Classification of Neurodegenerative Dementia by Gaussian Mixture Models applied to SPECT Images

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

***Abstract*—Gaussian mixture (GM) models can be applied for statistical classification of various types of dementia. As opposed to linear boundaries, they do not only provide the class membership of a case, but also a measure of its probability. This enables an improved interpretation and classification of neurodegenerative dementia datasets which comprise various stages of the disease, and also mixed forms of dementia.**

In this work, GM models are applied to a total number of 103 technetium-99methylcysteinatedimer (^{99m}Tc-ECD) SPECT datasets of asymptomatic controls (CTR), as well as Alzheimer's disease (AD) and frontotemporal dementia (FTD) patients in early or moderate stages of the disease. Prior to classification, multivariate analysis is applied: Principal component analysis (PCA) is used for dimensionality reduction, followed by a differentiation of the datasets via multiple discriminant analysis (MDA). A GM model on the resulting discrimination plane is constructed by computing the GM distribution associated with the underlying training set. The posterior probabilities of each case indicate its class membership probability. The performance of GM models for classification is assessed by bootstrap resampling and cross validation. Accuracy and robustness of the method are evaluated for different numbers of principal components (PCs), and furthermore the detection rate of dementia in early stages is calculated.

The GM model outperfomes classification with linear boundaries in both predicted accuracy and detection rate of early dementia, and is equally robust.

***Index Terms*—SPECT, Alzheimer's Disease, Frontotemporal Dementia, Multivariate Analysis, Gaussian Mixture Model, Probabilistic Classification.**

I. Introduction

Neurodegenerative dementia is one of the most expensive diseases in developed countries, and its prevalence is expected to double within the next 20 years [1]. Early detection and disease prediction with high accuracy is therefore needed, as early treatment can delay disease onset and can attenuate dementia's economic impact on society [2]. Currently, first pharmaceuticals are available that alleviate symptoms and there are even more under clinical trials.

Statistical analysis of medical image data has the potential to automatically extract significant features and patterns that characterize different types of dementia in a standardized way. Alzheimer's disease (AD) and frontotemporal dementia (FTD) are amongst the most prevalent neurodegenerative diseases [3], and have been successfully differentiated from healthy subjects (CTR) by multivariate analysis of single-photon emission computed tomography (SPECT) images [4]. To date, the subsequent classification by e.g. linear boundaries obtained from Fisher's discriminant analysis (FDA) as proposed by [5] provides a class membership for each case but contains no indication about the probability of its class membership or about the stage of disease.

The following sections present a probabilistic approach for classification using Gaussian mixture (GM) models applied to SPECT data of AD, FTD and CTR cases subsequent to performing multivariate analysis, i.e. dimensionality reduction via principal component analysis (PCA). Additionally, the method is validated by verifying the GM classification on image data of patients affected by early and mixed dementia, and by assessment of accuracy, detection rates and robustness using bootstrap resampling.

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.

II. Material and Methods

A. Image Data

The ^{99m}Tc-ECD SPECT datasets were aquired at the Clinic of Nuclear Medicine, University of Erlangen-Nuremberg. All datasets were preprocessed using affine registration, Gaussian smoothing with an FWHM of 12mm, and intensity normalization based on 25% brightest voxels within the whole-brain region according to the optimized protocol presented in [6]. Overall, 103 subjects (mean age 65.06±8.28, 59 females, 44 males) are included in the analysis:

An assured diagnosis is available for 26 cases with Alzheimer's disease (AD), 21 cases with frontotemporal dementia (FTD) and 26 asymptomatic controls (CTR).

Furthermore, datasets of patients with early stages or mixed forms of AD or FTD are available to further assess the proposed probabilistic classification. These include 9/2 cases with beginning AD/FTD, 8/7 cases where AD/FTD is suspected and 4 cases with mixed AD and FTD.

The training set for the multivariate analysis and subsequent probabilistic classification contains 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.

