Integrated Feature Extraction and Selection for Neuroimage Classification

Yong Fan and Dinggang Shen
IDEA Lab, Department of Radiology and BRIC, University of North Carolina,
Chapel Hill, NC 27599-7515

ABSTRACT

Feature extraction and selection are of great importance in neuroimage classification for identifying informative features and reducing feature dimensionality, which are generally implemented as two separate steps. This paper presents an integrated feature extraction and selection algorithm with two iterative steps: constrained subspace learning based feature extraction and support vector machine (SVM) based feature selection. The subspace learning based feature extraction focuses on the brain regions with higher possibility of being affected by the disease under study, while the possibility of brain regions being affected by disease is estimated by the SVM based feature selection, in conjunction with SVM classification. This algorithm can *not only* take into account the inter-correlation among different brain regions, *but also* overcome the limitation of traditional subspace learning based feature extraction methods. To achieve robust performance and optimal selection of parameters involved in feature extraction, selection, and classification, a bootstrapping strategy is used to generate multiple versions of training and testing sets for parameter optimization, according to the classification performance measured by the area under the ROC (receiver operating characteristic) curve. The integrated feature extraction and selection method is applied to a structural MR image based Alzheimer's disease (AD) study with 98 non-demented and 100 demented subjects. Cross-validation results indicate that the proposed algorithm can improve performance of the traditional subspace learning based classification.

Keywords: Feature extraction, feature selection, pattern recognition, statistical methods, neuroimage classification

1. INTRODUCTION

Neuroimaging techniques are being increasingly used in studies of neurological diseases to complement clinical assessments, such as performance on standard neuropsychological tests. Many neuropsychological diseases induce structural or functional changes that do not change local image characteristics as tumors or infarcts do; instead they have more subtle, spatially, and temporally diffuse pathology, which is often undetectable by direct human readings. For these diseases, including schizophrenia and early Alzheimer's disease (AD), a great deal of basic science research has been focusing on understanding how brain structure and function are affected in group level using univariate analysis methods [1-4]. Although these techniques have led to better understanding of neurological diseases, clinically useful diagnostic tools require more sophisticated image analysis methods. To this end, high-dimensional pattern classification techniques have built upon methods of computational anatomy and functional neuroimaging, demonstrating that classifications of individuals, in contrast to group analyses, can be achieved with relatively high classification accuracy in various studies [5-16].

To build robust neuroimage classification algorithms, it is necessary to extract a compact set of robust and informative features from 3D or 4D medical images. Since reliable prior knowledge about the problem under study is not always available in real applications, whole brain voxel-wise measures have been good choices for multivariate analysis of brain images. The typical voxel-wise measures include deformation fields, Jacobian determinants, and tissue density maps computed from structural brain images, and parametric maps computed from function images [2, 17-20]; however the sheer dimensionality of voxel-wise measurements must first be reduced to a small and optimal set of features. Most of the approaches that extract features from voxel-wise measures fall under three categories: 1) spatial filtering, including Gaussian smoothing and wavelet transformation, which are commonly used to improve the robustness of local features by averaging them using a Gaussian point spread function around each location or representing the voxel-wise features at multiple scales using wavelet techniques [2, 7]; 2) regional feature extraction, including adaptive regional feature extraction

Medical Imaging 2009: Image Processing, edited by Josien P. W. Pluim, Benoit M. Dawant,
Proc. of SPIE Vol. 7259, 72591U · © 2009 SPIE
CCC code: 1605-7422/09/$18 · doi: 10.1117/12.811781

Downloaded from SPIE Digital Library on 27 Oct 2010 to 152.19.89.63. Terms of Use: http://spiedl.org/terms

[11, 21], volume measures within specific anatomical regions of interest (ROIs), tessellation algorithms [16] describing shapes using small patches, and methods grouping adjacent image voxels [6, 8, 22]; 3) subspace learning methods, particularly principal component analysis (PCA) [23], which can be used to dramatically reduce the feature dimensionality [10, 15, 24, 25]. The extracted features are then either directly used to build classifiers or further selected by feature selection techniques.

