

Multiple Discriminant Analysis of SPECT Data for Alzheimer’s Disease, Frontotemporal Dementia and Asymptomatic Controls

Elisabeth Stühler, Günther Platsch, Markus Weih, Johannes Kornhuber, Torsten Kuwert and Dorit Merhof

Abstract—Multiple discriminant analysis (MDA) is a generalization of the Fisher discriminant analysis (FDA) and makes it possible to discriminate more than two classes by projecting the data onto a subspace. In this work, it was applied to technetium-99methylcysteinatedimer ($^{99m}\text{Tc-ECD}$) SPECT datasets of 10 Alzheimer’s disease (AD) patients, 11 frontotemporal dementia (FTD) patients and 11 asymptomatic controls (CTR). Principal component analysis (PCA) was used for dimensionality reduction, followed by projection of the data onto a discrimination plane via MDA. In order to separate the different groups, linear boundaries were calculated by applying FDA to two classes at a time (linear machine). By executing the F-test for different numbers of principal components and examining the corresponding classification accuracy, an optimal discrimination plane based on the first three principal components was determined. In order to further assess the method, another dataset comprising patients with early-onset AD and FTD (beginning or suspected disease) was projected by the same method onto this discrimination plane, resulting in a correct classification for most cases.

The successful discrimination of another dataset on the same plane indicates that the model is well suited to account for disease-specific characteristics within the classes, even for patients with early-onset AD and FTD.

Index Terms—SPECT, Alzheimer’s disease, frontotemporal dementia, principal component analysis, multiple discriminant analysis, linear machine, resampling.

I. INTRODUCTION

ALZHEIMER’S disease, the most common form of degenerative dementia, and frontotemporal dementia are amongst the most prevalent neurodegenerative diseases [1]. Disease progression in both AD and FTD results in specific patterns: AD extends from the entorhinal cortex to the hippocampus, the limbic system, and neocortical regions, whereas FTD mainly involves the frontal lobes and may extend to the temporal lobes.

Multivariate analysis of medical images has proven potential to extract significant features and statistical patterns from whole brain PET and SPECT datasets. To date, various methods have been proposed (e.g. [2]–[4]) to discriminate two classes of data, usually to distinguish a disease group from a control

group.

An important characteristic of a discrimination method is its capability to distinguish not only different types of dementia, but also early stages of the disease.

In this work, multivariate image analysis of SPECT datasets is performed. The goal of this work is to discriminate between patients with AD or FTD and asymptomatic controls (CTR), and to test the discrimination on patients with beginning or suspected disease.

The following sections present an approach to differentiate three or more groups of data. The generalization of the two-class case was achieved by principal component analysis and subsequent multiple discriminant analysis, as proposed in [5]. Subsequently, the data was classified by application of a linear machine [5].

II. MATERIAL AND METHODS

A. Image Data

The $^{99m}\text{Tc-ECD}$ SPECT datasets were acquired at the Clinic of Nuclear Medicine, University of Erlangen-Nuremberg. All data was preprocessed based on affine registration, Gaussian smoothing with an FWHM of 12mm, and intensity normalization according to the 25% brightest voxels within the whole-brain region [6]. Overall, 48 subjects (age over 50, mean age 66 ± 8.44 , 25 females, 23 males) were included in the analysis:

- Alzheimer dementia:
 - 10 cases of Alzheimer dementia (AD)
 - 8 cases of beginning AD (bAD)
 - 4 cases where AD was suspected (sAD)
- Frontotemporal dementia:
 - 11 cases of frontotemporal dementia (FTD)
 - 2 cases of beginning FTD (bFTD)
 - 2 cases where FTD was suspected (sFTD)
- 11 asymptomatic controls (CTR)

The training set for the principal component analysis and the following multiple discriminant analysis contained only the classes AD, FTD and CTR. The image data of the training set constitutes the data matrix X , where the rows represent the data values and the columns represent individual subjects. Subsequently, X was centered by setting the overall mean to zero.

