Neuroscience Letters 472 (2010) 99–103

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification

J. Ramírez*, J.M. Górriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. López, I. Álvarez, P. Padilla

Dept. of Signal Theory, Networking and Communications, University of Granada, Periodista Daniel Saucedo Aranda S/N, 18071 Granada, Spain

ARTICLE INFO

Article history:
Received 17 December 2009
Received in revised form 22 January 2010
Accepted 25 January 2010

Keywords:
Alzheimer disease
SPECT
Random forest
Partial least squares

ABSTRACT

This letter shows a computer aided diagnosis (CAD) technique for the early detection of the Alzheimer's disease (AD) by means of single photon emission computed tomography (SPECT) image classification. The proposed method is based on partial least squares (PLS) regression model and a random forest (RF) predictor. The challenge of the curse of dimensionality is addressed by reducing the large dimensionality of the input data by downscaling the SPECT images and extracting score features using PLS. A RF predictor then forms an ensemble of classification and regression tree (CART)-like classifiers being its output determined by a majority vote of the trees in the forest. A baseline principal component analysis (PCA) system is also developed for reference. The experimental results show that the combined PLS-RF system yields a generalization error that converges to a limit when increasing the number of trees in the forest. Thus, the generalization error is reduced when using PLS and depends on the strength of the individual trees in the forest and the correlation between them. Moreover, PLS feature extraction is found to be more effective for extracting discriminative information from the data than PCA yielding peak sensitivity, specificity and accuracy values of 100%, 92.7%, and 96.9%, respectively. Moreover, the proposed CAD system outperformed several other recently developed AD CAD systems.

Alzheimer's disease (AD) [12] is a progressive neurodegenerative disorder first affecting memory functions and then gradually affecting all cognitive functions with behavioral impairments and eventually causing death. As the number of AD patients has increased, its early diagnosis has received more attention for both social and medical reasons.

AD diagnosis is based on the information provided by a careful clinical examination, a thorough interview of the patient and relatives, and a neuropsychological assessment. An analysis of the regional cerebral blood flow (rCBF) by means of single photon emission computerized tomography (SPECT) is frequently used as a complimentary diagnostic tool in addition to the clinical findings [6,7,17]. However, in late-onset AD there are minimal perfusion deficits in the mild stages of the disease, and age-related changes, which are frequently seen in healthy aged people, have to be discriminated from the minimal disease-specific changes. These minimal changes in the images make visual diagnosis a difficult task that requires experienced explorers. Even with this problem still unsolved, the potential of computer aided diagnosis (CAD) has not been explored in this area.

Several approaches for designing CAD systems of the AD can be found in the literature [11,8,18]. The first family is based on the analysis of regions of interest (ROIs) by means of some discriminant functions. The second approach is the statistical parametric mapping (SPM) [4] software tool and its numerous variants. SPM is widely used in neuroscience. It was not developed specifically to study a single image, but for comparing groups of images. SPM has been designed as a univariate approach since the classical multivariate techniques such as MANCOVA [19] require the number of observations (i.e. scans) to be greater than the number of components (i.e. voxels) of the multivariate observation. The importance of multivariate approaches is that the effects due to activations, confounding effects and error effects are assessed statistically, both in terms of effects at each voxel, and interactions among voxels [4]. On the other hand, statistical learning classification methods have not been explored in depth for AD CAD, quite possibly due to the fact that images represent large amounts of data and most imaging studies have relatively few subjects (generally <100) [3,16,5]. Even with this problem still unsolved, the potentials of novel machine learning techniques have not been explored in depth for CAD [1,9,10].

Partial least squares (PLS) modeling [20,21] is an effective method for feature extraction that has shown improved results over other conventional feature extraction methods such as principal component analysis (PCA) in classification problems. On the other hand, random forests (RF) [2,13] based on an ensemble of decision trees, bagging and majority voting are more robust than a single decision tree classifier and yield improved results when

* Corresponding author.
E-mail address: javierrp@ugr.es (J. Ramírez).

