First publ. in: 2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC) / Ed. Bo Yu. IEEE, 2010, pp. 3721-3725

Classification Accuracy of Multivariate Analysis Applied to ^{99m}Tc-ECD SPECT Data in Alzheimer's Disease Patients and Asymptomatic Controls

Dorit Merhof [1], Pawel J. Markiewicz [2], Jérôme Declerck [1], Günther Platsch [4], Julian C. Matthews [3], and Karl Herholz [2]

***Abstract*—With increasing life expectancy in developed countries, there is a corresponding increase in the frequency of diseases typically associated with old age, in particular dementia. In recent research, multivariate analysis of Positron Emission Tomography (PET) datasets has shown potential for classification between Alzheimer's disease (AD) patients and asymptomatic controls. In this work, the feasibility of multivariate analysis using Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA) of Single Photon Emission Computed Tomography (SPECT) data is investigated. In order to obtain robust and reliable results, bootstrap resampling is applied and the robustness and classification accuracy of PCA/FDA are investigated. The robustness of the analysis is assessed by estimating the distribution of the angle between PCA/FDA discriminative vectors generated by bootstrap resampling, and the classification predictive accuracy is assessed using the .632 bootstrap estimator. The results indicate that PCA/FDA on SPECT data enables a robust differentiation between AD patients and asymptomatic controls based on three principal components, with a classification accuracy of 89%.**

***Index Terms*—Single photon emission computed tomography (SPECT), Alzheimer's disease (AD), Multivariate Analysis, Principal Component Analysis (PCA), Classification Accuracy.**

I. Introduction

ALZHEIMER'S DISEASE (AD) is the most common cause of dementia. The socioeconomic impact of dementia is extraordinarily large, with considerable effort being made to understand the pathophysiologic mechanisms of AD in order to further the development of effective treatment strategies for the disease.

According to pathologic studies, neurodegeneration in AD begins in the entorhinal cortex, progresses to the hippocampus, the limbic system and neocortical regions. It is characterized by accumulations of amyloid plaques and neurofibrillary tangles, which exert direct and indirect neurotoxic effects leading to neuronal loss. As a result, the affected cortical regions show reduced glucose metabolism on 2-fluoro-2-deoxy-D-glucose (FDG) Positron Emission Tomography (PET), and decreased brain perfusion as observed in technetium-99m-ethylcysteinatedimer (^{99m}Tc-ECD) Single Photon Emission Computed Tomography (SPECT).

In recent research, multivariate analysis has successfully been applied for classification between PET datasets of asymptomatic controls and AD patients [3], [4], [5]. Multivariate analyisis takes into account statistical relationships between all voxels simultaneously, and therefore has a greater statistical power and superior diagnostic performance compared to univariate analysis [4]. Multivariate analysis is therefore much better suited for applying group analysis results to new datasets for diagnostic evaluation.

In this work, multivariate analysis of Single Photon Emission Computed Tomography (SPECT) datasets of AD patients and asymptomatic controls is performed. The goals of this work are to assess whether multivariate analysis techniques are feasible in SPECT datasets, and to investigate the classification accuracy and robustness in asymptomatic controls and AD patients.

II. Image Data

A. Patient population

The ^{99m}Tc-ECD SPECT datasets used in this work comprise 28 patients with mild to medium AD (mixed population of female and male) with an age between 52 and 81 years (mean $\pm$ standard deviation: 67.4 $\pm$ 7.5), as well as 28 asymptomatic controls (mixed population of female and male) with an age between 50 and 78 years (mean $\pm$ standard deviation: 61.6 $\pm$ 8.0). The datasets were acquired between 2003 and 2008 at the Clinic of Nuclear Medicine, University of Erlangen-Nuremberg, Erlangen, Germany.

B. Acquisition parameters

Injection of 20mCi (740 MBq) of ^{99m}Tc labeled ECD was performed on subjects under rest conditions. The patients were lying with eyes closed in a quiet, dark or dimly lit environment from at least 10 minutes prior until 5 minutes post injection. For image acquisition, the patients were lying down supine in the scanner, with their arms down. The head was placed naturally so that the patients felt comfortable and motion could be minimized during the acquisition. The image data was acquired on a Siemens MultiSPECT3 scanner 30 minutes after injection of the tracer, with a scan duration of 30 minutes at

[1] Siemens Molecular Imaging, Oxford, UK
[2] Research School of Translational Medicine at the University of Manchester, Wolfson Molecular Imaging Centre, Manchester, UK
[3] School of Cancer, Enabling Sciences and Technology, Manchester Academic Health Science Centre, Wolfson Molecular Imaging Centre, The University of Manchester, Manchester, UK
[4] Siemens Molecular Imaging EU, Erlangen, Germany

most. The field of view of the image contained the entire brain and the cerebellum. The projection data was processed with filtered back projection, and Chang attenuation correction was applied.

