivariate relationships in the data [10]. More importantly, they have very limited diagnostic power, since in every single region there is typically significant overlap between healthy and diseased individuals. In order to overcome these limitations, high-dimensional pattern classification methods have been applied in computational anatomy [11]–[18], aiming at capturing multivariate relationships among various anatomical regions for more effectively characterizing group differences. Both linear classification approaches, such as linear discrimination analysis (LDA), and nonlinear classification approaches, such as nonlinear support vector machine (SVM), have been proposed in [11], [15]–[18]. While LDA may be optimal when the class distributions are Gaussian, SVMs are more effective in capturing complex nonlinear relationships.

Although HDMA methods can measure localized structural changes in unbiased (hypothesis-free) paradigms, they pose several challenges such as the sheer dimensionality of HDMA-based measurements, which is often in the millions, relative to the small number of training samples, which is, at best, in the hundreds and is often only a few dozen, and the inevitable measurement noise that might originate from registration inaccuracies or interindividual anatomical variations. Therefore, HDMA-based measurements must be distilled down to a relatively small number of most important features for classification. Moreover, these features must be robust for classification in order to support good generalization properties.

Manuscript received May 9, 2006; revised September 25, 2006. *Asterisk indicates corresponding author.*

*Y. Fan is with the Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: yong.fan@uphs.upenn.edu).

D. Shen and C. Davatzikos are with the Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA.

R. C. Gur and R. E. Gur are with the Brain Behavior Laboratory and Schizophrenia Research Center, Department of Psychiatry, University of Pennsylvania Medical Center, Philadelphia, PA 19104 USA.

Color versions of Figs. 2, 4–8 are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2006.886812

The wavelet transform has been previously proposed to capture morphological features in HDMA-based MR brain classification frameworks [15]. The advantage of using the wavelet decomposition is that it is able to powerfully represent data at multiple scales, from which the most pertinent for classification can be selected. However, the wavelet transform uses a fixed mother function that is not able to adapt to the arbitrary shape of potential abnormal regions, even after scaling, therefore informative features for classification might be missed. Furthermore, the wavelet-based feature representation method yields a large number of feature dimensions. Thus, additional feature selection and reduction methods must be applied to produce good generalization performance in small sample size problems.

Feature selection techniques have been adopted in many morphological brain analysis studies, in order to produce a small number of effective features for efficient classification and to increase the generalization performance of the classifier [13]–[15]. Feature ranking and feature subset selection are two types of typical feature selection methods [19]. Subset feature selection methods are generally time consuming, thus inapplicable when millions of morphological features are available, as in images in which each voxel is initially a feature. Ranking-based feature selection methods are subject to local optima. Therefore, these two feature selection methods are usually used jointly, i.e., using a ranking-based feature selection method for selecting an initial set of important features, and using a subset feature selection method for further selection. Notably, feature selection methods directly select features from the original feature set without any transformations. This makes the classification results easier to interpret, since features used for classification can be directly related to anatomical regions. Thus, feature selection methods are substantively different from the traditional linear feature dimensionality reduction methods, such as principal component analysis (PCA), which captures global features and are thus not always able to identify localized abnormal brain regions.

COMPARE is a classification method for identification of brain abnormality, which aims at overcoming certain limitations in previous morphological analysis methods. A main emphasis of this paper is the extraction of distinctive, but also robust features from high-dimensional morphological measurements obtained from brain MR images, which are used in conjunction with nonlinear SVMs for classification. In COMPARE, the morphological information used for group separation is obtained via high-dimensional deformation fields registering a template with an individual's image. Features that are subsequently used for classification are obtained in a way that is spatially adaptive to the data and does not depend on predefined anatomical regions. To achieve this, voxel-wise tissue density maps are first extracted for each individual brain in a mass-preserving shape transformation framework [5] that warps individuals to a template space via an image registration approach referred to as HAMMER [20], [21]. These morphological features are further clustered into regions by a watershed segmentation method [22], according to some measures of discriminative power and reliability. Finally, a volume increment algorithm produces robust features by grouping voxels that show similar relationships to the classification variable. Thus, the regional

volumetric features constructed this way form a feature vector, which is used as a “morphological signature” for each brain. The irrelevant and redundant features in the feature vector are further removed by a feature selection method, in order to improve the performance and generalization of a classifier in brain classification. With a set of discriminative and reliable features extracted and selected, a nonlinear SVM classifier is thus constructed to determine whether the morphological information derived from a particular brain implies abnormality. Furthermore, for the purpose of interpretation of group differences and separation, the group difference between two populations can also be estimated from a constructed classifier by a discriminative direction method [11], [15]. The performance of COMPARE has been tested on a morphometric study of schizophrenia using MR brain images.

The remainder of this paper is organized as follows. In Section II, we detail three important steps in our methodology, i.e., feature extraction, feature selection and SVM-based brain classification. In Section III, experimental results on clinical data are described. Conclusion and discussions are provided in Section IV.

II. METHODS

COMPARE involves three steps: feature extraction, feature selection, and nonlinear classification, which are detailed next.

A. Feature Extraction

As mentioned in Introduction, the features used for brain classification are extracted from automatically generated regions, which are determined from the training data. Several issues are taken into consideration here. First, morphological changes of brain structures resulting from pathological processes usually do not occur in isolated regions or in regions necessarily having regular shapes. Moreover, these regions are not known *a priori*. Second, noise, registration errors, and interindividual anatomical variations necessitate the collection of morphological information from regions much larger than the voxel size, which must additionally be distinctive of and adaptive to the pathology of interest. Third, multivariate classification methods are most effective and generalizable when applied to a small number of reliable and discriminative features. Accordingly, features irrelevant to classification must be eliminated.

In the following sections, we will detail the procedure of automatically generating spatially adaptive regions from a training dataset, by first introducing the method to extract local morphological features, then defining the criteria for adaptively clustering voxels into regions, and finally extracting overall features from each region.

1) *Construction of a Morphological Profile for each Brain:* Warping individual brain images into a template space is necessary for quantitative comparison of different individual brain images. By performing this image warping procedure, various morphological measurements, i.e., Jacobian determinant and tissue-density maps can be obtained in the same template space, thus facilitating the direct comparison of individual brains. Herein, we construct a morphological profile for each individual brain by following a mass-preserving shape transformation framework proposed in [5], which is related



Fig. 1. Cross-sectional views of representative tissue density maps (WM, GM, ventricular CSF, from left to right).

to “Jacobian modulation” widely used in the SPM software package.¹

Three steps are involved in the mass-preserving shape transformation framework [5]. Each skull-stripped MR brain image is first segmented into three tissues, namely gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), by a brain tissue segmentation method proposed in [23]. Afterwards, each tissue-segmented brain image is spatially normalized into a template space, by using a high-dimensional image warping method, called HAMMER [20]. The total tissue mass is preserved in each region during the image warping, which is achieved by increasing the respective density when a region is compressed, and vice versa. Finally, three tissue density maps, f^0, f^1, f^2 , are generated in the template space, each reflecting local volumetric measurements corresponding to GM, WM, and ventricular CSF, respectively. Representative tissue density maps are shown in Fig. 1. These tissue density maps give a quantitative representation of the spatial distribution of tissues in a brain, with brightness being proportional to the amount of local tissue volume before warping.