The feature extraction methods in the former two categories share same characteristics, i.e., the extracted features are still in original feature space, while the subspace techniques transfer the original features into a new linear subspace. The feature extraction methods in the former two categories focus on localized brain regions. Coupled with feature selection methods, they have the potential to identify disease specific brain regions. However, the inter-correlations among different brain regions are ignored during the feature extraction.

The subspace learning based feature extraction methods are good at capturing the inter-correlation among different brain regions. However, a large amount of brain regions remain intact, especially in the early stage of the disease. The features located in the intact brain regions are noisy with respect to the classification, due to the inter-subject variability of brain structure and function. Therefore, it is better to perform subspace learning only on those features within the disease-affected brain regions. Furthermore, the principal components are not necessarily the ones containing the most discriminative information.

Since the feature extraction methods cannot generate features that are all discriminative for classification, feature selection techniques have been increasingly used in neuroimage classification [7, 11, 16, 22, 26]. To achieve the best classification performance, it is advisable to use the subset feature selection methods that generally have better performance compared with the ranking based feature selection techniques [27]. However, the high computational cost of subset feature selection methods limits their application to problems with high feature dimensionality.

To leverage the advantages and overcome the limitations of above feature extraction and selection techniques, we propose an integrated feature extraction and selection method for neuroimage classification, which consists of two iterative steps: 1) constrained PCA feature computation, focusing on discriminative voxel-wise features which are identified by 2) support vector machine (SVM) based subset feature selection, working directly in the PCA feature space and indirectly in the original voxel-wise feature space. These two steps iteratively refine the feature extraction and selection until the best classification performance is achieved.

The proposed method has been applied to distinguish individuals with dementia from cognitively normal elders in individual level based on structural magnetic resonance (MR) images on a dataset with 98 early stage AD patients and 100 normal controls. The comparison results demonstrated that the proposed method can improve performance of the traditional subspace learning based classification.

2. METHODS

A neuroimage classification system typically consists of components of feature extraction, feature selection, and classifier construction. In general, these three components are independently developed and sequentially combined to build a classification system. This procedure is inherently circular, because the classifier construction requires a set of informative features, and the extraction and selection of a set of informative features requires a feedback about the performance of features from the classification result. To build a classification system with improved performance, we propose an integrated neuroimage classification method which iteratively performs a constrained subspace learning based feature extraction and a feature selection in conjunction with SVM classification to optimize the classification performance.

The constrained PCA feature computation is implemented by a weighted PCA [28], considering individual voxel-wise features selectively depending on their respective weights, to be determined by the feature selection step. Once the PCA features are computed, a SVM based feature selection approach is used to select the most informative PCA features to build SVM classifiers; the SVM classifiers are then used to determine the discriminative ability of each individual voxel-wise features. The integrated feature extraction and selection algorithm starts with the traditional PCA based feature extraction, and then iteratively perform soft feature selection and constrained PCA feature computation, until the best classification performance is achieved. These two iterative steps are detailed in the following.

Downloaded from SPIE Digital Library on 27 Oct 2010 to 152.19.89.63. Terms of Use: http://spiedl.org/terms

2.1 Computation of constrained PCA features

As mentioned in the introduction, whole brain voxel-wise measures have been good choices for multivariate analysis of brain images. For structural image classification, we focus on brain morphology measured by brain tissue density maps of grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF), which are computed via a tissue-preserving spatial normalization of MRI scans to a template brain space [17, 29]. The brain tissue density maps consists of millions of voxel-wise measures. The feature dimensionality has to be reduced to achieve a well-generalized classification.