0304-3940/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2010.01.056

compared to traditional methods in a number of applications. This letter shows a complete CAD system for the early detection of the AD by means of supervised SPECT image classification. The proposed method combining advanced PLS feature extraction schemes and RF prediction is developed with the aim of reducing the subjectivity in visual interpretation of SPECT scans by clinicians, thus improving the detection of the AD in its early stage.

Baseline SPECT data from 97 participants were collected from the "Virgen de las Nieves" hospital in Granada (Spain). The patients were injected with a gamma emitting ^{99m}Tc-ECD radiopharmaceutical and the SPECT raw data was acquired by a three-head gamma camera Picker Prism 3000. A total of 180 projections were taken with a 2° angular resolution. The images of the brain cross-sections were reconstructed from the projection data using the filtered backprojection (FBP) algorithm in combination with a Butterworth noise removal filter [14].

The SPECT images are then spatially normalized using the SPM software [4], in order to ensure that the voxels in different images refer to the same anatomical positions in the brain. This step allows us to compare the voxel intensities of the brain images of different subjects [15]. Then we normalize the intensities of the SPECT images with the maximum intensity I_{MAX}, which is computed for each image individually by averaging over the 3% of the highest voxel intensities. After the spatial normalization with the SPM software, one obtains a $95 \times 69 \times 79$ voxel representation of each subject, where each voxel represents a brain volume of 2.18 mm × 2.18 mm × 3.56 mm. The SPECT images were labeled by experts of the "Virgen de las Nieves" hospital using four different labels: (NOR) for patients without any symptoms of AD, and (AD1), (AD2) and (AD3) to distinguish between different levels of the presence of typical characteristics for AD. In total, the database consists of 41 NOR, 30 AD1, 22 AD2 and 4 AD3 patients.

PLS [20,21] is a wide class of methods for modeling relations between sets of observed variables by means of latent variables. It comprises of regression and classification tasks as well as dimension reduction techniques and modeling tools. The underlying assumption of all PLS methods is that the observed data is generated by a system or process which is driven by a small number of latent (not directly observed or measured) variables. In its general form, PLS creates orthogonal score vectors (also called latent vectors or components) by maximizing the covariance between different sets of variables. PLS can be naturally extended to regression problems. The predictor and predicted (response) variables are each considered as a block of variables. PLS then extracts the score vectors, which serve as a new predictor representation, and regresses the response variables on these new predictors. PLS can be also applied as a discrimination tool and dimension reduction method similar to principal component analysis (PCA). After relevant latent vectors are extracted, an appropriate classifier can be applied.

PLS is a linear algorithm for modeling the relation between two datasets $X \subset \mathbb{R}^N$ and $Y \subset \mathbb{R}^M$. After observing n data samples from each block of variables, PLS decomposes the $n \times N$ matrix of zero-mean variables $\mathbf{X}$ and the $n \times M$ matrix of zero-mean variables $\mathbf{Y}$ into the form:

$$\begin{aligned} \mathbf{X} &= \mathbf{T}\mathbf{P}^{\boldsymbol{T}} + \mathbf{E} \\ \mathbf{Y} &= \mathbf{U}\mathbf{Q}^{\boldsymbol{T}} + \mathbf{F} \end{aligned} \tag{1}$$

where $\mathbf{T}$ and $\mathbf{U}$ are $n \times p$ matrices consisting of the p extracted score vectors (components, latent vectors), the $N \times p$ matrix $\mathbf{P}$ and the $M \times p$ matrix $\mathbf{Q}$ represent matrices of loadings while the $n \times N$ matrix $\mathbf{E}$ and the $n \times M$ matrix $\mathbf{F}$ are the matrices of residuals. PLS is implemented by means of nonlinear iterative partial least squares (NIPALS) algorithm [21].