2) *Regional Grouping of Local Morphological Features by Watershed Segmentation*: The mass-preserving transformation procedure described above generates three variables for each brain voxel resulting in millions of variables. These measurements must be reduced to a relatively small set of measurements reflecting regional volumes, which the subsequent classifiers can handle successfully with relatively few training samples. Compared to voxel-wise features, regional measurements are always more robust to registration error and anatomical variation across individuals. Most of the approaches that group voxels into clusters fall under three categories: 1) Gaussian smoothing, which is commonly used to improve the robustness of local features by averaging them using a Gaussian point spread function around each location [3]; similar are methods that group adjacent image voxels in boxes [14], [24]; 2) wavelet transformation, which has been used to represent the morphological features at multiple scales; 3) template warping techniques [20], [21] measuring volumes within specific anatomical regions of interest (ROIs) and tessellation algorithms [13] describing shapes using small patches. However, most of these regional feature extraction approaches group local morphological features from fixed ROIs, based on Cartesian grids at various scales or anatomical regions, or based on tessellations that do not necessarily group

together regions displaying similar characteristics with respect to the classification variable. Therefore, these approaches are not necessarily adaptive to the particular disease under study, and they might generate volumetric measurements that blend abnormal and healthy tissue in a single ROI, or that split a single abnormal region into multiple ROIs, thereby enhancing the effects of noise and reducing predictive power.

The local morphological features should be adaptively grouped according to the problem under study, since different pathological processes might affect brain regions in different ways. Thus, the affected regions might have irregular shapes and are unknown in advance. Accordingly, a watershed segmentation algorithm is used herein for adaptively generating regions according to a discrimination and robustness measure (DRM) of each local morphological feature. The watershed segmentation algorithm is a traditional image segmentation approach, widely utilized in medical image analysis for partitioning images into different regions according to local intensity similarity [22], [25]. Here, the watershed segmentation method is used to partition a brain into different regions according to the similarity of DRM of local features. In the following, we define the DRM for local features.

For each voxel-wise morphological feature, its DRM is highly related to its discriminative power and its spatial consistency that is generally a measure of robustness as well as of spatial uniformity. The discriminative power of a feature can be quantitatively measured by its relevance to classification as well as its generalization ability. The relevance of a feature to classification can be measured by the correlation between this feature and the corresponding class label in a training dataset (i.e., normal -1 or pathological $+1$). In machine learning and statistical analysis, the correlation measures can be broadly divided into linear correlation and nonlinear correlation. Most nonlinear correlation measures are based on the information-theoretical concept of entropy, such as mutual information, computed by probability estimation. For continuous features, probability density estimation is a hard task especially when the number of available samples is limited. On the other hand, linear correlation measures are easier to compute even for continuous features and are robust to over-fitting, thus they are widely used for feature selection in machine learning [19]. Here, we used the Pearson correlation coefficient, whose use in feature selection is closely related to that of the t -test [19], to measure the relevance of each feature to classification. The larger the absolute value of Pearson correlation coefficient is, the more relevant to classification this feature is. Given a location, u , in the template space, the Pearson correlation coefficient between a feature, $f^i(u)$, of tissue i and class label y is defined as

$$\rho^i(u) = \frac{\sum_j (f_j^i(u) - \overline{f^i(u)}) (y_j - \overline{y})}{\sqrt{\sum_j (f_j^i(u) - \overline{f^i(u)})^2 \sum_j (y_j - \overline{y})^2}} \quad (1)$$

where j denotes the j th sample in the training dataset. Thus, $f_j^i(u)$ is a morphological feature of tissue i in the location u of j th sample, and $\overline{f^i(u)}$ is the mean of $f_j^i(u)$ over all samples. Similarly, y_j is a class label (normal -1 or pathological

¹<http://www.fil.ion.ucl.ac.uk/spm/software>

+1) of the j th sample, and \bar{y} is the mean of y_j over all samples. In addition to the relevance, the generalizability of a feature is equally important for classification, especially in applications with high dimensionality relative to the sample size, such as ours. A bagging strategy [26] is adopted to take the generalization ability into account, when measuring the discriminative power of a feature by Pearson correlation coefficient. That is, given n training samples, a leave-one-out procedure is used to measure the discriminative power of each feature, $f^i(u)$, by a conservative principle, i.e., selecting the worst discriminative power resulting from n leave-one-out measurements of the correlation coefficient. We can formulate this conservative definition for generalizability of a feature, $f^i(u)$, as

$$P^i(u) = \arg \min_{\{\rho_j^i(u) | 1 \leq j \leq n\}} |\rho_j^i(u)| \quad (2)$$

where $\rho_j^i(u)$ is the Pearson correlation coefficient between the feature $f^i(u)$ and the class label y at location u of tissue map i , from the j th leave-one-out case where the j th sample is excluded. The definition of $\rho_j^i(u)$ is similar to the definition of $\rho^i(u)$ in formula (1), except that the j th sample is excluded for correlation coefficient statistical calculation. The definition above helps exclude the outliers in the data. This is particularly important for dealing with the small sample size problem in our study, since outliers can be found by pure chance alone when examining so many voxels and their respective correlation coefficients. Other robust measurements, such as robust correlation [27], can be also used if these measurements can be computed very efficiently and are applicable to small sample size problems like ours.

The spatial consistency of voxel-wise features is another important issue in classification, which is directly related to the robustness of a feature, since voxel-wise morphological features are locally extracted and thus might not be reliable due to registration errors and inter-individual anatomical variations. A feature is spatially consistent if it is similar to other features in its spatial neighborhood, implying that small registration errors will not significantly change the value of this feature. The spatial consistency of a voxel-wise feature can be measured by the degree of agreement among all features in its spatial neighborhood, which can be computed by an intraclass correlation coefficient from the training samples [28]. For example, let us assume that we have n training samples, and consider the spatial consistency among m neighbors around the location, u , of the tissue-density map, i . If the immediate neighborhood is considered, the total number of neighbors is $m = 27$. Thus, for each location, u , in the tissue-density map, i , we can construct a $n \times m$ feature matrix, i.e., $[f_{j,k}^i(u)]$, $j = 1, \dots, n$, $k = 1, \dots, m$, where $f_{j,k}^i(u)$ is the tissue-density value of the j th training sample at the location of the k th neighbor around location u . In order to measure the amount of agreement among the features, a two-way random effect model is specified for this feature matrix, i.e.,

$$f_{j,k}^i = \mu + r_j + c_k + e_{j,k}, \quad j = 1, \dots, n, \quad k = 1, \dots, m,$$

where μ is the grand mean for all density values in the matrix. (For simplicity we have omitted the dependency of this random effect model on i , keeping in mind that spatial consistency is examined for all tissue density maps.) In the equation above, r_j , $j = 1, \dots, n$, are the row effect independent random variables, with mean 0 and variance σ_r^2 . c_k , $k = 1, \dots, m$, are the column effect independent random variables, with mean 0 and variance σ_c^2 . $e_{j,k}$, $j = 1, \dots, n$, $k = 1, \dots, m$, are the residual effect independent random variables, with mean 0 and variance σ_e^2 . Based on above model, the spatial consistency, $C^i(u)$, of a feature, $f^i(u)$, can be computed by

$$C^i(u) = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_c^2 + \sigma_e^2}$$

according to [28]. The spatial consistency, $C^i(u)$, can be estimated from the feature matrix, $[f_{j,k}^i(u)]$, $j = 1, \dots, n$, $k = 1, \dots, m$, by calculating a mean square for rows (MS_R), a mean square for columns (MS_C), and a mean square error for residuals (MS_E)

$$\begin{aligned} C^i(u) &= \frac{MS_R - MS_E}{MS_R + (m-1)MS_E + \frac{m}{n}(MS_C - MS_E)}, \\ MS_R &= m \left(\sum_{j=1}^n r_j - g \right)^2 / (n-1), \\ MS_C &= n \left(\sum_{k=1}^m c_k - g \right)^2 / (m-1), \\ MS_E &= \left(\sum_{j=1}^n \sum_{k=1}^m f_{j,k}^i(u)^2 - (n-1)MS_R - (m-1)MS_C \right. \\ &\quad \left. - m \cdot n \cdot g^2 \right) / ((m-1)(n-1)), \\ r_j &= \sum_{k=1}^m f_{j,k}^i / m, \\ c_k &= \sum_{j=1}^n f_{j,k}^i / n, \\ g &= \sum_{j=1}^n \sum_{k=1}^m f_{j,k}^i / (m \cdot n). \end{aligned} \quad (3)$$

In our applications, the value of $C^i(u)$ is constrained to be between 0 and 1.