To reduce the feature dimensionality, subspace learning methods, particularly principal component analysis (PCA), have been widely used in structural and functional neuroimage classification studies [10, 15, 24, 25]. The subspace learning-based feature extraction methods take into account the inter-correlation among voxel-wise features distributed in the whole brain space. This characteristic of the subspace learning techniques brings both advantages and limitations. Most neurological diseases have subtle, spatially, and temporally diffuse pathology, i.e., the brain is damaged in multiple regions which might highly interact with each other, rather than in one single isolated region. For the classification of these diseases, it is advantageous to jointly consider multiple brain regions in feature extraction. However, a large amount of brain regions remain intact, especially in the early stage of the diseases. The features located in intact brain regions are noise with respect to the classification due to the inter-subject variability of brain structure and function. To achieve better performance, it is necessary to perform subspace learning only on those voxel-wise features in the disease affected brain regions.

Several weighted versions of PCA have been proposed [28]. Unlike the standard approach, the weighted PCA considers individual voxels or images selectively, depending on the corresponding weights. In our application, the weighted PCA should focus on brain regions affected by the disease under study, i.e., brain regions with good discriminative ability. Since such brain regions are typically unknown in general applications, we need to estimate the discriminative ability of each voxel-wise feature from the training data. Although the discriminative ability of each feature can be measured individually by using univariate statistical analysis methods, such as Pearson correlation coefficient between the class label and the feature in the training data, we propose to estimate the relevance of features to classification directly from their contributions to the classification.

2.2 Estimation of features' relevance to classification

The SVM-based criteria have been demonstrated to have excellent performance to measure feature's discriminative ability in feature selection studies [27, 30, 31]. Such criteria based on linear SVM were first proposed in SVM-Recursive Feature Elimination (RFE) [30], then extended for nonlinear SVM [31]. For linear SVMs, the feature's discriminative ability is estimated as the square of the linear SVM classification function's coefficient corresponding to the feature under consideration; while for nonlinear SVMs it can be estimated as the square of the feature's corresponding component of the gradient of the classification function. Since the estimation has to be run recursively to select the best feature subset for classification, its expensive computation cost limits the direct application to the whole brain voxel-wise measures with millions of variables.

To circumvent these difficulties, we first apply SVM-RFE to the PCA features, and select the most informative PCA features to build linear SVM classifiers. The weights (or contributions) of the PCA features in the constructed SVM classification function are then mapped to the original voxel-wise feature space by using the inverse of PCA transformation matrix [10]. The weights on voxel-wise features can be used as an estimation of discriminative ability of features. Once the original features' weights are estimated, the PCA feature extraction can focus on those with higher weights by the weighted PCA algorithm [28].

In our application of separating individuals with AD pathology (class label +1) from normal controls (class label -1), positive weights correspond to the features with biologically implausible results, i.e., less atrophy in AD patients compared to normal controls. Therefore, only those features with negative weights should be involved in the PCA feature extraction. Rather than directly discarding those features with biologically implausible weights, we adopt logistic sigmoid functions to modulate the features in the weighted PCA computation. The logistic sigmoid function we used is defined as

$$y = \frac{1}{1+\exp(\sigma x)}, \quad (1)$$

Downloaded from SPIE Digital Library on 27 Oct 2010 to 152.19.89.63. Terms of Use: http://spiedl.org/terms

where σ is the steepness factor to suppress the unimportant or unreliable voxel-wise features. The steepness factor can be adjusted according to the classification performance (or confidence on the classification), starting with smaller values, to gradually suppress unwanted voxel-wise features. Other sigmoid functions can also be used to adapt to different applications. The logistic sigmoid functions with different steepness factors are shown in Fig. 1