C. Criteria for asymptomatic controls and AD patients

The asymptomatic datasets originate from patients who were referred to the Clinic of Nuclear Medicine for brain perfusion SPECT for diagnostic purposes other than dementia, but the results of the scans were normal. Further clinical investigations showed no evidence of any diseases which would lead to an altered brain perfusion pattern. A CT or MR was performed between four weeks before and four weeks after the SPECT examination and there were no clinical events between MR/CT and SPECT.

The AD datasets originate from daily clinical routine and were not collected as part of a prospective dementia study. For this reason, no neuropsychologic measures are available for these patients. All dementia patients were referred to the Clinic of Nuclear Medicine with questions related to diagnostic findings. The patients included in the AD population comprise mild and medium AD cases with uptake patterns typical for dementia of AD type.

D. Data preprocessing

The datasets were further processed in order to enable comparison on a voxel-by-voxel basis, as required by multivariate analysis. In a first step, all datasets were spatially registered to a common brain template. In order to account for intra-patient variations in gyri, injected dose, the resolution of the reconstruction of scans, and other factors, all of which cannot be compensated for by registration alone, the images were smoothed with a standard isotropic Gaussian filter with full width half maximum (FWHM) of 12 mm. Finally, intensity normalization according to the whole brain region was applied, which is required in order to allow for direct comparison of voxel values across the population. Since intensity values of the datasets may vary arbitrarily depending on factors such as injected dose and systemic tracer elimination, intensity normalization is needed in order to obtain meaningful statistics.

III. Methods

In this section, multivariate image analysis based on Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA) is outlined (Sections III-A and III-B). In order to assess the robustness and predictive accuracy of the approach, bootstrap resampling is applied (Section III-C), and the classification accuracy (Section III-D) as well as the robusness of classification (Section III-E) are investigated.

A. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) [6] is a multivariate analysis method that allows identifying the orthogonal directions of dominant variation in the data. The first principal component (PC) accounts for as much of the variability in the data as possible by a single component, and each succeeding component accounts for as much as possible of the remaining uncorrelated variability. PCA can thus project high-dimensional data onto a lower dimensional space represented by a subset of PCs, which can be more easily explored in order to analyze the underlying structure of the data. Even though each dataset can exactly be represented as a linear combination of all PCs, data analysis usually only retains a few PCs in order to focus on the main variations of the data and to take advantage of the dimensionality reduction effect obtained by PCA, considering the remaining PCs as noise or atypical variations.

The PCs are computed as the eigenvalues of the covariance matrix of the data matrix $\mathbf{X}$, which comprises all m datasets. Each column of $\mathbf{X}$ represents one dataset (i.e., contains the n voxels of the whole-brain region of this particular dataset). The PCs are computed using Singular Value Decomposition (SVD). However, since an individual dataset may be used multiple times in a bootstrap sample (Section III-C), the SVD may become numerically unstable. In these cases, a more stable but slightly slower method for calculating the PCs called Non-Linear Iterative Partial Least Squares (NIPALS) [7], [8] is used.

B. Fisher Discriminant Analysis (FDA)

Since the PCs represent variation within the whole population (i.e., asymptomatic controls and AD patients), a Fisher Discriminant Analysis (FDA) is performed in a second step in order to identify the direction of best separation between the two groups. The FDA provides a linear combination of the PCs so that both groups are well separated.

The goal of FDA is to identify a discrimination vector $\mathbf{w}$ such that projecting each dataset onto this vector provides the best possible separation between both groups. In order to obtain a good separation of the projected data, it is desirable that the difference between the means of each class is large relative to some measure of the variation in each class. For this reason, the criterion function J that is maximized in FDA [1]

$$J(\mathbf{w}) = \frac{\mathbf{w}^t \mathbf{S}_B \mathbf{w}}{\mathbf{w}^t \mathbf{S}_W \mathbf{w}}, \tag{1}$$

is based on the between-class scatter matrix $\mathbf{S}_B$ and the within-class scatter matrix $\mathbf{S}_W$

$$\mathbf{S}_B = \sum_c (\mathbf{m_c} - \bar{\mathbf{x}})(\mathbf{m_c} - \bar{\mathbf{x}})^T, \tag{2}$$

$$\mathbf{S}_W = \sum_c \sum_{i \in c} (\mathbf{x_i} - \mathbf{m_c})(\mathbf{x_i} - \mathbf{m_c})^T, \tag{3}$$

where $\bar{\mathbf{x}}$ is the mean image vector across subjects, c represents the classes to be separated, and $\mathbf{m_c}$ are the class means.

