PETRA: Multivariate analyses for neuroimaging data

F. Segovia[1,2], I. A. Illán[2], D. Salas-Gonzalez[2], F.J. Martinez-Murcia[2],
C. Phillips[1,3], C. G. Puntonet[4], J. Rámirez[2] and J. M. Górriz[2]

[1] Cyclotron Research Centre, University of Liège, Liège, Belgium
[2] Department of Signal Theory, Networking and Communications, University of Granada, Granada, Spain
[3] Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium
[4] Department of Computer's Architecture and Technology, University of Granada, Granada, Spain

Abstract. In last years, many research efforts in neurosciences have focused in multivariate approaches based on machine learning as an alternative to the use of Statistical Parametric Mapping and the univariate analyses that it provides. However, this relatively new field still lacks of a software framework that completely meets the needs of the scientific community. In this work we present a toolbox designed to facilitate the access to the recent advances in neuroimaging data analysis based on multivariate approaches. The toolbox, written on Matlab, is freely available and implements a Graphical User Interface that allows managing neuroimaging data in an easy way.

Keywords: Matlab toolbox, machine learning, neuroimaging, multivariate analyses, PET, SPECT

1   Introduction

Neuroimaging technology is advancing at an impressive pace and is having huge fallouts both at the research and at the practical clinical level. Both, structural and functional neuroimages are commonly used to assist the diagnosis of several neurodegenerative disorders, including Alzheimer's disease and Parkinsonism [19, 28, 8, 2, 23]. In this regard, the imaging techniques based on nuclear medicine have played an important role in the early diagnosis of dementia [25, 11, 18, 7].

Nowadays there exist many software alternatives to deal with neuroimages. According to their functionality, they can be divided in two branches: toolboxes mainly focused in the handling of neuroimaging data (MRIcron [5], QuantiSPECT or PMod [6] are some examples), and the software that not only allows to manage

---

[5] http://www.cabiatl.com/mricro/mricro/
[6] http://www.pmod.com/technologies/index.html

the data but also makes possible to perform statistical analyses. Toolboxes in the latter branch, in turn, can be classified according to the kind of analyses they perform.

Over the last two decades, the analysis of neuroimaging data have been successfully addressed through univariate approaches, which analyze separately each voxel of the brain volumes. In that sense, the General Linear Model (GLM) implemented in Statistical Parametric Mapping (SPM) [6] has become the standard approach for analyzing functional imaging data. Instead, structural imaging modalities such as Magnetic Resonance Imaging (MRI) are commonly studied by means of the Voxel-Based Morphometry (VBM) approach [1], which investigates focal differences in gray matter density between groups of subjects.

In recent years, an important part of neuroimaging community has focused on multivariate approaches motivated by the recent advances in machine learning algorithms and the development of new statistical classifiers with a high generalization ability. Accordingly, a new toolbox that facilitates performing multivariate analysis on neuroimaging have been recently presented [22]. PRoNTo [7] is an open-source toolbox that allows to perform machine learning based analysis in an easy way thanks to a Graphical User Interface (GUI) similar to SPM. Other examples of toolboxes that implement multivariate analyses are Matlab MVPA[8] or Sci-kit Learn[9], however these ones lack of a GUI and are used through the command line.

In this work, we present a recently developed toolbox to manage and analyze neuroimaging data by means of multivariate techniques. This toolbox is addressed to clinicians and other researchers and its main purpose is making accessible the recent advances in Computer Aided Diagnosis (CAD) systems based on neuroimaging data. It is based on Matlab and is compatible with SPM (since it support Analyze and NIFTI formats), allowing a easy adaptation to experienced clinicians and neuropsychologists.

The manuscript is organized as follows. After this introduction, section 2 draws some concepts about the application of machine learning based procedures to neuroimaging data. Section 3 presents the toolbox we have developed. And finally, the discussion and conclusions are gathered in the section 4. This section includes a comparison with previous approaches.

