Published in final edited form as:

Magn Reson Imaging. 2012 April ; 30(3): 446–452. doi:10.1016/j.mri.2011.11.001.

Partial Least Squares for Discrimination in fMRI Data

Anders H. Andersen^{1,4,*}, William S. Rayens², Yushu Liu², and Charles D. Smith^{3,4}

¹Anatomy and Neurobiology, University of Kentucky, Lexington, KY 40536

²Statistics, University of Kentucky, Lexington, KY 40536

³Neurology, University of Kentucky, Lexington, KY 40536

⁴Magnetic Resonance Imaging and Spectroscopy Center University of Kentucky, Lexington, KY 40536

Abstract

Multivariate methods for discrimination were used in the comparison of brain activation patterns between groups of cognitively normal women who are at either high or low Alzheimer's disease risk based on family history and apolipoprotein-E4 status. Linear discriminant analysis (LDA) was preceded by dimension reduction using either principal component analysis (PCA), partial least squares (PLS), or a new oriented partial least squares (OrPLS) method. The aim was to identify a spatial pattern of functionally connected brain regions that was differentially expressed by the risk groups and yielded optimal classification accuracy. Multivariate dimension reduction is required prior to LDA when the data contains more feature variables than there are observations on individual subjects. Whereas PCA has been commonly used to identify covariance patterns in neuroimaging data, this approach only identifies gross variability and is not capable of distinguishing among-groups from within-groups variability. PLS and OrPLS provide a more focused dimension reduction by incorporating information on class structure and therefore lead to more parsimonious models for discrimination. Performance was evaluated in terms of the cross-validated misclassification rates. The results support the potential of using fMRI as an imaging biomarker or diagnostic tool to discriminate individuals with disease or high risk.

Keywords

pattern; classification; neuro; imaging; biomarker; Alzheimer's disease

INTRODUCTION

Multivariate methods are likely to play an important role in the development of fMRI as an imaging biomarker or diagnostic tool to discriminate individuals with disease. The motivation is based on the potential of spatial patterns of functionally connected brain regions conveying more information than do individual regions. For instance, a brain region may not by itself exhibit a significant difference in its level of fMRI activation between, say,

© 2011 Elsevier Inc. All rights reserved

***Corresponding Author:** Anders H. Andersen, Ph.D. University of Kentucky College of Medicine Magnetic Resonance Imaging and Spectroscopy Center Room 37 Davis Mills Building (MRISC) 800 Rose Street Lexington, Kentucky 40536-0098 Phone: (859) 323-1108 Fax: (859) 323-1068 anders@mri.uky.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

a patient group and a control group. Yet, in combination with other brain areas, this region might become important for discrimination as an integral part of a distributed brain network. Neuroimaging data, however, typically has more features than there are observations on individual subjects, causing problems with the use of linear discriminant analysis (LDA). The most common approach has been to use principal component analysis (PCA) as a first step to reduce the dimension of the data [1–3]. Unfortunately, PCA only identifies gross variability and is not capable of distinguishing among-groups from within-groups variability. Partial least squares (PLS) for focused dimension reduction in discrimination was developed to circumvent this problem by incorporating information on class structure [4]. PLS was first used for spatial pattern analysis of functional brain images by McIntosh et al. [5]. Oriented partial least squares (OrPLS) is a new multivariate technique developed by our group [6]. It has been used in multivariate analysis of fMRI time series data as a way of incorporating information about the underlying experimental paradigm as well as noise and other confounds [7]. In the context of discrimination, OrPLS combines the tuning of PLS toward group separation with the ability to simultaneously orient away from within-group covariability [8,9].

