World Scientific
www.worldscientific.com

Empirical Functional PCA for 3D Image Feature Extraction Through Fractal Sampling

Andrés Ortiz* and Jorge Munilla†

Communications Engineering Department, University of Málaga
Málaga 29071, Spain
*aortiz@ic.uma.es
†aortiz@ic.uma.es

Francisco J. Martínez-Murcia‡, Juan M. Górriz§ and Javier Ramírez¶

Department of Signal Theory, Communications and Networking
University of Granada, Granada 18060, Spain
‡fjesusmartinez@ugr.es
§gorriz@ugr.es
¶javierrp@ugr.es

Medical image classification is currently a challenging task that can be used to aid the diagnosis of different brain diseases. Thus, exploratory and discriminative analysis techniques aiming to obtain representative features from the images play a decisive role in the design of effective Computer Aided Diagnosis (CAD) systems, which is especially important in the early diagnosis of dementia. In this work, we present a technique that allows using specific time series analysis techniques with 3D images. This is achieved by sampling the image using a fractal-based method which preserves the spatial relationship among voxels. In addition, a method called Empirical functional PCA (EfPCA) is presented, which combines Empirical Mode Decomposition (EMD) with functional PCA to express an image in the space spanned by a basis of empirical functions, instead of using components computed by a predefined basis as in Fourier or Wavelet analysis. The devised technique has been used to classify images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Parkinson Progression Markers Initiative (PPMI), achieving accuracies up to 93% and 92% differential diagnosis tasks (AD versus controls and PD versus Controls, respectively). The results obtained validate the method, proving that the information retrieved by our methodology is significantly linked to the diseases.

Keywords: List Hilbert curve; EEMD; empirical functional PCA; SVM; PET; Alzheimer's disease; Parkinson disease.

1. Introduction

Computer-based medical image analysis methods are currently attracting considerable research attention, as they usually determine the performance of the Computer Aided Diagnosis (CAD) tools. These techniques are especially important in neuroimaging techniques for the diagnosis of dementia, as three-dimensional and high-resolution images are often available. This is the case of neuroimaging modalities for the diagnosis of neurodegenerative diseases, which affect 50 Million people worldwide.[1] Moreover, it is expected that 82 Million people will suffer any type of dementia in 2030 increasing up to 152 Million people in 2050[2] according to the World Health Organization. Alzheimer's disease (AD) and Parkinson Disease (PD) are the two most common types of dementia. In the case of AD, it accounts for 60–70% of the cases.

Different image modalities are used in the *in vivo* diagnosis of these dementias. These include structural Magnetic Resonance Imaging (MRI) that can be used to assess the progression of the AD from the early stages of the disease,[2,3] as it provides a visual way to evaluate the brain artrophy that is a typical feature of neurodegeneration as a consequence of dendritic and neuronal losses. Thus, depending on the stage of the disease, the structure of different brain regions are progressively affected: in the early stages, neurodegeneration starts in the medial temporal lobe. Then, enthorhinal cortex in the hippocampus and the limbic system are progressively affected, and eventually, neocortical areas. This way, the typical study to assess the progression of AD is focused on the hippocampus, the enthorhinal cortex and the amygdala. However, cerebral atrophy is a result of neuronal damage but not entirely specific of AD.[2] On the other hand, studies using functional neuroimaging provide another source of information to prove the functional integrity of brain.[2] Single Photon Emission Computed Tomography (SPECT) or Positron Emission Tomography using the 18F-fluorodeoxyglucose (FDG) radiotracer (18F-FDG PET) are being intensively used[4–7] for the diagnosis of AD. Since glucose is the primary source of energy for the brain, FDG, a glucose-based radiotracer, can be used to measure metabolic activity at a specific region, as this consumption produces an emission that can be detected by a PET gammacamera. In a similar way, SPECT imaging using[123] I-ioflupane (DaTSCAN) radiotracer is widely used for the diagnosis of PD, as it provides a way to measure the amount of dopaminergic transporters (DaT) in the striatal region which is linked to the disease.[8,9] Alternatively, there are other methods such as Electroencephalography (EEG)[10–14] or Magneto-Encephalography (MEG)[15] to capture relevant information to evaluate the functional state of the brain, providing useful information of the disease progression, especially in the early stages. These methods can also be used to assess the disease by reconstructing the brain network, indicating brain regions that are coupled in the development of specific tasks[16–22] or during resting state.[23] The analysis of these images not only can help physicians to obtain helpful information related to the neurodegeneration that occurs, but also to the early diagnosis of the disease and the improvement in the accuracy

of the diagnosis.[8,24] However, the large amount of data provided by these images makes it necessary to develop specific processing techniques. In fact, feature extraction in medical image processing still remains a challenge since, as with other real-world data, the expected number of available samples is considerably lower than the dimension of the feature space. Thus, the development of effective techniques to reduce the number of features while preserving the information plays a decisive role, as they avoid the use of raw data (e.g. VAF technique[25]) sidestepping the *curse of Dimensionality* problem.[26] Hitherto, two main alternatives, although complementary, have been generally employed to reduce the dimensionality of the feature space. The first consists in selecting the most discriminative features which can be addressed by filtering or wrapper techniques.[27–30] The second lies in the computation of a reduced set of new features from the raw data (i.e. the original feature space), obtaining a new, low-dimensional feature space.[31–35]

These techniques have been previously used in functional 18F-FDG PET imaging[36–39] to build CAD systems. This is, however, a difficult task since structural and functional changes in the early stages of AD are similar to those that appear as a consequence of the natural aging process.

There exist different approaches to extract relevant information for predictive diagnosis in different image modalities. Thus, in Ref. 40 and similarly in Ref. 41 for MRI images, Total Variation (TV) as a regularization and structured sparsity is used for classifying fMRI images. Moreover, a review of machine learning-based methods for feature selection in MRI is presented in Ref. 42, where classification experiments using images from the ADNI database are shown to compare the methods exposed. Other approaches[43] improve the Support Vector Machine (SVM) classifier by introducing spatial and anatomical *a priori* information in SVM and Ref. 44 presents a robust framework for feature selection in MRI that uses different types of priors introduced in Support Vector Classifiers (SVCs), which is assessed using a database composed of 137 patients with AD and 162 elderly controls.

Other recent works propose the use of fractals to compute discriminative features from images or EEG signals. For instance, in Refs. 45 and 46, the fractal dimension is used to study the complexity

and dynamical changes in autistic spectrum disorder in the brain and Ref. 47 shows the application to investigate the frontal brain of major depressive disorder. Other EEG-based applications such as Brain Computer Interfaces (BCI) have also used fractals to extract features for motor imagery tasks classification.

In this paper, we propose a method based on EMD decomposition to extract relevant features. However, instead of using multidimensional EMD, we convert the images into a sequence of voxels by sampling the 3D images using space-filling fractal curves which preserve the voxel neighborhood to avoid the loss of spatial information. As a result, a sequence of voxel intensities is obtained for each image, and then, temporal and spectral features are computed from the components obtained by Empirical Mode Decomposition (EMD). Space-filling curves have been employed in Ref. 48 to avoid using the bidimensional extension of EMD, which is computationally expensive, although in that work, pixel neighbohood preservation was not an essential requirement. Unlike Fourier or Wavelet-based methods, which use a predefined basis to present the original signal (e.g. in the case of Fourier analysis, signals are expresses as a combination of sine/cosine functions), we use EMD to decompose the original signal into a number of components named Intrinsic Mode Functions (IMF). These IMF components are empirical, which allows reconstructing precisely the original signal. In a previous work,[49] we used the specific extension of EMD to bidimensional signals to extract EMD components directly to PPMI images which requires applying the 2D EMD method iteratively to some relevant slices of the image. This method, besides being computationally less efficient, does not use all the spatial information contained in the image, since SPECT images are 3D by nature. Our proposal uses a way to decompose 3D images using 1D EMD without losing spatial information (i.e. exploiting the three-dimensional neighborhood properties). Hence, it is not necessary to select only the most relevant slices but features from all the 3D structure containing the striatum are processed, contributing to improve the classification performance.

Thus, in this work we present two main contributions. The first is the use of multidimensional space-filling curves to sample a multidimensional signal converting it to a one-dimensional time varying signal. Specifically, we used Peano–Hilbert curves to fill the $\mathbb{R}^d$ space, since others methods, such as, for example, the simple rearranging of the array components by columns or by rows would not have into account the underlying neighborhood properties of the signals being sampled. This is the main motivation for sampling by means of space-filling curves, as in medical imaging is well known that relevant information is not only in the pixel/voxel individual values but also in the relationships of these with their vicinity. In fact, second order features such as texture features are based on this.[33,37,38,50–53] Moreover, this is one of the key points in some Deep Learning applications[54] such as Convolutional Neural Networks,[55] where features are hierarchically extracted by convolving the images with a kernel learned by backpropagation. The second contribution is the implementation of functional Principal Component Analysis (fPCA)[56] by means of EMD components. fPCA is a PCA-based technique specifically developed to extract high variance components from time-varying signals. However, fPCA is normally implemented by applying classical PCA on the signal represented on a basis of smooth functions, typically obtained by means of Fourier or Wavelet analysis.[57–60] Due to the limitation of using Fourier analysis or other methods that use a predefined set of basis functions, we propose a new technique called EfPCA, empirical functional PCA. It represents the original signal in a basis composed of empirical components, which are the IMF computed by EMD. This requires some extra algebra to the procedure in order to find the best representation of the signal. In our proposal, Basis-Pursuit is used to this end, which allows finding the best representation of the signal in an over-complete basis, in terms of minimum ℓ_2 error. The proposed method has been applied to 18F-FDG PET and DaTSCAN SPECT image classification, from the ADNI and PPMI databases respectively, to demonstrate that features extracted are representative enough and provide discriminative information to a SVC. In the case of PET images from the ADNI database, they were split into regions according to the Automated Anatomical Labeling (AAL) atlas[61] to reduce the computational complexity, as it allows processing each region independently.

The rest of the paper of the paper is organized as follows. First, in Sec. 2, the methodology is presented. This section describes the different stages in our proposal: we first describe the sampling method by means of fractal curves and then the implementation of fPCA using empirical components. The pursuit-based method to represent the original signal as a combination of IMFs is also described. Afterwards, in Sec. 3, the described method is applied to extract representative features from 18F-FDG PET images of the ADNI database and from DaTSCAN SPECT images from the PPMI database to show the applicability of the proposed technique to the diagnosis of neurodegenerative disorders. Details regarding the databases are provided along with an analysis of the capabilities of the proposed method including, among others, the classification results when feeding SVC with the features computed by the proposed method. Moreover, the discussion of the results is provided in Sec. 4 and finally, the conclusions of this work are drawn in Sec. 5.