B. Multivariate Analysis

1) Dimensionality reduction: The data matrix X containing all training data is mean-centered and principal component analysis (PCA) is performed followed by dimensionality reduction and to exclude redundant information as well as noise. PCA performs a singular value decomposition of X, where the

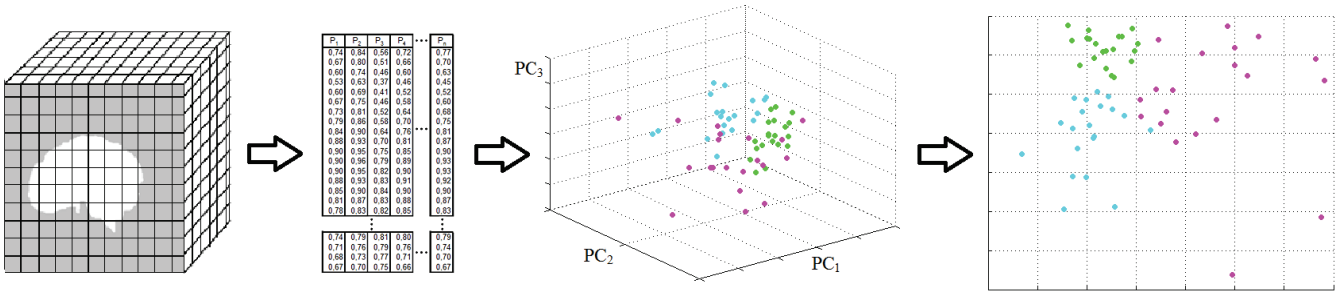

Fig. 1. Example of multivariate analysis applied to a stratified training set (without boostrap): SPECT images of the whole-brain regions are transformed into vectors and stored columnwise in the data matrix X, then PCA and subsequently MDA are applied.

principal components (PCs) are the left singular vectors of X. The very high dimensional datasets (more than 10^5 voxels of the whole brain region, i.e. rows of matrix X) can then be represented in the PC space by a linear combination of PCs. The dimensionality of a dataset can be reduced by maintaining only the first n PCs. An example for $n = 3$ PCs is depicted in Figure 1, where the originally high dimensional variable space *(second picture, from left)* is projected by the PCs into a three-dimensional subspace *(third picture, from left)*.

In general, the first few PCs are considered to be the most dominant directions for separating SPECT datasets belonging to different disease groups, whereas the inclusion of more PCs might over-train the classification [5]. Nevertheless, all subsequent analysis is performed for any number of the first 25 PCs at a time to find the best possible trade-off between accuracy and robustness.

2) Discrimination plane: Multiple discriminant analysis (MDA) is a generalization of the Fisher's discriminant analysis (FDA), e.g. as described in [7]. In this work, it is applied to determine a hyperplane within the previously defined PC space, where all data is optimally separated [4], [7]. Only after projecting and separating the classes in this way, a classifier can be applied.

This hyperplane is of dimension $c-1$, where c is the number of classes within the training data. In this work it is two-dimensional, as only three classes are used for training the classification. An example for a plane determined by MDA is depicted in Figure 1 *(right)*, where an optimal separation of three classes can be seen. The basic idea is to find an optimal projection of the data onto this discrimination plane, i.e. such that the ratio of the general between-class scatter and the general within-class scatter of all data points is maximized. The maximization of this ratio, the so-called generalized Rayleigh-quotient, is achieved by solving a generalized eigenvalue problem. The resulting eigenvectors constitute the optimized projection vectors.

C. The Gaussian Mixture (GM) Model

A GM model is applied for statistical classification on the discrimination plane [8], [9], which was previously determined by multivariate analysis (Section II-B). In this work, its mixture contains three density functions associated to each class, i.e. the underlying probability density function is calculated by a convex combination of these mixture components. An example is shown in Figure 2 *(left)*, where the three different mixture components associated to AD, FTD and CTR are still distinguishable within the resulting probability density function.

The mixture weights are chosen to be equal, as in this work the probability of all mixture components is assumed to be equal. After projecting the data onto the discrimination plane, the means and covariances of each class are calculated to specify the associated Gaussian distributions. The GM distribution and its probability density function is then constructed using the MATLAB Statistics Toolbox. Posterior probabilities are evaluated for each case within the training set and for each class. They indicate the probability that a particular case belongs to a specific class, i.e. CTR, AD or FTD.

Furthermore, all datasets of patients with early stages or mixed forms of AD or FTD are projected onto the discrimination plane by the vectors determined during multivariate analysis of the training set (as described in Section II-B), and their posterior probabilities are evaluated analogously.