In addition to helping alleviate the impact of the potential registration error, the spatial consistency is an important step in the whole process for additional reasons that will become clearer below. Briefly, it serves the purpose of a “statistical group-wise edge detector,” in that it analyzes the tissue density maps of a number of individuals, and identifies regions displaying different behavior, in a statistical sense. These regions will be used in the following section by a watershed segmentation algorithm in order to form boundaries of clusters of voxels showing similar properties.

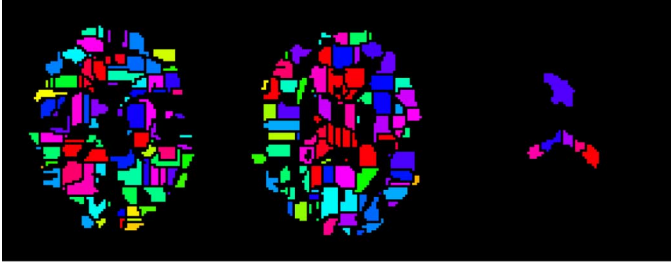


Fig. 2. Cross-sectional views of automatically generated brain regions from WM, GM, and CSF tissue density maps (from left to right).

As described above, the discriminative power, measured by the absolute value of $P^i(u)$, and the spatial consistency, measured by $C^i(u)$, have nonnegative values with high score indicating a better feature for classification. We combine these two measurements into one by the following equation:

$$s^i(u) = P^i(u)C^i(u) \quad (4)$$

thus, obtaining a single score, $s^i(u)$, for each feature, $f^i(u)$. The absolute value of $s^i(u)$ reflects the suitability of a feature, $f^i(u)$, for classification. Three score maps are produced for GM, WM, and CSF, respectively.

By calculating the gradient map of a score map, $s^i(u)$, and using it in conjunction with a watershed segmentation algorithm, we partition a brain into R^i different regions, i.e., $\{r_l^i, 1 \leq l \leq R^i\}$. In order to avoid over-segmentation, Gaussian smoothing is applied to the score map before computing its gradient map. By applying the watershed segmentation algorithm to the score map of each tissue, we finally obtain separate partitions for each tissue. Typical brain region partition results, with regions generated from each specific tissue-density map, are shown in Fig. 2.

3) *Extraction of Regional Features From Adaptively Generated Regions:* A simple way to use the watershed-derived regional volumetric elements would be to sum all tissue density values in each region, yielding a volumetric measure, $v_{l,j}^i$, corresponding to l th region in j th sample. Such volumetric measures from all WM, GM, and CSF regions could constitute a feature vector to represent morphological information of the brain, which is robust to noise, registration error, and interindividual anatomical variation. However, using all voxels in the region to compute the overall volumetric measure might decrease the discriminative power of this region for classification, since the watershed segmentation provides a rough partitioning of the space, without directly optimizing classification performance obtained from the derived regions. On the other hand, the watershed-based region partition can provide a good initialization for the next step of selecting a subvolume from each watershed region by optimizing the classification power of the features extracted from the selected subvolume. For this purpose, we use a selective volumetric increment method based on a similar idea to that of the forward feature selection method. We first select a voxel with the highest discriminative power in each region under consideration. Then, we start to include each neighboring voxel,

under the condition that inclusion of this neighboring voxel will not decrease the discriminative power of regional feature calculated from the voxels currently selected. This procedure is iterated, similarly to a traditional region growing method, until no more voxels can be added to the set of selected voxels. The discriminative power of a regional feature is measured by the absolute value of the Pearson correlation coefficient between this regional feature and the class label, similar to (2)

$$P(V_l^i) = \min_{1 \leq j \leq n} |\rho_j(V_l^i)| \quad (5)$$

where V_l^i is a regional feature generated from the l th region of tissue map i , r_l^i . $\rho_j(V_l^i)$ is the Pearson correlation coefficient between regional feature V_l^i and class label y in the j th leave-one-out case where the j th sample is excluded. This regional feature extraction algorithm is summarized next.

Volume increment feature extraction algorithm:

Input: Regions $\{r_l^i, l = 1, 2, \dots, R^i, i = 1, 2, 3\}$ respectively generated for three brain tissues by a watershed segmentation algorithm; and tissue-density maps $\{f_j^i, i = 1, 2, 3, j = 1, 2, \dots, n\}$ corresponding to n training samples.

Output: Regional features $\{(V_l^i)_j, l = 1, 2, \dots, R^i, i = 1, 2, 3, j = 1, 2, \dots, n\}$, where $(V_l^i)_j$ is a regional feature calculated from the l th region of tissue density map i of the j th sample; and a set of partial regions $\{U_l^i, l = 1, 2, \dots, R^i, i = 1, 2, 3\}$ used to extract a regional feature in each region. (Note that we use $(V_l^i)_j$ as a regional feature value of V_l^i in the j th sample.)

Begin:

For each region $r_l^i, i = 1, 2, 3, l = 1, \dots, R^i$:

1. Select a voxel-based morphological feature with the highest DRM in the region r_l^i , i.e., at location u . Thus, the regional feature for each sample is $\{(V_l^i)_j = f_j^i(u), j = 1, 2, \dots, n\}$, and the set of selected voxels is $U_l^i = \{u\}$.

2. Repeat

For each voxel \tilde{u} in the region r_l^i , if it is not included in U_l^i and it is a neighbor of one voxel in U_l^i , add this voxel \tilde{u} into U_l^i , obtaining a new set of selected voxels $\tilde{U}_l^i = U_l^i \cup \{\tilde{u}\}$, and then update the regional features for each sample by $(\tilde{V}_l^i)_j = \sum_{w \in \tilde{U}_l^i} f_j^i(w) / \text{card}(\tilde{U}_l^i), j = 1, 2, \dots, n$.

```

    If  $P(\tilde{V}_l^i) \geq P(V_l^i)$ ,
        Then, set  $\{(V_l^i)_j = (\tilde{V}_l^i)_j, j = 1, 2, \dots, n\}$  and  $U_l^i \leftarrow \tilde{U}_l^i$ .
    End if
  End for
Until no more voxels in the region  $r_l^i$  can be added into  $U_l^i$ .
End for
End

```

B. Feature Selection Using SVM-Based Criteria

Although the number of the above generated regions is much smaller than the number of original voxels, measures from some regions are less effective, irrelevant and redundant for classification. This requires a feature selection method to select a small set of features in order to improve generalization and performance of classification.

Generally, feature selection methods can be divided into feature ranking methods and feature subset selection methods; the latter can be further divided into filters, wrappers and embedded methods [19]. The feature ranking methods compute a ranking score for each feature according to its discriminative power, and then simply select top ranked features as final features for classification. These feature selection methods are preferable for high dimensional problems, due to their computational scalability. However, the subset of features selected by the feature ranking methods might contain a lot of redundant features, since the ranking score is computed independently for each feature, by completely ignoring its correlation with others. On the contrary, the feature subset selection methods focus on selecting a subset of features that jointly have better discriminative power. In general, sophisticated subset selection methods have better classification performance than feature ranking methods, but their high computational cost usually limits their applications to the high dimensional problems. We present a method which combines the advantages of both the feature ranking method and the feature subset selection method, as detailed next.