The emphasis of the present work is on eigenstructure-based dimension reduction, particularly focused dimension reduction as compared to the variance summaries of PCA, in combination with linear discriminant analysis for the classification of multivariate fMRI data from groups of subjects. In a setting of the large variability in level of activation between subjects frequently observed in fMRI studies, PCA for dimension reduction may fail to capture the succinct information on group separation essential for classification within a linear subspace spanned by components accounting for even a large proportion of total variability. Other approaches have used support vector machines (SVM) for learning and classification of fMRI data in conjunction with an initial step of PCA-based dimension reduction [10]. Also, the literature on genetic algorithms contains a host of routines for variable selection and dimension reduction. As input we may use ROI-, parcellation- [2], or voxel-based [3] functional neuroimaging data. The present work uses a parcellation of the brain into regions based on the Talairach atlas. The aim is to identify a spatial pattern of functionally connected brain regions that is differentially expressed by groups of subjects as measured by the subject-wise discriminant scores and which yields optimal classification accuracy.

MATERIALS AND METHODS

Data Acquisition

fMRI was used in the present study to observe cortical activation during a confrontation naming task in 13 women with high Alzheimer's disease (AD) risk and 11 with low risk, based on family history and apolipoprotein-E4 status [11]. A blocked experimental design was used, and activation to the naming task compared to the rest periods was evaluated by the general linear model. The EPI time series data was preprocessed using standard methods including slice timing correction, motion correction, spatial smoothing, masking, and image intensity normalization. Multiple linear regression analysis included motion parameter estimates in the baseline model. Fractional signal change values were calculated from each of three runs and collapsed to yield a single measure of activation per voxel per subject. Individual subject data was registered across sessions and normalized to Talairach coordinate space with AFNI software using an automated 12-parameter affine transformation [12]. Region-of-interest (ROI) parcellation data was subsequently extracted from 50 non-overlapping brain regions in each hemisphere based on the Talairach atlas labels of Lancaster and Fox [13]. The mean value of the fractional signal change within each region was retained as the composite activation measure. A second data set was available, as the identical scan protocol was repeated in the same subjects after four years [14].

Data Analysis

ROI data in the left hemisphere was analyzed by linear discriminant analysis (LDA) preceded by dimension reduction using either PCA, PLS, or oriented partial least squares (OrPLS) methods. OrPLS is a broad new class of techniques developed by our group [6–9]. When applied in the setting of discrimination, within-class structure is accounted for thus enabling the analysis to orient away from within-class covariability. OrPLS combines this property with the inherent ability of PLS to orient the dimension reduction toward patterns that exhibit group separation [4]. In principle, multiple OrPLS components may be derived and used in a subsequent LDA [8]. In the present work, however, we consider only the first OrPLS component for classification. The methods that will be compared are then PCA-LDA, PLS-LDA and OrPLS.

In the specific context of the two-group classification problem, the first OrPLS direction vector $\mathbf{a}_{OrPLS}$ becomes the eigenvector associated with the largest eigenvalue of $(\mathbf{E} + k \cdot \mathbf{I})^{-1} \mathbf{H}$, where $\mathbf{H}$ is Fisher's among-groups sums-of-squares and cross-products matrix and $\mathbf{E}$ is the pooled within-groups sums-of-squares and cross-products matrix. Its elements are weights, one for each of the original ROI feature variables. In the two-group problem we now have the much simplified relationship for the OrPLS direction vector

$$\mathbf{a}_{OrPLS} \propto (\mathbf{E} + k \cdot \mathbf{I})^{-1} \mathbf{a}_{PLS}$$

