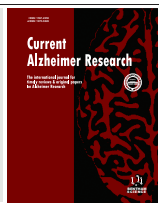
Current Alzheimer Research, 2016, *13*, 831-837



Combining Feature Extraction Methods to Assist the Diagnosis of Alzheimer's Disease

F. Segovia[1,2,*], J. M. Górriz[2], J. Ramírez[2], and C. Phillips[1] for the Alzheimer's Disease Neuroimaging Initiative[#]

[1]*Cyclotron Research Centre, University of Liège, Liège, Belgium;* [2]*Department of Signal Theory, Networking and Communications, University of Granada, Granada, Spain*

F. Segovia

Abstract: Neuroimaging data as ^{18}F-FDG PET is widely used to assist the diagnosis of Alzheimer's disease (AD). Looking for regions with hypoperfusion/ hypometabolism, clinicians may predict or corroborate the diagnosis of the patients. Modern computer aided diagnosis (CAD) systems based on the statistical analysis of whole neuroimages are more accurate than classical systems based on quantifying the uptake of some predefined regions of interests (ROIs). In addition, these new systems allow determining new ROIs and take advantage of the huge amount of information comprised in neuroimaging data. A major branch of modern CAD systems for AD is based on multivariate techniques, which analyse a neuroimage as a whole, considering not only the voxel intensities but also the relations among them. In order to deal with the vast dimensionality of the data, a number of feature extraction methods have been successfully applied. In this work, we propose a CAD system based on the combination of several feature extraction techniques. First, some commonly used feature extraction methods based on the analysis of the variance (as principal component analysis), on the factorization of the data (as non-negative matrix factorization) and on classical magnitudes (as Haralick features) were simultaneously applied to the original data. These feature sets were then combined by means of two different combination approaches: i) using a single classifier and a multiple kernel learning approach and ii) using an ensemble of classifier and selecting the final decision by majority voting. The proposed approach was evaluated using a labelled neuroimaging database along with a cross validation scheme. As conclusion, the proposed CAD system performed better than approaches using only one feature extraction technique. We also provide a fair comparison (using the same database) of the selected feature extraction methods.

Keywords. Alzheimer's disease, Computer aided diagnosis systems, Dimensionality reduction, Machine learning, Support Vector Machine, ^{18}F-FDG PET.

Received: May 04, 2015 **Revised: October 16, 2015** **Accepted: November 11, 2015**

1. INTRODUCTION

During last decades, neuroimaging technology has become an usual clinical practice to assist the diagnosis of neurodegenerative disorders. It led a radical change, from the traditional visual inspection of films and printouts to accurate quantification of indices of interest such as the brain metabolism or the regional cerebral blood flow. A number of studies also suggests [1, 2] that pathological manifestations of Alzheimer's disease (AD) are detectable through neurogimaging data even before that patients become symptomatic. This led clinicians to use neuroimages not only to confirm the diagnosis but also to anticipate it.

*Address correspondence to this author at the Cyclotron Research Centre, University of Liège, Liège, Belgium; Department of Signal Theory, Networking and Communications, University of Granada, Granada, Spain;
E-mail: fsegovia@ugr.es

[#]*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/ uploads/how to apply/ADNI Acknowledgement List.pdf*

Nowadays, visual inspection of the neuroimages is not recommended because it often misses crucial information. Computer systems instead allow to quantify small differences invisible to the human eye and remove the subjectivity inherent to visual analysis. Over the last years, different computer-based approaches have been presented to analyse neuroimaging data. First systems were based on the analysis of regions of interest (ROIs) previously defined and were therefore focused on a specific disorder [3, 4]. The main criticism made to these systems is that clinicians should manually delimit the region, which is complicated and prone to error. Moreover, revealing unknown ROIs is only possible by using an inefficient try and error scheme.

On the other hand, the Statistical Parametric Mapping (SPM) [5] software, created by Prof. Karl Friston, facilitates the application of univariate analyses, i.e. each voxel is considered individually, over brain images, becoming a standard in the neuroimaging research community. In the early 2000's Signorini *et al.* demonstrated that SPM could be used to model the pattern of cerebral functional neurodegeneration. Recently, SPM has been successfully used to assist de diagnosis of Alzheimer's disease [6], including the differentiation between Mild Cognitive Impairment (MCI), Alz-

1875-5828/16 $58.00+.00

heimer's disease (AD) and Frontotemporal Lobar Degeneration (FTLD) patients [7].