1) *Correlation Based Feature Ranking*: The Pearson correlation coefficient has been successfully employed as a feature ranking criterion in a number of applications [19]. In order to achieve better generalization, the absolute value of leave-one-out Pearson correlation coefficient computed by formula (5) is used here to rank features. Since the correlation among regional features has been completely ignored in this feature ranking method, some redundant features can be inevitably selected, which ultimately affects the classification results as demonstrated in the Results section. Therefore, in COMPARE this feature selection method provides a good set of initial features, to be further optimized by the feature subset selection method, as explained below.

2) *SVM-Based Algorithm for Selecting a Subset of Features From Top-Ranked Features*: SVM-based feature selection

methods have been successfully applied in a variety of problems. One good example is the support vector machine-recursive feature elimination (SVM-RFE) algorithm which was initially proposed for a cancer classification problem [29], and was later extended by introducing SVM-based leave-one-out error bound criteria in [30]. The goal of SVM-RFE is to find a subset of size n among d features ($n < d$) that optimizes the performance of the classifier. This algorithm is based on a backward sequential selection method that removes one feature at a time. At each time, the removed feature makes the variation of SVM-based leave-one-out error bound smallest, compared to removing other features. In order to apply this subset selection method to our problem in reasonable computation time and to avoid local optima, we first use the proposed correlation-based feature ranking method to select the most relevant features, and then apply the SVM-RFE algorithm on the set of selected features. However, such a procedure may miss informative features whose rank is low, but which perform well jointly with the top ranked features for classification. To partially solve this problem, starting from the same initial feature subset, a forward sequential feature selection method is applied, which adds one feature at a time. At each time, the added feature makes the variation of SVM-based leave-one-out error bound smallest, compared to adding other features. The search space of the forward selection is limited to a predefined feature subset in order to obtain a solution with reasonable computation cost.

C. SVM-Based Classification

The nonlinear support vector machine is a supervised binary classification algorithm [31]. SVM constructs a maximal margin linear classifier in a high dimensional feature space, by mapping the original features via a kernel function. The Gaussian radial basis function kernel is used in COMPARE, which is defined as

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (6)$$

where x_1 and x_2 are two feature vectors, and σ controls the size of the Gaussian kernel.

SVM is not only empirically demonstrated to be one of the most powerful pattern classification algorithms, but also has provided many theoretical generalization bounds to estimate its capacity, for example, the radius/margin bound, which could be utilized in feature selection. Another reason for us to select the SVM as a classifier is its inherent sample selection mechanism, i.e., only support vectors affect the decision function, which may help us find subtle differences between groups.

III. EXPERIMENTAL RESULTS

A. Testing Classification Performance

We tested the performance of COMPARE on two datasets of MR T1 brain images, with the goal of comparing the brain differences between schizophrenia patients and healthy controls. The ‘‘dataset A’’ is on female subjects, with 23 schizophrenia patients (SC) and 38 normal controls (NC), while ‘‘dataset B’’ is on male subjects, with 46 SC and 41 NC, which were described in [32]. For each of these MR T1 images, tissue density maps

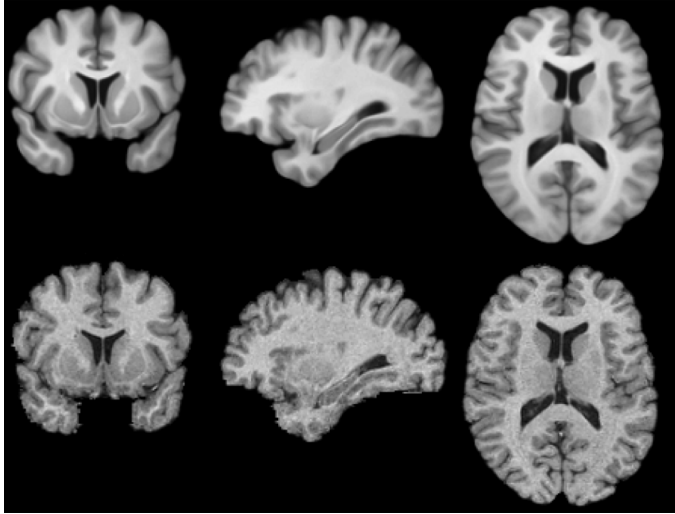


Fig. 3. Average image (top row) of 148 images in datasets A and B after they were spatially normalized via elastic warping to the template (bottom row). Clarity of the average image visually reflects good coregistration of different tissues.

were generated via tissue classification [23] and elastic transformations [20], [21]. In order to improve the signal-to-noise ratio of the tissue density maps of these brain images, and to account for potential local registration errors, a 9-mm full-width half-maximum (FWHM) Gaussian smoothing was applied to all tissue density maps. Since the regional features are generated from tissue density maps in the template space, the accuracy of the registration is important for the accuracy of regional features. Registration accuracy of the HAMMER algorithm has been extensively evaluated elsewhere [20], [21]. In order to visually confirm registration performance in the datasets used in this study, we formed an average image of all 148 spatially normalized images in datasets A and B after they were warped to the template. Three different views of the average image along with the respective sections of the template image are shown in Fig. 3, indicating good registration achieved for each of the three tissues, i.e., GM, WM, and CSF.

Although a three-way split validation is the best way to estimate the classification accuracy, the high computational cost and the relative small number of available samples prevent us from such a validation test. Instead, a leave-one-out cross-validation was performed in our experiments to test the performance of COMPARE. In each leave-one-out validation experiment, one subject was first selected as a testing subject, and the remaining subjects were used for the entire adaptive regional feature extraction, feature selection, and training procedure, as described in Section II. Then, the classification result on the testing subject using the trained SVM classifier was compared with the ground-truth class label, to evaluate the classification performance. By repeatedly leaving each subject out as a testing subject, we obtained the average classification rate from all of these leave-one-out experiments. Finally, these experiments were repeated for different numbers of features used for classification, in order to test the stability of classification results with respect to the number of features used.

To determine the suitable kernel size for our classification problem, we tested different kernel sizes ranging from 0.1 to 100, and found that kernel sizes ranging from 1 to 10 generally yield better results. As for the value of “C,” the tradeoff parameter between training error and SVM margin, used in the SVM classifiers, we tested 1, 10, 50, and 100, and found that 10 and 50 were the better choices for our problems. Notably, when “C” was too big, the classifiers were typically overtrained and had poor generalization ability. By using a SGI Origin 300 workstation with 4 GB memory, with parameters fixed, it takes around 75 h to finish a leave-one-out cross-validation for dataset A, and around 115 hours for dataset B.

The best average correct classification rates were 91.8% (with 20 out of 23 SCs and 36 out of 38 NCs correctly classified) by using 39 features for dataset A, and 90.8% (with 44 out of 46 SCs and 35 out of 41 NCs correctly classified) by using 44 features for dataset B, as shown in right column of Fig. 4. Although a reasonably good performance was achieved by only using a ranking-based feature selection method, as shown in left column of Fig. 4, more stable performance was achieved by incorporating the SVM-based feature selection method (right column of Fig. 4), since the relatively simpler ranking-based feature selection method does not consider correlations among features. Furthermore, these plots indicate that the described algorithm is quite robust with respect to the size of Gaussian kernel used in SVM.

To further show the performance of COMPARE, the receiver operating characteristics (ROC) curves of the leave-one-out classifiers that yield the best classification results are constructed, and are shown in Fig. 5. The ROC curves indicate that our classification scheme has large area under ROC curve (0.88 for dataset A, and 0.92 for dataset B).

From these results, it can be observed that our method has different performances on dataset A and dataset B. These differences might be due to the brain structural difference between male and female in both normal controls and schizophrenia patients [33]–[36]. A relatively small number of training samples for dataset A might be another reason that makes our classification method have relative worse performance on female.