In mathematical physics, Equation 1 is well known as the generalized Rayleigh quotient. It can be shown that the solution $\mathbf{w}$ that optimizes J is [1]

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2), \tag{4}$$

whereas $\mathbf{m}_1$ and $\mathbf{m}_2$ indicate the n-dimensional sample means of the two populations.

The resulting discriminant image (or discriminant vector) is used for projecting individual images (vectors) onto it, yielding a scalar value that discriminates between the two groups.

C. Bootstrap resampling

In order to evaluate the predictive accuracy of the PCA/FDA model trained on a limited sample, .632 bootstrap resampling is applied. In statistics, resampling techniques are commonly used to validate models and to assess their statistical accuracy by using random subsets (bootstrapping, cross validation) [2].

For a total number of 500 replications, 28 asymptomatic controls and 28 AD patients are randomly drawn from each group, resulting in a new bootstrap sample per replication. For each bootstrap sample, PCA is performed, as well as FDA for different numbers of PCs. Bootstrap resampling followed by PCA and FDA is performed for different numbers of PCs, and the classification accuracy is calculated (see Section III-D) in order to evaluate the different data preprocessing methods.

D. Classification Accuracy

Since the bootstrap samples are drawn with replacement, the probability of any given instance not being part of the n instances in the sample is

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{n} \approx e^{-1} \approx 0.368\,. \qquad (5)$$

The expected number of instances from the original dataset appearing in each training sample is thus 0.632.

The classification accuracy based on .632 bootstrap resampling is estimated as follows: (1) The subjects not included in a given bootstrap sample are used to estimate the predictive accuracy of the PCA/FDA discrimination analysis determined for the bootstrap sample. Since on average only 63.2% subjects are used for training, the resulting estimation of the predictive accuracy is biased downward and hence has to be (2) 'corrected' by a term which is found based on the whole sample treated as a training and testing set. Thus according to the .632 bootstrap, the predictive accuracy of the discrimination analysis is estimated as:

$$\mathrm{acc}_{boot} = \frac{1}{b}\sum_{i=1}^{b}\left(0.632\cdot \mathrm{acc}_i + 0.368\cdot \mathrm{acc}_{training}\right), \qquad (6)$$

where b is the number of bootstrap replications, acc_i is the accuracy of the individual bootstrap sample that corresponds to replication i, and $\mathrm{acc}_{training}$ is the accuracy of the training set [2]. In order to provide a more accurate estimate of $\mathrm{acc}_{training}$, again bootstrap resampling based on 200 iterations is applied, whereas the same instances used in the bootstrap sample are also used for calculating the accuracy, which is averaged across all samples.

E. Angle between FDA vectors

The robustness of classification is indicated by the sampling distribution of the angle between FDA vectors. For each bootstrap replication, the angle between the FDA vector based on the whole sample and the FDA vector based on the bootstrap sample is calculated for different numbers of PCs. A small angle and a narrow distribution of angles are signs for a robust classification. For an increasing number of PCs, the angle is expected to increase, since the PCA and FDA analysis adapts to features specific of the sample, rather than features that differentiate between both classes, which indicates a decrease of robustness.

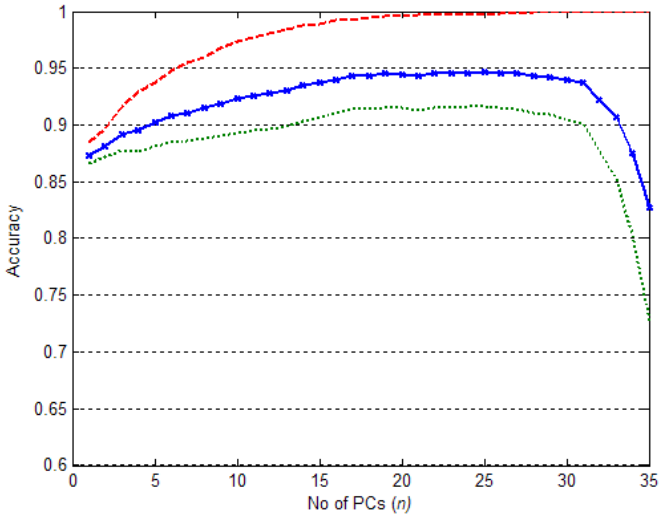

Fig. 1. Classification accuracy of the training set *(red, dashed)*, for individual bootstrap samples (averaged) *(green, dotted)* and for the bootstrap .632 estimator (corrected, averaged) *(blue, solid)*, depending on the number of PCs included into the analysis.