D. Resampling

1) Stratification of training set: As the training classes AD, CTR and FTD are of different size, a fixed number k of cases is drawn randomly from each class *without replacement*. Afterwards, mean-centering and PCA are applied to the voxels of the whole-brain region.

2) Bootstrap resampling: After projecting the stratified training set onto the discrimination plane, k cases are drawn randomly from each class *with replacement* and a GM distibution is constructed as outlined in Section II-C. By this procedure, 63.2% of all cases are selected on average for training [10]. The remaining cases constitute the cross validation set.

E. Accuracy

Each case is projected onto the discrimination plane and is assigned to the class with the highest associated posterior probability. The prediction accuracy $Acc^{.632}$ of classification via GM models is calculated by the .632 bootstrap estimator introduced by [10] using bootstrap resampling of the previously stratified training set:

$$Acc^{.632} = 1 - (.632 \cdot err_{app} + .368 \cdot err_{exp}), \qquad (1)$$

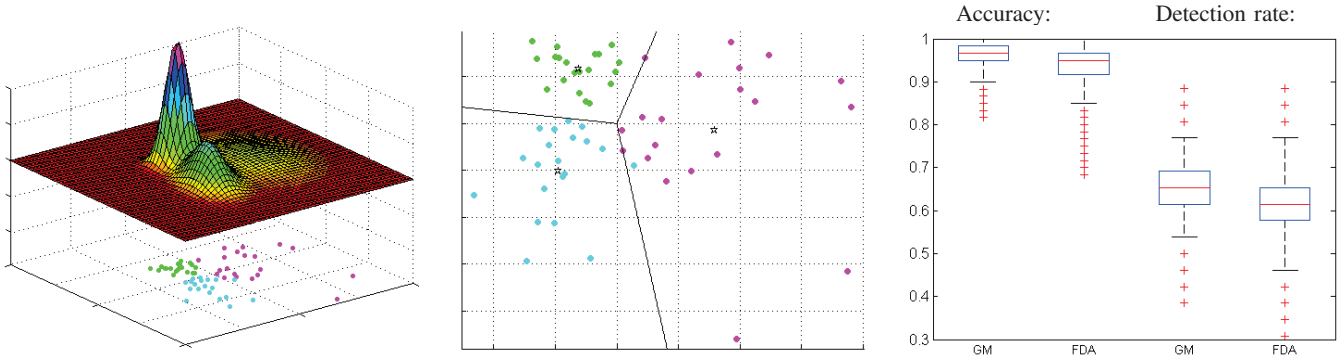

Fig. 2. Comparison of probabilistic classification via GM and classification with linear boundaries achieved by FDA applied to a stratified training set after multivariate analysis based on three PCs.
Left: Probability density function of the GM distribution; the associated discriminant plane and projection of the stratified training set (without bootstrap) is indicated below the surface plot, where green markers denote CTR cases, magenta markers AD cases and cyan markers FTD cases. *Middle:* The same discriminant plane with linear boundaries. *Right:* Boxplots of apparent accuracy and detection rates in all bootstrap iterations for both classification methods.

where err_{app} denotes the apparent and err_{exp} the expected error rate, referring to the percentage of misclassified cases within the bootstrapped training set and the cross validation set, respectively [4]. The constants of this formula refer to the resulting average percentage of training and cross validation cases, as described above in Section II-D.

F. Detection Rate of Beginning and Suspected Dementia

As described in Section II-C, 27 cases of beginning or suspected AD or FTD are additionally projected onto the discrimination plane in each iteration of the bootstrap resampling, and their posteriors are evaluated for the constructed GM-distribution. The detection rate expresses the correct assignment of testcases affected by an early stage of dementia to AD and FTD: If a case belongs with higher probability to a dementia class (AD or FTD, respectively), the dementia is considered to be detected.

G. Robustness

In addition to estimating accuracy, which also takes into account the performance of an independet cross validation set within each resampling iteration, robustness analysis is performed to detect overtraining of the classification, e.g. by inclusion of too many PCs during multivariate analysis:

1) Deliberate mislabeling: In each bootstrap resampling iteration, the labels of two cases randomly drawn from different classes are swapped, and the percentage of correct classification of those cases in spite of mislabeling is evaluated [4], [5].