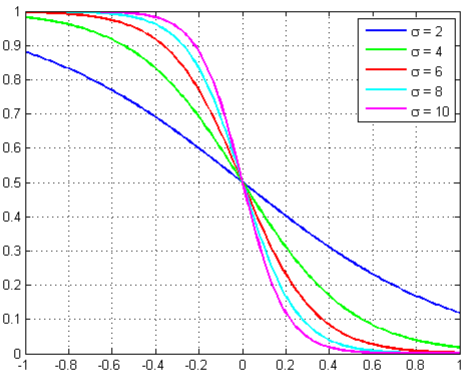

Fig.1. Logistic sigmoid functions for suppressing unreliable or less important voxel-wise features, thus achieving soft feature selection. The curves of different colors denote the weighting function with respect to different parameters. With the progress of iterative feature extraction and selection, the feature selection gradually converts to be '*hard* feature selection' from '*soft* feature selection'.

To facilitate robust feature selection and optimal selection of classification parameters, a bootstrapping resampling method is used to generate multiple versions of the training set [32]. In particular, a leave-one-out sampling scheme is implemented in this study. For a training set with *m* samples, *m* different bootstrapping training sets are generated and each of them has *m-1* samples. Then, a linear SVM classifier, referred to as a base classifier, is built for each training set based on PCA features selected by SVM-RFE. The number of features used in classification and the SVM parameters are determined by optimizing the classifier's performance estimated by applying it to the left-out training sample excluded from the bootstrapping training set. For a specific set of classification parameters, the overall performance of these base classifiers trained on different bootstrapping training sets can be measured by the *a*rea *u*nder the ROC (Receiver Operating Characteristics) *c*urve (*AUC*), a measurement that has been demonstrated more effective than accuracy in evaluating the learning algorithms [33]. By searching the parameter space, we can find classifiers with the maximal AUC. Once the classifiers are built based on different training subsets, an ensemble classifier can be constructed by combining their outputs to apply to the test samples. In this study, we combine these classifiers by averaging SVM scores.

2.3 Classification of early stage AD subjects based on their structural MR images

The proposed method is applied to the classification of early stage AD subjects based on their structural MR images. The image data were obtained from the Open Access Series of Image Studies (OASIS), including 98 healthy elders and 100 early-stage AD patients [34]. Each subject has three to four individual T1-weighted MR images acquired on a 1.5T scanner. These images were averaged to create a single image with high contrast to noise ratio. Detailed information about the clinical and demographic data for all subjects is available at the OASIS website (http://www.oasis-brains.org). The subject demographics and dementia status are summarized in table 1.

Table 1. Characteristics of the participants in this study.

Group	Healthy Elder	Early-stage AD
No. of subjects	98	100
Percent of males	26.5%	41.0%
Age (year), mean±std	75.92±8.99	76.76±7.12
MMSE, mean±std	28.96±1.21	24.32±4.17
No. of subjects with CDR 0/0.5/1/2	98/0/0/0	0/70/28/2

Based on the structural MR images, brain morphology is measured by regional brain tissue density maps of GM, WM and CSF, which are computed via a tissue-preserving spatial normalization of the MRI scans for each subject to a

Downloaded from SPIE Digital Library on 27 Oct 2010 to 152.19.89.63. Terms of Use: http://spiedl.org/terms

template brain space [17, 29]. To account for variations in head size, the morphological representations are normalized by the total intra-cranial volume (ICV), and smoothed using 8 mm full-width at half-maximum (FWHM) smoothing kernel. In this study, we focus on GM tissue density maps.

3. EXPERIMENTAL RESULTS

3.1 Classification performance

To get an unbiased estimation of the classification performance, a full cross-validation experiment was performed. The demented and non-demented groups were randomly divided into two sets with matched MMSE scores, respectively, resulting in two demented sub-groups and two non-demented sub-groups, each of demented sub-groups having 50 subjects and each of non-demented subgroups having 49 subjects. The algorithm was then trained using a combination of one demented and one non-demented sub-groups, and tested on the remaining sub-groups. This process was repeated 4 times by picking different combination of demented and non-demented sub-groups as a training dataset and estimating the classification performance of the algorithm on the subjects of other sub-groups by comparing their estimated results with clinically measured labels. The parameters of the classifiers were optimized within the integrated feature extraction and selection framework, based on the training data.