2   Machine learning on neuroimaging

Neuroimaging data are usually stored as a sequence of voxels intensities, which interpretation depends on the modality used. For example, in the case of Positron Emission Tomography (PET) with ^{18}FDG radiotracer, each voxel intensity represents the glucose intake at its location in the brain. Instead for structural MRI, after segmentation of the image (into grey matter, white matter and cerebrospinal fluid), the voxel intensities represent the local density of each tissue.

---

[7] http://www.mlnl.cs.ucl.ac.uk/pronto/
[8] https://code.google.com/p/princeton-mvpa-toolbox/
[9] http://scikit-learn.org/stable/

A labeled dataset consists of pairs of a neurimage (as a sequence of voxels intensities), $\mathbf{X}_i \in \mathbb{R}^d$, and its label, y_i, that define the category to which that neuroimage belongs to. Given such neuroimage dataset, $D = \{\mathbf{X}_i, y_i\}$, $i = 1, ..., N$, the goal of a machine learning algorithm is to learn a function f that can accurately predict the labels of new unseen neuroimages [15]:

$$f(\mathbf{X}_i) = y_i \tag{1}$$

The classification can be perform using all the voxels in the brain volumes or specific regions of interest. Scanning the brain volume with a series of ROIs, such as in the *searchlight approach* [14]. Because of their prediction capacity, machine learning approaches are frequently used in the development of CAD systems for several disorders.

The main challenge of applying machine techniques to neuroimaging data is dealing with the dimensionality of the data, i.e. the huge number of voxels in a brain image (typically of few tens of thousand) compared to the limited number of samples (typically a few tens).

Due to this difference, the machine learning based studies usually suffer from the small sample size problem [4]. In order to address it, a dimensionality reduction technique can be applied. In that sense, methods based on Principal Component Analysis, Partial Least Squares and Independent Components Analysis have been successfully used [11, 25, 11]. These techniques, implemented in PETRA toolbox, are briefly described bellow.

2.1 Principal Component Analysis

Principal Component Analysis (PCA) is a simple, non-parametric method to extract relevant information from large data sets [12]. Mathematically, PCA performs a linear transformation that projects the data on a different orthogonal coordinate system such that the largest variance by any projection of the data comes to lie on the first dimension, the second largest variance on the second dimension, and so on. It can be used to reduce the imaging data dimensionality as follows:

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$ be a set of n brain volumes with each vector $\mathbf{x}_i$ being a sequence of voxel intensities. First the images are transformed to have zero mean and unity norm, resulting in a new set $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n]$. The covariance matrix, $\mathbf{C}$, is then computed:

$$\mathbf{C} = \frac{1}{N}\mathbf{Z}\mathbf{Z}^t \tag{2}$$

Subsequently, the eigenvector Γ and eigenvalue Λ matrices are calculated as $\mathbf{C}\Gamma = \Gamma\Lambda$. Since the image size is greater than the number of images, diagonalizing $\mathbf{Z}^t\mathbf{Z}$ instead of $\mathbf{Z}\mathbf{Z}^t$ reduces the computational burden and the eigenvectors/eigenvalues decomposition is reformulated as [26]:

$$(\mathbf{Z}^t\mathbf{Z})\Phi = \Phi\Lambda^* \tag{3}$$

$$\Gamma^* = \mathbf{Z}\Phi \tag{4}$$

where $\Lambda^* = diag(\lambda_1, \lambda_2, ..., \lambda_n)$ and $\Gamma^* = [\Gamma_1, \Gamma_2, ...\Gamma_n]$ are the first n eigenvalues and eigenvectors respectively. Finally, the images are modeled by projecting them over those eigenvectors (a.k.a. principal components).