2) Deviation of GM distribution parameters: As the class-means of the mixture model are dependent on scale and rotation of the discrimination plane, direct comparison of the GM distribution centers accross resampling iterations and number of PCs is not possible. However, this can be achieved by standardizing: The Euclidian distance of the bootstrapped class means to the original class means $d_{i,btp} = m_{i,btp} - m_i$ is substituted by the ratio $d_{i,btp}/\bar{d}_i$, where $\bar{d}_i$ denotes the mean distance of all cases of class i from the class mean m_i. A high mean value of the ratio indicates that the method is not sufficiently robust against mislabeled data.

H. Comparison to Linear Machine

The performance of the GM model for statistical classification is compared to classification with linear boundaries obtained by application of Fishers's discriminant analysis (FDA) to two classes at a time [7].
Accuracy rates, detection rates and robustness via mislabel of the FDA are evaluated analogously to Sections II-E to II-G within the same bootstrap iterations and for any number of the first 25 PCs.
Robustness of linear boundaries is assessed by measuring the angle between the bootstrapped and original FDA-vectors [4], [5]. Increasing angles between those vectors indicate an increasing instability of the classification.

III. Results and Discussion

A. Accuracy and Detection Rates

The accuracy and detection rates as well as the robustness assessment were calculated in 20.000 iterations. The training set was randomly stratified in 200 resampling iterations and bootstrap resampling was subsequently applied 100 times, as described in Section II-D.

1) Accuracy: The predicted accuracy generated by decision for the highest class membership probability in a GM model rises considerably (from 85.13% to 93.39%) when a third PC is included before discriminant analysis, and increases slowly for increasing number of PCs (up to 96.51%). Figure 3 *(left)* shows the predicted accuracy rates for the GM model (denoted by red lines) and FDA (black) classification by the bootstrap estimator, and the associated apparent and expected accuracy rates. The classification via GM model outperformes the FDA classification for any number of PCs. Another advantage over linear boundaries (not yet taken into account in this analysis) is the enhanced interpretability of the results provided by the class membership probabilities, as discussed in Section III-D.

2) Detection rates: The detection rate of suspected and beginning dementia rises clearly when a third PCs is added for further analysis (see Figure 3, *right*), similarily to the previously calculated accuracy rates. It reaches 66.28% for three PCs and its maximum at 72.89% for the inclusion of

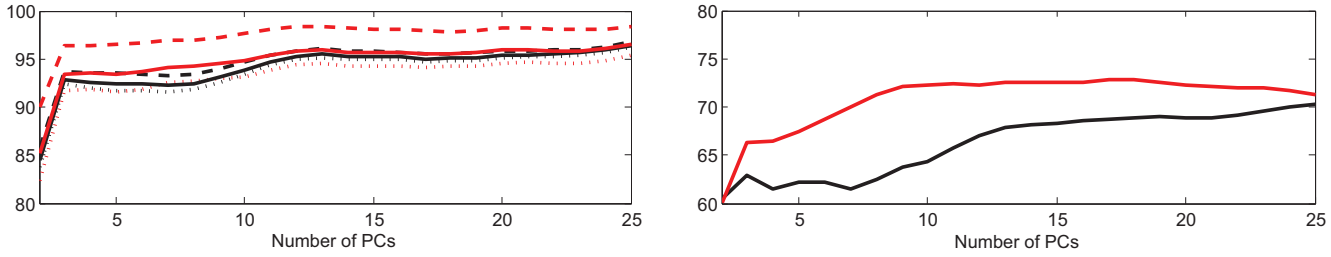

Fig. 3. Accuracy *(left)* and detection rates *(right)* in percent for any number of PCs, where red coloring denotes rates achieved by GM classification, and black coloring denotes rates achieved by evaluating linear boundaries. *Left:* Accuracy rates are expressed by predicted accuracy (solid), expected accuracy (dotted) and apparent accuracy (dashed).

18 PCs. The improvement on the detection of early dementia compared to linear boundaries is more pronounced than in the analysis of accuracy, and thereby underlines the benefits regarding the increased sensitiveness of a GM model to small deviations of the data from normality.