The difference between PLS and PCA is that the former creates orthogonal weight vectors by maximizing the covariance between elements in $\mathbf{X}$ and $\mathbf{Y}$. Thus, PLS not only considers the variance of the samples but also considers the class labels. Fisher Discriminant Analysis (FDA) is, in this way, similar to PLS. However, FDA has the limitation that after dimensionality reduction, there are only $c-1$ meaningful latent variables, where c is the number of classes being considered. Additionally, when the number of features exceeds the number of samples, the covariance estimates when applying PCA do not have full rank and the weight vectors cannot be extracted.

Fig. 1 shows a block diagram of the feature extraction process. Once the SPECT images have been acquired, reconstructed and normalized, the $95 \times 69 \times 79$ voxel representation of each subject is downsampled by a factor of 2. The resulting $35 \times 48 \times 40$ volumes of the 41 normal controls are averaged. A binary mask is then defined by considering only those voxels with a mean intensity above 50% of the maximum intensity. Thus, the dimensionality of the classification task is reduced from 517,845 voxels to 20,638 voxels. Finally, score PLS features are extracted based on the regression model in Eq. (1) for a further dimensionality reduction and classification of a given subject by means of a random forest predictor. Fig. 2 illustrates the feature extraction process and the matrices involved in PLS regression of $\mathbf{X}$ and $\mathbf{Y}$. It is shown how the PLS model accurately matches the given data by means of the extracted score vectors ($\mathbf{T}$ and $\mathbf{U}$) and loadings ($\mathbf{P}$ and $\mathbf{Q}$).

Various ensemble classification methods have been proposed in recent years for improved classification accuracy [2]. In ensemble classification, several classifiers are trained and their results are combined through a voting process. Perhaps, the most widely used such methods are boosting and bagging. Boosting is based on sample re-weighting but bagging uses bootstrapping. The random forest classifier [2] uses bagging, or bootstrap aggregating, to form an ensemble of classification and regression tree (CART)-like classifiers $h(\mathbf{x}, T_k)$, $k = 1, \ldots$, where the T_k is bootstrap replica obtained by randomly selecting N observations out of N with replacement, where N is the dataset size, and $\mathbf{x}$ is an input pattern [11]. For classification, each tree in the random forest casts a unit vote for the most popular class at input $\mathbf{x}$. The output of the classifier is determined by a majority vote of the trees. This method is not sensitive to noise or overtraining, as the resampling is not based on weighting. Furthermore, it is computationally more efficient than methods based on boosting and somewhat better than simple bagging.

Several experiments were conducted to evaluate the PLS feature extraction process and the random forest classifier. First of all, Fig. 3 shows the out-of-bag error rate for PCA and PLS which is used to analyze the convergence of the random forest and the benefits of PLS over PCA. In this analysis, the random forest is trained with the first PCA or PLS scores. Note that, the generalization error converges to a limit as the number of trees in the forest becomes large.

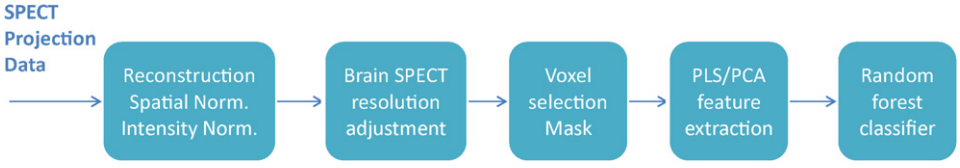

Fig. 1. Brain SPECT feature extraction by means of PLS/PCA and random forest classification.