2. Methodology

The method devised in this work describes the way for processing multidimensional signals using 1D signal analysis tools, providing the arena to explore patterns with non-specific image (i.e. 2D or 3D) techniques. The fractal sampling method used, enable us to convert any multidimensional signal to 1D signal, while spatial information is preserved. In this work, the method has been applied to 3D medical images, but it can be extended to $\mathbb{R}^D$ since the space-filling curves can be extended to any dimension. The limit on the dimensionality is imposed by the computing requirements for the generation of the curve. Thus, when extending to $D > 3$, parallel techniques are advisable to compute the curves in a reasonable time.[62] The idea of using space-filling curves to sample multidimensional signals allows applying typical 1D multivariate techniques such as fPCA[56] which aims to find functional components in time varying data. Roughly, fPCA can be seen as the application of classical PCA over a smooth version of the signal. In practice, this can be addressed by representing the original signal in a basis of smooth functions and then applying PCA to the coefficients, which are coordinates of the original signal in that basis. In fact, it is frequent to use Fourier or Wavelet-based components to this end. In this work, we use EMD instead, as it uses an empirical basis rather than a predefined set of functions (as in Fourier or Wavelet analysis). The computation of the corresponding coefficients is accomplished by Basis Pursuit (BP), providing a sparse representation of a signal in the basis composed of a large set of EMD components. The overall method is shown in Fig. 1, which is applied independently for each brain region, obtaining a per-region set of features that are eventually classified by an SVM. Finally, the predictions obtained for each brain region are combined by a majority voting mechanism. The different stages
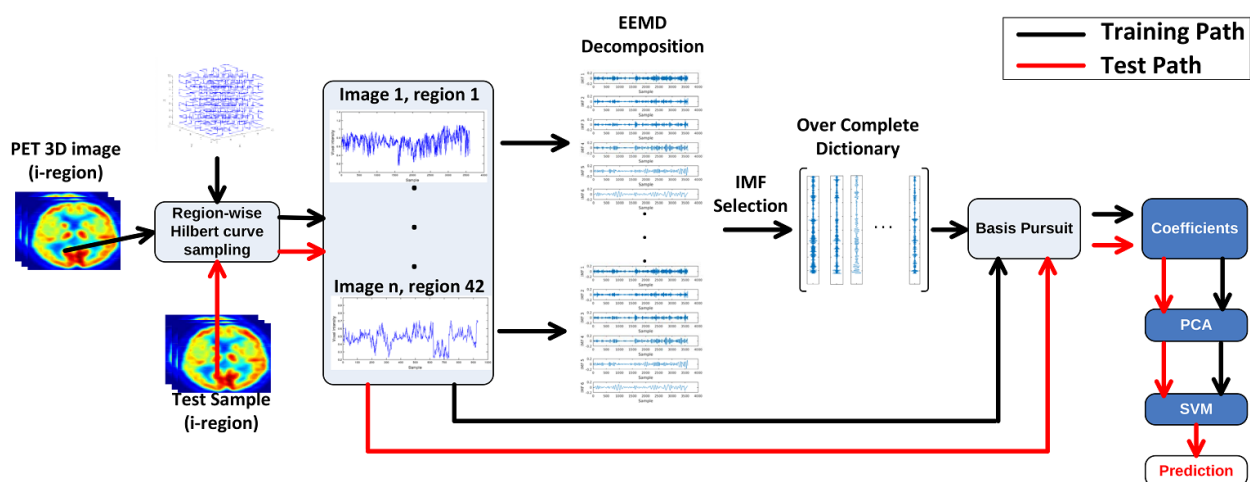

Fig. 1.   Block diagram of the overall method for 18F-FDG PET image classification (ADNI database). In the first part, fractal sampling is used to transform each 3D region (According to AAL atlas) into 1D signal. In the second part, functional PCA is implemented by EMD decomposition.
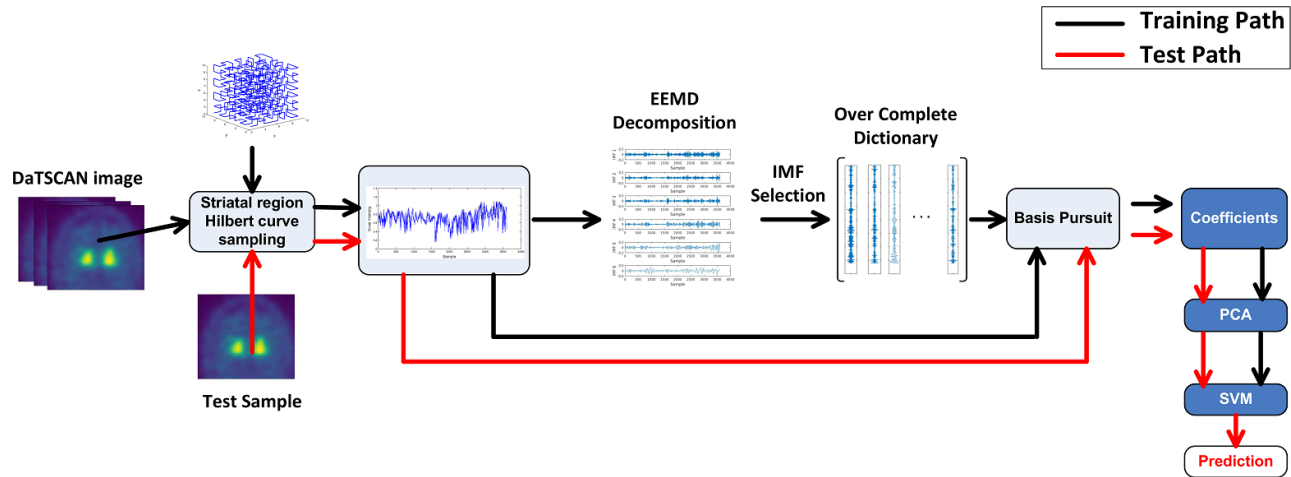
Fig. 2.   Block diagram of the overall method for DaTSCAN image classification (PPMI database). In the first part, fractal sampling is used to transform the striatal 3D region into 1D signal. In the second part, functional PCA is implemented by EMD decomposition.

shown in Fig. 1 are described in detail in the following subsections.

2.1.   *Fractal sampling using 3D homogeneous Peano–Hilbert curves numbering and spacing*

A Peano–Hilbert curve is a continuous fractal space-filling curve geometrically described by Hilbert[63] as a variant of the Peano's curve.[64] It can be defined as a continuous function whose domain is the unit interval [0,1] and its range is in a 2D Euclidean space, formally: $f : \mathbb{R} \rightarrow \mathbb{R}^d$. Thus, for any point t on the unit line segment [0,1], this function assigns the corresponding point (x, y) in the unit square [0, 1].[2] However, the range of the Hilbert curves can be extended to d-dimensions. The $\mathbb{R} \rightarrow \mathbb{R}^d$ mapping provided by Hilbert curves have the following properties[65] (for simplicity, $d = 2$):

- Continuity is preserved: values close in the [0, 1] line have similar values in the $[0, 1]^2$ unit square (adjacency condition).
- The mapping is quasi-invertible: the construction of Hilbert curves tends to correspond similar values of (x, y) coordinates to similar t values. This is particularly important when there is neighborhood information.
- The curve is uniquely defined by fixing the mapping of the initial and final subintervals, as well as a rotation matrix.

- They can be generated by the iterative application of affine transformations to a starting mapping, and can be implemented by recursive algorithms.

Figure 3 shows an example of Hilbert curves in 2D and 3D.

2.2.   *Empirical functional component analysis*

Functional Component Analysis (fPCA) is a statistical method that aims to find principal components in functional data (i.e. time series).[56] Thus, as in classical PCA the target is to obtain the eigen-time series[66] (being of the same length of the original time series) representing the most important directions of variation of the signal. A usual way of implementing fPCA consists in approximating each original time series of dimension d with k basis functions. Thus, the representation of the original signal in that basis reduces the dimension from d to k.

Formally, a signal $x(t)$ can be expanded in terms of a pre-defined basis ϕ as:

$$x(t) = \sum_{i=1}^{k} \phi(t)c_i(t) = \phi\mathbf{c}, \qquad (1)$$

where $\phi(\mathbf{t})$ is a basis for $\mathbf{x}$. Due to stability reasons, using a Fourier basis is a common choice in signal processing applications, consisting of *sine* and *cosine* functions of increasing frequency: $\{\sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \dots, \sin(m\omega t), \cos(m\omega t)\}$.
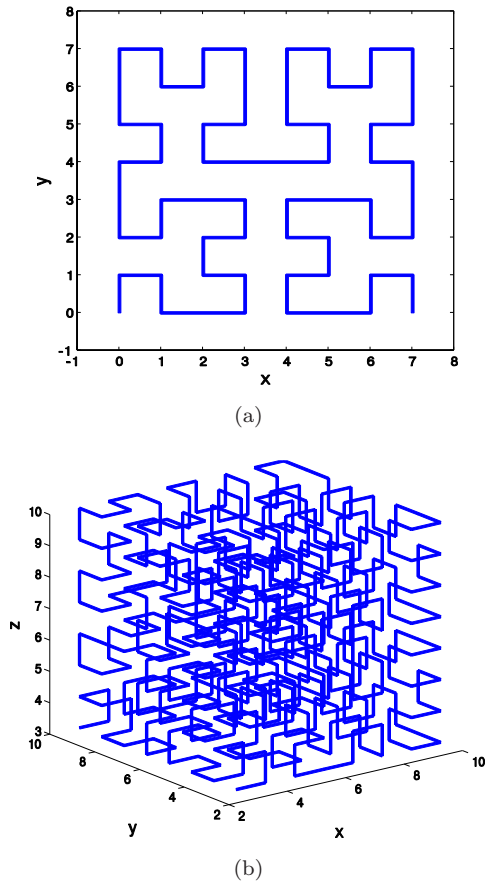
A. Ortiz et al.

Fig. 3. Example of 2D (a) and 3D (b) Hilbert curves.

This represents an alternative to polynomial functions that are commonly used to describe natural data. Nevertheless, due to the periodic properties of the basis functions, the representations obtained by the Fourier basis are periodic. Consequently, the resulting representations may not be accurate if the original signal is not periodic. Other approaches also use pre-defined functions to represent the original signal, such as Wavelet Decomposition which depends on the mother wavelet function chosen, and the number of computed levels.

In this work, we use a different approach that does not use any pre-defined basis to decompose the original signal but an empirical decomposition by means of EMD.[67] Once the original signals are expressed as coordinates in the subspace spanned by the components (ϕ), PCA is applied in the classical way on that subspace. The implemented process, which can be considered as a particular implementation of the functional PCA, can be described in a more formal way as follows. Let $\mathbf{c}$ be the coordinates

of $\mathbf{x}$ in the subspace spanned by the basis ϕ, we implement fPCA as follows:

(i) Decompose each data sample $\mathbf{x}$ into a set of basis signals ϕ.

(ii) Obtain the representation $\mathbf{c}$ of $\mathbf{x}$ in the subspace spanned by the basis ϕ.

(iii) Then apply the standard PCA method to $\mathbf{c}$:

   (a) Subtract the mean from each sample $\mathbf{y} = \mathbf{c} - \bar{\mathbf{c}}$.

   (b) Compute the population covariance matrix $\mathbf{Q}$ of $\mathbf{y}$: $\mathbf{Q} = \mathbf{y}\mathbf{y^T}/(n-1)$, where n is the number of samples.

   (c) Compute for the eigen time-series their corresponding eigenvalues of $\mathbf{Q}$ by means of SVD.

   (d) Since ($\mathbf{Q}$) is diagonal: $\mathbf{Q} = \mathbf{USV}^\top \cdot \mathbf{V}$ is the eigenvectors matrix and $\lambda_i = s_i^2/(n-1)$ is the eigenvalue that account for the variance explained by the eigen time-series $\mathbf{v_i}$.

   (e) Compute the projection of new data on the space spanned by the k first eigen time-series (sorted by decreasing eigenvalue).

   (f) Compute $\mathbf{P} = \mathbf{U_k S_k}$, where $\mathbf{U_k}$ and $\mathbf{S_k}$ represent the k first columns of U, and the $k \times k$ upper-left part of S, respectively ($\mathbf{P}$ is a $n \times k$ matrix containing k eigen time-series of length n).

Since all of these signals form, in general, a nonorthogonal and nonlinearly independent dictionary, the representation of each signal in the subspace spanned by the basis is obtained by means of *Basis Pursuit*, explained later in Sec. 2.4.

As in standard PCA, the variance explained by the principal components determines the most important variation directions of that data. Hence, the number of components used in the projection depends on the threshold for the variance explained and vice-versa. In this work, we carried out experiments for different values of the variance explained, in order to compare the performance obtained by the classical PCA approach with that provided by the proposed EfPCA.

2.3. *Empirical mode decomposition*

EMD is a method to decompose any complicated signal into a sum of finite number of oscillatory components, named IMFs. By definition, an IMF is

any *function with the same number of extrema and zero crossings, whose envelopes are symmetric with respect to zero.*[67] EMD is a highly adaptive decomposition method that works in time domain, since the decomposition is based on the local characteristic time scale of the data. Although there are many ways to decompose a signal into components (e.g. Fourier analysis or Wavelet decomposition), EMD does not make any assumption about the stationarity or linearity of the data and stays in the time domain. Moreover, the decomposition performed by EMD implies *completeness*; that is, the original signal can be exactly recovered by summing up the components. The basic idea behind EMD decomposition is to consider a signal $x(t)$ as a superposition of high $d_i(t)$ and low $r_N(t)$ frequency oscillations. Thus, the method, called *sifting*,[67] iterates on the low oscillations component considered as a new signal to be decomposed:

$$x(t) = \sum_{i=1}^{N-1} d_i(t) + r_N(t), \qquad (2)$$

where d_i are the IMFs and r_N is the residual signal that represents the overall trend. The sifting process is repeated until the stopping condition, consisting in reaching a threshold in the Standard Deviation (SD) over two consecutive sifting results.