IV. Results

According to the classification accuracy (average for all bootstrap replications) in Figure 1, the discriminant image obtained from PCA/FDA allows differentiating between AD

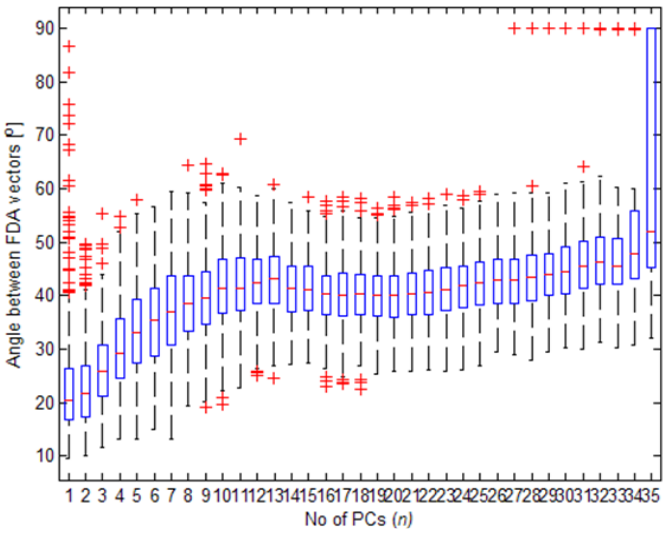

Fig. 2. Angle between the discriminant vectors resulting from 500 boostrap replications, for different numbers of PCs.

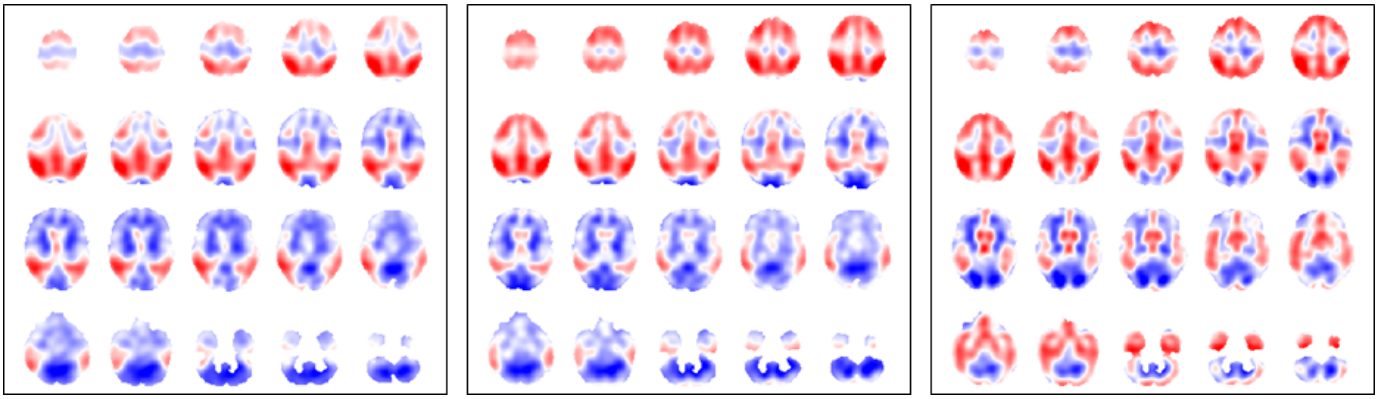

Fig. 3. Discriminant image based on one *(left)*, three *(middle)* and six *(right)* PCs and FDA.

patients and asymptomatic controls with an accuracy between 87% and 94%.