Conversely to SPM analyses, multivariate approaches based on machine learning examine all the voxels from an image as a whole, considering the relations between them. In 2004, it was suggested that a statistical classifier trained using the voxels as features is a satisfactory method to differentiate between healthy subjects and AD patients. Nevertheless, the sensitivity of the classifiers is usually limited by the huge dimensionality of the data, significantly larger than the number of training samples used in most of studies [8]. In order to address this issue, some authors propose to include a feature extraction step before the classification [9-12]. In this step the neuroimages are summarized into a reduced set of features that contains the information useful to differentiate between groups. Methods based on Principal Component Analaysis (PCA) [13], Non-negative Matrix Factorization (NNMF) [14] and the co-occurrence-based method of Haralick (HF) [15] have been successfully used [10, 16-18].

Other researchers, instead, proposed combining the ROIs selection typical of the first systems and a machine learning method. In [19], the author used an atlas to select some previously defined ROIs and then a statistical classifier was used to differentiate between controls and AD patients. This method inherits the disadvantages of ROIs-based methods: the ROIs should be defined *a priori* using a data independent procedure.

Finally, during the last few years, some authors proposed using an ensemble of classifiers in order to deal with the high dimensionality of the neuroimaing data. For example, in [20] a local patch-based subspace ensemble method was demonstrated. The authors built several classifiers using different subsets of the neuroimaing data and then, the classifiers were combined.

In this manuscript we propose a new approach consisting on the combination of several feature extraction techniques to summarize neuroimaging data in order to improve the performance of the subsequent classification procedure. This will result in more accurate computer aided diagnosis (CAD) system for neurodegenerative disorders such as AD. The major difficulty of this approach is to find a way to combine the feature extraction methods. We demonstrated two methodologies: i) using a multiple kernel learning classifier and ii) using one classifier per feature extraction method and selecting the final output by majority voting. For evaluation purposes, a database with 210 ^{18}F-FDG-PET images from healthy subjects, MCI patients and AD patients was used. Our experiments suggest that combining several feature extraction methods provides more accurate CAD systems than using only one.

The second methodology proposed in this manuscript and consisting on using one classifier per feature extraction approach is similar to the method described in [20]. However, they differ in the way in which classifiers are built: While we used a feature extraction technique to obtain a discriminative representation of the neuroimaing data, in [20] the authors used a subset of the neuroimaging data to train each classifier.

2. MATERIALS AND METHODS

2.1. Image Database

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5- year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-todate information, see www.adni-info.org.

In this work, a set 210 ^{18}F-FDG-PET images was used. These images were categorized into three groups according to the patient's diagnosis: neurologically healthy, MCI or AD (some demographic details of the patients are gathered in Table **1**). All the images were normalized before the statistical analysis. The spatial normalization was performed by means of the template matching algorithm provided by SPM8. Additionally, the intensity values were normalized by scaling them with respect to the intensity values obtained in the cerebellum (this method was demonstrated as superior to global normalization in identifying dementia patients in comparison to control subjects [21]). To this end, the cerebellar region was delimited by means of the automatic anatomical labelling atlas (AAL) [22], in a way similar to the procedure performed in [23]. Normalized images matched the Montreal Neuroimaging Institute (MNI) space and had a voxel size of $1.5 \times 1.5 \times 1.5$ mm^3.

2.2. Feature Extraction Based on Principal Component Analysis

Principal Component Analysis (PCA) [13] is a mathematical procedure that rotates the axes of data space along the lines of maximum variance. The axis of greatest variance

Table 1. Demographic details of the PET dataset used in this work. μ and σ stand for the average and the standard deviation respectively.

	#	Sex			Age	
		M	**F**	μ	σ	**range**
Healthy subjects	70	35	35	73.33	6.48	57-89
MCI patients	70	32	38	71.40	7.56	55-85
AD patients	70	39	31	74.49	7.93	56-90

are called principal components. The dimensionality reduction of 3D images based on PCA may be performed as follows [16]: Let $\mathbf{X} = [\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N]$ be a set of N functional brain images in vector form. After normalizing the images to unity norm and subtracting the mean, a new set $\mathbf{Y} = [\mathbf{y}_1,\mathbf{y}_2,...,\mathbf{y}_N]$ is obtained. The covariance matrix of the normalized vectors set is defined as:

$$\mathbf{C} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathbf{t}} \quad (1)$$