Based on the EMD method, the Ensemble EMD approach (EEMD),[68] is an improved version that makes EMD more robust to noisy signals. The core idea of EEMD is to add white noise to the original signal, composing a number n of trials:

$$x_i(t) = x(t) + w_i(t), \quad i = \{1, \ldots, n\}. \qquad (3)$$

Then, EMD decomposition is applied to $x_i(t)$, obtaining a set of n *noisy* IMFs. Finally, the (ensemble) means of the corresponding *noisy* IMFs of the decompositions are computed and used as final IMFs. The use of EEMD in this work aims to deal with the intra-class inherent variability in actual 18F-FDG PET image data.

2.3.1. *Empirical mode decomposition*

When EMD decomposition is applied, it is usual, particularly when noisy signals are processed,[69] to find IMFs that present a low correlation with the original signal. On the other hand, to reduce the computational complexity, it is convenient to reduce the

number of IMFs for the construction of the dictionary to be used in BP. Thus, only the most relevant IMFs regarding their correlation to the original signal have been selected to reduce the computation time associated to the calculation of the coefficients by means of BP. In particular, we compute the correlation between each IMF and the original signal from which these IMFs were extracted, and preserve only those IMFs in which correlation is above a threshold. For the sake of clarity, let us define a correlation threshold λ. Assuming the least relevant components have relatively poor correlation with the original signal, (i.e. 10 times lower), we can define the threshold:

$$\lambda = \frac{\max(\rho_i)}{10}, \quad i = \{1, \ldots, N\}, \qquad (4)$$

where $max(\rho_i)$ is the maximum Pearson correlation coefficient ρ_i observed, computed as

$$\rho_i = \frac{Cov\{d_i, x\}}{\sigma_{d_i} \sigma_x}, \qquad (5)$$

where σ_{d_i} and σ_x are the SDs of the ith IMF and the original signal, respectively. Consequently, only the IMFs that fulfill the correlation threshold are kept to be included in the over-complete dictionary.

2.4. *Basis pursuit*

After decomposing the original data set into a number of components (in our case, IMFs), it is necessary to express them as coordinates in the subspace spanned by the IMFs. In other words, it is necessary to find the best linear combination of IMFs that effectively reconstruct the original signal. Under the assumption that similar classes may share similar IMF components, we define a basis composed by the IMFs extracted from the training images and use it to express an image.

However, the basis arranged by the IMFs components is non-orthogonal and overcomplete, and requires specific method to compute the coefficients. BP[70] deals with this problem. Assuming that $\boldsymbol{\Psi}$ is a dictionary of functions in a Hilbert space, BP is intended to solve the optimization problem defined as:

$$\min_x \|\mathbf{x}\|_1 \quad \text{s.t. } \mathbf{y} = \boldsymbol{\Psi}\mathbf{x}, \qquad (6)$$

where $\mathbf{x}$ is a $N \times 1$ solution vector (coefficients), $\mathbf{y}$ is a $M \times 1$ vector (signal), $\boldsymbol{\Psi}$ is a $M \times N$ over-complete basis matrix $M < N$ (see Fig. 4). Thus, BP
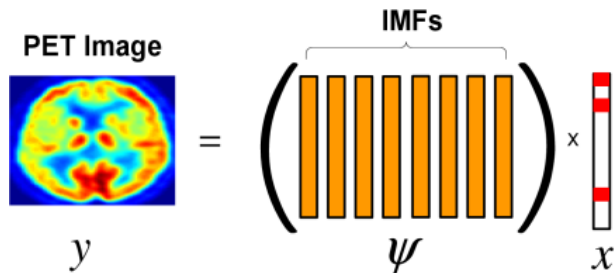
A. Ortiz et al.

Fig. 4.   BP method for sparse representation. In the Figure, a 18F-FDG PET image (y) is represented as a sparse linear combination (x) of IMFs extracted from the images composing the training set (Ψ).

allows obtaining the sparsest solution $\mathbf{x}$ in terms of ℓ_1-norm. In other words, it provides the coefficients for a sparse linear combination of the elements in $\boldsymbol{\Psi}$ that best reconstructs the signal $\mathbf{y}$.

BP can be extended for dealing with noisy data. Thus, Chen and Donoho[70] proposed to obtain an approximate decomposition of $\mathbf{y}$:

$$\mathbf{y} = \boldsymbol{\Psi}\mathbf{x} + \mathbf{r}, \tag{7}$$

where $\mathbf{r}$ is a residual term that accounts for the difference between the original signal and the reconstructed by means of the sparse representation driven by $\mathbf{x}$. This way, dealing with a noise level $\sigma > 0$ can be addressed by solving

$$\min_{x} \|\boldsymbol{\Psi}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda_n \|\mathbf{x}\|_1, \tag{8}$$

which is known as Denoising BP, with $\lambda_n = \sqrt{2 \log \#D}$ vectors in the dictionary $\boldsymbol{\Psi}$.

3.   Experimental Results

In the following subsections, the proposed methodology is assessed in detail using 18F-FDG PET images from the Alzheimer's disease Neuroimaging Initiative (ADNI) database. Then experiments using images ^{123}I-Ioflupane (DatSCAN) SPECT images were also conducted (described in Sec. 3.2) to show the applicability of the proposal in the diagnosis of Parkinsonian Syndromes.

3.1.   *Experimental results using the ADNI database*

3.1.1.   *Database description*

The database used in this work contains multimodal 18F-FDG PET image data from 138 subjects, comprising 68 Controls (CN) and 70 AD patients from

Table 1.   Demographic data of patients in the ADNI database used in the experiments.

Diagnosis	Age	Gender #M/#F	MMSE
Control	75.81 ± 4.93	$43/25$	29.06 ± 1.08
AD	75.33 ± 7.17	$46/24$	22.84 ± 2.91

the ADNI database.[71] This repository was created to study the advance of the AD, collecting a vast amount of MRI and PET images as well as blood biomarkers and cerebrospinal fluid analyses. The main goal of this database is to provide a way to the early diagnosis of the AD. Patient's demographics are shown in Table 1. 18F-FDG PET data have been used in this work.

3.1.2.   18*F-FDG PET image preprocessing*

AQ: Please check edit.

18F-FDG PET images from the ADNI database were normalized through a general affine model, with 12 parameters,[72,73] using the SPM8 software.[74] After the affine normalization, the resulting image was registered using a more complex nonrigid spatial transformation model. The nonlinear deformations to the Montreal Neurological Imaging (MNI) PET Template were parametrized by a linear combination of the lowest-frequency components of the three-dimensional cosine transform bases. A small-deformation approach was used, and regularization was applied by the bending energy of the displacement field. This process ensures that each image voxel corresponds to the same anatomical position. Thus, after image registration, all the PET images were resized to $79 \times 95 \times 68$ voxels with voxel-size of $3\,\mathrm{mm}$ (Sagittal) $\times 3\,\mathrm{mm}$ (Coronal) $\times 3\,\mathrm{mm}$ (Axial). Subsequently, PET images are also normalized in intensity in order to compute comparable levels among the images. Intensity normalization is performed by means of the mean image, which is used as a normalization template. Specifically, the normalization value applied to each image is calculated as the mean of the 1% of the voxels with a higher activation level in the template. This helps to homogenize the activation levels, using the same scale and making them comparable. Moreover, we used the 116-regions Automated Anatomical Labelling Atlas (AAL) to extract the voxels corresponding to these areas. Voxels outside the atlas-defined areas are considered as background. On the other hand, only 42

Table 2.   Names and the corresponding indexes of the regions used in the classification experiments carried out in this work. These regions are associated to the development of AD according to Ref. 75.

	Frontal lobe		Parietal lobe		Occipital lobe		Temporal lobe
1	Frontal_Sup_L	13	Parietal_Sup_L	21	Occipital_Sup_L	27	Temporal_Sup_L
2	Frontal_Sup_R	14	Parietal_Sup_R	22	Occipital_Sup_R	28	Temporal_Sup_R
3	Frontal_Med_L	15	Parietal_Inf_L	23	Occipital_Mid_L	29	Temporal_Pole_Sup_L
4	Frontal_Med_R	16	Parietal_Inf_R	24	Occipital_Mid_R	30	Temporal_Pole_Sup_R
5	Frontal_Sup_Medial_L	17	Precuneus_L	25	Occipital_Inf_L	31	Temporal_Mid_L
6	Frontal_Sup_Medial_R	18	Precuneus_R	26	Occipital_Inf_R	32	Temporal_Mid_R
7	Frontal_Mid_Orb_L	19	Cingulum_Pos_L			33	Temporal_Pole_Mid_L
8	Frontal_Mid_Orb_R	20	Cingulum_Pos_R			34	Temporal_Pole_Mid_R
10	Rectus_L					35	Temporal_Inf_L 8301
11	Rectus_R					36	Temporal_inf_R 8302
12	Cingulum_Ant_L					37	Fusiform_L
	Cingulum_Ant_R					38	Fusiform_R
						39	Hippocampus_L
						40	Hipocampus_R
						41	ParaHippocampal-L
						42	ParraHippocampal-R

regions out of the 116 included in the AAL atlas, distributed in the frontal, parietal, occipital and temporal lobes, have been selected here for brain connectivity modeling, as they are considered the most relevant for AD diagnosis.[75] The use of a reduced number (but relevant) of regions allows to shorten the computation time. These regions are detailed in Table 2.

3.1.3.   *EMD components from fractal sampled images*

As explained in previous sections, we propose the use of a sampling method based on fractal curves to transform the original 3D image into a time varying signal, in which consecutive points corresponds to neighbor voxels in the 3D space. This allows the use of 1D signal processing techniques to extract patterns from the 3D image. In our case, EMD is used to decompose the original image into a number of signals containing different frequency components. As an example, Fig. 5 shows the decomposition of the left hippocampus region (AAL atlas region 36) into 6 IMFs.

Differences between the IMFs for controls and AD patients can be visually assessed in both cases, while correlation between signals corroborates it. In particular, correlation between IMF 4 signals are 0.63 and 0.58 for Control/AD and MCI/AD, respectively, and correlation between IMF 6 signals

are 0.73 and 0.69 for Control/AD and MCI/AD groups, respectively. In addition, activation of different brain regions can be reconstructed from the IMFs to show the activation levels represented by each one. Figure 6 shows the reconstruction of the regions over a structural brain image, revealing differences in the metabolic activity levels among groups.

3.1.4.   *Classification*

In the experiments performed, 6 IMFs are extracted from each brain region using 20 stages in the EEMD method and adding 1% of noise. These IMFs are arranged by columns to compose an over-complete dictionary. Images are then expressed as a sparse linear combination of these IMFs and the coefficients are used to compute the eigenvectors by means of PCA. Subsequently, the Hilbert transform is used to compute the analytic version of each IMF and the coordinates in the complex plane are used as features. Such features are then used to train a SVM for each region. These SVM classifiers act as *weak classifiers* that are combined using the majority voting rule, as shown in Fig. 7.