where $\mathbf{a}_{PLS}$ is the first partial least squares direction vector. The PLS solution is the eigenvector associated with the largest eigenvalue of $\mathbf{H}$. In the two-group problem, the PLS direction vector has a particularly simple and intuitively appealing form; it is proportional to the difference in group means, i.e., $\mathbf{a}_{PLS} \propto \mathbf{m}_1 - \mathbf{m}_2$. The role of the matrix $\mathbf{E}$ in the two-group case is to affect a subsequent rotation away from within-group covariability as well as to exploit existing correlations among the original feature variables. This rotation becomes progressively more important with increasing non-sphericity of data scatter in the respective ROI variables [9]. Note that, when the random-noise component $k \cdot \mathbf{I}$ dominates, $\mathbf{a}_{OrPLS} \rightarrow \mathbf{a}_{PLS}$. We can think of k as a regularization parameter that controls the amount of shrinkage toward the identity $\mathbf{I}$ [8,9,15]. While resembling Fisher's linear discriminant (based on the eigenstructure of $\mathbf{E}^{-1} \mathbf{H}$), this regularization or ridging is necessary because $\mathbf{E}$ is not invertible as it is. The sample covariance matrix is singular due to the number of variables (brain regions) exceeding the number of observations (subjects); in this particular setting, $\mathbf{E}$ has rank 22. From a purely mathematical perspective, any value of k greater than zero will render $\mathbf{E} + k \cdot \mathbf{I}$ non-singular and thus invertible. However, small values of k will tend to exaggerate the importance of direction vectors in the low-variance subspace spanned by the eigenvectors corresponding to the smallest sample eigenvalues [15]. The regularization parameter k effectively “shrinks” the contribution from feature variables of low sample variance in the measurement space. In a practical setting, cross-validation may be used to determine the optimal value of the parameter “ k ” for a particular sample data set [8]. Here we have used a value of k determined by the mean of the trace of $\mathbf{E}$, i.e., the average eigenvalue.

When PCA is used as an initial step for dimension reduction, the first direction vector $\mathbf{a}_{PCA}$ will be determined as the eigenvector associated with the largest eigenvalue of $\mathbf{E} + \mathbf{H}$, the combined within- and between-groups variability. As with PLS, the matrix $\mathbf{E}$ may be used in turn to affect a subsequent rotation of the principal component away from within-group covariability. This approach is known as oriented PCA and can be stated as a generalized eigenvalue problem [7,16]. However, it should be noted that the initial first step of PCA is missing the ability of distinguishing $\mathbf{H}$ from $\mathbf{E}$. This distinction becomes particularly critical

in situations where the between-groups variability is considerably less than the within-groups variability (traces of $\mathbf{H}$ and $\mathbf{E}$, respectively), as it is frequently seen with fMRI data.

Classification of a subject data vector $\mathbf{x}$ of dimension 50×1 is based on the associated discriminant score $\mathbf{a}^T \mathbf{x}$, where the direction vector $\mathbf{a}$ of spatial pattern weights is obtained by dimension reduction using either PCA, PLS or OrPLS methods. The baseline scan data was used as a training set for the linear classifier. Stability of the classifier to small perturbations was assessed by the standard method of leave-one-out cross-validation [17]. Additionally, the four-year follow-up scans were serving as a separate test data set for further validation.

The initial dimension reduction preceding discrimination may also be used to identify a subspace spanned by several direction vectors $\{\mathbf{a}_k\}$ rather than merely a single direction. In that case, the subsequent discrimination analysis will involve the calculation of an appropriate linear combination of direction vector component scores $\mathbf{a}_k^T \mathbf{x}$, which yields the best separation of risk groups. This approach has been used recently with voxel-wise and ROI data from functional neuroimaging studies to identify brain networks that optimally discriminate between young and old subjects [2] or between groups of Alzheimer's disease patients and cognitively normal controls [1]. In those works, PCA was used in the initial step of dimension reduction to capture spatial patterns reflecting combined between- and within-group variability. This was followed by linear discriminant analysis with a reduced/selected number of principal components and their associated scores used as the predictor variables. In keeping with Barker and Rayens [4], we have carried out similar analyses for discrimination among AD risk groups employing PCA as well as PLS for initial dimension reduction of the 50-region ROI atlas parcellation data representing brain activation to the confrontation naming task. Classification of a subject data vector $\mathbf{x}$ of dimension 50×1 is now based on the associated composite discriminant score

$$\sum_k \beta_k \mathbf{a}_k^T \mathbf{x},$$

where $\{\mathbf{a}_k\}$ for $k=1, \dots, K$ denotes a subset of the PC or PLS direction vectors of unit norm..