B. Principal component analysis

Due to the vast amount of variables (226985 voxels of the whole brain region), it is necessary to perform a dimensionality

E. Stühler is with the University of Konstanz, Konstanz, Germany.

G. Platsch is with Siemens Molecular Imaging EU, Erlangen, Germany.

M. Weih is with the Department of Psychiatry and Psychotherapy, University of Erlangen-Nuremberg, Erlangen, Germany.

J. Kornhuber is with the Department of Psychiatry and Psychotherapy, University of Erlangen-Nuremberg, Erlangen, Germany.

T. Kuwert is with the Clinic of Nuclear Medicine, University of Erlangen-Nuremberg, Erlangen, Germany.

D. Merhof is with the University of Konstanz, Konstanz, Germany.

reduction prior to the discriminant analysis. Principal component analysis (PCA) reduces the data efficiently via singular value decomposition of X , where the principal components (PCs) are the left singular vectors of X . PCs replace the variables, and PC scores are calculated for each subject [4]. An important question to resolve is the number of PCs needed in order to represent the structure and variability of the data. This was addressed by executing the F-test. For every number of PCs, the residual sum of squares associated to the achieved PC scores and the corresponding F-values were calculated according to [4].

C. Multiple discriminant analysis

Multiple discriminant analysis (MDA) [5] is a generalization of the Fisher's discriminant analysis (FDA) and can be applied to differentiate between more than two classes. The basic idea is to project the data onto a discrimination hyperplane (of dimension k , which is the number of classes minus one), where the ratio of the general between-class scatter S_b and the general within-class scatter S_w is maximized.

To avoid elaborate calculations, it is useful to define the total scatter S_t [5], whose computation (after performing PCA) simplifies to:

$$S_t = (PC \cdot X) \cdot (PC \cdot X)^T = \text{diag}(\sigma^2), \quad (1)$$

where σ are the singular values of X , and PC are the left singular vectors of X . S_b can be derived by

$$S_b = \sum_c n_c \cdot (M_c - M)(M_c - M)^T, \quad (2)$$

where M is the overall mean, n_c is the number of elements in class c , and M_c is the mean vector of the class c . As the data was centered during PCA, $M = 0$ and the equation reduces to:

$$S_b = \sum_c n_c \cdot M_c M_c^T. \quad (3)$$

As $S_t = S_w + S_b$, the general within-class scatter S_w results to $S_w = S_t - S_b$.

The eigenvectors w_i of the generalized eigenvalue problem $S_b w_i = \lambda S_w w_i$ are the columns of the projection Matrix W . As S_b is of rank $c - 1$ (or less), there will be only $c - 1$ eigenvectors, and the data will be projected onto a hyperplane of dimension $c - 1$.

D. Linear Machine

A linear machine [5] divides the discrimination hyperplane into decision regions by linear boundaries, that can be determined by various methods. In this work, they are determined using an FDA on two classes at a time. This results in an FDA-vector, which is orthogonal to the direction vector for each boundary. The overall mean $M = 0$ of the training set is taken as reference for each boundary. As a result, the linear classifier between FTD and AD can be calculated according to Condition 4 (analogous for the two other boundaries), where B_{c_1, c_2} denotes the boundary between classes c_1 and c_2 , $S_{w, c}$ the scatter within class c , and M_c the mean vector of class c :

$$B_{FTD, AD} \perp (S_{w, FTD} + S_{w, AD})^{-1} \cdot (M_{FTD} - M_{AD}). \quad (4)$$

E. Accuracy

1) *Prediction accuracy via the .632 bootstrap estimator:* Accuracy of the MDA can be ascertained by resampling methods, in this case bootstrapping: For a bootstrap sample, random but stratified subjects are drawn with replacement from the training set, then PCA, MDA and the linear machine are applied to the sample.