3.1.5.   *Computing the score of the ensemble of classifiers*

The output score of a classifier is computed by taking into account the scores provided for each classifier and for each class independently. For the sake of
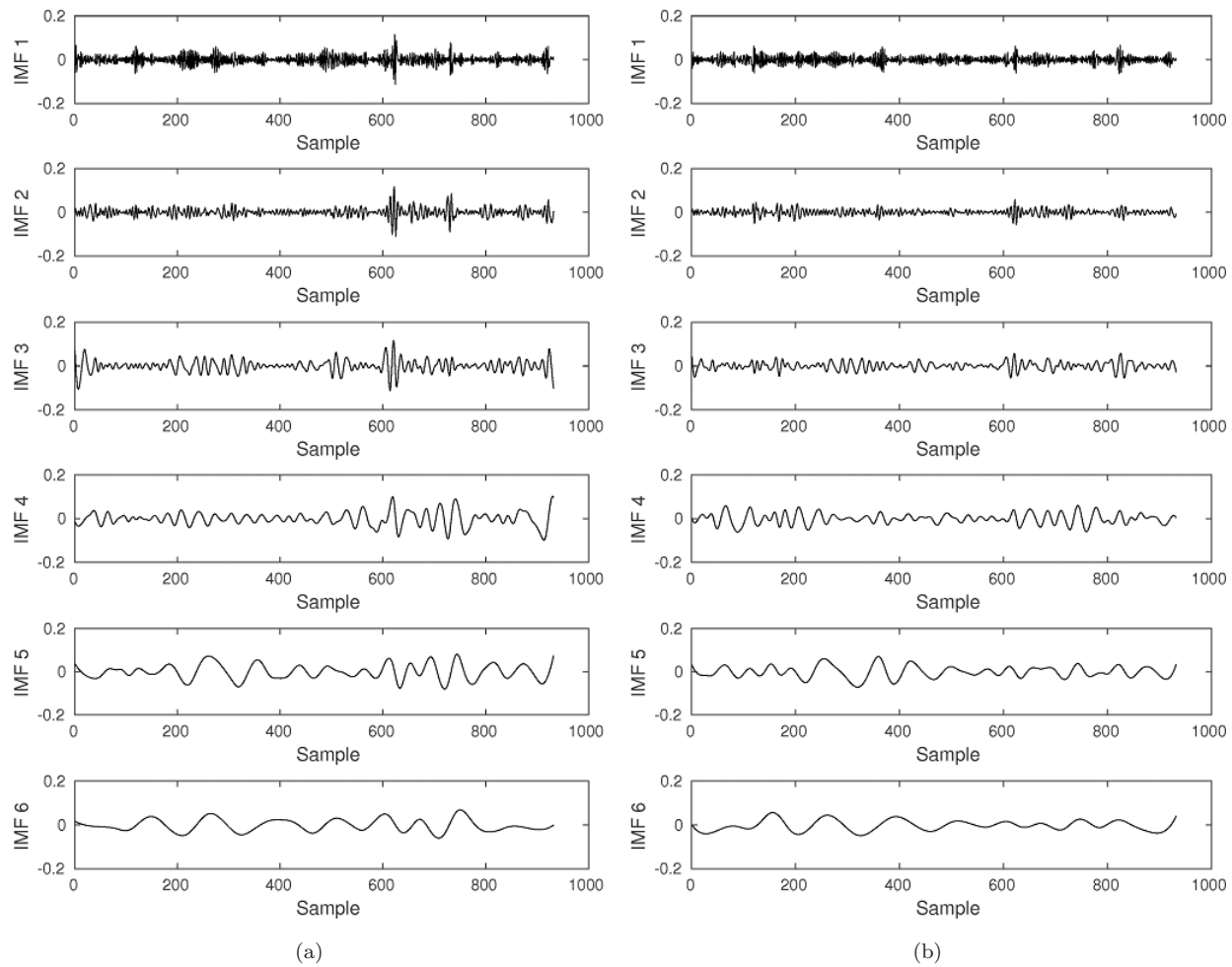
A. Ortiz et al.

(a)   (b)

Fig. 5.   Example IMFs computed for the left hippocampal region of (a) Control subject and (b) AD patient.

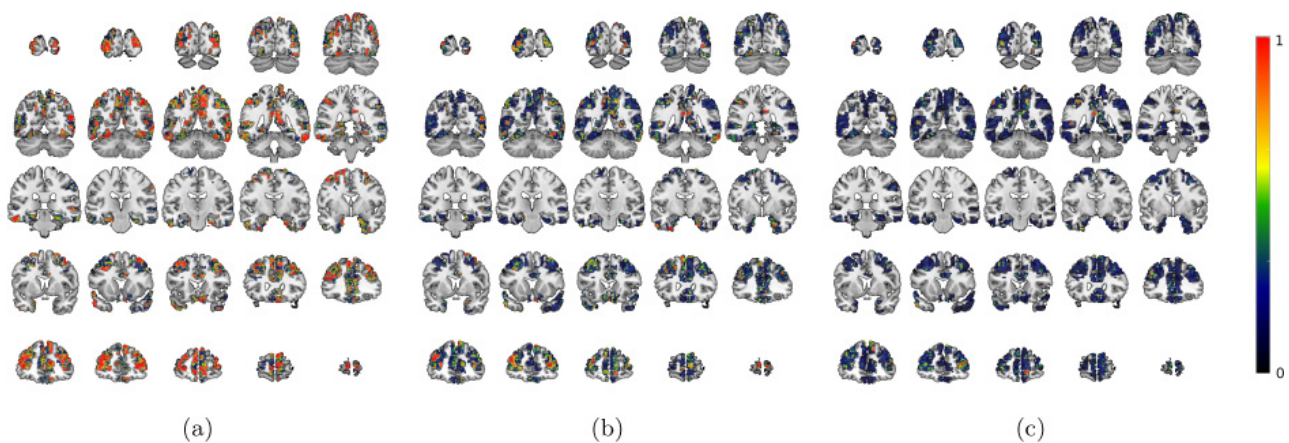(a)   (b)   (c)

Fig. 6.   Representative slices of the coronal plane for the IMF 2, selected as relevant according to Eq. (4) for the 42 brain regions indicated in Table 2 over a structural brain image. Control subject is shown in (a), MCI in (b) and AD patient in (c).
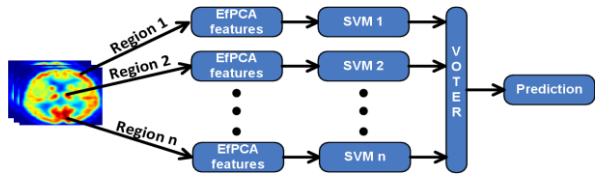
Fig. 7. Classification is addressed by an ensemble of SVMs.

Table 3. Parameter values used for Control/AD and MCI/AD classification experiments.

Experiment	#Regions	σ^2 explained	Sparsity
Control versus AD	42	0.8	80
MCI versus AD	42	0.95	150

clarity, let us define the procedure in a formal way. Let s_i be the score provided by the ith binary classifier, and let s_i^0 and s_i^1 be the score of the ith classifier obtained for the 0 and 1 class, respectively. The score S of the ensemble of n classifiers can be computed as:

$$S = \sum_i s_i^0 - \sum_j s_j^1 \quad \text{for } i, j = \{1, \ldots, n\}, \ \ i \neq j. \tag{9}$$

In the case of SVC-based classifiers, s_i^k is the distance to the hyperplane when the sample is classified as belonging to class k. This way, S is the score used to construct the ROC curve exposed in the next section.

3.1.6. *Assessment generalization capabilities of the models*

The limited number of available samples, an issue known as *small sample size problem* and common when working with biomedical data, makes it necessary to properly assess the results in order to ensure that they are database independent; that is, that the models are not overfitted and they can generalize to new samples. Thus, the use of a resampling method is necessary to assess the generated models. In this work we used resampling by *k-fold* cross-validation ($k = 10$) to estimate the prediction error.

3.1.7. *Parameter selection*

There are two main parameters to have into account in the experiments. The first one controls the threshold on the variance explained, which limits the number of eigenvectors used in the fPCA projections. The other parameter regards the sparsity; that is, the maximum number of coefficients that are not zero in the linear combination computed by BP. In the experiments, as expected, we noticed an important dependence of the performance with the variance explained. Thus, we swept the variance explained

to find the best value. On the contrary, the selection of the sparsity is not as critical as the variance explained, but helps to fine-tune the performance. The values used in Control/AD and MCI/AD classification, see Table 3, were found by experimentation with the training set (i.e. test samples were never used for tuning the parameters).

The first classification experiment using the methodology described above consists in classifying between Controls and AD patients. In order to determine the best value of the variance explained, we performed different experiments that are summarized in Fig. 8. In this figure, Accuracy (a), Sensitivity (b) and Specificity (c) values for different values of the variance explained by the functional PCA components are shown.

Taking the accuracy as a reference, we determined that its maximum value is obtained when using the number of components that explain 80% of the variance. Thus, the ROC curve shown in Fig. 9 was computed for that value, obtaining an Area under ROC Curve (AUC) of 0.95. Receiving Operating Characteristic (ROC) curve shows the trade-off between sensitivity and specificity, computed using the score of the ensemble of classifiers as previously defined in Sec. 3.1.5. The AUC indicates the probability that a positive sample does not have a more extreme value than a negative one.

Additionally, classification experiments between MCI subjects and AD were also carried out, obtaining the results summarized in Figs. 10(a)–(c), corresponding to accuracy, sensitivity and specificity, respectively, for different values of the variance explained threshold chosen in the fPCA step. Figure 11, on its side, shows the ROC curve computed using the number of components that explain 95% of the variance, which is the point where the accuracy reaches its maximum value. The corresponding AUC is 0.81.

The coefficients computed from the overcomplete basis composed of the IMFs extracted from
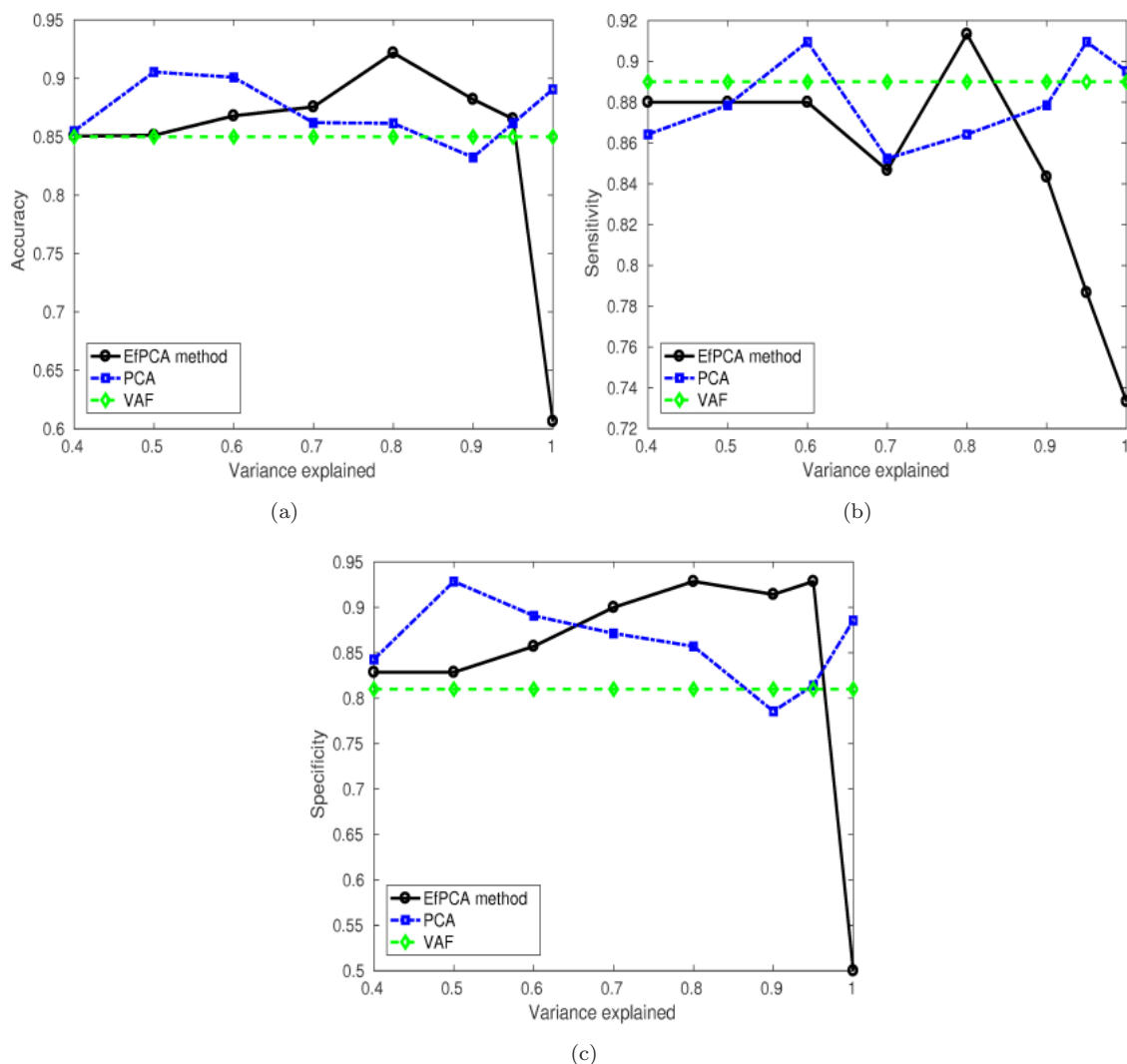
A. Ortiz et al.