2.2 Partial Least Squares

Partial Least Squares (PLS) is a statistical method that model the data in function of latent (i.e. not directly observed or measured) variables [27]. Therefore it is similar to PCA however PLS carries out the transformation by maximizing the covariance between the data and some properties of the data. Mathematically, PLS is a linear algorithm for modeling the relation between two data sets $\mathbf{X} \subset \mathbb{R}^N$ and $\mathbf{Y} \subset \mathbb{R}^M$. 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

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$$
$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \tag{5}$$

where N and M are respectively the number of features and the number of properties of the observed variables. $\mathbf{T}$ and $\mathbf{U}$ are $n \times p$ matrices of the p extracted score vectors (also known as components or latent vectors), the $N \times p$ matrix $\mathbf{P}$ and the $M \times p$ matrix $\mathbf{Q}$ are the matrices of loadings and the $n \times N$ matrix $\mathbf{E}$ and the $n \times M$ matrix $\mathbf{F}$ are the matrices of residuals (or error matrices).

For dimensionality reduction purposes, matrices $\mathbf{X}$ and $\mathbf{Y}$ are formed with the imaging data and their labels respectively [24]. After the decomposition, the x-scores in $\mathbf{T}$ are linear combinations of the variables in $\mathbf{X}$ and can be considered a good summary of the neuroimaging data.

2.3 Independent Components Analysis

Independent Components Analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals [3]. Its main applications is blind source separation, which, in its linear form, consists in finding the sources $\mathbf{S}$ which, when mixing using a weight matrix $\mathbf{A}$, provide the vector $\mathbf{X}$ of observed variables:

$$\mathbf{X} = \mathbf{AS} \tag{6}$$

where the sources $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n)$ are assumed to be statistically independent. In order to estimate both the mixing matrix $\mathbf{A}$ and the sources $\mathbf{S}$, ICA adaptively calculates the matrix $\mathbf{W} = \mathbf{A}^{-1}$ which either maximizes the nongaussianity or minimizes the mutual information. This technique has been successfully applied to dimension reduction problems by projecting the data into its independent components, performing that way the reduction [11].

2.4 Neuroimaging preprocessing

In order to make comparable neuroimaging data from different subjects, a preprocessing procedure should be performed. This procedure is usually divided in two steps: spatial normalization and intensity normalization. The former ensures that the same position in the volume coordinate system in different images corresponds to the same anatomical position. Several procedures have been proposed to perform that normalization. A comparison between them can be found in [13].

The intensity normalization step, in turn, removes the differences in the voxels intensity due to the application of different acquisition protocols. It is specially suitable in studies that involve nuclear imaging modalities, where the voxels intensity also depends on the radiopharmaceutical uptake. Several algorithms have been proposed to carry out this normalization [20, 5, 21] but, in general, they proceed to equalize the intensity of voxels not affecting by the disorder under study.

3 PETRA toolbox

The main purpose of this toolbox is to serve as a diagnosis assistant system whose decisions are based on the automatic analysis of neuroimaging data. Since the potential user is not necessarily familiar with the use of command line interfaces, the toolbox presented here is equipped with a complete GUI. This interface allows accessing all the features provided by the toolbox. Its main characteristics are listed below along with a small description:

- **Support for the main neuroimaging file formats.** The toolbox is able to load and save images stored in Dicom (.dcm), Nifti/Analyze (.hdr/.img and .nii) and other terms.
- **Multiple image modalities.** At the moment, the toolbox supports the following image modalities: SPECT ECD, FDG-PET and DaTSCAN (SPECT).
- **Different ways of displaying the data.** Although the most interesting feature are related to the statistical analysis of the data, the toolbox provides several options to visually analyze and compare neuroimages. Specifically, once loaded, the brain volumes can be shown in three different modes: i) in a list, with three slices (one per dimension) per volume, ii) as a sequence, showing only one volume at a time and, iii) in a grid with a slice from each volume, allowing the visual comparison of the volumes. Figure 1 shows three screenshots of the toolbox display facilities, one per each display mode.
- **Automatic orientation of the brain volumes.** Spatial normalization requires that both, image and template have equal orientation. In addition, the multivariate analyses also suppose that all the images have the same orientation. In order to ensure that, the neuroimaging data are automatically oriented to a predefined orientation using the procedure described in [10].
- **Image normalization.** Both intensity and spatial normalization may be performed through the GUI. Regarding the intensity normalization, two methods are provided: based on the maximum [21] and based on the global
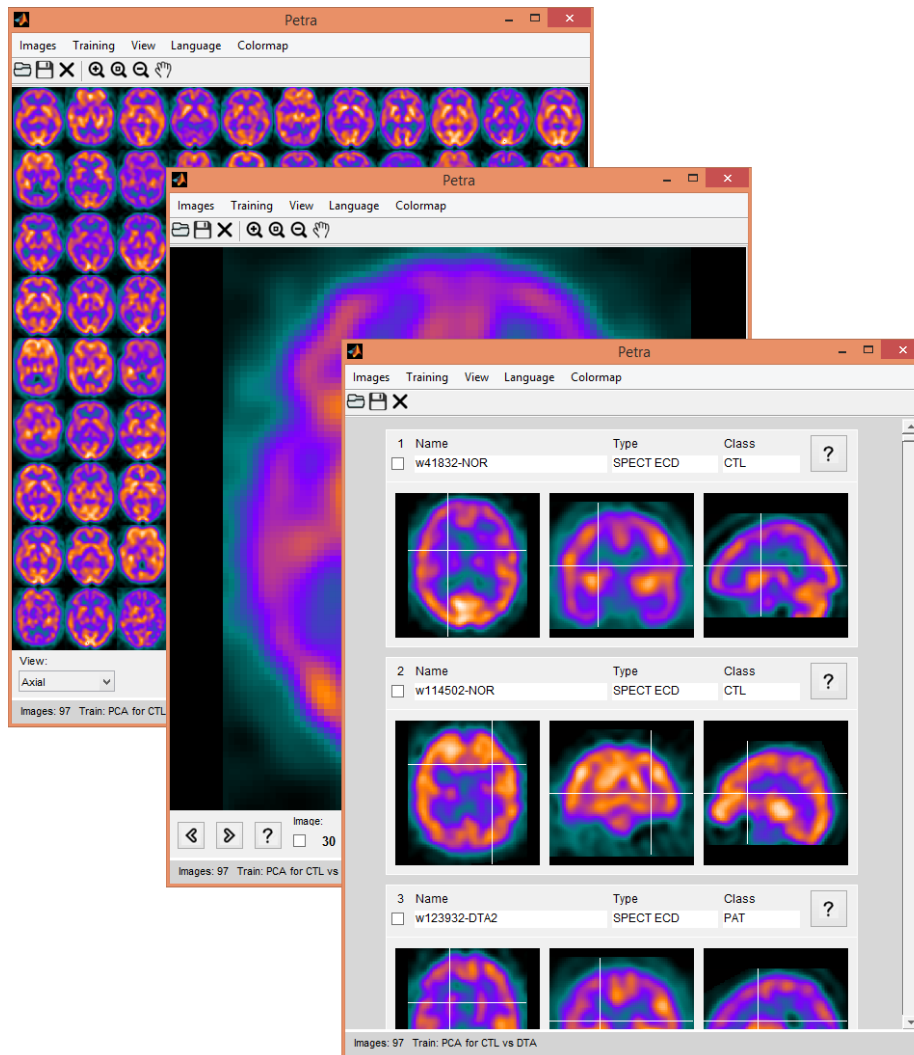
Fig. 1. Display facilities of the PETRA toolbox

intensity [5]. However, spatial normalization is not implemented in the toolbox at the moment. Instead, it uses the routines included in SPM (and therefore this function requires SPM is installed on our system).

– **Freely available and multi-platform** Petra toolbox can be freely download from http://http://tst1.ugr.es/petra/. Since it is developed on Matlab[10], is can be run in Windows and Unix-based systems (such as Linux or Mac OS). The widespread use of Matlab in the neuroscience community and its high-level language makes it one of the most appropriate option to analyze neuroimaging data.