Then, the eigenvector Γ and eigenvalue Λ matrices are computed as $\mathbf{C}\Gamma = \Gamma\Lambda$. Since the image size is greater than the number of images, diagonalizing $\mathbf{Y}^t\mathbf{Y}$ instead of $\mathbf{Y}\mathbf{Y}^t$ reduces the computational burden and the eigenvectors/eigenvalues decomposition is reformulated as [24]:

$$(\mathbf{Y}^t\mathbf{Y})\Phi = \Phi\Lambda^* \quad (2)$$

$$\Gamma^* = \mathbf{Y}\Phi \quad (3)$$

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. The eigenvectors are ordered in a decreasing order of the variance expressed by themselves. Therefore, by selecting a minimum percentage of variance to be captured, only the first N eigenvectors can be selected. Finally, the features are extracted by projecting the images over the selected eigenvectors (a.k.a. principal components).

2.3. Feature Extraction Based on NonNegative Matrix Factorization

Non-negative Matrix Factorization (NNMF) is a factorization technique intended to decompose nonnegative data [14]. Given a non-negative matrix **X**, NNMF computes two non-negative matrices, **W** and **H** such that

$$\mathbf{X} \approx \mathbf{W}\mathbf{H} \quad (4)$$

In contrast to other decomposition methods such as PCA or independent component analysis (ICA), NNMF ensure representations purely additive (not allowing subtractions) due to the non-negativity condition. This factorization can be applied to reduce the dimensionality of neuroimaging data in the following manner. Let suppose **X** is a $M \times N$ matrix, in which the columns represent different neuroimages, the rows are voxel positions and $M >> N$. After the factorization, the matrix **X** is approximated by a $M \times k$ matrix **W** and a $k \times N$ matrix **H**. Usually k is chosen to be smaller than M or N, so that **W** and **H** are smaller than **X**, resulting in a compressed version of the original data matrix. An appropriate decision on the value of k is critical in practice, but the choice is very often problem dependent.

The matrix **W** can be considered as a new representation space for the data, while the matrix **H** is the representation of the data in the transformed space. Thus, the NNMF factorization yields a reduced matrix **H** which represents **X** in terms of **W**.

2.4. Feature Extraction Based on the Haralick Features

In 1973 Prof. Robert M. Haralick defined a set of easily computable textural features based on spatial dependencies intended for object detection and/or classification using imaging data. The features for a given image are based on its co-occurrence matrix:

$$\mathbf{C} = \begin{pmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,I_N} \\ C_{2,1} & C_{2,2} & \cdots & C_{2,I_N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{I_N,1} & C_{I_N,2} & \cdots & C_{I_N,I_N} \end{pmatrix}$$

It consists in a $I_N \times I_N$ matrix where I_N is the number of gray levels considered and each C_{ij} is generated by counting the number of times a pixel with value i is adjacent to a pixel with value j divided by the total number of such comparisons made. The concept of adjacency depends on the direction we define. Usually, all possible directions (north-south, east-west, northeast-southwest and northwest-southeast in case of 2D images) are taken into account. Additionally it is possible to consider as adjacent two pixel at distance greater than 1.

Co-occurrence matrices were initially defined for 2 dimensional images but the idea can be easily extended to 3 dimensional data by adding additional directions. However, considering several distances and all the possible directions for neuroimaging data can result in a relatively high number of redundant features. In order to address this issue, the Fisher Discriminant Ratio (FDR) [25] is used to rate the features according to its ability to separate two groups:

$$FDR = \frac{(\mu_1 - \mu_2)^2}{{\sigma_1}^2 + {\sigma_2}^2} \quad (5)$$

where μ_i and σ_i denote, respectively, the mean and the variance for the i-th group samples. Only the most discriminant features (the ones with highest FDR) will be used in the subsequent analysis.