The .632 bootstrap estimator, as presented in [7], is employed to estimate the prediction error $Err^{.632}$ of the MDA and the linear machine:

$$Err^{.632} = .632 \cdot err_{btp} + .368 \cdot err_t. \quad (5)$$

For any given number of PCs, the apparent error rate err_t of the training set is calculated by

$$err_t = \frac{|\text{misclassified subjects}|}{|\text{training set}|}. \quad (6)$$

As above, the training set contains only cases with assured diagnosis (AD, FTD and CTR).

err_{btp} describes the bootstrap expected error rate for the subjects not appearing in the bootstrap sample but in the validation set, and is determined as follows:

$$err_{btp} = \frac{1}{B} \cdot \sum_{b=1}^B \frac{|\text{misclassified subjects in valid. set}|}{|\text{subjects in valid. set}|} \quad (7)$$

For every bootstrap sample, MDA is performed and new boundaries are computed for the linear machine. Every subject of the validation set is projected onto the discriminant plane and its classification according to the linear machine is checked.

2) *Misclassification of one subject at a time:* The relabeling of one subject at a time simulates the possibility of misdiagnosis and the successive loss of accuracy of the MDA (and, subsequently, of the linear machine). For a fixed number of PCs, the classification of one subject at a time is deliberately changed in every iteration. MDA is performed for the whole training set including the relabeled case, and new boundaries are calculated for the linear machine. The impact of the misclassification on the relabeled subjects was measured by the apparent error rate $err_{relabel}$ [7]:

$$err_{relabel} = \frac{|\text{forced misclassifications}|}{2 \cdot |\text{training set}|} \quad (8)$$

Basically, $err_{relabel}$ describes the proportion of those relabeled cases, which have been forced out of their original class by the following execution of MDA.

III. RESULTS AND DISCUSSION

A. Number of PCs

PCA is performed and the F-test is executed. As a result, the first three PCs represent significant variance of the data, as can be seen in Figure 2. This is in line with the findings on PET data [4].

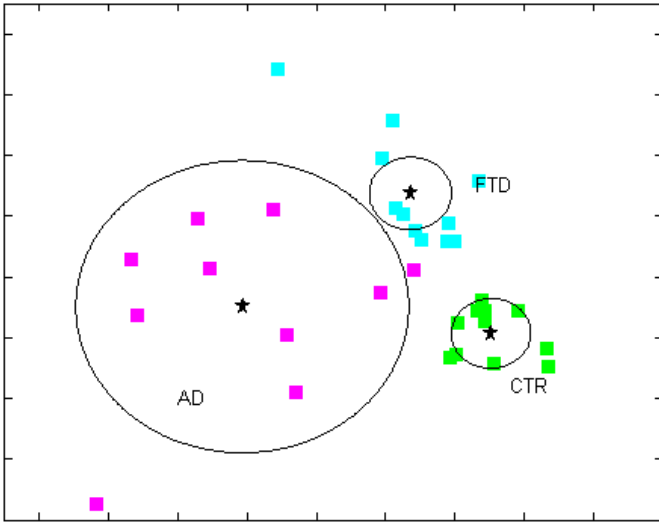


Fig. 1. The discrimination plane with class means and standard deviations of the training set.