In summary, the processing steps involved in the three different methods considered are:

PCA-LDA (1) An initial dimension reduction by PCA yielding a set of direction vectors $\{\mathbf{a}_{k,PCA}\}$ is followed by (2) linear discriminant analysis based on the scores $\{\mathbf{a}_{k,PCA}^T \mathbf{x}\}$ of the first K principal components.

PLS-LDA (1) An initial dimension reduction by PLS yielding a set of direction vectors $\{\mathbf{a}_{k,PLS}\}$ is followed by (2) linear discriminant analysis based on the scores $\{\mathbf{a}_{k,PLS}^T \mathbf{x}\}$ of the first K PLS components.

OrPLS (1) The OrPLS direction vector is determined from the first PLS component by multiplication with $(\mathbf{E} + k \cdot \mathbf{I})^{-1}$ followed by (2) simple discrimination using a threshold determined from the group means of the scores $\mathbf{a}_{OrPLS}^T \mathbf{x}$.

The multivariate analyses of dimension reduction for determining the direction vectors and their associated scores for each of the PCA, PLS, and OrPLS approaches were done in MATLAB v. 7.2 (The MathWorks, Inc., Natick, MA) software using code developed in house. The LDA analyses on scores were done in SPSS v. 18.0 (IBM SPSS).

RESULTS

Table 1 shows the comparison of principal component and partial least squares regression as an increasing number of components are retained and included in the model. The individual component scores $\mathbf{a}_k^T \mathbf{x}$ were used as predictor variables with group membership encoded as an indicator variable being the dependent response measure. Linear regression has been shown to yield the same linear combination (the β_k coefficients) of predictor variables as linear discriminant analysis for the two-group problem and offers some further insights [18]. The table illustrates the well-known trade-off between variance summary and achieved correlation for PLS [8,19]. The correlation values for PLS along with the significance of the regression are consistently higher than those for PCA for the same number of model components retained, whereas the amount of variance summarized is consistently higher for PCA. In fact, while PCA yields the best possible fit to the original data for any given number of components (in the sense of orthogonal least squares), the PC regression with risk group as the response measure never reaches significance, even with 10 components retained accounting for 95% of total variability in the fMRI data. In comparison, the PLS regression model, focusing on group separation and incorporating information on class structure as part of the initial dimension reduction, is significant ($p < 0.05$) with just a single component even though only 15.6% of total variance is summarized. The associated Akaike information criterion (AIC) decreases monotonically as the residual sum-of-squares error continues to decrease at a faster rate than the concomitant increase in complexity of the regression model, as more PLS components are included. For PC regression, the AIC attains its minimum value with just a single component.

Table 2 illustrates the classification accuracy for dimension reduction by PCA or PLS followed by linear discriminant analysis. From 1 to 5 components, ranked in order of the amount of variability summarized, were used to form the composite subject-wise discriminant scores. Additionally a step-wise analysis was performed using standard criteria for entering and/or removing individual predictor variables. Cross-validated misclassification rates were computed by Lachenbruch's hold-one-out method [17]. This approach yields an unbiased estimate of the true (expected) prediction error. In contrast to the AIC, which assesses the in-sample prediction error for the training set, cross-validation methods are direct estimates of the extra-sample error [20]. Like the AIC, the area under the curve (AUC) from an ROC analysis provides a measure of the accuracy of the model to correctly classify samples from the training set. As anticipated, PLS-LDA consistently performs better than PCA-LDA as indicated by lower misclassification rates. These results illustrate several of the known issues associated with the use of PCA for dimension reduction. Since PCA summarizes gross variability, the first principal component may not in itself be a direction vector of spatial pattern weights suitable for discrimination among groups, as it appears to be the case here. The first PC fails to capture hardly any of the between-group variability. And further rotations as in oriented PCA approaches would prove ineffective. In addition to being incapable of distinguishing among-groups from within-groups variability, any benefit from including additional principal components into the subsequent LDA is quickly swamped by a sharp drop in the cross-validated rate of correctly classified cases due to a lack of generalizeability from over-fitting. Also, the step-wise approach fails to identify any suitable linear combination of PC scores for discrimination in this data. Whereas the PLS approach does show a leveling-off in the cross-validated classification rate as the number of components is increased and may even drop off, the rate of correctly classified cases is always higher than for dimension reduction by PCA. The stepwise discriminant analysis identifies a parsimonious linear model based on the first two PLS components. The improved performance of partial least squares comes about because the initial step of dimension reduction is informed by the ultimate goal of discrimination