2.5. Background on Supervised Learning Based on Support Vector Machine

Support Vector Machine (SVM) is a supervised learning method derived from the statistical learning theory, which was developed by Vladimir Vapnik in the late 90s [26]. In supervised learning, the classification is performed in two steps: training and test. In the first one, the classifier is defined according to the feature vectors of a given training set, where the category of each vector (its class label) is known. Once defined, the classifier can be used to categorize unseen samples (the test stage). Using training data consisting on k-dimensional patterns $\mathbf{x}_i$ and their class labels, y_i:

$$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\ldots,(\mathbf{x}_N,y_N) \in (\mathbb{R}^k \times \{\pm 1\}) \quad (6)$$

a SVM classifier compute an hyperplane, $g(\mathbf{x})$, that has the largest distance to the closest training data point of any class:

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0, \quad (7)$$

where $\mathbf{w}$ is the weight vector, orthogonal to the decision hyperplane, and w_0 is the threshold. This hyperplane is called maximal margin hyperplane. During the test stage, the classifier assigns a class label to each new data point according to the side of the hyperplane where the data point is.

When no linear separation of the training data is possible, SVM can work effectively in combination with kernel techniques so that the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space.

2.6. Proposed Methodology

In this work, we are presenting a new feature extraction methodology for neuroimaging data consisting on combining some classical feature extraction methods in order to improve the performance of a classification procedure that separate the neuroimages into groups. This methodology allows to enlarge the feature space, resulting on a more heterogeneous and rich feature set. We selected three classical techniques of different *nature* (based on the variance of the data, on the factorization of the data and on classical magnitudes) that had been previously used for dimensionality reduction purposes: PCA, NNMF and HF. They were applied to neuroimaging data as described in previous sections.

Two strategies were followed to combine the feature sets generated by the three feature extraction methods. On the one hand, we used a multiple kernel learning [27] algorithm. Specifically we computed one linear kernel per feature set and then the resulting kernels where combined by using the following function:

$$k(x_i, x_j) = \sum_{m=1}^{3} w_m k_m(x_i^m, x_j^m) \quad (8)$$

where $k_m(\mathbf{x}^m{}_i, \mathbf{x}^m{}_j)$ and $\mathbf{x}^m{}_i$ are, respectively, the kernel and the features corresponding to method m. w_m is a weight related to the accuracy achieved by using only k_m in the subsequent classification procedure. Specifically it is computed as follows:

$$w_m = \frac{a_m}{\sum_{m=1}^{3} a_m} \quad (9)$$

where $a_m \in \{0,0.5\}$ is the percentage of accuracy over 50%, calculated by subtracting 0.5 to the obtained accuracy.

On the other hand we used a SVM classifier per feature extraction method and the final output prediction was estimated by combining the outputs of all the classifiers. This combination was performed as follows: Let suppose e_i and c_i with $i = 1,2,3$ and $e_i \in \{-1,1\}$ are respectively the estimated category and a measure of the confidence for each classifier. The combined category, e, was computed as

$$e = sign\left(\sum_{i=1}^{3} e_i c_i\right) \quad (10)$$

Since we used SVM-based classifiers, the confidence of a classification procedure is estimated as the distance to the decision hyperplane.

3. RESULTS AND DISCUSSION

The methodology proposed above was evaluated and compared with previous approaches using the database described in section 2.1. The experiments were carried out in two steps. First we classified healthy subjects vs AD patients and then MCI vs AD patients. A SVM classifier and linear kernels were used in all the experiments. The accuracy, sensitivity and specificity were estimated by means of a 5-fold cross-validation scheme. The results are shown in Table **2**.

Parameters for feature extraction methods are usually problem dependent. In order to select the most suitable ones for AD diagnosis purposes using ^{18}FFDG PET data, we considered several thresholds for the percentage of total variance (PCA), the parameter k (NNMF) and the number of Haralick features (based on the FDR). Fig. (**1**) shows the obtained accuracies. Note that the thresholds providing highest accuracy are almost the same for both classifications (healthy subjects vs AD patients and MCI vs AD patients). That suggests that parameters depend on the problem, i.e. AD diagnosis, but not on the data. PCA and NNMF allow to analyse the regions focused by the feature extraction algorithm through the examination of the new representation space. ROIs for PCA (shown in Fig. **2** top) was computed by averaging the first few principal components (the ones used to project the neuroimages and extract the features) and reshaping the resulting vector into brain form. Observe that a few regions (mainly the posterior cingulate gyrus, paracentral lobule and supplementary motor area) are assigned a high importance. Those regions have been previously related with AD in the literature [28-31]. In case of NNMF, the average of matrix $\mathbf{W}$, which defines the representation space, is shown (Fig. **2** bottom). Both methods, PCA and NNFM, perform the feature extraction in a different way: PCA focuses in a few regions while NNMF assigns a high weight to large number of regions in the brain cortex. The advantage of the method we are proposing is that the classification algorithm can take advantage of both approaches, yielding higher accuracy rates.