3.1 PETRA as a computer aided diagnosis system

In order to estimate the class (healthy or pathological) of a given image, a classification function (see Equation 1) should be used. This kind of functions, known as *training* in the toolbox interface, has a specific form (depending on the classification algorithm) and thus relies on a set of parameters. These parameters need to be estimated in order to provide an accurate prediction. For supervised learning approaches, this is called *training phase* and requires a *training dataset* that includes the images with their labels.

The following classification methods are implemented at the moment[11]: Spatial Component Analysis [9], PCA-based system [17], PCA- and LDA-based system [16], PCA on class means, ICA on class mean and PLS-based system [24]. Please, see the corresponding references for further information.

Classification function can be saved and applied later on new images to predict their class. Figure 2 shows the window displayed after estimating the class for a new image. Notice that not only the category is provided, but it shows additional information that can be used to measure the estimation confidence. For example, when a method based in the singular value decomposition (such as PCA of PLS) is used, the results window also shows the first *eigenbrains* [11, 17], i.e. the first eigenvectors in 3D-volume form. The *eigenbrains* can be viewed as elementary brain volumes from which all the images are built and are useful to realize the regions focused by the dimensionality reduction.

4   Discussion and conclusions

We have presented a software toolbox designed to assist clinicians in the diagnosis of neurological disorders that can be predicted though neuroimaging data, such as Alzheimer's disease or Parkinsonism. This toolbox benefits from recent advances in machine learning based algorithms and facilitates their application in the clinical practice.

From a technical point of view, the main novelty introduced by this toolbox is the combination of advanced dimensionality reduction techniques and recent

---

[10] A comercial software created by Mathworks
[11] Running most of these methods requires using the Matlab statistical toolbox, thus a specific license of it is need.
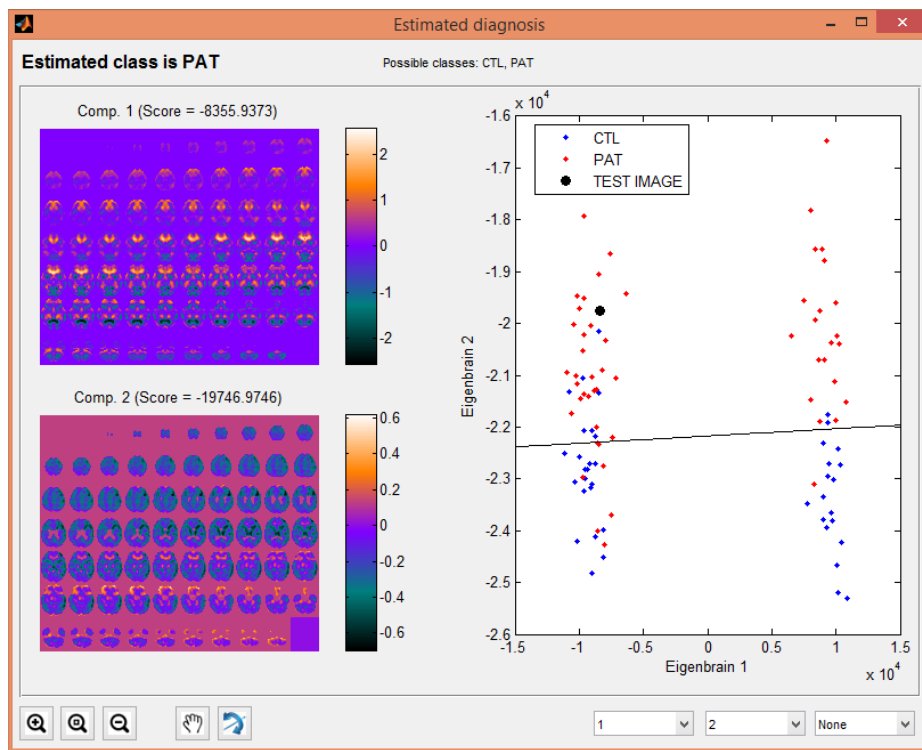
Fig. 2. Petra GUI window showing the results of estimating the class for a new neuroimage.

machine learning algorithms. All the methods implemented in the software have been published in peer review-journals and therefore accepted by the scientific community. In addition, the GUI has been meticulously implemented to be accessible to people not familiarized with pattern recognition algorithms.