among risk groups. Oriented PLS in turn outperforms both of the two-step PCA-LDA and PLS-LDA approaches and does not require any explicit dimension reduction up front.

Figure 1 illustrates the distribution of activation to the naming task by Talairach brain atlas region and risk group along with the resulting weights of the OrPLS direction vector for each region. Since only the left-hemisphere ROI measures were considered for analysis here, all brain regions can be shown. This pattern of varying weights across brain regions represents a topography that is differentially expressed across risk groups in the same sense as the brain networks discussed previously in the neuroimaging literature [1,2]. In addition to strong weighting on activation in the fusiform gyrus and superior parietal lobule, regions for which we have previously reported differences between risk groups using voxel-based analyses [11,14], the multivariate discrimination analysis with OrPLS also points to other regions such as the cingulate gyrus, precuneus, and frontal cortex. These latter regions are part of the so-called default-mode network, which has been implicated in MCI and mild AD [21,22].

The fMRI scans from the four-year follow-up study on the same subjects were used as a separate test data set to further validate the OrPLS technique. The direction vector of spatial ROI weights and the optimal threshold criterion on scores were established from the training data of the original study. Here, 3 of the 13 high-risk subjects (76.9% sensitivity) and 1 of the 11 low-risk subjects (90.9% specificity) are classified incorrectly, for an overall misclassification rate of 16.7% [23]. Table 3 shows a comparison of all of the methods when applied to the follow-up test data set. As expected, PCA-LDA exhibits an erratic behavior in terms of classification accuracy due to its lack of focus during training. However, the PLS-LDA and OrPLS approaches both yield classification accuracies that are consistent with the values obtained from cross-validation on the training data.

DISCUSSION

Principal component analysis has a long history as the workhorse for multivariate coordinate transformations and dimension reduction in statistical pattern recognition [24]. Yet, the variance summaries of PCA may not provide the best possible approach when discrimination is the goal and dimension reduction is required due to more feature variables than there are observations on individual subjects. We have illustrated here how PCA for dimension reduction of functional neuroimaging data may fail to capture important between-group variability, even as multiple components are retained, and with the PCs ultimately failing as predictors to yield a good classifier for discrimination among groups of subjects. Habeck et al. [25] observed fairly small values for the inverse coefficient of variation in a bootstrap resampling analysis of the stability of the spatial pattern of weights used to compute the discriminant scores. This may be seen as a reflection of the large variability between subjects in their level of fMRI activation. Thus, whereas PCA for dimension reduction has been used successfully as the initial step in conjunction with linear discriminant analysis for classification of functional neuroimaging data from PET studies, it is possible that methods of focused dimension reduction incorporating information on class structure such as PLS and OrPLS may be required in the multivariate analysis of distributed brain networks for group discrimination from fMRI data.

The use of a parcellation scheme fixed to the Talairach atlas for the reduction of voxel-wise activation measures into ROI mean values for analysis may have the undesired effect of diluting the activation and in turn reducing the sensitivity and specificity of the classification. This effect could be minimized by considering an intersection of parcellation-based and functionally defined ROI methods, where only those voxels that exhibit significant activation above some threshold at the group level are included in the calculation