CONCLUSION

In this work, we demonstrated a feature extraction approach that improve the computer-assisted diagnosis of AD based on machine learning. This method consists in combining several classical feature extraction techniques in one system. The combination was performed in two ways: i) using a

Table 2. Accuracy, sensitivity and specificity achieved by the proposed methodology when separating healthy subjects from AD patients and MCI from AD patients.

				Positive	Negative
	Accuracy	**Sensitivity**	**Specificity**	**Likelihood**	**Likelihood**
Healthy subjects vs AD patients					
PCA	82.86%	77.14%	88.57%	6.75	0.26
NNMF	84.29%	82.86%	85.71%	5.80	0.20
Haralick	80.00%	78.57%	81.43%	4.23	0.26
Proposed method (multikernel)	85.00%	88.57%	81.43%	4.77	0.14
Proposed method (3 classifiers)	85.00%	88.57%	81.43%	4.77	0.14
MCI vs AD patients					
PCA	75.71%	78.57%	72.86%	2.90	0.29
NNMF	76.43%	78.57%	74.29%	3.06	0.29
Haralick	72.14%	67.14%	77.14%	2.94	0.43
Proposed method (multikernel)	79.29%	75.71%	82.86%	4.42	0.29
Proposed method (3 classifiers)	78.57%	75.71%	81.43%	4.08	0.30

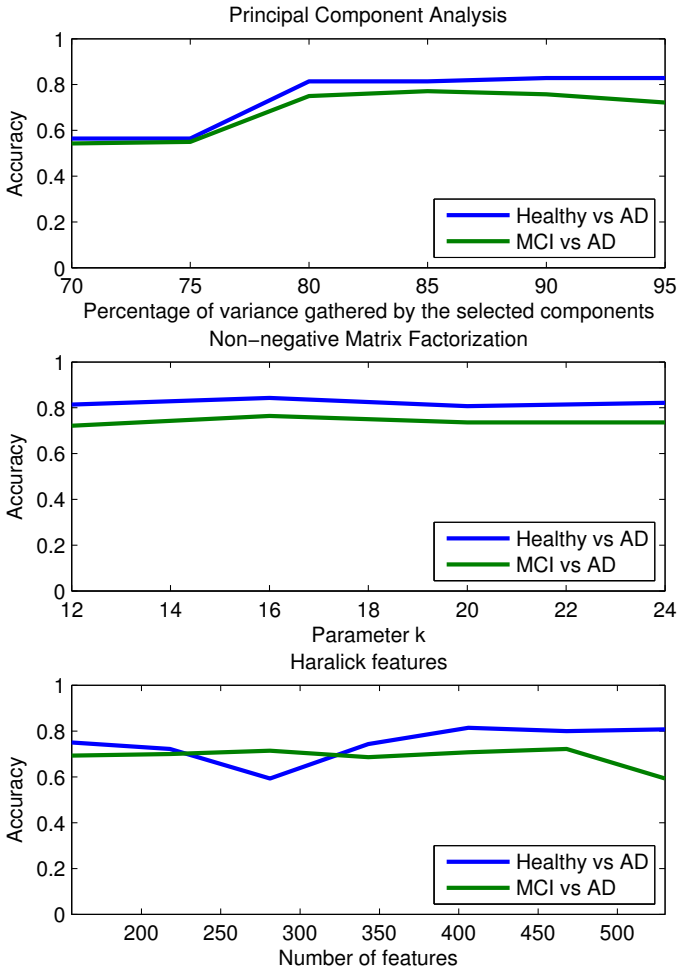


Fig. (1). Accuracy obtained by the three feature extraction methods studied according to the selected parameters.