B. Robustness

The GM distribution accross bootstrap iterations is in general very robust. The percentage of correct classification of two cases within the boostrapped training set despite a deliberate mislabel is over 91% for any number of PCs greater than two, as depicted in Figure 4 by the red line. It can also be seen that GM classification clearly outperformes linear boundaries for any number of PCs, with respect to this type of robustness analysis.

With respect to the ratio of classmeans-distance, resampling analysis produces very similar robustness of the classifiers: the generalized distance of cases to their classmean is nearly constant at 0.23 for any number of PCs, and the mean angles between the linear boundaries computed for comparison are very small and do not exceed 11° for any number of PCs.

However, it should be noted that the robustness of the preceding multivariate analysis has a major impact on the final classification result. Results of prior studies suggest, that PCA and subsequent MDA applied to SPECT data is reasonable robust only for the first few PCs. As a distinct increase of accuracy and detection rates can be observed especially between the inclusion of the second and third PC, and significant drops of robustness can be observed by inclusion of the first five PCs [5], a good trade-off between robustness and accuracy is found for three PCs. This is also consistent with the results of [5], where robustness of PCA applied to functional datasets was assessed systematically.

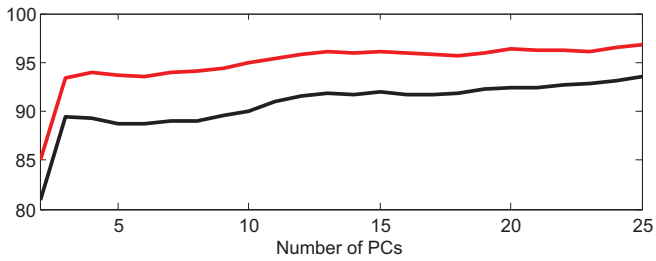

Fig. 4. Percentage of correctly classified cases despite mislabel for any number of PCs. Red coloring denotes rates achieved by GM classification, and black coloring denotes rates achieved by evaluating linear boundaries.

C. Trade-off

This section summarizes all results for the trade-off determined above, i.e. all results based on three PCs:

The predicted accuracy of probabilistic classification with GM models reaches an average of 93.39%, an average detection rate of 68.54% is achieved, and the method is fairly robust with 93.41% correctly classified cases despite mislabel and a ratio of classmean-distance of 0.214 (CTR-class), 0.225 (AD) and 0.229 (FTD).

Figure 2 *(right)* depicts a comparison via boxplots of both classification by GM models and by linear boundaries (via FDA) for accuracy and detection rate performance in all bootstrap iterations, where the superiority of probabilistic classification by GM models becomes apparent.

Furthermore, the difference between those two classification methods was statistically significant for three PCs with respect to both accuracy and detection rates ($p = 0.05$).

D. Probabilistic information of the GM model

So far, the validation of the GM classification after multivariate analysis is based on the maximum class membership probability but does not take into account further probabilistic information conveyed by the GM model. Whereas the probability density function (as depicted in Figure 2, *left*) describes group characteristics (e.g. the more complex disease pattern of AD results in a wider and less decisive mixture component), posterior probabilities are used to evaluate the classfication of each case within the training set. Below, an example for a GM model, which was constructed based on a stratified but not bootstrapped training set, is discussed:

1) Posteriors of training set: On the left side of Figure 5, the posterior probabilities of all training set cases (x-axis) are depicted, where class membership (y-axis) and the confidence of classification (color encoded) can be deduced. All CTR cases were classified correctly, with a mean probability of 91.11%. Two cases were classified with a noticeable low confidence of less than 70%, and an increased probability of FTD