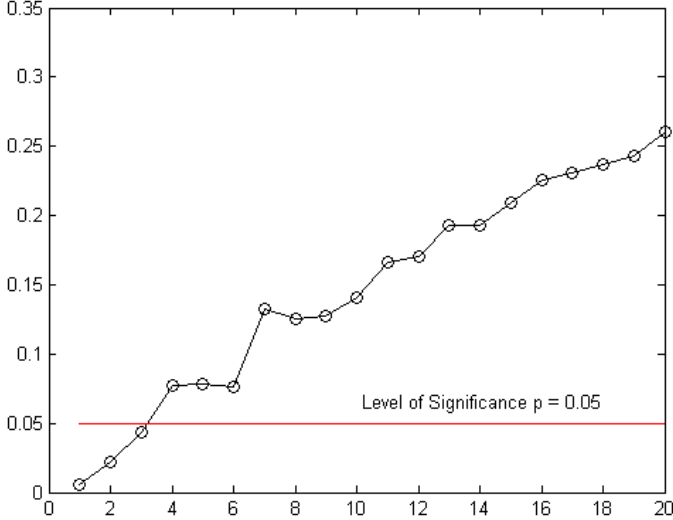


Fig. 2. F-test result: Only the first three PCs add significant variance to the analysis (for $p \leq 0.05$).

B. Multiple discriminant analysis

The training set comprising the three classes AD, FTD and CTR is projected onto a discrimination plane of dimension two, as shown in Figure 1, where class means and standard deviations of each class are indicated by circles.

C. Linear machine

As can be seen in Figure 3, all subjects of the training set are well separated by the boundaries calculated according to II-D, except for two AD cases. This is not surprising as the standard deviation of the AD subjects on the discrimination plane is comparatively large (in comparison to classes CTR and FTD, see Figure 1).

D. Accuracy

1) *Prediction accuracy via .632 bootstrap estimator*: Results of the estimated prediction accuracy $1 - Err^{.632}$, the apparent accuracy $1 - err_t$ and the bootstrap expected accuracy $1 - err_{btp}$ are depicted in Figure 4. The accuracy calculated

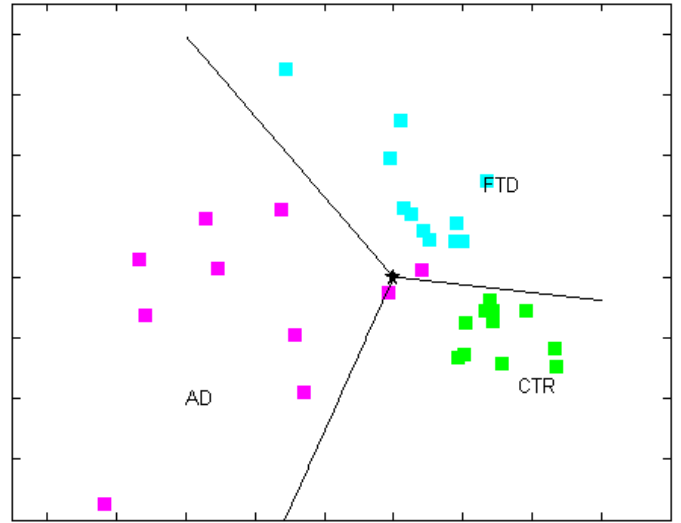


Fig. 3. FDA-boundaries on discrimination plane, classification of training set subjects (linear machine).

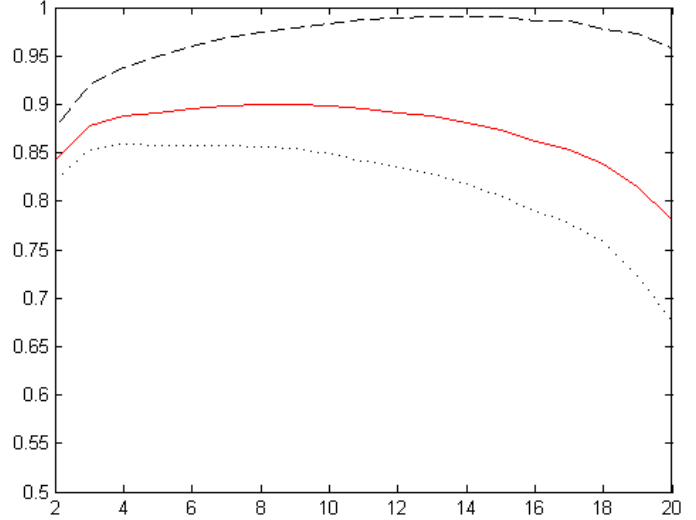


Fig. 4. Prediction accuracy for number of PCs indicated by the solid red line. The dotted and the dashed line represent respectively the bootstrap expected accuracy and the apparent accuracy.

by means of the .632 bootstrap estimator reaches its maximum for eight PCs, unlike the previous conclusion that three PCs represent the data in sufficient detail (F-test with significance level 0.05). According to [4], this apparent inconsistency is caused by the presumption that the diagnosis for every subject is correct. In the following section, accuracy analysis via deliberate misclassification demonstrates the impact of this presumption.