Fig. 8.   Classification performance obtained in the Control/AD classification experiment (a) Accuracy, (b) Sensitivity and (c) Specificity are shown for different values of variance explained used in the fPCA stage.
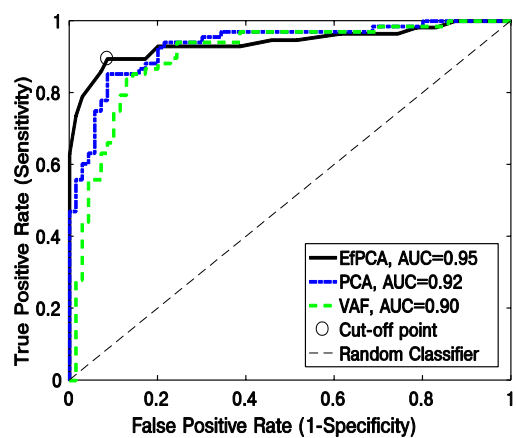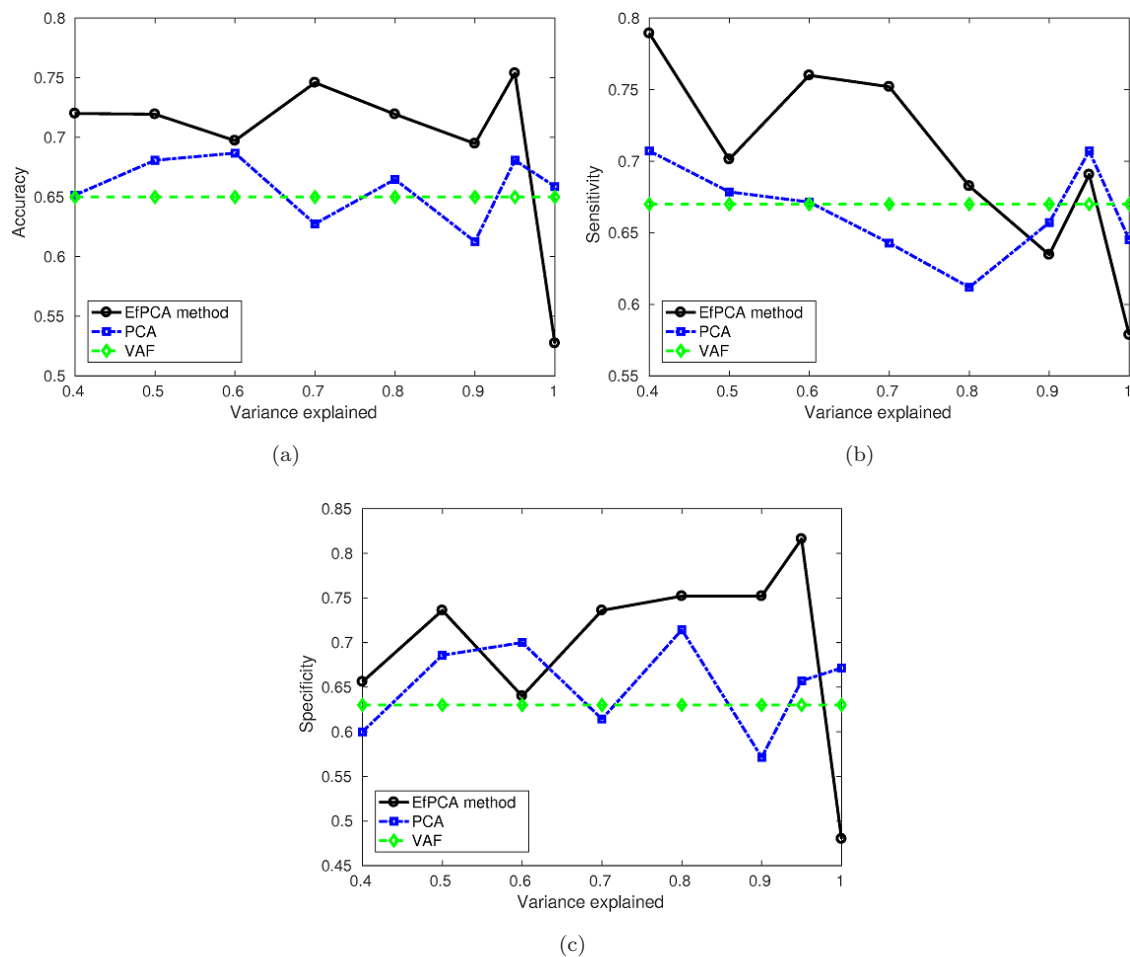

Fig. 9.   ROC curve obtained for Control/AD classification.

the training set of images can be shown graphically by reconstructing an image. Thus, the sparse representation of Control and AD images in terms of the available IMFs are shown in Fig. 12. These coefficients represent the contribution of each selected IMF to the computed features. In other words, marked areas correspond to voxels used for classification whereas the color indicates the weight of each contribution to the reconstruction of NOR and AD images.

Statistical significance of the results has been assessed by hypothesis test, to compare the accuracy values provided by the EfPCA method to the other methods shown in Table 4. This has been addressed using two-sample t-test for equal means

Fig. 10.   Classification performance obtained in the MCI/AD classification experiment (a) Accuracy, (b) Sensitivity and (c) Specificity are shown for different values of variance explained used in the fPCA stage.
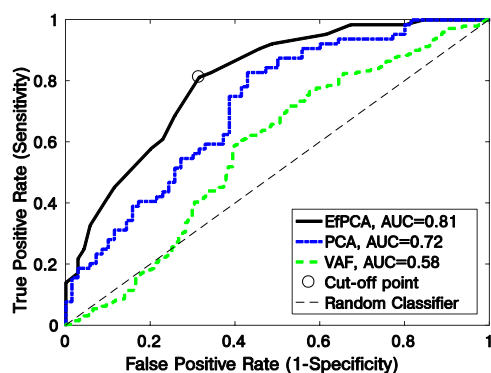

Fig. 11.   ROC curve obtained for MCI/AD classification.

and unknown variances. Hypothesis testing shows a statistical significance of 5% in the means difference for the VAF method in the case of Control versus AD classification. In those cases where p-value is above 0.05 it is not possible to ensure the mean difference. In the case of MCIs versus AD classification experiments, it is possible to assess the statistical significance of 5% in the accuracy mean differences between EfPCA, VAF, PCA and HHT methods.

3.2.   *Experimental results using DatSCAN images (*PPMI *database)*

3.2.1.   *Database description*

Data used in this section were obtained from the Parkinson's Progression Markers Initiative (PPMI) database. For up-to-date information on the study, visit *www.ppmi-info.org/data*. The images in this database were imaged $4 + 0.5\,\mathrm{h}$ after the injection of between 111 and 185 MBq of DaTSCAN. Raw projection data are acquired into a 128×128 matrix
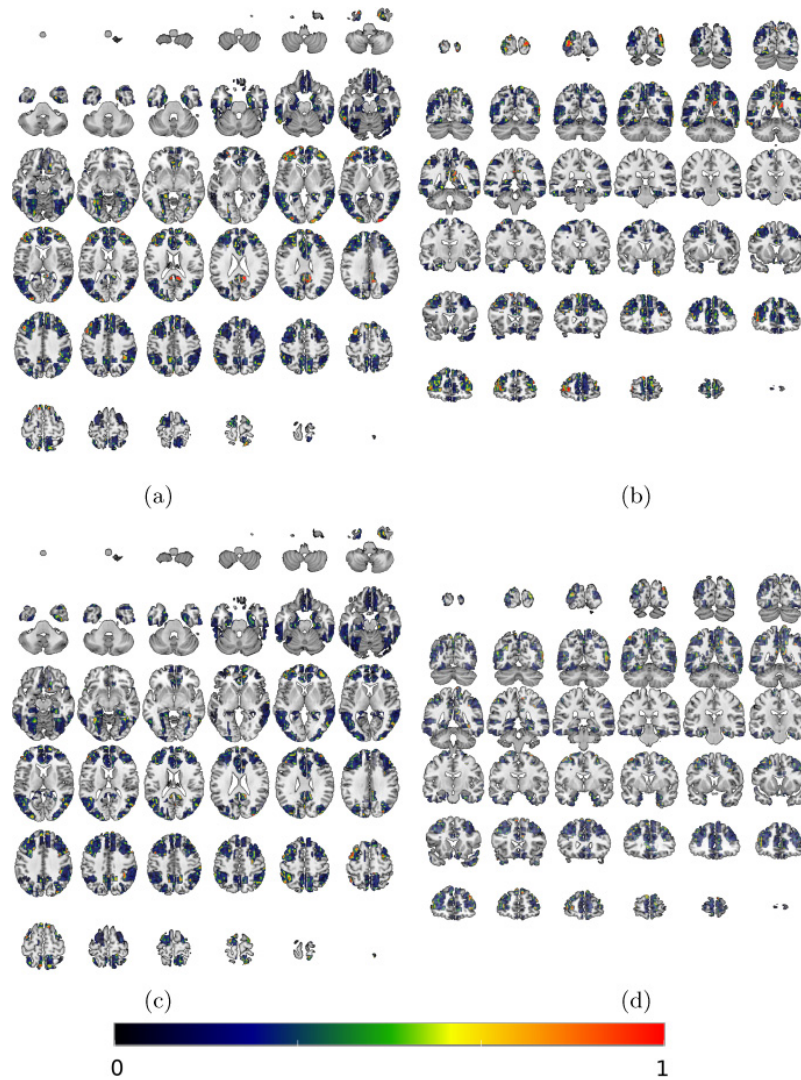
A. Ortiz et al.

Fig. 12.   Sparse representation of Control (a, b) and AD images (c, d). Axial and coronal planes shown differences in brain regions associated to AD such as the left and right hippocampus and left and right parahippocmpal regions.

Table 4.   Classification results for VAF, PCA, EMD and EfPCA methods. Mean values along with the difference between maximum and minimum values during k-fold evaluation are shown.

Method	Accuracy	Sensitivity	Specificity	AUC
Control versus AD				
VAF	0.85 ± 0.05	0.89 ± 0.13	0.81 ± 0.12	0.91
PCA	0.90 ± 0.09	0.87 ± 0.10	0.92 ± 0.10	0.93
HHT-fs	0.92 ± 0.06	0.93 ± 0.10	0.92 ± 0.11	0.95
EfPCA	$\mathbf{0.93 \pm 0.05}$	$\mathbf{0.93 \pm 0.09}$	$\mathbf{0.92 \pm 0.10}$	**0.95**
MCI versus AD				
VAF	0.65 ± 0.13	0.67 ± 0.17	0.63 ± 0.17	0.58
PCA	0.68 ± 0.09	0.70 ± 0.15	0.65 ± 0.20	0.72
HHT	0.71 ± 0.11	0.66 ± 0.13	0.75 ± 0.11	0.75
EfPCA	$\mathbf{0.75 \pm 0.09}$	$\mathbf{0.70 \pm 0.14}$	$\mathbf{0.81 \pm 0.19}$	**0.81**

Table 5. Statistical significance of the comparison between EfPCA and other methods in Table 4. Results corresponding to p-values below 0.05 are indicated with $*$.

Experiment	Method	p-value
Control versus AD	VAF	0.001$*$
Control versus AD	PCA	0.4
Control versus AD	HHT	0.7
MCI versus AD	VAF	0.001$*$
MCI versus ADTH	PCA	0.007$*$
MCI versus AD	HHT	0.008$*$

Table 6. Demographic data of patients in the PPMI database used in the experiments.