of the subject-specific ROI mean values. Starting with a parcellation-based approach, researchers working with neuroimaging data from PET studies have already been successful in deriving stable and generalizeable covariance network patterns for group discrimination directly from higher-resolution voxel-based data and using PCA-LDA [1,2,26]. However, others working with a multivariable support vector machine (SVM) pattern classification approach found the cross-validated classification accuracy to be dropping for over 70 regions after an initial increase as more region features were included in the analysis [27]. Our own preliminary experiences with PLS-type approaches from different fMRI data sets and a parcellation scheme for a larger number of regions as well as with structural MRI network data have indicated a similar U-shaped dependence of the classification accuracy on the number of regions being considered. With voxel-wise fMRI data resampled through interpolation on a 2 mm grid in Talairach space, we observed a cresting of the cross-validated classification accuracy for a 5-component PLS model based on 2,500 voxel variables (unpublished results). Nevertheless, the very high classification accuracy of close to 100% afforded by the focused dimension reduction of PLS in the context of discrimination could not have been achieved with a similar PCA-based approach.

Our results support multivariate methods for discrimination and the potential of using fMRI as a biomarker for diagnosis in individuals with high risk of future AD. In particular, we see the superior performance of OrPLS for dimension reduction in linear discriminant analysis as compared to the PCA-LDA approaches for identification of brain functional networks that have appeared in the recent neuroimaging literature. The OrPLS direction vector indicates which topography of Talairach regions or which linear combination of regional activation responses best separates the two groups. In this neuroimaging data from our study of presymptomatic Alzheimer's disease, the OrPLS vector implicates a brain network which includes not only the ventral temporal lobe, Brodmann areas 19 and 37, but also the precuneus and cingulate gyrus. The multivariate discriminant analysis techniques, initially trained on the baseline data, were applied to a separate fMRI test data set from follow-up scans acquired on the same subjects and using the same protocol but four years later. The OrPLS method yielded the same low misclassification rate as with cross-validation on the original baseline data. In the future, analysis of an independent sample will be used to further validate this technique. Furthermore, with the increased complexity as we seek to apply OrPLS in a setting of voxel-wise rather than ROI-based neuroimaging data, computational approaches for dealing with the inversion of the very large covariance matrices of dimension voxels by voxels will have to be developed.

Acknowledgments

This work was supported by a grant from the National Institute of Neurological Disorders and Stroke (R01-NS036660).

REFERENCES

- [1]. Scarmeas N, Habeck CG, Zarahn E, Anderson KE, Park A, Hilton J, Pelton GH, Tabert MH, Honig LS, Moeller JR, Devenand DP, Stern Y. Covariance PET patterns in early Alzheimer's disease and subjects with mild cognitive impairment but no dementia: utility in group discrimination and correlations with functional performance. *NeuroImage*. 2004; 23(1):35–45. [PubMed: 15325350]
- [2]. Stern Y, Habeck C, Moeller J, Scarmeas N, Anderson KE, Hilton HJ, Flynn J, Sackeim H, van Heertum R. Brain networks associated with cognitive reserve in healthy young and old adults. *Cereb Cortex*. 2005; 15(4):394–402. [PubMed: 15749983]
- [3]. Habeck C, Foster NL, Pernheczky R, Kurz A, Alexopoulos P, Koeppe RA, Drzezga A, Stern Y. Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. *NeuroImage*. 2008; 40(4):1503–1515. [PubMed: 18343688]