Although the classification accuracy increases with increasing number of PCs included into the analysis, only a limited number of PCs contribute to a robust classification. This is shown in Figure 2, where the angle between the discriminant vectors resulting from the 500 bootstrap replications is displayed for different numbers of PCs included into the analysis. For an increasing number of PCs, this angle increases, indicating that the model adapts to noise and features specific to the individual sample, rather than features characteristic of the two populations (AD patients and asymptomatic controls). It can therefore be concluded that the increase of accuracy for an increasing number of PCs occurs at the cost of robustness. For this reason, it can be assumed that a small number of PCs provides a good trade-off between robust analysis and high accuracy. More precisely, three PCs seems to be a particularly good choice, which is consistent with previous investigations based on FDG PET data [5].

Figure 3 shows the discrimination image for one, three and six PCs, with more intense colors (blue, red) indicating a higher contribution to the discrimination. It should be noted that areas of high hypo-metabolism (red) are located in the temporal, parietal and frontal lobes, as expected in AD patients. The central region is not affected and only shows very small values (white, and light blue and red), which is according to expectations.

V. Discussion

In this work, the feasibility of multivariate analysis applied to SPECT data is investigated. In order to obtain robust and reliable results, bootstrap resampling is used and the robustness and classification accuracy of PCA/FDA are assessed.

A limitation of the results relates to the fact that no neuropsychologic measure is available as the datasets originate from daily clinical routine and were not collected as part of a prospective dementia study. As a result, the accuracy and robustness measures reported in this work are based on visual reading by an expert nuclear medicine physician as ground truth, i.e. the system has been optimized to achieve classification results which are comparable to visual reading by a medical expert.

The classification accuracy provides evidence that multivariate analysis of SPECT data is feasible. A good trade-off between accuracy and robustness can be achieved for a low number of PCs (e.g., three PCs seem to be a good choice). The results are very promising and are actually comparable to the results based on multivariate analysis is PET datasets [5]. Due to the substantially less cost of SPECT compared to FDG PET, it is worthwhile to optimise the analysis of SPECT for broader clinical use.

The visualization of axial slices of the discriminant image shows patterns typically expected in dementia of AD type, with high values in the temporal, parietal and frontal lobes indicating a higher contribution to the discrimination. In a way, these visualizations confirm that PCA/FDA is a viable approach in SPECT datasets.

VI. Conclusion

The accuracy results show that multivariate analysis of SPECT data is feasible, with a predictive accuracy of 89% if three PCs are used. According to the results, high classification accuracy is already obtained with a low number of PCs, which indicates the discriminative power of the approach.

VII. Acknowledgment

This work is funded by Siemens Molecular Imaging, Oxford, UK. We thank Prof. Torsten Kuwert, Clinic of Nuclear Medicine, University of Erlangen-Nuremberg, Erlangen, Germany, for providing the ^{99m}Tc-ECD SPECT data.

References

[1] R.O. Duda and P.E. Hart and D.G. Stork, *Pattern classification*, John Wiley & Sons, Inc., 2001

[2] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall/CRC, 1993

[3] J. Fripp and P. Bourgeat and O. Acosta and P. Raniga and M. Modat and K.E. Pike and G. Jones and G. O'Keefe and C.L. Masters and D. Ames and K.A. Ellis and P. Maruff and J. Currie and V.L. Villemagne and C.C. Rowe and O. Salvado, S. Ourselin, *Appearance modeling of 11C PiB PET images: characterizing amyloid deposition in Alzheimer's disease, mild cognitive impairment and healthy aging*, Neuroimage 43(3), pp. 430-439, 2008

[4] C. Habeck and N.L. Foster and R. Perneczky and A. Kurz and P. Alexopoulos and R.A. Koeppe and A. Drzezga and Y. Stern, *Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease*, Neuroimage 40(4), pp. 1503-1515, 2008

[5] P.J. Markiewicz and J.C. Matthews and J. Declerck and K. Herholz, *Robustness of multivariate image analysis assessed by resampling techniques and applied to FDG-PET scans of patients with Alzheimer's disease* Neuroimage 46(2), pp. 472-485, 2009
[6] K. Pearson, *On Lines and Planes of Closest Fit to Systems of Points in Space*, Philosophical Magazine 2(6), pp. 559-572, 1901
[7] H. Wold, *Estimation of principal components and related models by iterative least squares*, In: Multivariate Analysis (Ed., P.R. Krishnaiah), Academic Press, 1966
[8] H. Wold, *Path models with latent variables: the NIPALS approach*, In: Quantitative Sociology: International perspectives on mathematical and statistical model building (Ed., H.M. Blalock et al.), Academic Press, 1975