2) *Misclassification of one subject at a time*: The apparent error rate $err_{relabel}$ for one relabeled subject at a time is calculated as in Equation 8. The resulting accuracy $1 - err_{relabel}$ of the MDA reaches its maximum for three PCs and is still close to 90% (in case of three PCs: $acc_{relabel} \approx 87.5\%$) despite of misclassification. It steadily decreases for additional PCs, which indicates increasing overtraining of the MDA.

E. Test on early-onset cases

In the previous section, the optimal discrimination plane (with linear classifiers) was found by reducing the dataset to

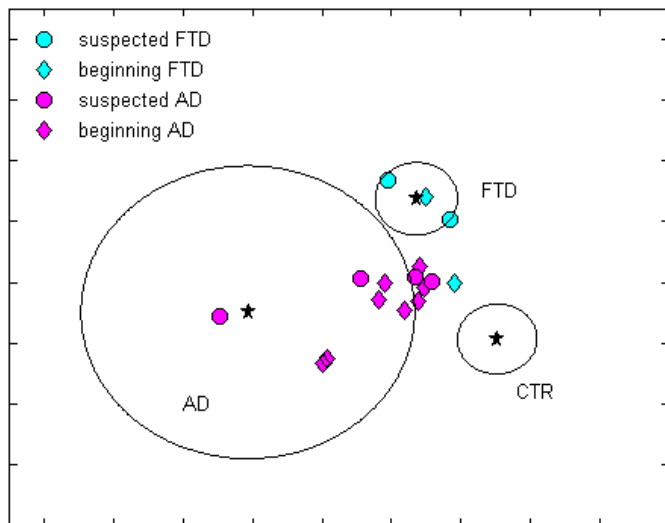


Fig. 5. Cases of beginning and suspected disease on the discrimination plane, with class means and standard deviations of the original training set.

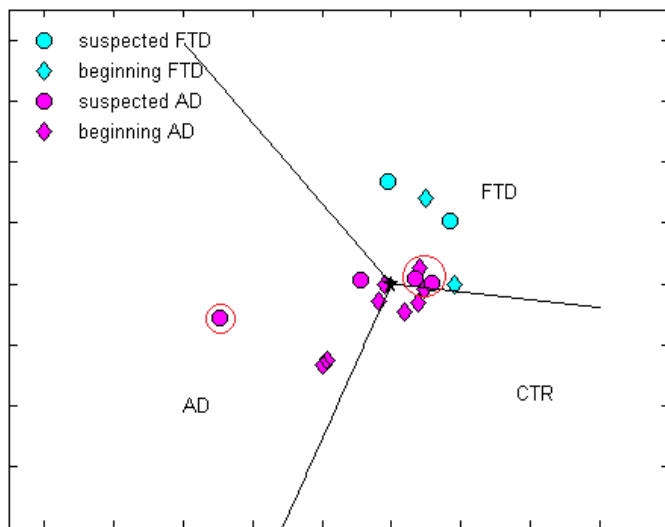


Fig. 6. FDA-boundaries on discrimination plane (linear machine for training set), classification of subjects with beginning/ suspected disease.

three dimensions (PCA) and subsequently applying MDA and linear machine.

The projection of the bAD, bFTD, sAD and sFTD cases (which were previously centered by subtraction of the training set mean, and reduced in dimension by the three PCs determined in paragraph III-A) on the discrimination plane results in Figure 5, where it can be seen that in general those subjects are projected closer to the asymptomatic controls. In Figure 6, boundaries for the class regions are included, for further and more precise classification of individual subjects.