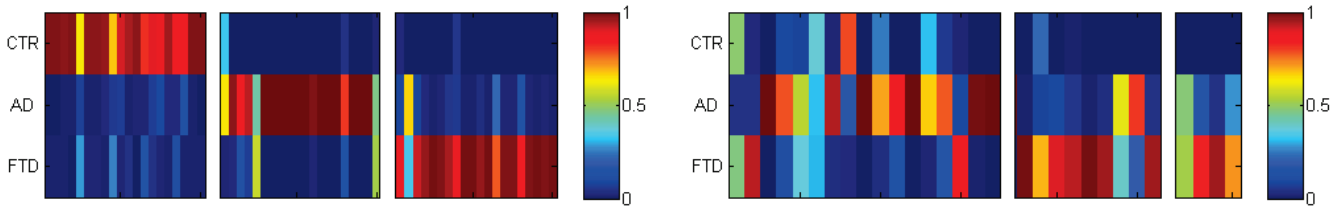

Fig. 5. Posterior probabilities of a stratified training set (*left*) and testcases affected by early or mixed dementia (*right*); The training set contains (from left to right) CTR, AD cases and FTD cases, testcases include suspected or beginning AD, FTD and four cases of mixed AD/FTD (from left to right).

membership (more than 25%). This suggests in both cases a slight or beginning disease, which was confirmed in one of the cases. AD was classified overall with a mean probability of 90.58%. Two of the AD cases were misclassified to FTD, but both with less than 60% confidence. In a rereading of these datasets it transpired that in both cases the left temporal lobe was affected by the dementia. The first AD case has a CTR membership probability of 33.75% indicating a mild form of dementia. The mean of FTD membership probability was 91.86%, and only one case was misclassified with 65.84% probability to AD. Rereading of this case confirmed a mixed form of AD and FTD.

2) Posteriors of early and mixed dementia: On the right, Figure 5 shows the probabilities of each set of testcases (early and mixed dementia) belonging to CTR, AD or FTD. In this example, both early AD and FTD were detected with high confidence, with exception of five cases. Three AD cases were classified to FTD, and two FTD cases to AD. In four of these cases, clinical rereading resulted in a confirmation of a mixed form of AD and FTD, or a hindered classification due to atrophy.

Mixed cases of AD and FTD were all classified correctly but only for the first and last case within the set the class membership probabilities showed the ambivalence of the disease, the two other cases were classified with high probability and according to more prevalent FTD-specific characteristics.

IV. Conclusion

The predicted accuracy of probabilistic classification based on GM models increases significantly if the variable space of the original datasets is reduced to three dimensions during multivariate analysis, i.e. by projecting the data into a subspace using only the first three PCs before performing discriminant analysis. This is concordant with prior results regarding the differentiation via MDA.

The performance of the probabilistic classifier was compared to linear boundaries achieved by FDA with respect to robustness, accuracy and detection rates: The robustness of a GM distribution is sufficiently high for any number of 25 PCs, and outperforms FDA with respect to mislabeled cases within the training set. Furthermore, the probabilistic classification via GM models does not only exceed FDA significantly in accuracy and detection of dementia in early stages, but conveys also valuable information regarding mixed forms of neurodegenerative dementia and stages of disease.

All results were confirmed using bootstrap resampling and cross validation applied to datasets with assured diagnosis of dementia and asymptomatic controls, and furthermore by testing the GM models evaluated within each iteration by application to a set of test cases composed by image data of patients affected by suspected, beginning or mixed forms of dementia.

V. Acknowledgment

This work was supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK Baden-Württemberg; grant number 655.042/Merhof/1) (E. Stühler, D. Merhof).

References

[1] Abbott, A.: Dementia: A problem for our age. Nature **475** (2011) S2–S4
[2] Petrella, J.R., Coleman, R.E., Doraiswamy, P.M.: Neuroimaging and early diagnosis of Alzheimer disease: A look to the future. Radiology **226** (2003) 315–336
[3] Jefferies, K., Agrawal, N.: Early-onset dementia. Advances in Psychiatric Treatment **15** (2009) 380–388
[4] Stühler, E., Platsch, G., Weih, M., Kornhuber, J., Kuwert, T., Merhof, D.: Multiple discriminant analysis of SPECT data for Alzheimer's disease, frontotemporal dementia and asymptomatic controls. In: IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). (2011) 4398–4401
[5] 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
[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] Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification. Wiley-Interscience (2001)
[8] Bishop, C.M.: Pattern recognition and machine learning (Information Science and Statistics). 1st ed. 2006. corr. 2nd printing edn. Springer (2007)
[9] McLachlan, G., Peel, D.: Finite mixture models. John Wiley & Sons, Inc. (2005)
[10] Efron, B., Tibshirani, R.J.: An introduction to the bootstrap. Chapman & Hall/CRC (1993)