Table 2. Performance of the traditional PCA based feature extraction and the integrated feature extraction and selection algorithm (IPCA) in classification of early stage AD subjects.

	Classification Rate	AUC
PCA	66.9%	0.723
IPCA	69.7%	0.743

We compared our method with traditional PCA based feature extraction in the same experiment setting [10, 25]. The SVM-RFE algorithm is used to select the most discriminative features for building linear SVM classifiers, and the classification parameters are optimized within the same framework. The classification performances of these methods on testing subjects are summarized in table 2. Their corresponding ROC curves are shown in Fig. 2.

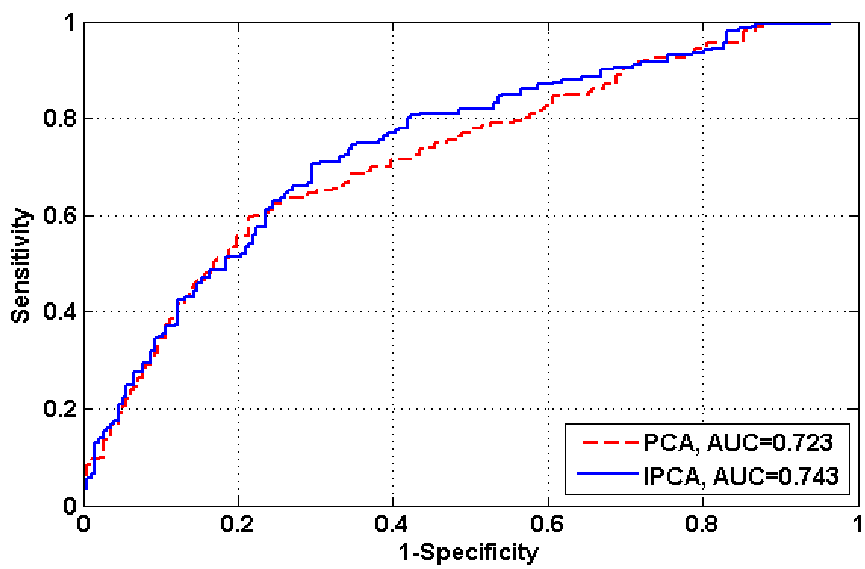

Fig.2. ROC curves of the classifiers estimated from the SVM scores of testing subjects.

3.2 Dementia-specific spatial pattern of brain structure

To generate a visual picture of the brain regions that collectively contributed to the classification, the average of weighting maps of voxel-wise features in these four cross-validation experiments are shown in Fig. 3, along with the *t*-map obtained by the widely used VBM method via voxel-by-voxel *t*-test analysis on the smoothed tissue density maps [2, 18], which was implemented using the SPM5 software (http://www.fil.ion.ucl.ac.uk/spm/software/spm5). The false

Downloaded from SPIE Digital Library on 27 Oct 2010 to 152.19.89.63. Terms of Use: http://spiedl.org/terms

discovery rate (FDR) method was utilized for correction of multiple comparisons [35], as implemented in the SPM software. The voxel-wise weighting maps obtained by the traditional PCA feature based classification and the proposed method, as well as the *t*-statistic map, are shown in Fig. 3, for visual inspection. As shown in these maps, the propose method has the potential to better localize brain atrophy than the traditional PCA method.