The linear machine works very well for the classification of subjects with beginning or suspected FTD (Figure 6), and for the following cases of beginning and suspected AD (bAD, sAD):

- Four cases of bAD are classified as AD but are located very close to the boundary to CTR. Three further cases of bAD are classified as CTR, indicating an early stage of the disease.
- One case of sAD is classified as AD, but is very close to

the center (i.e., overall mean) of the dataset.

There are four cases that seem to be misclassified (indicated by a red frame in Figure 6):

- 1) One subject labeled bAD is classified as FTD.
- 2) Two sAD cases are assigned to the FTD region.
- 3) One more subject with sAD is closely located to the class mean of AD (see Figure 5) indicating an assured case of AD.

However, rechecking these four cases in terms of diagnosis reveals that

- 1) Is of uncertain diagnosis and could be both bAD and FTD, hence the classification as FTD.
- 2) One of the cases was rediagnosed as mixed AD and FTD and is therefore classified correctly. The other one was rediagnosed as healthy and was thus misclassified, but is located very close to the boundary to the CTR region.
- 3) Is a case of certain AD and was mislabeled.

This indicates that the model is well adapted to cope even with subtle disease-specific characteristics within the classes.

IV. CONCLUSION

The discrimination plane obtained for the AD, FTD and CTR cases is most significant if the original data is reduced to three PCs before performing MDA. MDA of SPECT data is well suited to discriminate between AD, FTD and CTR. Even though MDA and the linear machine seem to be most accurate for eight PCs, further examinations taking into account the possibility of misdiagnosis show that MDA is overtrained for eight PCs, and most resistant to deliberate misclassification in case of three PCs. This is in line with the result of the F-test. Projecting another dataset of cases with early-onset diagnoses onto the discrimination plane shows how well the MDA accounts for disease-specific characteristics within the classes. This makes it possible to detect mislabeled cases and (to some extent) provides a suggestion in cases of uncertain diagnosis.

REFERENCES

- [1] Jefferies, K., Agrawal, N.: Early-onset dementia. *Advances in Psychiatric Treatment* **15** (2009) 380–388
- [2] Habeck, C., Foster, N.L., Permecky, R., Kurz, A., Alexopoulos, P., Koeppe, R.A., Drzezga, A., Stern, Y.: Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. *NeuroImage* **40** (2008) 1503–1515
- [3] Frapp, J., Bourgeat, P., Acosta, O., Raniga, P., Modat, M., Pike, K.E., Jones, G., O'Keefe, G., Masters, C.L., Ames, D., Ellis, K.A., Maruff, P., Currie, J., Villemagne, V.L., Rowe, C.C., Salvado, O., Ourselin, S.: Appearance modeling of 11C PiB PET images: Characterizing amyloid deposition in Alzheimer's disease, mild cognitive impairment and healthy aging. *NeuroImage* **43** (2008a) 430 – 439
- [4] Markiewicz, P.J., Matthews, J.C., Declerck, J., Herholz, K.: Robustness of multivariate image analysis assessed by resampling techniques and applied to FDG-PET scans of patients with Alzheimer's disease. *NeuroImage* **46** (2009) 472 – 485
- [5] Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley-Interscience (2001)
- [6] Merhof, D., Markiewicz, P.J., Platsch, G., Declerck, J., Weih, M., Kornhuber, J., Kuwert, T., Matthews, J.C., Herholz, K.: Optimized data preprocessing for multivariate analysis applied to 99mTc-ECD SPECT data sets of Alzheimer's patients and asymptomatic controls. *J Cereb Blood Flow Metab* **31** (2011) 371–383
- [7] Efron, B., Tibshirani, R.J.: *An Introduction to the Bootstrap*. Chapman & Hall/CRC (1993)