As mentioned in the introduction, there already exist several toolboxes to deal with neuroimaging data but, to our knowledge, neither of them contains the features provided by PETRA toolbox: tools for visually analyze neuroimages, ability to perform statistical analyses (including the normalization of the data), user-friendly GUI, etc. Probably, the most similar alternatives (from a functionality point of view) are SPM and PRoNTo. The other ones lack of either statistical analysis abilities or a GUI that makes them suitable to user not used to command line interfaces. Nevertheless, it is worth noting that both SPM and PRoNTo are richer toolboxes which provide more image utilities and application possibilities. Indeed, these softwares have several years of development (decades in the case of SPM) whereas Petra is still in an initial stage. Nonetheless, Petra provides, in our opinion, some advantages. Compared to SPM, PETRA is able to perform multivariate analyses, which provide higher sensitivity and are more suitable for the assisted diagnosis. In general, SPM is more suitable for group comparisons whereas Petra is more suitable for the assisted diagnosis. Compared to PRoNTo, our toolbox implements dimensionality reduction techniques, which allow to mitigate the small sample size problem. In addition, it provides a way to save classification functions (i.e. a training procedure) that makes possible to categorize new images in an easy way and enforce its capabilities as CAD system.

Acknowledgement

References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry–The methods. NeuroImage 11(6), 805–821 (2000)
2. Booij, J., Tissingh, G., Boer, G.J., Speelman, J.D., Stoof, J.C., Janssen, A.G., Wolters, E.C., van Royen, E.A.: [123I]FP-CIT SPECT shows a pronounced decline of striatal dopamine transporter labelling in early and advanced Parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry 62(2), 133–140 (1997)
3. Comon, P.: Independent component analysis, a new concept? Signal Processing 36(3), 287–314 (Apr 1994)
4. Duin, R.P.W.: Classifiers in almost empty spaces. In: Proceedings 15th International Conference on Pattern Recognition. vol. 2, pp. 1–7. IEEE (2000)

5. Dukart, J., Mueller, K., Horstmann, A., Vogt, B., Frisch, S., Barthel, H., Becker, G., Mller, H.E., Villringer, A., Sabri, O., Schroeter, M.L.: Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies. NeuroImage 49(2), 1490–1495 (2010)

6. Friston, K., *et al.* (eds.): Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press (2007)

7. Górriz, J.M., Segovia, F., Ramírez, J., Lassl, A., Salas-Gonzalez, D.: GMM based SPECT image classification for the diagnosis of Alzheimer's disease. Applied Soft Computing 11(2), 2313 – 2325 (2011)

8. Hamilton, D., O'Mahony, D., Coffey, J., Murphy, J., O'Hare, N., Freyne, P., Walsh, B., Coakley, D.: Classification of mild alzheimer's disease by artificial neural network analysis of SPET data. Nuclear medicine communications 18(9), 805–810 (Sep 1997)

9. Illán, I.A., Górriz, J.M., López, M.M., Ramírez, J., Salas-Gonzalez, D., Segovia, F., Chaves, R., Puntonet, C.G.: Computer aided diagnosis of alzheimer's disease using component based SVM. Appl. Soft Comput. 11(2), 23762382 (2011)

10. Illán, I.A., Górriz, J.M., Ramírez, J., Salas-Gonzlez, D., Martínez-Murcia, F.J., Segovia, F., Puntonet, C.G.: Automatic orientation of functional brain images for multiplataform software. In: Vicente, J.M.F., Sánchez, J.R.A., López, F.d.l.P., Moreo, F.J.T. (eds.) Natural and Artificial Models in Computation and Biology, pp. 406–411. No. 7930 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (2013)

11. Illán, I., Górriz, J., Ramírez, J., Salas-Gonzalez, D., López, M., Segovia, F., Chaves, R., Gómez-Rio, M., Puntonet, C.: 18F-FDG PET imaging analysis for computer aided Alzheimers diagnosis. Information Sciences 181(4), 903–916 (2011)