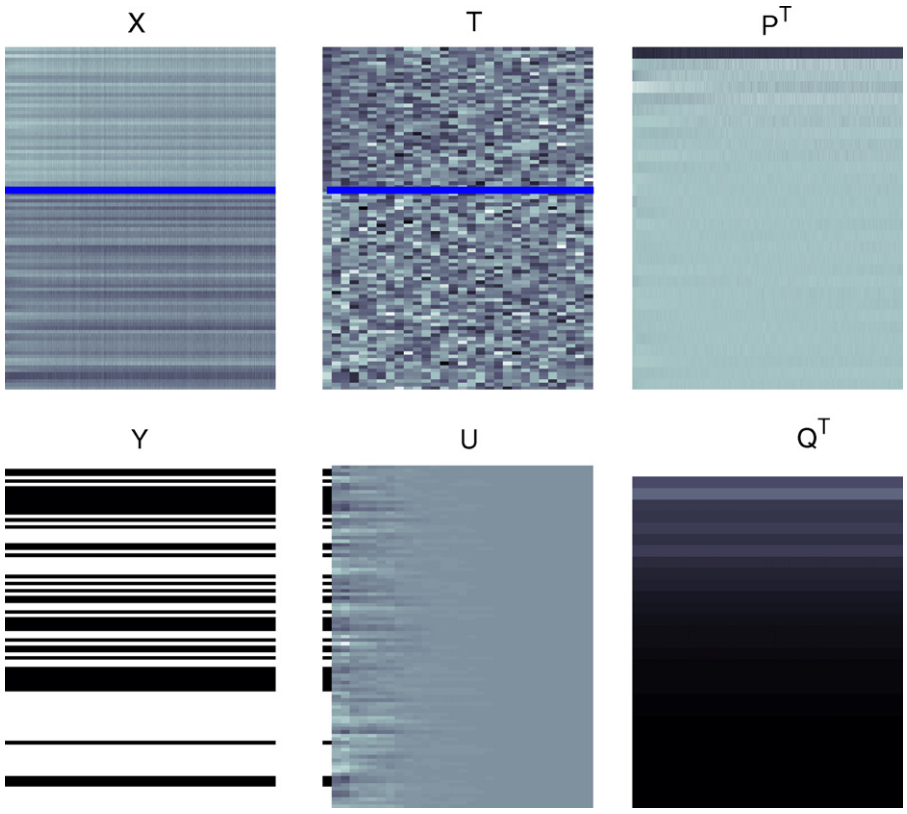

Fig. 2. Feature extraction process and matrices involved in the PLS regression model.

Moreover, the generalization error depends on the strength of the individual trees in the forest and the correlation between them. It can be concluded that:

(i) the random forest classifier converges for about 20–30 trees grown, and
(ii) PLS yields a significant improvement in the out-of-bag error rate when compared to PCA.

The importance of a feature variable for classification can be also estimated by randomly permuting all the values of the variable in the out of bag samples for each classifier (thereby missing the information provided by that feature). An increased out-of-bag error is an indication of the importance of that feature. Thus, it is not needed to supply test data for bagged ensembles because reliable estimates of the predictive power and feature importance are obtained in the process of training, which is an attractive feature of bagging.

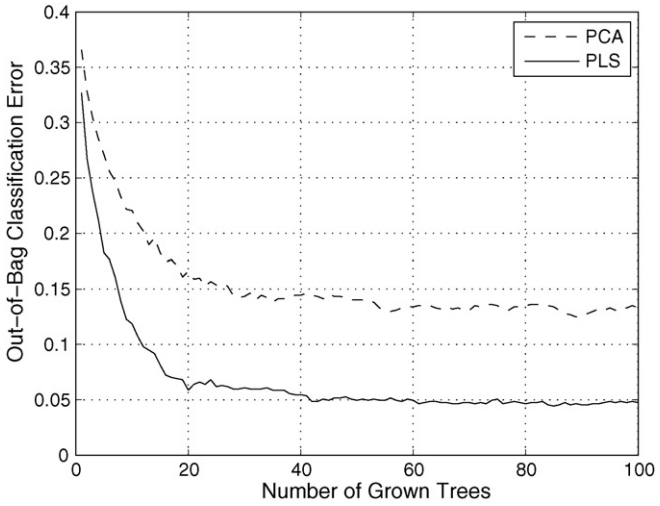

Fig. 3. Out-of-bag rate as a function of the number of grown trees.