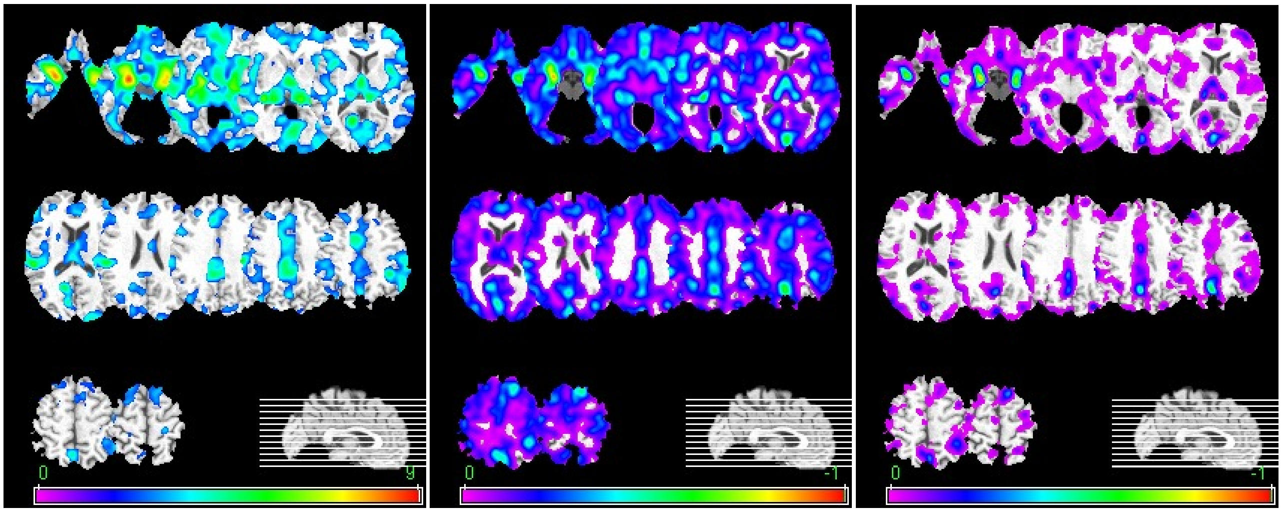

Fig.3. Brain atrophy detected in the GM of early-stage AD subjects, by performing t-test (left), PCA based classification (middle), and the integrated feature extraction and selection (right). The multiple comparison in t-test based analysis is corrected by false discover rate (p<0.05, FDR corrected). The color bars indicate the t-value (left column), or normalized weights in SVM classification (middle and right). Images are shown in radiology convention.

4. CONCLUSION

This paper presents an integrated feature extraction and selection framework for neuroimage classification, which has been evaluated in the separation of early stage AD from non-demented subjects based on their structural MR brain images. The experiment results indicate that the proposed method has the potential to overcome the limitation of traditional PCA based feature extraction method, and achieve better performance in identification of early stage AD based on structural MR brain images.

The proposed approach shares some similarities with 'feature shaving' which was developed to identify significant differences between patients and controls [36]. However they are significantly different in that 1) rather than discarding less discriminative features as did in 'feature shaving', the proposed method performs a soft feature selection by assigning different weights for voxel-wise features, which may bring back the voxel-wise features discarded in the early stage of the iterative feature extraction and selection; 2) a more sophisticated subset feature selection technique, i.e., SVM-recursive feature elimination [30], is used in the proposed approach, rather than a simple forward or backward feature ranking technique; 3) the proposed approach can achieve better classification performance, while 'feature shaving' can lead to worse classification performance compared with the classification with original PCA features [36].

As observed in the experimental results, both the PCA based feature extraction method and the proposed IPCA have moderate classification accuracy due to the following three reasons. *First*, the patient group has diverse dementia degree as reflected by their CDR scores, which makes it difficult to find an optimal linear classification function to separate the diverse patient group from the normal control group. *Second*, the relationship between class label and morphological features might be nonlinear, therefore linear PCA, coupled with linear SVM, cannot capture such nonlinear relationship. *Third*, the voxel-wise features might be not robust to registration error, which limits the generalization ability of the extracted features. Previously, we have developed an adaptive regional feature extraction method [11, 21], which can reduce the feature dimensionality and improve the robustness of features. However, the clustering of voxel-wise features is based on a univariate discriminative measure, which cannot take into consideration the joint correlation among different brain regions. We are working to integrate the current method and our previous regional feature extraction method, for taking the advantages of both methods.