In order to interpret the classification results, we utilize a discriminative direction method, as used in [10] and [14], to estimate the group differences between schizophrenia patients and normal controls. Since a leave-one-out validation is performed in our experiments for testing the generalizability of COMPARE, the group difference will be constructed by averaging all group difference maps obtained from all leave-one-out cases. For each leave-one-out case, the group difference map is estimated by three steps as described next. First, for each support vector, look for its corresponding projection vector on the other side of separation hypersurface, by following the steepest gradient of classification function. The difference between this support vector and its corresponding projection vector on other side reflects changes on the selected regional features when a normal brain changes to the respective configuration in the patient group, or vice versa. Second, by summing up all regional differences calculated from all support vectors, an overall group difference vector can be obtained for the current leave-one-case under study. Finally, the group difference vector is mapped to

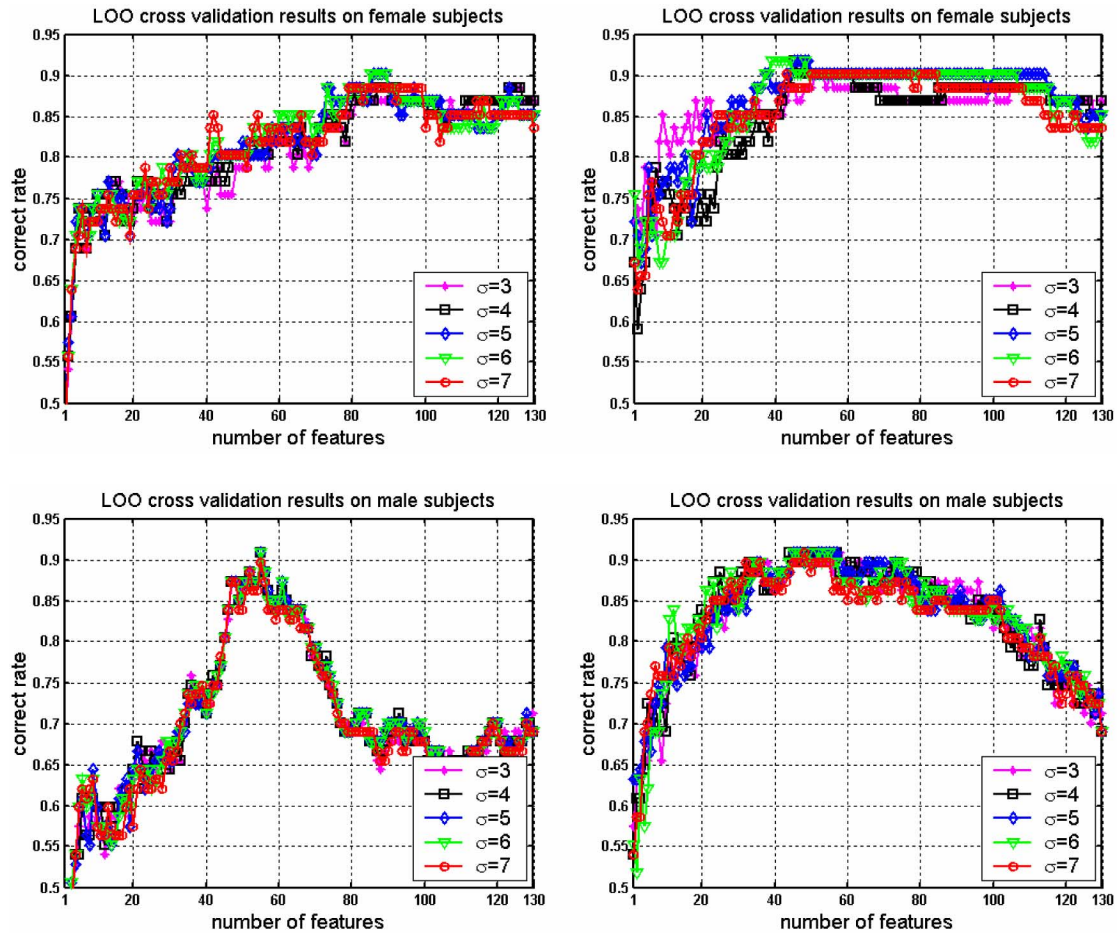


Fig. 4. Classification performance of ranking-based feature selection (left column) and SVM-based feature selection (right column) for female subjects (top row) and male subjects (bottom row). Plotted are the average classification rates with respect to different kernel sizes and different feature numbers used in the SVM. SVM-based algorithm performs a robust selection of features and leads to relatively more stable performance.

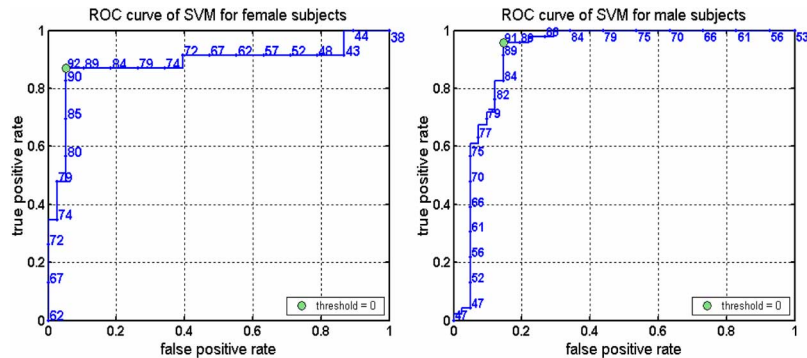


Fig. 5. ROC curves of classifiers for female subjects (left plot) and for male subjects (right plot). Numbers around the curves are the correct classification rates (%). Circled points on the curves correspond to the classification results with zero as the classification threshold.

its corresponding brain regions in the template space and subsequently added to other leave-one-out repetitions. The group difference maps for dataset A and B are shown in Fig. 6, which highlight the most significant and frequently detected group differences in our leave-one-out experiments. Most of the group difference locations are consistent with previous VBM findings, such as hippocampi [32].

It is worth noting that the proposed method provides a statistical map for group difference. In the leave-one-out experi-

ments, the feature selection method might discard some significant features that are redundant with respect to classification, if they are highly related to other significant features that the feature selection method has already selected. Notably, one significant feature which is regarded as redundant in a leave-one-out case might be selected as an effective feature in other leave-one-out cases. Thus, in all leave-one-out experiments, all significant features can be possibly selected as effective features, thereby contributing for the construction of group difference maps.

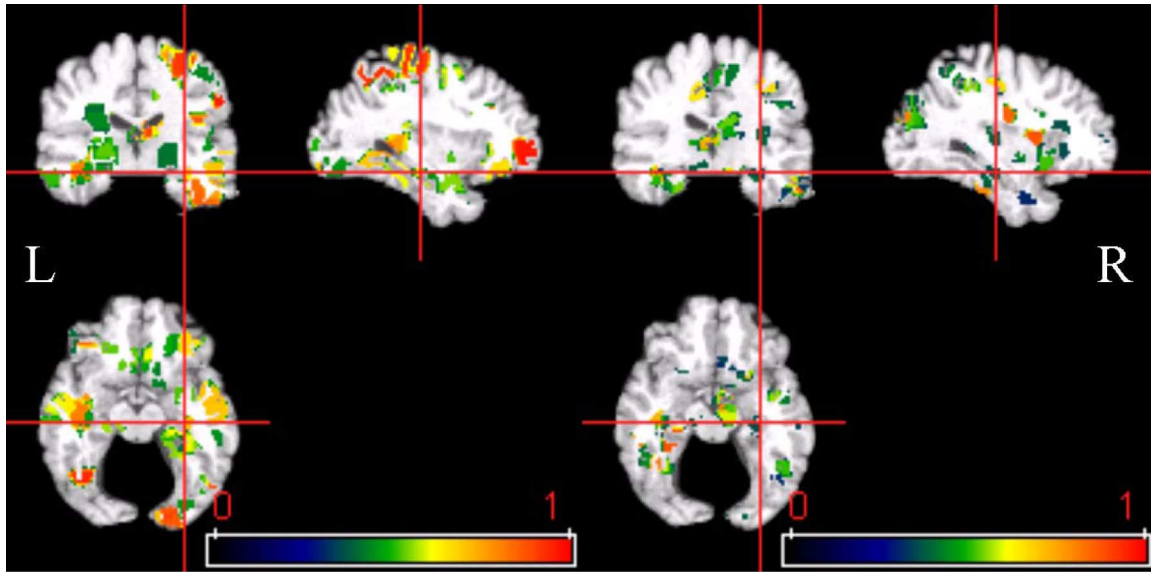


Fig. 6. Regions most representative of the group differences (left from female subjects and right from male subjects), found via decision function gradient (high value indicates more significant).