12. Jolliffe, I.T.: Principal Component Analysis. Springer, 2nd ed. 2002 edn. (Oct 2002)

13. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46(3), 786–802 (2009)

14. Kriegeskorte, N., Bandettini, P.: Analyzing for information, not activation, to exploit high-resolution fMRI. NeuroImage 38(4), 649–662 (Dec 2007)

15. Lemm, S., Blankertz, B., Dickhaus, T., Mller, K.R.: Introduction to machine learning for brain imaging. NeuroImage 56(2), 387–399 (2011)

16. López, M., Ramírez, J., Górriz, J.M., Álvarez, I., Salas-Gonzalez, D., Segovia, F., Chaves, R.: SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA. Neuroscience Letters (2009)

17. López, M., Ramírez, J., Górriz, J., Salas-Gonzalez, D., Álvarez, I., Segovia, F., Puntonet, C.G.: Automatic tool for alzheimer's disease diagnosis using PCA and bayesian classification rules. Electronics Letters 45(8), 389–391 (2009)

18. Ramírez, J., Górriz, J.M., Gómez-Río, M., Romero, A., Chaves, R., Lassl, A., Rodríguez, A., Puntonet, C.G., Theis, F., Lang, E.: Effective emission tomography image reconstruction algorithms for SPECT data. In: Bubak, M., Albada, G.D.v., Dongarra, J., Sloot, P.M.A. (eds.) Computational Science ICCS 2008, pp. 741–748. No. 5101 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (2008)

19. Román, G., Pascual, B.: Contribution of neuroimaging to the diagnosis of alzheimers disease and vascular dementia. Archives of Medical Research 43(8), 671–676 (2012)

20. Salas-Gonzalez, D., Górriz, J.M., Ramírez, J., Illán, I.A., Lang, E.W.: Linear intensity normalization of FP-CIT SPECT brain images using the -stable distribution. NeuroImage 65, 449–455 (2013)
21. Saxena, P., Pavel, D.G., Quintana, J.C., Horwitz, B.: An automatic threshold-based scaling method for enhancing the usefulness of tc-HMPAO SPECT in the diagnosis of alzheimers disease. In: Wells, W.M., Colchester, A., Delp, S. (eds.) Medical Image Computing and Computer-Assisted Intervention MICCAI98, pp. 623–630. No. 1496 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (1998)
22. Schrouff, J., Rosa, M.J., Rondina, J.M., Marquand, A.F., Chu, C., Ashburner, J., Phillips, C., Richiardi, J., Mouro-Miranda, J.: PRoNTo: pattern recognition for neuroimaging toolbox. Neuroinformatics (2013)
23. Segovia, F., Górriz, J.M., Ramírez, J., Álvarez, I., Jiménez-Hoyuela, J.M., Ortega, S.J.: Improved parkinsonism diagnosis using a partial least squares based approach. Medical physics 39(7), 4395–4403 (2012)
24. Segovia, F., Górriz, J., Ramírez, J., Salas-Gonzalez, D., Álvarez, I.: Early diagnosis of alzheimers disease based on partial least squares and support vector machine. Expert Systems with Applications 40(2), 677–683 (2013)
25. Segovia, F., Górriz, J.M., Ramírez, J., Chaves, R., Illán, I.Á.: Automatic differentiation between controls and parkinson's disease datscan images using a partial least squares scheme and the fisher discriminant ratio. In: Advances in Knowledge-Based and Intelligent Information and Engineering Systems - 16th Annual KES Conference. pp. 2241–2250 (2012)
26. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
27. Varmuza, K., Filzmoser, P.: Introduction to Multivariate Statistical Analysis in Chemometrics. Taylor and Francis - CRC Press, Boca Raton, FL (2009)
28. Vernooij, M.W., Smits, M.: Structural neuroimaging in aging and alzheimer's disease. Neuroimaging Clinics of North America 22(1), 33–55 (2012)