Downloaded from SPIE Digital Library on 27 Oct 2010 to 152.19.89.63. Terms of Use: http://spiedl.org/terms

5. ACKNOWLEDGEMENT

This work is partly supported by NIH grants 1R03EB008760, 7R03MH076970, and 1R01EB006733.

REFERENCES

[1] C. D. Good, R. I. Scahill, N. C. Fox *et al.*, "Automatic differentiation of anatomical patterns in the human brain: Validation with studies of degenerative dementias," Neuroimage, 17(1), 29-46 (2002).

[2] J. Ashburner, and K. J. Friston, "Voxel-based morphometry: the methods," Neuroimage, 11(6), 805-821 (2000).

[3] N. R. Giuliani, V. D. Calhoun, G. D. Pearlson *et al.*, "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).

[4] K. J. Friston, A. P. Holmes, K. Worsley *et al.*, "Statistical parametric maps in functional imaging: a general linear approach," Human Brain Mapping, 2(4), 189-210 (1995).

[5] S. Kloppel, C. M. Stonnington, C. Chu *et al.*, "Automatic classification of MR scans in Alzheimer's disease," Brain, 131(3), 681-689 (2008).

[6] P. Vemuri, J. L. Gunter, M. L. Senjem *et al.*, "Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies," Neuroimage, 39(3), 1186-1197 (2008).

[7] Z. Lao, D. Shen, Z. Xue *et al.*, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," Neuroimage, 21(1), 46-57 (2004).

[8] C. Davatzikos, K. Ruparel, Y. Fan *et al.*, "Classifying spatial patterns of brain activity for lie-detection," Neuroimage, 28(3), 663-668 (2005).

[9] D. D. Cox, and R. L. Savoya, "Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex," NeuroImage, 19(2), 261–270 (2003).

[10] J. Mourao-Miranda, A. L. W. Bokde, C. Born *et al.*, "Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data," Neuroimage, 28(4), 980-995 (2005).

[11] Y. Fan, D. Shen, R. C. Gur *et al.*, "COMPARE: Classification Of Morphological Patterns using Adaptive Regional Elements," IEEE Transactions on Medical Imaging, 26(1), 93-105 (2007).

[12] E. Duchesnay, A. Cachia, A. Roche *et al.*, "Classification based on cortical folding patterns," IEEE Transactions on Medical Imaging, 26(4), 553-565 (2007).

[13] Y. Fan, N. Batmanghelich, C. M. Clark *et al.*, "Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline," Neuroimage, 39(4), 1731-1743 (2008).

[14] P. Golland, Fischl,B., Spiridon,M., Kanwisher,N., Buckner,R.L., Shenton,M.E., Kikinis,R., Dale,A., and Grimson,W.E.L, "Discriminative Analysis for Image-based Studies." LNCS 2488, 508-515, MICCAI, (2002).

[15] S. Duchesne, A. Caroli, C. Geroldi *et al.*, "MRI-Based Automated Computer Classification of Probable AD Versus Normal Controls," IEEE Transactions on Medical Imaging, 27(4), 509-520 (2008).

[16] P. Yushkevich, S. Joshi, S. M. Pizer *et al.*, "Feature selection for shape-based classification of biological objects." 2732 of LNCS, 114-125, Information Processing in Medical Imaging, (2003).

[17] C. Davatzikos, A. Genc, D. Xu *et al.*, "Voxel-Based Morphometry Using the RAVENS Maps: Methods and Validation Using Simulated Longitudinal Atrophy," NeuroImage, 14(6), 1361-1369 (2001).

[18] J. Ashburner, J. G. Csernansky, C. Davatzikos *et al.*, "Computer-assisted imaging to assess brain structure in healthy and diseased brains," The Lancet (Neurology), 2(2), 79-88 (2003).