- [4]. Barker M, Rayens WS. A partial least squares paradigm for discrimination. *J Chemometrics*. 2003; 17(3):166–173.
- [5]. McIntosh AR, Bookstein FL, Haxby JV, Grady CL. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*. 1996; 3(3):143–157. [PubMed: 9345485]
- [6]. Rayens WS, Andersen A. Oriented partial least squares. *Riv Stat Appl—Ital J Appl Stat, RCE Edizioni, Napoli*. 2003; 15(3):367–388.
- [7]. Rayens WS, Andersen AH. Multivariate analysis of fMRI data by oriented partial least squares (OrPLS). *Magn Reson Imag*. 2006; 24(7):953–958.
- [8]. Liu Y, Rayens W. PLS and dimension reduction for classification. *Comp Statist*. 2007; 22:189–208.
- [9]. Rayens W, Liu Y, Andersen A, Smith C. Using OrPLS to identify asymptotic women at risk for Alzheimer's disease. *J Chemometrics*. 2008; 22(9):522–527.
- [10]. Mourão-Miranda J, Bokde AL, Born C, Hampel H, Stetter M. Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. *NeuroImage*. 2005; 28(4):980–995. [PubMed: 16275139]
- [11]. Smith CD, Andersen AH, Kryscio RJ, Schmitt FA, Kindy MS, Blonder LX, Avison MJ. Altered brain activation in cognitively intact individuals at high risk for Alzheimer's disease. *Neurology*. 1999; 53(7):1391–1396. [PubMed: 10534240]
- [12]. Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996; 29:162–173. [PubMed: 8812068]
- [13]. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT. Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp*. 2000; 10(3):120–131. [PubMed: 10912591]
- [14]. Smith CD, Kryscio RJ, Schmitt FA, Lovell MA, Blonder LX, Rayens WS, Andersen AH. Longitudinal functional alterations in asymptomatic women at risk for Alzheimer's disease. *J Neuroimaging*. 2005; 15(3):271–277. [PubMed: 15951411]
- [15]. Friedman J. Regularized discriminant analysis. *J Am Stat Assoc*. 1989; 84(405):165–175.
- [16]. Diamantaras, KI.; Kung, SY. *Principal Component Neural Networks*. Wiley; New York: 1996.
- [17]. Lachenbruch PA, Mickey MR. Estimation of error rates in discriminant analysis. *Technometrics*. 1968; 10:1–11.
- [18]. Flury B, Riedwyl H. T2 tests, the linear two-group discriminant function, and their computation by linear regression. *The American Statistician*. 1985; 39(1):20–25.
- [19]. de Jong S. PLS fits closer than PCR. *J Chemometrics*. 1993; 7:551–557.
- [20]. Hastie, T.; Tibshirany, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. Springer; New York: 2009.
- [21]. Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc Natl Acad Sci USA*. 2004; 101(13):4637–4642. [PubMed: 15070770]
- [22]. Rombouts SA, Barkhof F, Goekoop R, Stam CJ, Scheltens P. *Hum Brain Mapp*. 2005; 26(4): 231–239. [PubMed: 15954139]
- [23]. Rayens, WS.; Liu, Y.; Andersen, AH.; Smith, CD. *Multivariate Methods for Discrimination in the Analysis of fMRI Data*. 15th Scientific Meeting of the International Society for Magnetic Resonance in Medicine; Berlin. May 2007; p. 1851
- [24]. Fukunaga, K. *Introduction to Statistical Pattern Recognition*. Academic Press; New York: 1972.
- [25]. Habeck C, Hilton J, Zarahn E, Flynn J, Moeller J, Stern Y. Relation of cognitive reserve and task performance to expression of regional covariance networks in an event-related fMRI study of nonverbal memory. *NeuroImage*. 2003; 20(3):1723–1733. [PubMed: 14642482]
- [26]. Habeck C, Stern Y. Multivariate data analysis for neuroimaging data: Overview and application to Alzheimer's disease. *Cell Biochem Biophys*. 2010; 58:53–67. [PubMed: 20658269]
- [27]. Fan Y, Resnick SM, Wu X, Davatzikos C. Structural and functional biomarkers of prodromal Alzheimer's disease: A high-dimensional pattern classification study.

Figure 1. Illustrating left-hemisphere activation in the two risk groups and the OrPLS weights for discrimination across Talairach atlas brain regions (cerebellar regions not shown).

Figure 2. Illustrating the OrPLS weights overlaid onto sagittal sections at positions 48, 26, and 4 mm. The color scale for the signed weights is linear with 100% depicting the fusiform gyrus region.

Table 1

Comparison of PCA- and PLS-regression. The Akaike AIC values are shown on a relative scale as differences with the AIC for a model based on one PLS component set to zero.