Diagnosis	Gender (#M/#F)	Age M	Age F
Control	66/45	59.68 ± 11.48	55.37 ± 10.97
PD	113/45	62.94 ± 8.70	61.20 ± 10.18

stepping each $3°$ for a total of 120 projection into two 20% symmetric photopeak windows centered on 159 KeV and 122 KeV with a total scan duration of approximately 30–45 min.[77]

A total of $N = 269$ DaTSCAN images from this database were used in the preparation of the article. Specifically, the baseline acquisition from 158 subjects suffering from PD and 111 normal controls was used. For more details on the demographics of this dataset, please check Table 6.

3.2.2. *DaTSCAN image preprocessing*

The DaTSCAN images from the PPMI dataset are roughly realigned. We will refer to this as nonnormalized (given that it is only a similarity transformation that preserves shape). We further preprocessed the images using the SPM12 new normalize procedure with default parameters, which applied affine and local deformations to achieve the best warping of the images and a custom DaTSCAN template defined in Ref. 78. The resulting images have a final size of $95 \times 69 \times 79$ voxels. On the other hand, the images have to be normalized in intensity to allow comparing the uptake value in areas of specific activity (related to dopaminergic transporters) and areas of nonspecific activity (vascular activity) between subjects. The normalization value I_n is computed using

the Integral Normalization approach,[79] which sets the normalizing value to the average of all values in a certain volume of the image, in an approximation of the integral. In PD, this is often set to the average of the brain without the specific areas: the striatum; although the influence of these areas is often small, and it can be approximated by the mean of the whole image.

3.2.3. *Striatum volume selection*

Since relevant information for PD diagnosis is within a specific brain region (i.e. the striatum), the subvolume containing it is extracted to avoid processing noninformative voxels.

To perform this selection, we use the algorithm described in Ref. 80. The method first averages all the images in the database to create a mean image I_{mean}. Then, a threshold I_{th} is set depending on the characteristics of the images. Due to the nature of the DaTSCAN images, the highest intensities as well as the major differences between affected patients and controls are located at the striatum. Due to the DatSCAN imaging properties, ROIs in are located in the regions with high intensity voxels. Hence, it is possible to determine the threshold as the central value of intensity in the whole image, and computed as

$$I_{\mathrm{th}} = \frac{1}{2}(\max(I_{\mathrm{mean}}) - \min(I_{\mathrm{mean}})) + \min(I_{\mathrm{mean}}),$$

$$(10)$$

where I_{mean} is the mean of all the images in the training set. Finally, we define a box-shaped that contains only brain voxels with an intensity level higher than the computed threshold I_{th} This way we select the ROIs as well as their surrounding areas in each image for further processing. Figure 13 shows an example
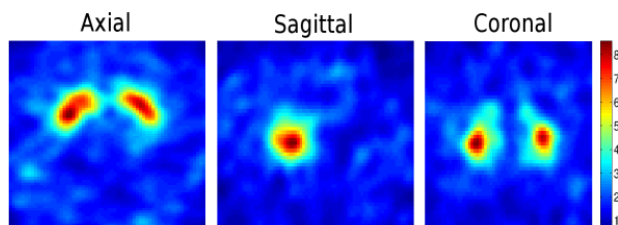

Fig. 13. Example of extracted volume containing regions of interest related to dopamine activity for an image from the PPMI database.

of the extracted volume for an image from the PPMI database using the method explained above.

3.2.4.   *IMF selection*

Using the same methodology previously described and applied to 18F-FDG PET images, IMF functions are computed over the striatum region in DaTSCAN images, once they are sampled using the path indicated by the fractal curve. Then, the IMFs are selected using the correlation-based criterion explained in Sec. 2.3.1. The resulting IMFs calculated for each image are then used to compose an over-complete dictionary in the same way that explained for the PET images. An example of selected IMFs is shown in Figs. 14 and 15 for CN and PD subjects, respectively.

3.2.5.   *Classification results*

The results of the experiments conducted using the PPMI database are shown in Fig. 16, where the performance in terms of accuracy, sensitivity and specificity are shown using different thresholds of the variance explained to select components. They have been obtained by k-fold stratified cross-validation ($k = 10$), which ensures the same label distribution in training and testing subsets. The final performance is obtained by averaging. Moreover, the results are summarized in Table 8 and ROC curve is provided in Fig. 17 to allow comparing the performance of
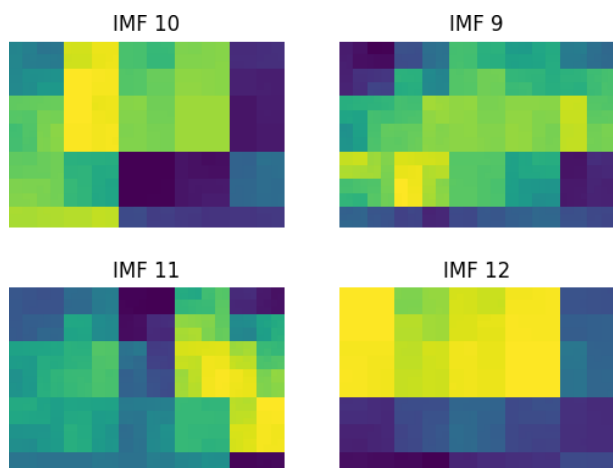

Fig. 14.   Example of the four most relevant IMFs in a Control subject according to the correlation-based criterion. Above figures shown the reconstruction of the indicated IMFs.
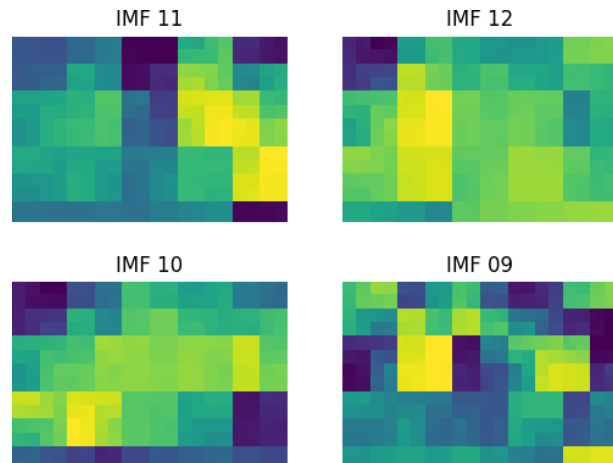

Fig. 15.   Example of the four most relevant IMFs in a PD subject, according to the correlation-based criterion. Above figures shown the reconstruction of the indicated IMFs.

the EfPCA with respect to other approaches used as baseline (i.e. VAF and PCA).

The results have been statistically validated using a t-test hypothesis test, showing a statistical significance of 5% in the means difference for the VAF, PCA and HHT methods for the CN versus PD classification experiments.

4.   Discussion

Table 4 summarizes the classification results of 18F-FDG PET images from the ADNI database obtained by different methods for comparison. Specifically, the method has been compared to VAF, PCA and a recently proposed method based on HHT features.[76] VAF method[25] consists in using the individual voxel values as features without any transformation. This is possible by classifying with SVCs, as they can handle high dimensional feature vectors.[81] PCA method[30,82,83] has been implemented to compress the activation data in order to reduce the dimensionality of the feature space. This is achieved by projecting the original data onto the principal directions (i.e. those that account for the most part of the variance). Experimental results show the superiority of the proposed method over VAF and PCA, providing a higher accuracy, sensitivity and specificity. Specifically, the accuracy value provided by the EfPCA method is up to 0.93, while VAF and PCA methods show values up to 0.85 and 0.90, respectively. This is also exposed, as explained above,
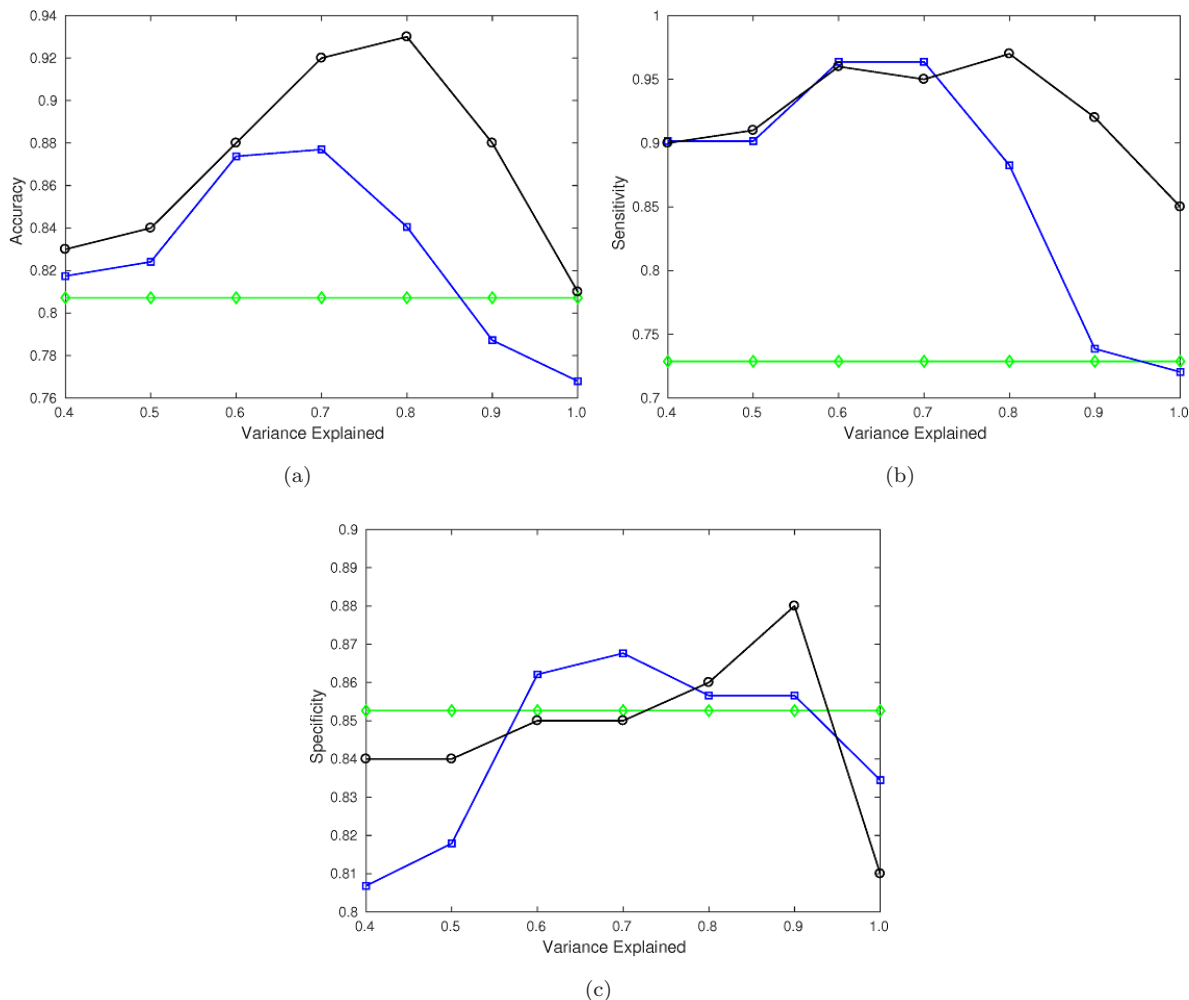
September 18, 2018   7:42   1850040

Fig. 16.   Classification performance obtained in the Control/PD classification experiment using DaTSCAN images from the PPMI database. (a) Accuracy, (b) Sensitivity and (c) Specificity are shown for different values of variance explained used in the fPCA stage.

with the AUCs of the ROC curves. We obtain AUC values of 0.91, 0.93 and 0.95 for VAF, PCA and EfPCA methods, respectively. The performance values obtained demonstrate the capability of the

Table 7.   Classification results for PPMI DatSCAN images from PPMI database using VAF, PCA, HHT and EfPCA methods. Mean values along with the difference between maximum and minimum values obtained during k-fold evaluation are shown.