[19] P. M. Thompson, J. N. Giedd, R. P. Woods *et al.*, "Growth patterns in the developing brain detected by using continuum mechanical tensor maps," Nature, 404, 190-193 (2000).

[20] M. K. Chung, K. J. Worsley, T. Paus *et al.*, "A unified statistical approach to deformation-based morphometry," Neuroimage, 14(3), 595-606 (2001).

[21] Y. Fan, D. Shen, and C. Davatzikos, "Classification of Structural Images via High-Dimensional Image Warping, Robust Feature Extraction, and SVM," Lecture Notes in Computer Science. 3749 1-8, MICCAI, (2005).

[22] Y. Liu, L. Teverovskiy, O. Carmichael *et al.*, "Discriminative MR Image Feature Analysis for Automatic Schizophrenia and Alzheimer's Disease Classification." LNCS 3216, 393-401, MICCAI, (2004).

[23] I. T. Jolliffe, [Principal Component Analysis] Springer, New York(2002).

Downloaded from SPIE Digital Library on 27 Oct 2010 to 152.19.89.63. Terms of Use: http://spiedl.org/terms

[24] S. Duchesne, N. Bernasconi, A. Bernasconi *et al.*, “MR-based neurological disease classification methodology: application to lateralization of seizure focus in temporal lobe epilepsy,” Neuroimage, 29(2), 557-566 (2006).
[25] S. Duchesne, A. Caroli, C. Geroldi *et al.*, "Predicting clinical variable from MRI features: application to MMSE in MCI," Lecture Notes in Computer Science. 3749, 392-399, MICCAI, (2005).
[26] Y. Fan, H. Rao, H. Hurt *et al.*, “Multivariate examination of brain abnormality using both structural and functional MRI,” Neuroimage, 36(4), 1189-1199 (2007).
[27] I. Guyon, and A. Elisseeff, “An introduction to variable and feature selection,” Journal of Machine Learning Research, 3, 1157-1182 (2003).
[28] D. Skočaj, A. Leonardis, and H. Bischof, “Weighted and robust learning of subspace representations ” Pattern Recognition Letters, 40(5), 1556-1569 (2007).
[29] D. Shen, and C. Davatzikos, “HAMMER: Hierarchical attribute matching mechanism for elastic registration,” IEEE Transactions on Medical Imaging, 21(11), 1421-1439 (2002).
[30] I. Guyon, J. Weston, S. Barnhill *et al.*, “Gene Selection for Cancer Classification using Support Vector Machines ” Machine Learning, 46(1-3), 389-422 (2002).
[31] A. Rakotomamonjy, “Variable Selection using SVM-based criteria,” Journal of Machine Learning Research, 3, 1357-1370 (2003).
[32] Y. Fan, S. M. Resnick, and C. Davatzikos, "Bagging SVM classifiers for brain differential diagnosis using information jointly extracted from MRI/PET images." SPIE medical imaging, (2008).
[33] J. Huang, and C. X. Ling, “Using AUC and Accuracy in Evaluating Learning Algorithms,” IEEE Transactions on Knowledge and Data Engineering, 17(3), 299-310 (2005).
[34] D. S. Marcus, T. H. Wang, J. Parker *et al.*, “Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults,” Journal of Cognitive Neuroscience, 19, 1498–1507 (2007).
[35] D. Yekutieli, and Y. Benjamini, “Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics,” Journal of Statistical Planning and Inference, 82(1-2), 171-196 (1999).
[36] M. W. A. Caan, K. A. Vermeer, L. J. v. Vliet *et al.*, “Shaving diffusion tensor images in discriminant analysis: A study into schizophrenia”, Medical Image Analysis 10(6), 841-849 (2006).

Downloaded from SPIE Digital Library on 27 Oct 2010 to 152.19.89.63. Terms of Use: http://spiedl.org/terms