Decision surfaces are useful for illustrating the operation of the classifier on reduced dimension feature spaces. They were built for comparing PCA and PLS as feature dimensionality reduction techniques when just the three first extracted components were used. Fig. 4 shows the 3D input PCA and PLS feature spaces and the ability of the random forest classifier to separate the two classes (normal controls in circles vs. patients affected by AD in squares).

The performance of the random forest CAD system was further evaluated as a tool for the early detection of the AD. The experiments considered an increasing number of features for designing the classifier. The success rate, sensitivity and specificity were estimated by means of leave-one-out cross-validation. Fig. 5 compares the classification performance of the PLS and PCA random forest CAD systems. Note that, the performance of the CAD system improves with the number of PCA or PLS components used as input features for classification up to a maximum stable value. PLS outperforms PCA as a feature extraction technique yielding peak values of sensitivity = 100%, specificity = 92.7% and accuracy = 96.9% when compared to PCA that just yields sensitivity = 94.6%, specificity = 80.5% and accuracy = 88.7%.

Finally, Fig. 6 shows the accuracy of the proposed PLS-RF method when compared to recently proposed AD CAD systems. It is clearly shown that the proposed PLS-RF system not only outperforms PCA-RF but also CAD systems combining PCA and support vector machines (SVM) [1], PCA and Bayesian classification rules [10], Gaussian mixture models (GMM) and SVM [5] as well as the voxel-as-features (VAF) approach [18] that yields just a poor 83% classification accuracy by considering the voxel intensities of the SPECT images as input data for a SVM classifier.

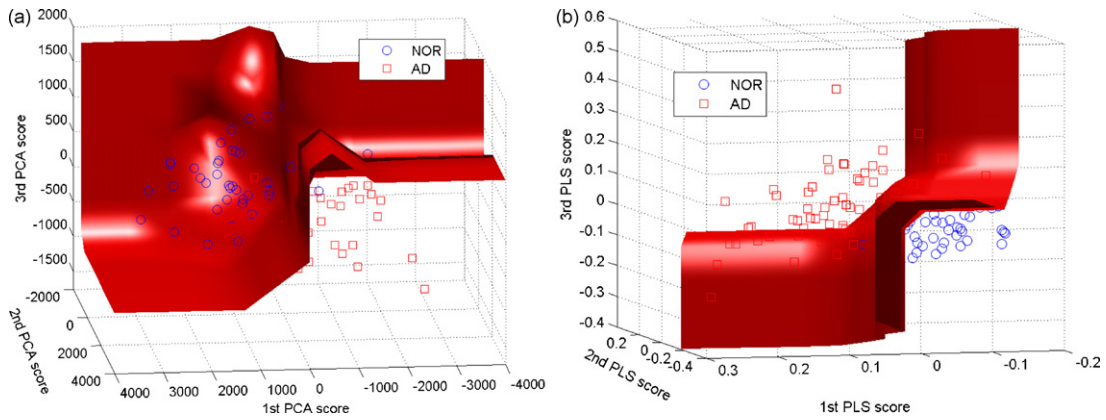

Fig. 4. Decision surfaces for random forest classifiers defined on: (a) PCA and (b) PLS scores.