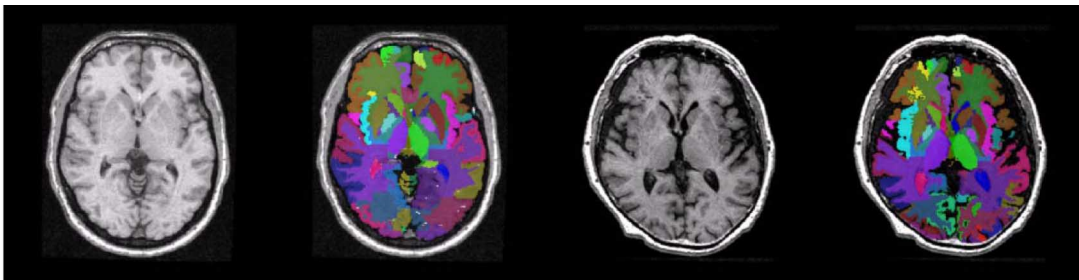


Fig. 7. From left to right: Template image, ROIs manually defined in the template, individual image, ROIs in the individual image automatically labeled by atlas warping.

B. Comparison With Other Feature Extraction Methods

For comparison purposes, we also applied three other feature extraction methods to the same data, i.e., ROI-based volumetric feature extraction, PCA-based feature extraction, and wavelet-based feature extraction. As for step of feature selection, traditional eigenvalue-based feature ranking is adopted for PCA-based feature extraction method, while the absolute value of leave-one-out Pearson correlation coefficient was used for both wavelet-based feature extraction method and ROI-based volumetric feature extraction method. The same 9-mm FWHM Gaussian smoothing was applied to all tissue density maps in all experiments, except for experiments on the wavelet-based feature extraction method, since wavelet transformation is able to represent data in a multiscale fashion. A nonlinear SVM with Gaussian radial basis function kernel was chosen as a classifier for all experiments. The brief description of these feature extraction methods are given below.

- In ROI-based volumetric feature extraction method, the ROIs were first defined in the template space. We adopted a labeled atlas developed by Noor Kabani at the Montreal Neurological Institute, which includes 100 ROIs over the entire brain. Representative cross-sectional views of the atlas along with the atlas warping based parcellation of an individual's brain image are shown in Fig. 7. Three

volumes are computed for each ROI, respectively for WM, GM, and CSF, thereby yielding 300 volumetric features. (Many of these features are 0, since most structures are composed of a single tissue). Finally, these features are ranked in descending order according to the absolute values of their leave-one-out Pearson correlation coefficients computed by (5). The top-ranked features are used for classification.

- In PCA-based feature extraction method, the measures from WM, GM and CSF tissue-density maps are concatenated into a long vector. Then, the eigenspace is constructed by a set of training samples, which is used to compute the features for a new sample. The corresponding features are ranked in descending order according to eigenvalues.
- As for wavelet-based feature extraction, we follow the procedure described in [15]. Daubechies wavelet decomposition is first applied to the tissue density maps, leading to a scale-space representation for each tissue-density map. Then, the features are ranked in descending order according to the absolute value of their leave-one-out Pearson correlation coefficients computed by formula (5). Finally, the top-ranked wavelet features are used for brain classification.

TABLE I
COMPARISON ON DIFFERENT FEATURE EXTRACTION METHODS IN BRAIN CLASSIFICATION. THE NUMBER OF CORRECTLY CLASSIFIED SCHIZOPHRENIA PATIENTS, THE NUMBER OF CORRECTLY CLASSIFIED NORMAL CONTROLS, AND THE OVERALL CLASSIFICATION RATE ARE PROVIDED FROM LEFT TO RIGHT, RESPECTIVELY

Methods Datasets	ROI-based Features	PCA	Wavelet-based features	Proposed method
Dataset A (23 SCs and 38 NCs)	16, 33, (80.3%)	17, 33, (82.0%)	17, 36, (86.9%)	19, 36, (90.2%)
Dataset B (46 SCs and 41 NCs)	38, 32, (80.5%)	32, 32, (73.6%)	41, 33, (85.1%)	44, 35, (90.8%)

TABLE II
COMPARISON ON DIFFERENT FEATURE EXTRACTION METHODS IN BRAIN CLASSIFICATION USING LINEAR SVM. THE NUMBER OF CORRECTLY CLASSIFIED SCHIZOPHRENIA PATIENTS, THE NUMBER OF CORRECTLY CLASSIFIED NORMAL CONTROLS, AND THE OVERALL CLASSIFICATION RATE ARE PROVIDED FROM LEFT TO RIGHT, RESPECTIVELY

Methods Datasets	ROI-based Features	PCA	Wavelet-based features	Proposed method
Dataset A (23 SCs and 38 NCs)	15, 32, (77.1%)	19, 30, (80.3%)	16, 37, (86.9%)	18, 36, (88.5%)
Dataset B (46 SCs and 41 NCs)	32, 30, (71.3%)	32, 30, (71.3%)	39, 31, (80.5%)	42, 35, (88.5%)

Nonlinear SVM classifiers were trained and tested with full leave-one-out cross-validation procedures. Different feature numbers and SVM kernel sizes were tested for determining the best parameters for classification. The best classification results for these feature extraction methods are compared in Table I, along with classification results by COMPARE, which is based only on ranking-based feature selection.

Specifically, a search over all possible numbers of principal components has been done to determine the best number of principal components for classification. The coefficient corresponding to each component has been normalized by its variance. The best result for dataset A was obtained by using the first 39 components, and the best result for dataset B was obtained by using the first nine components. For the wavelet-based method, a search has been done over the 600 top-ranked wavelet features. The best result for dataset A was obtained by using 550 features and the best result for dataset B was obtained by using 210 features. For the ROI based method, a full search has been done over all ROI features. The best result for dataset A was obtained by using 5 features and the best result for dataset B was obtained by using 17 features. To further test the performance of ROI-based feature extraction, the larger ROIs were regularly split into eight or four subregions, thus generating totally 747 small ROIs in our atlas. Based on this set of smaller ROIs, the best results were 83.6% for female subject classification and 80.5% for male subject classification. Note that, for female subject classification, its performance was slightly better than using 100 original ROIs, i.e., 80.3%, while for male subject classification, its performance was the same as using 100 original ROIs. We believe that the number of regions available for classification is not very important. The real important point is how to generate adaptive regions for extracting robust and discriminative regional features for brain classification, as we proposed in the paper.

To better understand the performance of these different feature extraction methods, linear SVM classifiers were also tested with the same leave-one-out cross-validation procedure, as we described above. The best number of features used for classification was also obtained based on ranking-based feature selection method. The best classification result for each of these methods is summarized in Table II.