Number of Components	Principal Component Regression					Partial Least Squares Regression				
	Variance summarized	Correlation with class structure	Residual sum of squares	Akaike AIC	p-value	Variance summarized	Correlation with class structure	Residual sum of squares	Akaike AIC	p-value
1	0.4344	0.0011	5.9583	4.63	0.9961	0.1565	0.4189	4.9127	0	0.0416
2	0.6095	0.2454	5.5995	5.14	0.5209	0.2273	0.6010	3.8065	-4.12	0.0091
3	0.7112	0.2750	5.5076	6.74	0.6568	0.4073	0.6798	3.2046	-6.25	0.0054
4	0.7900	0.2950	5.4398	8.45	0.7692	0.7152	0.7179	2.8879	-6.75	0.0061
5	0.8396	0.3569	5.1993	9.36	0.7540	0.7725	0.7849	2.2877	-10.34	0.0024
6	0.8706	0.3661	5.1595	11.18	0.8429	0.8221	0.8366	1.7884	-14.25	0.0010
7	0.8977	0.4043	4.9842	12.35	0.8582	0.8650	0.8751	1.3955	-18.2	0.0004
8	0.9184	0.4434	4.7869	13.38	0.8662	0.8851	0.9070	1.0568	-22.88	0.0002
9	0.9348	0.4602	4.6966	14.92	0.9045	0.8995	0.9383	0.7124	-30.34	0.0000
10	0.9491	0.5323	4.2700	14.64	0.8517	0.9159	0.9583	0.4869	-37.48	0.0000

Table 2

Classification accuracy in % from dimension reduction by PCA or PLS followed by linear discriminant analysis (LDA). Sensitivity reflects correctly classified high-risk subjects; equal prior probabilities for groups was assumed. The AUC denotes the area under the curve from ROC analysis. The cross-validated misclassification rates (%) for the baseline training data set were computed using Lachenbruch's hold-one-out method.

		PCA-LDA					PLS-LDA					OrPLS		
Number of Components		1	2	3	4	5	Step wise ¹	1	2	3	4	5	Step wise ²	1
Original Data	Sensitivity	46.2	69.2	61.5	61.5	61.5		69.2	84.6	76.9	84.6	100.0	84.6	92.3
	Specificity	54.5	45.5	63.6	54.5	63.6		63.6	72.7	81.8	81.8	90.9	72.7	81.8
	Correctly Classified	50.0	58.3	62.5	58.3	62.5	NA	66.7	79.2	79.2	83.3	95.8	79.2	87.5
	AUC	0.52	0.65	0.64	0.64	0.70		0.73	0.84	0.90	0.94	0.96	0.84	0.89
Cross-Validated by leave one out method	Sensitivity	0.0	61.5	53.8	46.2	30.8		69.2	76.9	76.9	69.2	76.9	76.9	84.6
	Specificity	9.1	45.5	45.5	36.4	18.2		63.6	63.6	63.6	72.7	72.7	63.6	81.8
	Correctly Classified	4.2	54.2	50.0	41.7	25.0	NA	66.7	70.8	70.8	70.8	75.0	70.8	83.3

¹ Step-wise LDA using the first five PCA components fails as no variables are qualified for analysis by standard selection criteria.

² Step-wise LDA using the first five PLS components enters variables 1 and 2.

Table 3

Classification accuracy in % for the follow-up test data set. The spatial patterns of weights used to determine the subject-wise scores for classification are those determined from the training data in the corresponding Table 2. The respective classification threshold values are held fixed from the training data as well.

		PCA-LDA					PLS-LDA					OrPLS		
Number of Components		1	2	3	4	5	Step wise	1	2	3	4	5	Step wise	1
Repeat Test Data	Sensitivity	92.3	61.5	61.5	61.5	69.2		61.5	84.6	84.6	69.2	76.9	84.6	76.9
	Specificity	54.5	54.5	63.6	63.6	45.5		63.6	54.5	63.6	54.5	63.6	54.5	90.9
	Correctly Classified	75.0	58.3	62.5	62.5	58.3	NA	62.5	70.8	75.0	62.5	70.8	70.8	83.3