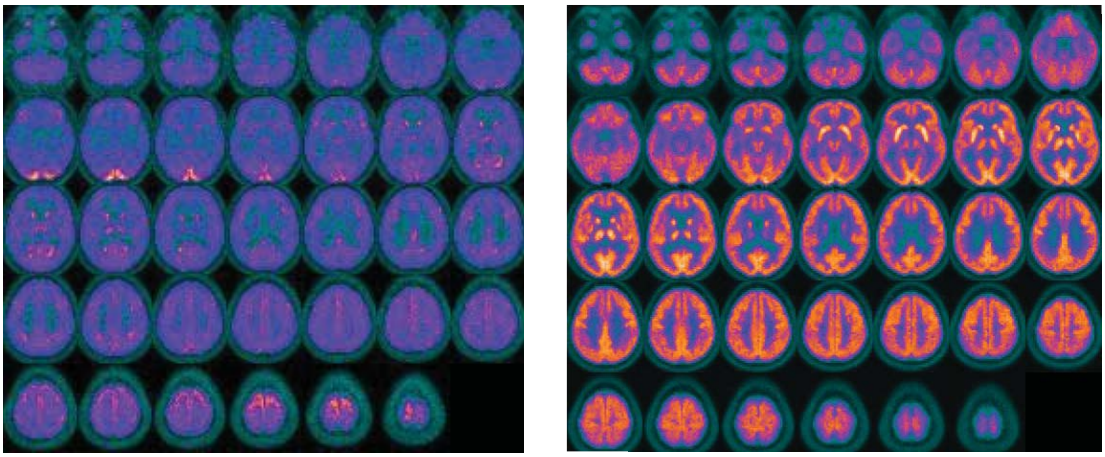


Fig. (2). Regions focused by the NNMF (left) and the PCA (right) algorithms to perform the feature extraction.

multiple kernel learning with one kernel per feature extraction technique and, ii) using one classifier per feature extraction technique and then combining the classifier outputs according to its confidence.

The proposed methodology was evaluated using a SVM classifier and 210 ^{18}F-FDG PET neuroimages from the ADNI database. We considered three feature extraction techniques: Principal Component Analysis, Non-Negative Matrix Factorization and the Haralick features. Our experiments suggest that combining several techniques provide higher accuracy than using only one. In addition, the manuscript provides a fair comparison between three classical feature extraction algorithms for neuroimaging data.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by: i) the MICINN under the projects TEC2008-02113 and TEC2012-34306; ii) the Ministry of Economy, Innovation, Science and Employment of the Junta de Andalucía under the Excellence Projects P09-TIC-4530 and P11-TIC-7103 and a Talent Hub project granted to FS (project approved by the Andalucía Talent Hub Program launched by the Andalusian Knowledge Agency, co-funded by the European Union's Seventh Framework Program, Marie Sklodowska-Curie actions (COFUND – Grant Agreement no 291780) and the Ministry of Economy, Innovation, Science and Employment of the Junta de Andalucía) and, iii) the Universities of Liège (Belgium) and Granada (Spain).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimers Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; ; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

[1] Canu E, McLaren DG, Fitzgerald ME, Bendlin BB, Zoccatelli G, Alessandrini F, *et al.* Microstructural Diffusion Changes are Independent of Macrostructural Volume Loss in Moderate to Severe Alzheimer's Disease. J Alzheimers Dis 19(3): 963-976 (2010).

[2] Whitwell JL, Przybelski SA, Weigand SD, Knopman DS, Boeve BF, Petersen RC, *et al.* 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. Brain 130(7): 1777-1786 (2007).

[3] Hamilton D, O'Mahony D, Coffey J, Murphy J, O'Hare N, Freyne P, *et al.* Classification of mild Alzheimer's disease by artificial neural network analysis of SPET data. Nucl Med Comm 18(9): 805-810 (1997).

[4] Dawson MR, Dobbs A, Hooper HR, McEwan AJ, Triscott J, Cooney J. Artificial neural networks that use single-photon emission tomography to identify patients with probable Alzheimer's disease. Eur J Nucl Med 21(12): 1303-1311 (1994).

[5] Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Amsterdam ; Boston: Academic Press 2006.

[6] Yakushev I, Hammers A, Fellgiebel A, Schmidtmann I, Scheurich A, Buchholz H, *et al.* SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnestic mild cognitive impairment from healthy aging. NeuroImage 44(1): 43-50 (2009).

[7] Perani D, Della Rosa PA, Cerami C, Gallivanone F, Fallanca F, Vanoli EG, *et al.* Validation of an optimized SPM procedure for

FDG-PET in dementia diagnosis in a clinical setting. NeuroImage Clin 6: 445-454 (2014).

[8] Duin R. Classifiers in almost empty spaces. In: 15th International Conference on Pattern Recognition, 2000. Proceedingss. Vol. 2; pp. 1-7 (2006)

[9] Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in neuroimaging. Neuroinformatics Vol. 2: 12(2): 229-244 (2014).