Method	Accuracy	Sensitivity	Specificity	AUC
VAF	0.80 ± 0.05	0.72 ± 0.17	0.85 ± 0.14	0.87
PCA	0.87 ± 0.04	0.96 ± 0.03	0.86 ± 0.04	0.90
HHT-fs	0.90 ± 0.07	0.92 ± 0.04	0.90 ± 0.05	0.91
EfPCA	$\mathbf{0.93 \pm 0.05}$	$\mathbf{0.97 \pm 0.08}$	$\mathbf{0.88 \pm 0.05}$	**0.94**

Table 8.   Statistical significance of the comparison between EfPCA and other methods in Table 7. Results corresponding to p-values below 0.05 are indicated with *.

Experiment	Method	p-value
Control versus PD	VAF	$< 10^{-7}*$
Control versus PD	PCA	$2.10^{-5}*$
Control versus PD	HHT	$0.01*$

features extracted to represent the image data. Additionally, the experiments carried out to explore the effect of the variance explained on the fPCA method show that part of the variance is related to noise. In fact, the best value is obtained for $\sigma^2_{\exp} = 0.80$. Consequently, 20% of the variance in the data is
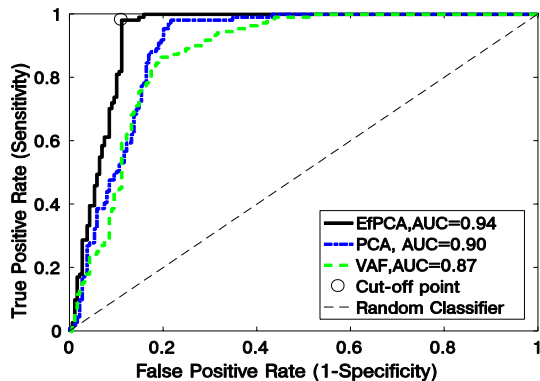
Fig. 17.   ROC curve obtained for Control/PD classification. The EfPCA method provides an AUC of 0.94.

expected to correspond to noise (or at least, it does not contain discriminative information); in Fig. 8, the performance clearly diminishes when projecting the data onto all the eigenvectors (explaining 100% of the variance). As mentioned, we also compare the obtained results with those provided by another recently proposed method based on HHT features.[76] In the case of CN/AD classification, similar values are provided by the HHT-based method and by the EfPCA method. Nevertheless, the EfPCA method outperforms the HHT-based method in MCI/AD classification, as shown with the accuracy and the AUC values in Table 4. The main difference with the proposed method lies in the use of EMD decomposition, which allows representing an image as a linear combination of the components extracted from all the training images. It is worth noting that this method produces an over-complete basis containing a higher number of vectors than usually used in the SRC method.[84] In the proposed EfPCA method, the atoms composing the over-complete basis describe the images in more detail and the resulting linear combination may contain only part of an image. In addition, it is worth noting that the whole method (3D fractal sampling, EEMD computation, dictionary generation, OMP execution and SVM training) takes about 1h using 24 Xeon E5-2640 cores running at 2.40 GHz.

In order to assess the method using a different image modality, we carried out classification experiments using ^{123}I-Ioflupane (DaTSCAN) images from the PPMI database. The results, summarized in Table 7 and Fig. 16, demonstrate the applicability of the proposal to the diagnosis of Parkinsonian

syndromes, outperforming the baseline methods used as reference. Overall, the classification performance achieves up to 93% of accuracy and an AUC of 0.94.

5.   Conclusions and Future Work

In this paper, we propose a method to process 3D images using time-series data analysis techniques. This is addressed by converting the 3D image data into a time series by sampling the voxels using a fractal curve-based method which preserves the spatial relationship. Once the 3D images are transformed into 1D signals, these are then analyzed using functional PCA. Functional PCA is implemented by representing the original signals in an over-complete basis composed of EMD components. Unlike decomposition based on a predefined set of functions such as Fourier Analysis or Wavelet Analysis, the empirical nature of the IMFs provides a more flexible way to represent the original signals. Subsequently, the projection of the original signals into the eigensignals computed by means of fPCA are used as features to describe each brain region. These features feed a one-per-region SVM classifiers which are eventually combined as an ensemble to leverage the classification performance. Different experiments have been conducted to analyze the performance provided by the proposed method in two classification tasks using data from the ADNI database. The first was the classification between CN and AD images obtaining an accuracy of 0.93 and AUC of 0.95, which is close to the limit imposed by the ADNI clinical labels and outperforming the results obtained by VAF and PCA. On the other hand, experiments performed with MCI/AD subjects provided an accuracy of 0.75 and AUC of 0.81. Additionally, experiments using DaTSCAN SPECT images from the PPMI database show a classification accuracy of 0.93 and AUC of 0.94, showing the applicability of the proposed approach for the diagnosis of Parkinsonian syndromes. The experiments performed validate the method in two different ways: first, they demonstrated that fractal sampling is an effective method to convert 3D image data into 1D signals. And second, the use of EMD-based functional PCA allows expressing the original signals as a sparse linear combination of the IMFs, which represents different features of each image in the training set. Besides,

the proposed methodology pave the way to use time signal analysis or other processing techniques with 3D image data, which can be exploited in other image modalities such as MRI.

References

1. World Health Organization, Rates of dementia, Available at http://www.who.int/en/news-room/fact-sheets/detail/dementia (2018).
2. K. A. Johnson, N. C. Fox, R. A. Sperling and W. E. Klunk, Brain imaging in Alzheimer disease, *Cold Spring Harbor Perspect. Med.* **2**(4) (2012).
3. L. Khedher, I. A. Illán, J. M. Górriz, J. Ramírez, A. Brahim and A. Meyer-Baese, Independent component analysis-support vector machine-based computer-aided diagnosis system for Alzheimers with visual support, *Int. J. Neural Syst.* **27**(3) (2017) 1650050.
4. R. Romero-Garcia, M. Atienza and J. L. Cantero, Different scales of cortical organization are selectively targeted in the progression to Alzheimers disease, *Int. J. Neural Syst.* **26**(2) (2016) 1650003.
5. G. Mirzaei, A. Adeli and H. Adeli, Imaging and machine learning techniques for diagnosis of Alzheimers disease, *Rev. Neurosci.* **27**(8) (2016) 857–870.
6. S. Hulbert and H. Adeli, EEG/MEG- and imaging based diagnosis of Alzheimer's disease, *Rev. Neurosci.* **24**(6) (2013) 563–576.
7. H. Adeli, S. Ghosh-Dastidar and N. Dadmehr, Alzheimers disease and models of computation: Imaging, classification, and neural models, *J. Alzheimers Dis.* **7**(3) (2005) 188–197.
8. S. Badoud, D. V. D. Ville, N. Nicastro, V. Garibotto, P. R. Burkhard and S. Haller, Discriminating among degenerative Parkinsonisms using advanced 123i-ioflupane spect analyses, *NeuroImage Clin.* **12** (2016) 234–240.
9. F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, I. A. Illán and A. Ortiz, Automatic detection of Parkinsonism using significance measures and component analysis in datscan imaging, *Neurocomput.* **126** (2014) 58–70.
10. N. Mammone, L. Bonanno, S. D. Salvo, S. Marino, P. Bramanti, A. Bramanti and F. C. Morabito, Permutation disalignment index as an indirect, EEG-based, measure of brain connectivity in MCI and AD patients, *Int. J. Neural Syst.* **27**(5) (2017) 1750020.
11. D. López-Sanz, P. Garcés, B. Álvarez, M. L. Delgado-Losada, R. López-Higes and F. Maestú, Network disruption in the preclinical stages of Alzheimers disease: From subjective cognitive decline to mild cognitive impairment, *Int. J. Neural Syst.* **27**(8) (2017) 1750041.
12. M. Ahmadlou, H. Adeli and A. Adeli, Fractality and a wavelet-chaos-methodology for EEG-based diagnosis of alzheimer disease, *Alzheimer Dis. Assoc. Disord.* **25**(1) (2011) 85–92.

AQ: Please provide page range.

A. Ortiz et al.

13. M. Ahmadlou, A. Adeli, R. Bajo and H. Adeli, Complexity of functional connectivity networks in mild cognitive impairment subjects during a working memory task, *Clin. Neurophysiol.* **125**(4) (2014) 694–702.

14. N. Mammone, C. Ieracitano, H. Adeli, A. Bramanti and F. C. Morabito, Permutation Jaccard distance-based hierarchical clustering to estimate EEG network density modifications in MCI subjects, *IEEE Trans. Neural Netw. Learning Syst.* (2018) 1–14.

15. J. P. Mezquita-Sanchez, A. Adeli and H. Adeli, A new methodology for automated diagnosis of mild cognitive impairment (MCI) using magnetoencephalography (MEG), *Behav. Brain Res.* **305** (2016) 174–180.

16. Y. Yang, T. Solis-Escalante, J. Yao, A. Daffertshofer, A. C. Schouten and F. C. T. van der Helm, A general approach for quantifying nonlinear connectivity in the nervous system based on phase coupling, *Int. J. Neural Syst.* **26**(1) (2016) 1550031.

17. S. A. Akar, S. Kara, F. Latifolu and V. Bilgi, Analysis of the complexity measures in the EEG of schizophrenia patients, *Int. J. Neural Syst.* **26**(2) (2016) 1650008.

18. E. Giraldo-Suarez, J. D. Martinez-Vargas and G. Castellanos-Dominguez, Reconstruction of neural activity from EEG data using dynamic spatiotemporal constraints, *Int. J. Neural Syst.* **26**(7) (2016) 1650026.

19. Y. Yang, T. Solis-Escalante, J. Yao, F. C. T. van der Helm, J. P. A. Dewald and A. C. Schouten, Nonlinear connectivity in the human stretch reflex assessed by cross-frequency phase coupling, *Int. J. Neural Syst.* **26**(8) (2016) 1650043.

20. E. O'Sullivan-Greene, L. Kuhlmann, E. S. Nurse, D. R. Freestone, D. B. Grayden, M. Cook, A. Burkitt and I. Mareels, Probing to observe neural dynamics investigated with networked Kuramoto oscillators, *Int. J. Neural Syst.* **27**(1) (2017) 1650038.

21. C. Liu, B. Abu-Jamous, E. Brattico and A. K. Nandi, Towards tunable consensus clustering for studying functional brain connectivity during affective processing, *Int. J. Neural Syst.* **27**(2) (2017) 1650042.

22. Z.-K. Gao, Q. Cai, Y.-X. Yang, N. Dong and S.-S. Zhang, Visibility graph from adaptive optimal Kernel time-frequency representation for classification of epileptiform EEG, *Int. J. Neural Syst.* **27**(4) (2017) 1750005.

23. M. Graña, L. Ozaeta and D. Chyzhyk, Resting state effective connectivity allows auditory hallucination discrimination, *Int. J. Neural Syst.* **27**(5) (2017) 1750019.

24. M. L. Schroeter, T. Stein, N. Maslowski and J. Neumann, Neural correlates of Alzheimer's disease and mild cognitive impairment: A systematic and quantitative meta-analysis involving 1351 patients, *Neuroimage* **47** (2009) 1196–1206.

25. J. Stoeckel and G. Fung, SVM feature selection for classification of spect images of Alzheimer's disease using spatial information, in *Proc. Fifth IEEE Int Data Mining Conf* (2005).

26. R. P. Duin, Classifiers in almost empty spaces, in *Proc. 15th Int. Conf. Pattern Recognition* (2000), pp. 1–7.

27. D. Kimovski, J. Ortega, A. Ortiz and R. Baños, Parallel alternatives for evolutionary multi-objective optimization in unsupervised feature selection, *Expert Syst. Appl.* **42** (2015) 4239–4252.

28. E. de la Hoz, E. de la Hoz, A. Ortiz, J. Ortega and A. Martnez-lvarez, Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps, *Knowl.-Based Syst.* **71** (2014) 322–338.

29. S. Theodoridis and K. Koutroumbas, *Pattern Recognition* (Academic Press, 2009).

30. J. Górriz, A. Lassl, J. Ramírez, D. Salas-Gonzalez, C. Puntonet and E. Lang, Automatic selection of ROIs in functional imaging using Gaussian Mixture Models, *Neurosci. Lett.* **460**(2) (2009) 108–111.

31. H. Adeli and S. Ghosh-Dastidar, *Automated EEG-Based Diagnosis of Neurological Disorders*: *Inventing the Future of Neurology* (CRC Press, Taylor & Francis, Boca Raton, Florida, 2017).