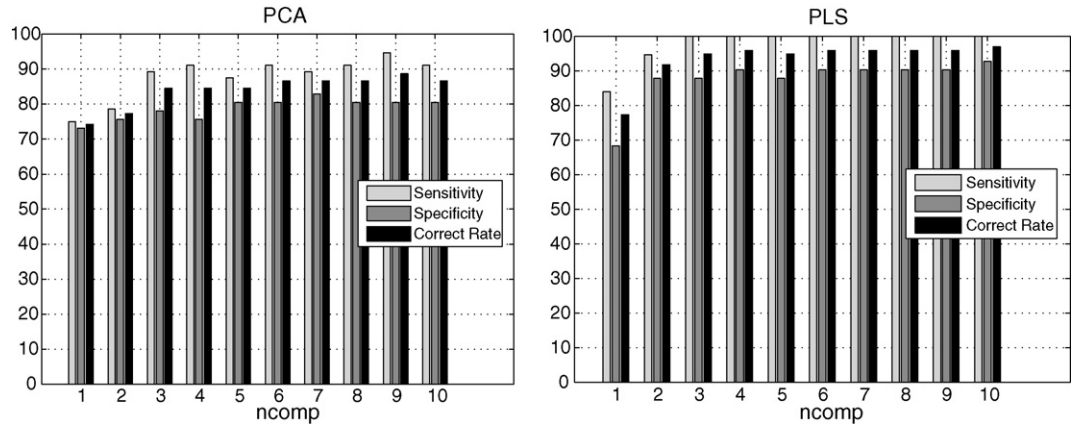

Fig. 5. Classifier performance (accuracy, specificity and sensitivity) for random forest classifiers defined on: (a) PCA and (b) PLS scores, as a function of the number of components.

This paper showed a computer aided diagnosis system for the early detection of the Alzheimer's disease. The proposed system is based on a partial least square regression model for feature extraction and a random forest predictor. It was shown that the generalization error of the random forest classifier converges to a limit as the number of trees in the forest increases. Moreover, the generalization error depends on the strength of the individual trees in the forest and the correlation between them. PLS outperformed PCA as a feature extraction technique yielding peak values of sensitivity = 100%, specificity = 92.7% and accuracy = 96.9% when compared to PCA that just yielded sensitivity = 94.6%, specificity = 80.5% and accuracy = 88.7%. Finally, the proposed PLS-RF outperformed several recently developed AD CAD systems combining PCA and SVM, PCA and Bayesian classification rules, GMM and SVM as well as the VAF approach.

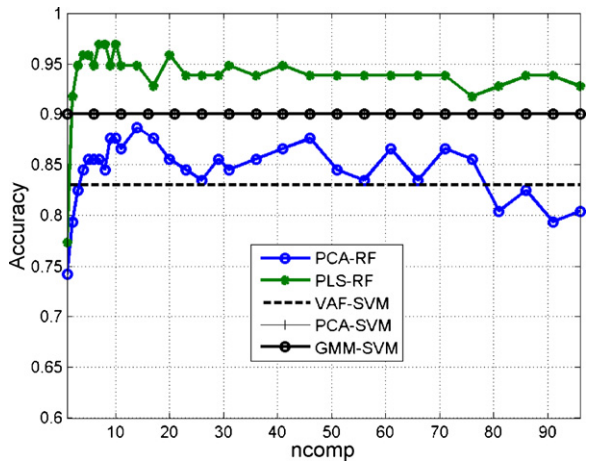

Fig. 6. Accuracy as a function of the number of components for the proposed PLS-RF CAD system compared to PCA-RF, VAF-SVM and GMM-SVM methods.

Acknowledgments

This work was partly supported by the MICINN of Spain under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030-C02-01) and HD2008-0029 projects and the *Consejería de Innovación, Ciencia y Empresa* (Junta de Andalucía, Spain) under the Excellence Projects TIC-02566 and TIC-4530.

References

[1] I. Álvarez, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, C.G. Puntonet, F. Segovia, Alzheimer's diagnosis using eigenbrains and support vector machines, Electronics Letters 45 (7) (2009) 342–343.
[2] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.
[3] R. Chaves, J. Ramírez, J.M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, SVM-based computer aided diagnosis of the Alzheimer's disease using *t*-test NMSE feature selection with feature correlation weighting, Neuroscience Letters 461 (3) (2009) 293–297.
[4] K.J. Friston, J. Ashburner, S.J. Kiebel, T.E. Nichols, W.D. Penny, Statistical Parametric Mapping: The Analysis of Functional Brain Images, Academic Press, 2007.
[5] J.M. Górriz, A. Lassl, J. Ramírez, D. Salas-Gonzalez, C.G. Puntonet, E.W. Lang, Automatic selection of ROIs in functional imaging using Gaussian mixture models, Neuroscience Letters 460 (2) (2009) 108–111.