[10] Padilla P, Górriz JM, Ramírez J, Lang EW, Chaves R, Segovia F, *et al.* Analysis of SPECT brain images for the diagnosis of Alzheimer's disease based on NMF for feature extraction. Neurosci Lett 479(3): 192-196 (2010).

[11] Segovia F, Górriz JM, Ramírez J, Salas-Gonzalez D, Álvarez I, López M, *et al.* A comparative study of feature extraction methods for the diagnosis of Alzheimer's disease using the ADNI database. Neurocomputing 75(1): 64-71 (2012).

[12] Segovia F, Górriz JM, Ramírez J, Salas-Gonzalez D, Álvarez I, López M, *et al.* Classification of functional brain images using a GMM-based multi-variate approach. Neurosci Lett 474(1): 58-62 (2010).

[13] Jolliffe IT. Principal Component Analysis. New York: Springer 2002.

[14] Lee DD, Seung HS. Algorithms for Non-negative Matrix Factorization. In: Leen TK, Dietterich TG, Tresp V, Eds. Advances in Neural Information Processing Systems 13. USA: MIT Press pp. 556-562 (2001).

[15] Haralick R, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Transact Syst Man Cybernetics SMC-3(6): 610-621 (1973).

[16] López M, Ramírez J, Górriz J, Salas-Gonzalez D, Álvarez I, Segovia F, *et al.* Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules. Electron Lett 45(8): 389-391 (2009).

[17] Freeborough P, Fox N. MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease. IEEE Transact Med Imaging 17(3): 475-478 (1998).

[18] Liu S, Cai W, Wen L, Feng DD, Pujol S, Kikinis R, *et al.* Multi-Channel neurodegenerative pattern analysis and its application in Alzheimer's disease characterization. Comput Med Imaging Graphics 38(6): 436–444 (2014).

[19] Dukart J, Mueller K, Barthel H, Villringer A, Sabri O, Schroeter ML. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. Psychiatr Res: Neuroimaging 212(3): 230-236 (2013).

[20] Liu M, Zhang D, Shen D. Ensemble sparse classification of Alzheimer's disease. NeuroImage 60(2): 1106-1116 (2012).

[21] Dukart J, Mueller K, Horstmann A, Vogt B, Frisch S, Barthel H, *et al.* Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDGPET studies. NeuroImage 49(2): 1490-1495 (2010).

[22] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject. Brain NeuroImage 15(1): 273-289 (2002).

[23] Dukart J, Perneczky R, Frster S, Barthel H, Diehl-Schmid J, Draganski B, *et al.* Reference Cluster Normalization Improves Detection of Frontotemporal Lobar Degeneration by Means of FDG-PET. PLoS One 8(2): e55415 (2013).

[24] Turk M, Pentland A. Eigenfaces for Recognition. J Cogn Neurosci 3(1): 71-86 (1991).

[25] Webb AR. Statistical Pattern Recognition, 2nd Ed. USA: Wiley 2002.

[26] Vapnik VN. Statistical Learning Theory. New York: Wiley-Interscience, 1998.

[27] Gönen M, Alpaydın E. Multiple Kernel Learning Algorithms. J Mach Learn Res 12: 2211-2268 (2011).

[28] Ding B, Ling HW, Zhang Y, Huang J, Zhang H, Wang T, *et al.* Pattern of cerebral hyperperfusion in Alzheimer's disease and amnestic mild cognitive impairment using voxel-based analysis of 3D arterial spin-labeling imaging: initial experience. Clin Interven Aging 9: 493-500 (2014).

[29] Vidoni ED, Thomas GP, Honea RA, Loskutova N, Burns JM. Evidence of altered corticomotor system connectivity in early-stage Alzheimer's disease. J Neurol Phys Ther: JNPT 36(1): 8-16 (2012).

[30] Huang C, Wahlund L, Svensson L, Winblad B, Julin P. Cingulate cortex hypoperfusion predicts Alzheimer's disease in mild cognitive impairment. BMC Neurol 2(1): 9 (2002).

[31] Valla J, Berndt JD, Gonzalez-Lima F. Energy hypometabolism in posterior cingulate cortex of Alzheimer's patients: superficial laminar cytochrome oxidase associated with disease duration. J Neurosci 21(13): 4923-4930 (2001).