The results, from both nonlinear SVM and linear SVM, indicate the importance of extracting regional features from the regions adaptively generated according to the problem under study. Generally, the ROI-based feature extraction method produces robust morphological features with respect to noise. However, these ROI-based features seem to have relatively lower discriminative powers according to the classification results, since the predefined anatomical ROIs do not necessarily coincide with the shapes of the brain regions affected by a particular pathological process. On the other hand, it seems that PCA-based feature extraction method does not work well in our small sample size problem, since PCA is unable to capture sufficient information from a relatively small set of samples with extremely high dimensional features. Wavelet-based feature extraction has better performance, compared to ROI-based and PCA-based methods; however, it is still worse than COMPARE. Finally, although these comparisons are carried out based on a ranking-based feature selection methods for simplicity purposes, they suggest that a learning-based regional feature extraction is more suitable for schizophrenia brain classification.

C. Evaluation on Different Regional Feature Extraction Ways

We also evaluated our regional feature extraction step, by comparing it with two other possible regional feature extraction ways, i.e., 1) removing the spatial consistency criterion in computing classification score map, 2) extracting the best voxel-wise

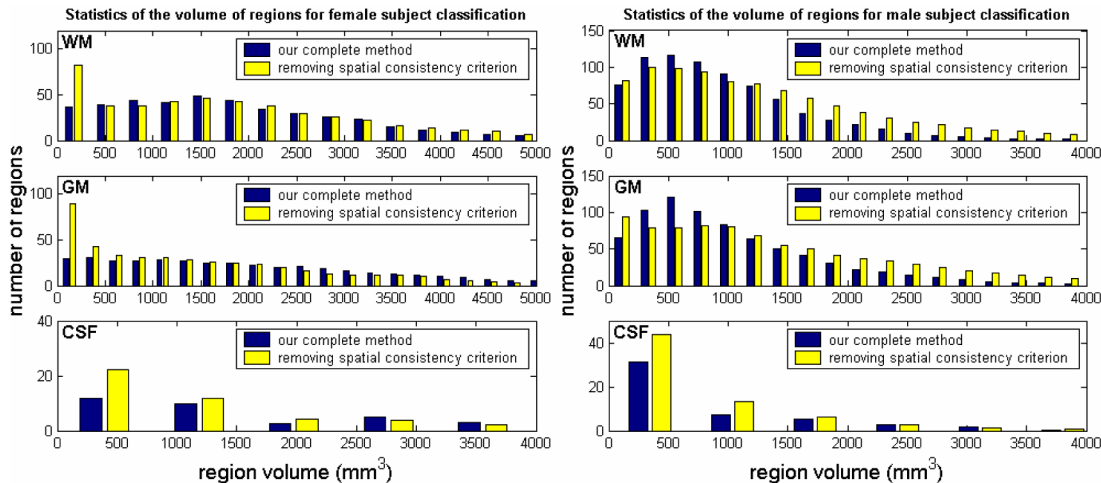


Fig. 8. From left to right: statistics of the volume of automatically generated regions for female and male subject classification.

TABLE III
BRAIN CLASSIFICATION PERFORMANCE OF DIFFERENT IMPLEMENTATION OPTIONS. THE CORRECTLY CLASSIFIED NUMBER OF SCHIZOPHRENIA PATIENTS, THE CORRECTLY CLASSIFIED NUMBER OF NORMAL CONTROLS, AND OVERALL CLASSIFICATION RATES ARE PRESENTED FROM LEFT TO RIGHT, RESPECTIVELY

Methods \ Datasets	Without spatial consistency	Voxel-wise features	Proposed method
Dataset A (23 SCs and 38 NCs)	18, 36, (88.5%)	17, 33, (82.0%)	19, 36, (90.2%)
Dataset B (46 SCs and 41 NCs)	38, 33, (81.6%)	43, 29, (82.8%)	44, 35, (90.8%)

feature from each generated region instead of extracting average feature from a subvolume in each generated region. For the first case, we found that, without using this spatial consistency criterion, a lot of small regions are generated (Fig. 8), since the score map produced by using only the Pearson correlation criterion is noisy. On the other hand, the spatial consistency criterion can help alleviate the impact of noise in computing the classification score map and thus generating reasonable sizes of regions, as shown in Fig. 8. Based on the regional features extracted from these newly generated regions, we can examine the performances of brain classification by the same leave-one-out cross-validation procedure as described above. As shown in Table III, for dataset A, the best classification rate was 88.5%, which was slightly worse than the result obtained by our complete method (90.2%); for dataset B, the best classification rate was 81.6%, which was much worse than the result obtained by our complete method (90.8%).

For the second case, we directly selected the best voxel-wise feature in each generate region of each brain tissue, as a regional feature used for brain classification. By using the same leave-one-out cross-validation procedure, we found that the best classification rates were 82.0% for dataset A and 82.8% for dataset B, which are much worse than those obtained by our complete classification method. These results are again shown in Table III.

These comparison results indicate that our complete regional feature extraction method is more effective to alleviate the ad-

verse impact of noise, which is prominent in MR images as well as in deformation-based methods, especially in very high-dimensionality data that amplify the likelihood of selecting features by pure chance.

IV. DISCUSSION AND CONCLUSION

We presented a pattern classification method for identification of structural brain abnormalities based on regional tissue volumetric information. The experimental results indicate that COMPARE can achieve high classification rate in a schizophrenia study. Further studies are necessary to investigate whether this methodology can assist in the early diagnosis of schizophrenia, as well as of other diseases. COMPARE also provides an alternative approach for constructing spatial maps of structural group differences, as discussed in more detail in [11], [15] and shown in Fig. 6.

The nonlinear SVM classifier is built on a morphological representation of the brain, adaptive extraction of regional features, and robust selection of features. In particular, the morphological information is represented by brain tissue-density maps that are derived by warping an individual brain into a template space in a mass-preserving framework using high-dimensional image warping. Based on this morphological representation, the regional features are adaptively extracted and further selected for brain classification.

The adaptive regional feature extraction method employed by COMPARE aims at overcoming the limitations of the traditional

ROI method that are often based on prior knowledge of what specific regions might be affected by disease, and the limitations of voxel-based morphometric analysis methods [3], [6] that use an identical isotropic Gaussian filter to collect regional morphological information in all brain locations. Also, our regional feature extraction method is different from the PCA-based feature extraction method that collects global features, and wavelet-based feature extraction methods that group local features in various regions with fixed shapes and sizes determined by the respective wavelets. In contrast, in COMPARE, volumetric measurements are obtained from brain regions that are adaptively generated by grouping local morphological features with similar DRM through a watershed segmentation algorithm. In order to achieve robustness to outliers, the DRM of each feature is measured by a leave-one-out cross-validation framework. The resulting regional features are stable, thereby leading to good generalization performance of the classifier.

A limitation of our current study has been the relatively limited sample size, compared to the dimensionality of the structural measurements. Although the leave-one-out cross-validation accuracy obtained may be optimistic, the limited sample size did not allow us to explore other cross-validation techniques, since we would under-train COMPARE. Our sample was quite diverse, and it included both sexes and all ages between 18 and 49. However, our results must be replicated in the future with larger datasets. Moreover, prospective hypothesis-based studies, using the regions detected in this sample, can be designed.

In the future, we plan to perform more sophisticated feature grouping and feature selection methods, for further improving classification performance. In particular, we plan to investigate the use of both floating searching based and stochastic search based feature selection methods [37], [38]. Moreover, we plan to develop an integrated framework to simultaneously extract and select effective regional features for classification, since feature extraction, feature selection, and classification are currently implemented independently. Finally, we are testing the performance of COMPARE on other patient groups.