[6] R. Higdon, N.L. Foster, R.A. Koeppe, C.S. DeCarli, W.J. Jagust, C.M. Clark, N.R. Barbas, S.E. Arnold, R.S. Turner, J.L. Heidebrink, S. Minoshima, A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging, Statistics in Medicine 23 (2004) 315–326.
[7] J.M. Hoffman, K.A. Welsh-Bohmer, M. Hanson, FDG PET imaging in patients with pathologically verified dementia, Journal of Nuclear Medicine 41 (11) (2000) 1920–1928.
[8] K. Ishii, A.K. Kono, H. Sasaki, N. Miyamoto, T. Fukuda, S. Sakamoto, E. Mori, Fully automatic diagnostic system for early- and late-onset mild Alzheimer's disease using FDG PET and 3D-SSP, European Journal of Nuclear Medicine and Molecular Imaging 33 (5) (2006) 575–583.
[9] M. López, J. Ramírez, J.M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA, Neuroscience Letters 464 (4) (2009) 233–238.
[10] M. López, J. Ramírez, J.M. Górriz, D. Salas-Gonzalez, I. Álvarez, F. Segovia, C.G. Puntonet, Automatic tool for the Alzheimer's disease diagnosis using PCA and Bayesian classification rules, Electronics Letters 45 (8) (2009) 389–391.
[11] S. Minoshima, K.A. Frey, R.A. Koeppe, N.L. Foster, D.E. Kuhl, A diagnostic approach in Alzheimer's disease using three dimensional stereotactic surface projections of fluorine-18-FDG PET, Journal of Nuclear Medicine 36 (7) (1995) 1238–1248.
[12] J.R. Petrella, R.E. Coleman, P.M. Doraiswamy, Neuroimaging and early diagnosis of Alzheimer disease: a look to the future, Radiology 226 (2003) 315–336.
[13] J. Ramírez, J.M. Górriz, R. Chaves, M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, SPECT image classification using random forests, Electronics Letters 12 (45) (2009) 604–605.
[14] J. Ramírez, J.M. Górriz, M. Gómez-Río, A. Romero, R. Chaves, A. Lassl, A. Rodríguez, C.G. Puntonet, F. Theis, E. Lang, Effective emission tomography image reconstruction algorithms for SPECT data, Lecture Notes in Computer Science 5101 (2008) 741–748.
[15] D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, A. Lassl, C.G. Puntonet, Improved Gauss–Newton optimization methods in affine registration of SPECT brain images, Electronics Letters 44 (22) (2008) 1291–1292.
[16] D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, M. López, I. Álvarez, F. Segovia, C.G. Puntonet, M. Gómez-Río, Analysis of SPECT brain images for the diagnosis of Alzheimer's disease using moments and support vector machines, Neuroscience Letters 461 (1) (2009) 60–64.
[17] D.H. Silverman, G.W. Small, C.Y. Chang, Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome, Journal of the American Medical Association 286 (17) (2001) 2120–2127.
[18] J. Stoeckel, G. Fung, SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information, in: Proc. of the Fifth International Conference on Data Mining (ICDM05), 2005, pp. 410–417.
[19] B.G. Tabachnick, L.S. Fidell, Computer-assisted Research and Design Analysis, Pearson Education, 2000.
[20] S. Wold, H. Ruhe, H. Wold, W.J. Dunn, The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverse, Journal of Scientific and Statistical Computations 5 (1984) 735–743.
[21] H. Wold, Quantitative Sociology: International Perspectives on Mathematical and Statistical Model Building, Chapter Path models with Latent Variables: The NIPALS Approach, Academic Press, 1975, pp. 307–357.