32. N. Houmani, G. Dreyfus and F. B. Vialatte, Epoch-based entropy for early screening of Alzheimer's disease, *Int. J. Neural Syst.* **25**(8) (2015) 1550032.

33. A. Ortiz, J. M. Górriz, J. Ramírez, F. J. Martínez-Murcia, A. D. N. Initiative *et al.*, Automatic ROI selection in structural brain MRI using SOM 3D projection, *PLOS One* **9**(4) (2014) e93851.

34. M. Turk and A. Pentland, Eigenfaces for recognition, *J. Cogn. Neurosci.* **3**(1) (1991) 71–86.

35. I. Á. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, C. G. Puntonet and F. Segovia, Alzheimer's diagnosis using eigenbrains and support vector machines, *IET Electron. Lett.* **45**(7) (2009) 342–343.

36. A. Ortiz, F. J. Martínez-Murcia, M. J. García-Tarifa, F. Lozano, J. M. Górriz and J. Ramírez, Automated diagnosis of Parkinsonian syndromes by deep sparse filtering-based features, in *Innovation in Medicine and Healthcare* (Springer, 2016), pp. 249–258.

37. F. Martinez-Murcia, J. Górriz, J. Ramírez and A. Ortiz, The Alzheimer's Disease Neuroimaging Initiative *et al.*, A spherical brain mapping of MR images for the detection of Alzheimer's disease, *Curr. Alzheimer Res.* **13**(5) (2016) 575–588.

38. F. J. Martinez-Murcia, J. M. Grriz, J. Ramírez and A. Ortiz, A structural parametrization of the brain using hidden Markov models-based paths in Alzheimer's disease, *Int. J. Neural Syst.* **26**(7) (2016) 1650024.

39. C. Plant, C. Sorg, V. Riedl and A. Wohlschläger, Homogeneity-based feature extraction for

classification of early-stage Alzheimer's disease from functional magnetic resonance images, in *Proc. 2011 Workshop on Data Mining for Medicine and Healthcare, DMMH '11* (ACM, New York, NY, USA, 2011), pp. 33–41.

40. V. Michel, A. Gramfort, G. Varoquaux, E. Eger and B. Thirion, Total variation regularization for FMRI-based prediction of behavior, *IEEE Trans. Med. Imaging* **30** (2011) 1328–1340.

41. M. Eickenberg, E. Dohmatob, B. Thirion and G. Varoquaux, Total variation meets sparsity: Statistical learning with segmenting penalties, in *Medical Image Computing and Computer Aided Intervention* (*MICCAI*), *Proceedings of MICCAI 2015* München, Germany (2015).

42. J. Tohka, E. Moradi and H. Huttunen, Comparison of feature selection techniques in machine learning for anatomical brain MRI in dementia, *Neuroinformatics* **14** (2016) 279–296.

43. R. Cuingnet, J. A. Glauns, M. Chupin, H. Benali and O. Colliot, Spatial and anatomical regularization of SVM: A general framework for neuroimaging data, *IEEE Trans. Pattern Anal. Mach. Intell.* **35** (2013) 682–696.

44. L. Grosenick, B. Klingenberg, K. Katovich, B. Knutson and J. E. Taylor, Interpretable whole-brain prediction analysis with graphnet, *NeuroImage* **72** (2013) 304–321.

45. M. Ahmadlou, A. Adeli, R. Bajo and H. Adeli, Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder, *J. Clin. Neurophysiol.* **27**(5) (2010) 328–333.

46. M. Ahmadlou and H. Adeli, Visibility graph similarity: A new measure of generalized synchronization in coupled dynamic systems, *Physica D Nonlinear Phenomena* **241**(4) (2012) 326–332.

47. M. Ahmadlou, H. Adeli and A. Adeli, Fractality analysis of frontal brain in major depressive disorder, *Int. J. Psychophysiol.* **85**(2) (2012) 206–211.

48. P. Costa, J. Barroso, H. Fernandes and L. J. Hadjileontiadis, Using Peano–Hilbert space filling curves for fast bidimensional ensemble EMD realization, *EURASIP J. Adv. Signal Process.* **2012**(1) (2012) 181.

49. A. Rojas, J. Górriz, J. Ramírez, I. Illán, F. Martínez-Murcia, A. Ortiz, M. G. Río and M. Moreno-Caballero, Application of empirical mode decomposition (EMD) on datscan spect images to explore Parkinson's disease, *Exp. Syst. Appl.* **40**(7) (2013) 2756–2766.

50. A. Ortiz, J. Górriz, J. Ramírez and D. Salas-González, Two fully-unsupervised methods for MR brain image segmentation using SOM-based strategies, *Appl. Soft Computing* **13**(5) (2013) 2668–2682.

51. A. Ortiz, J. Górriz, J. Ramírez and D. Salas-González, The Alzheimer's disease neuroimaging initiative, improving MRI segmentation with probabilistic GHSOM and multiobjective optimization, *Neurocomputing* **114**(19) (2013) 118–131.

52. A. Ortiz, J. Gorriz, J. Ramirez and D. Salas-Gonzalez, Improving MR brain image segmentation using self-organising maps and entropy-gradient clustering, *Inform. Sci.* **262**(Supplement C) (2014) 117–136.

53. S. Ghosh-Dastidar, H. Adeli and N. Dadmehr, Voxel-based morphometry in Alzheimers patients, *J. Alzheimer's Dis.* **10**(4) (2006) 445–447.

54. A. Ortiz, J. Munilla, J. M. Grriz and J. Ramrez, Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease, *Int. J. Neural Syst.* **26**(7) (2016) 1650025.

55. F. J. Martinez-Murcia, A. Ortiz, J. M. Górriz, J. Ramírez, F. Segovia, D. Salas-Gonzalez, D. Castillo-Barnes and I. A. Illán, A 3D convolutional neural network approach for the diagnosis of Parkinson's disease, *Natural and Artificial Computation for Biomedicine and Neuroscience*: *International Work-Conf. Interplay Between Natural and Artificial Computation, IWINAC 2017*, Coruña, Spain, 19–23 June 2017, Proceedings, Part I (Springer International Publishing, Cham, 2017), pp. 324–333.

56. J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, 1st edn. (Springer Publishing Company, Incorporated, 2005).

57. A. Ortiz-Rosario, H. Adeli and J. A. Buford, Wavelet methodology to improve single unit isolation in primary motor cortex cells, *J. Neurosci. Methods* **246** (2015) 106–118.

58. O. Faust, U. Acharya, A. H. and A. A., Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis, *Eur. J. Epilepsy* **26**(1) (2006) 56–64. ← AQ: Please provide complete author names.

59. R. Martis, U. R. Acharya, H. Adeli, H. Prasad, J. H. Tan, K. Chua, C. Loon Too, S. W. J. Yeo and L. Tong, Computer aided diagnosis of atrial arrhythmia using dimensionality reduction methods on transform domain representation, **13** (2014) 295–305. ← AQ: Please provide journal title.

60. S. Ghosh-Dastidar, H. Adeli and N. Dadmehr, Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection, *IEEE Trans. Biomed. Eng.* **55**(2) (2008) 512–518.

61. S. Ghosh-Dastidar, H. Adeli and N. Dadmehr, Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection, *IEEE Trans. Biomed. Eng.* **55**(2) (2008) 512–518.

62. M. Bader, *Space-Filling Curves*: *An Introduction with Applications in Scientific Computing* (Springer Publishing Company, Incorporated, 2012).

A. Ortiz et al.

63. D. Hilbert, über die stetige abbildung einer linie auf ein flächenstck, *Math. Ann.* **38**(1) (1891) 459–460.

64. Peano, Sur une courbe, qui remplit toute une aire plane, *Math. Ann.* **36** (1890) 157–160.

65. E. Estevez-Rams and I. Brito-Reyes, Arithmetic properties of homogeneous Hilbert curves, arXiv e-prints (2013) 1–14.

66. F. Lozano, A. Ortiz, J. Munilla, A. Peinado and for the Alzheimers Disease Neuroimaging Initiative, Automatic computation of regions of interest by robust principal component analysis. Application to automatic dementia diagnosis, *Knowl.-Based Syst.* **123**(1) (2017) 229–237.

67. N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung and H. H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. London A Math. Phys. Eng. Sci.* **454**(1971) (1998) 903–995.

68. Z. Wu and N. E. Huang, Ensemble empirical mode decomposition: A noise-assisted data analysis method, *Adv. Adaptive Data Anal.* **1**(1) (2009) 1–41.

69. A. Ayenu-Prah and N. Attoh-Okine, A criterion for selecting relevant intrinsic mode functions in empirical mode decomposition, *Adv. Adaptive Data Anal.* **2**(1) (2010) 1–24.

70. S. S. Chen, D. L. Donoho and M. A. Saunders, Atomic decomposition by basis pursuit, *SIAM Rev.* **43** (2001) 129–159.

71. Alzheimer's Disease Neuroimaging Initiative, Available at http://adni.loni.ucla.edu/ (accessed 10 March 2018).

72. R. P. Woods, S. T. Grafton, C. Holmes, S. Cherry and J. C. Mazziotta, Automated image registration: I. general methods and intrasubject, intramodality validation, **22** (1998) 139–152.

73. I. A. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. M. López, F. Segovia, R. Chaves, M. Gómez-Rio and C. G. Puntonet, 18f-FDG PET imaging analysis for computer aided Alzheimer's diagnosis, *Inform. Sci.* **181** (2011) 903–916.

74. J. Ashburner and T. Group, SPM8. Functional Imaging Laboratory, Institute of Neurology, 12, Queen Square, London WC1N 3BG, UK (2011).

75. S. Huang, J. Li, L. Sun, J. Liu, T. Wu, K. Chen, A. Fleisher, E. Reiman and J. Ye, Learning brainconnectivity of Alzheimer's disease from neuroimaging data, in *Advances in Neural Information Processing Systems* eds. Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams and A. Culotta (Curran Associates, Inc., 2009), pp. 808–816.

76. A. Ortiz, F. Lozano, A. Peinado, M. García-Tarifa, J. Górriz and J. Ramírez, Pet image classification using HHT-based features through fractal sampling, *Nat. Artif. Comput. Biomed. Neurosci.* **1** (2017) 314–323.

77. The Parkinson Progression Markers Initiative, Imaging technical operations manual, 2 edn (June 2010). Available at http:// http://www.ppmi-info.org/wp-content/uploads/2017/06/PPMI-TOM-V8_09-March-2017.pdf

78. D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán, P. Padilla, F. J. Martínez-Murcia and E. W. Lang, Building a FP-CIT spect brain template using a posterization approach, *Neuroinformatics* **13** (2015) 391–402.

79. I. A. Illán, J. M. Górriz, J. Ramírez, F. Segovia, J. M. JiménezHoyuela and S. J. O. Lozano, Automatic assistance to Parkinsons disease diagnosis in DaTSCAN spect imaging, *Med. Phys.* **39**(10) 5971–5980.

80. F. J. Martinez-Murcia, J. M. Górriz, J. Ramírez, M. Moreno-Caballero and M. Gómez-Río, Parametrization of textural patterns in ^{123}I-Ioflupane imaging for the automatic detection of Parkinsonism, *Med. Phys.* **41** (2014) 012502.

81. V. N. Vapnik, *Statistical Learning Theory* (Wiley-Interscience, 1998).

82. I. Álvarez, J. Górriz, L. M. M., J. Pérez, D. Salas-González, S. F. R. Chaves and C. Puntonet, Computer aided diagnosis of Alzheimer's disease using component based SVM, *Appl. Soft Computing* **11** (2011) 2376–2382.

83. I. Álvarez, J. Górriz, J. Ramírez, D. Salas-González, M. Lopez, F. Segovia, R. Chaves, M. Gomez-Río and C. García-Puntonet, 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis, *Inform. Sci.* **184**(4) (2011) 903–196.

84. A. Ortiz, F. Lozano, J. Górriz, J. Ramrez and F. J. Martnez-Murcia, Discriminative sparse features for Alzheimer's disease diagnosis using multimodal image data, *Curr. Alzheimer Res.* **15**(1) (2018) 1–24.

AQ: Please provide journal title.

AQ: Please provide year.