REFERENCES

- [1] N. R. Giuliani, V. D. Calhoun, G. D. Pearlson, A. Francis, and R. W. Buchanan, "Voxel-based morphometry versus region of interest: A comparison of two methods for analyzing gray matter disturbances in schizophrenia," *Schizophrenia Res.*, vol. 74, pp. 135–147, 2005.
- [2] P. M. Thompson, D. MacDonald, M. S. Mega, C. J. Holmes, A. Evans, and A. W. Toga, "Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces," *J. Comput. Assisted Tomogr.*, vol. 21, pp. 567–581, 1997.
- [3] J. Ashburner and K. J. Friston, "Voxel-based morphometry: The methods," *NeuroImage*, vol. 11, pp. 805–821, 2000.
- [4] M. K. Chung, K. J. Worsley, T. Paus, C. Cherif, D. L. Collins, J. N. Giedd, J. L. Rapoport, and A. C. Evans, "A unified statistical approach to deformation-based morphometry," *NeuroImage*, vol. 14, pp. 595–606, 2001.
- [5] C. Davatzikos, A. Genc, D. Xu, and S. M. Resnick, "Voxel-based morphometry using the RAVENS maps: Methods and validation using simulated longitudinal atrophy," *NeuroImage*, vol. 14, pp. 1361–1369, 2001.
- [6] C. D. Good, R. I. Scahill, N. C. Fox, J. Ashburner, K. J. Friston, D. Chan, W. R. Crum, M. N. Rossor, and R. S. J. Frackowiak, "Automatic differentiation of anatomical patterns in the human brain: Validation with studies of degenerative dementias," *NeuroImage*, vol. 17, pp. 29–46, 2002.
- [7] J. Ashburner, J. G. Csernansky, C. Davatzikos, N. C. Fox, G. B. Frisoni, and P. M. Thompson, "Computer-assisted imaging to assess brain structure in healthy and diseased brains," *Lancet (Neurology)*, vol. 2, pp. 79–88, 2003.
- [8] J. G. Csernansky, S. Joshi, L. Wang, J. W. Haller, M. Gado, J. P. Miller, U. Grenander, and M. I. Miller, "Hippocampal morphometry in schizophrenia by high dimensional brain mapping," *Proc. Natl. Acad. Sci. USA*, vol. 95, pp. 11406–11411, 1998.
- [9] L. Wang, S. C. Joshi, M. I. Miller, and J. G. Csernansky, "Statistical analysis of hippocampal asymmetry in schizophrenia," *NeuroImage*, vol. 14, pp. 531–545, 2001.
- [10] C. Davatzikos, "Why voxel-based morphometric analysis should be used with great caution when characterizing group differences," *NeuroImage*, vol. 23, pp. 17–20, 2004.
- [11] P. Golland, W. E. L. Grimson, M. E. Shenton, and R. Kikinis, "Deformation analysis for shape based classification," *Lecture Notes Comput. Sci.*, vol. 2082, pp. 517–530, 2001.
- [12] G. Gerig, M. Styner, and J. Lieberman, "Shape versus Size: Improved understanding of the morphology of brain structures," presented at the Med. Image Comput. Comput.-Assisted Intervention (MICCAI) 2001, Utrecht, The Netherlands, 2001.
- [13] P. Yushkevich, S. Joshi, S. M. Pizer, J. G. Csernansky, and L. E. Wang, "Feature selection for shape-based classification of biological objects," presented at the Information Processing in Medical Imaging, Ambleside, U.K., 2003.
- [14] Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M. Shenton, C. S. Carter, V. A. Stenger, S. Davis, H. Aizenstein, J. Becker, O. Lopez, and C. Meltzer, "Discriminative MR image feature analysis for automatic schizophrenia and alzheimer's disease classification," presented at the 7th Int. Conf. Med. Image Comput. Comput.-Assisted Intervention (MICCAI) 2004: , Saint-Malo, France, 2004.
- [15] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," *NeuroImage*, vol. 21, pp. 46–57, 2004.
- [16] C. E. Thomaz, J. P. Boardman, D. L. G. Hill, J. V. Hajnal, D. D. Edwards, M. A. Rutherford, D. F. Gillies, and D. Rueckert, "Using a maximum uncertainty LDA-based approach to classify and analyse MR brain images," presented at the 7th Int. Conf. Med. Image Comput. Comput.-Assisted Intervention (MICCAI) 2004: , Saint-Malo, France, 2004.
- [17] M. Miller, A. Banerjee, G. Christensen, S. Joshi, N. Khaneja, U. Grenander, and L. Matejic, "Statistical methods in computational anatomy," *Statist. Meth. Med. Res.*, vol. 6, pp. 267–299, 1997.
- [18] E. Duchesnay, A. Roche, D. Riviere, D. Papadopoulos, Y. Cointepas, and J.-F. Mangin, "Population classification based on structural morphometry of cortical sulci," presented at the IEEE Int. Symp. Biomed. Imag.: Macro to Nano, 2004., Arlington, VA, 2004.
- [19] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.
- [20] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1421–1439, Nov. 2002.
- [21] D. G. Shen and C. Davatzikos, "Very high resolution morphometry using mass-preserving deformations and HAMMER elastic registration," *NeuroImage*, vol. 18, pp. 28–41, 2003.
- [22] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–589, Jun. 1991.
- [23] D. Pham and J. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 737–752, Sep. 1999.
- [24] C. Davatzikos, M. Acharyya, K. Ruparel, D. G. Shen, J. Loughhead, R. C. Gur, and D. Langleben, "Classifying spatial patterns of brain activity for lie-detection," *NeuroImage*, vol. 28, pp. 663–668, 2005.
- [25] V. Grau, A. U. J. Mewes, M. Alcañiz, R. Kikinis, and S. K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 447–458, Apr. 2004.
- [26] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1996.
- [27] R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing*. New York, NY: Academic, 1997.
- [28] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychological Meth.*, vol. 1, pp. 30–46, 1996.

- [29] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, pp. 389–422, 2002.
- [30] A. Rakotomamonjy, "Variable selection using SVM-based criteria," *J. Mach. Learn. Res.*, vol. 3, pp. 1357–1370, 2003.
- [31] V. N. Vapnik, *The Nature of Statistical Learning Theory (Statistics for Engineering and Information Science)*, 2nd ed. New York: Springer-Verlag, 1999.
- [32] C. Davatzikos, D. G. Shen, X. Wu, Z. Lao, P. Hughett, B. I. Turetsky, R. C. Gur, and R. E. Gur, "Whole-brain morphometric study of schizophrenia reveals a spatially complex set of focal abnormalities," *JAMA Arch. General Psychiatry*, vol. 62, pp. 1218–1227, 2005.
- [33] J. Goldstein, L. Seidman, L. O'Brien, N. Horton, D. Kennedy, N. Makris, V. J. Caviness, S. Faraone, and M. Tsuang, "Impact of normal sexual dimorphisms on sex differences in structural brain abnormalities in schizophrenia assessed by magnetic resonance imaging," *Arch. Gen. Psychiatry*, vol. 59, pp. 154–164, 2002.
- [34] P. Nopoulos, M. Flaum, and N. C. Andreasen, "Sex differences in brain morphology in schizophrenia," *Am. J. Psychiatry*, vol. 154, pp. 1648–1654, 1997.
- [35] P. Nopoulos, M. Flaum, S. Arndt, and N. Andreasen, "Morphometry in schizophrenia revisited: Height and its relationship to pre-morbid function," *Psychological Med.*, vol. 29, pp. 655–663, 1998.
- [36] M. Frederikse, A. Lu, E. Aylward, P. Barta, T. Sharma, and G. Pearlson, "Sex differences in inferior parietal lobule volume in schizophrenia," *Am. J. Psychiatry*, vol. 157, pp. 422–427, 2000.
- [37] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.
- [38] J.-S. L. II-Seok Oh and B.-R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1421–1437, Nov. 2004.