NeuroImage 222 (2020) 117209

Contents lists available at ScienceDirect

NeuroImage

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

Detecting and visualizing differences in brain structures with SPHARM and functional data analysis☆

L. Ferrando[a], N. Ventura-Campos[a,b,*], I. Epifanio[c]

[a] Grup Neuropsicologia i Neuroimatge Funcional, Universitat Jaume I, Spain
[b] Dept. Educació i Didàctiques Específiques. Universitat Jaume I, Spain
[c] Dept. Matemàtiques-IF. Universitat Jaume I, Spain

ARTICLE INFO

Keywords:
Reversal error
Magnetic resonance imaging
Neuroeducation
Functional discriminant analysis
Functional data analysis
Shape analysis

ABSTRACT

A new procedure for classifying brain structures described by SPHARM is presented. We combine a dimension reduction technique (functional principal component analysis or functional independent component analysis) with stepwise variable selection for linear discriminant classification. This procedure is compared with many well-known methods in a novel classification problem in neuroeducation, where the reversal error (a common error in mathematical problem solving) is analyzed by using the left and right putamens of 33 participants. The comparison shows that our proposal not only provides outstanding performance in terms of predictive power, but it is also valuable in terms of interpretation, since it yields a linear discriminant function for 3D structures.

1. Introduction

Nowadays three-dimensional (3D) magnetic resonance imaging (MRI) with high spatial resolution enables the visualization of different brain structures. After their segmentation, anatomical structures of interest, or a region of interest (ROI), are analyzed, since structural abnormality might explain and help to detect certain conditions (Chung et al., 2010). Volumetry is a common marker in many studies, such as those involved in the diagnosis of Alzheimer's disease (Gerardin et al., 2009). However, analysis of a structure's shape can report richer information than volumetry, because ROI-based volumetric measurements do not make explicit if the volume difference occurs over the whole ROI or it is localized within specific zones of the ROI. Gaining insight into morphological changes can provide researchers with a better understanding of the condition (Epifanio and Ventura-Campos, 2014). This is the reason why shape analysis plays an important role in neuroimaging nowadays (Styner et al., 2003). This is also the case in neuroeducation. For example, Sandman et al. (2014) showed that shape analysis may be more sensitive than volumetric analysis when we want to associate brain differences with performance, and they found that deformity of the basal ganglia may be a neurophenotype associated with risk of developmental impairment.

1.1. Motivation

Neuroeducation is another active field of research. There has been growing interest in the support that neuroscience can provide to education. In the specific case of problem solving, some examples are Hanakawa et al. (2003) and Anderson et al. (2012), who observed activation of different areas of the brain while the participants were doing a problem-solving task.

Let us focus on mathematical modeling and the difficulties of translating a practical situation into mathematical notation. Behavioral studies such as Clement (1982), Clement et al. (1981) and Clement et al. (1980) found that most students made mistakes when translating sentences from natural language into algebraic language. When students know the information from the statement, but they are not able to build a correct equation, this is known as reversal error (RE). Clement (1982) showed that the structure of sentences where RE was present was as follows: "Write an equation using the variables S and P to represent the following statement: There are six times as many students as professors at this university. Use S for the number of students and P for the number of professors" (Clement, 1982, p. 17). Most of the wrong answers were $P = 6 \cdot S$, while the correct answer is $S = 6 \cdot P$. Numerous behavioral studies have analyzed this error (Cooper, 1986; González-Calero et al., 2015; Wollman, 1983), but they did not take into account the importance of people's brain development by using MRI.

Ferrando (2019) studied the differences in gray matter (GM) volume that may exist between subjects that make REs versus those who do not. An increase in the volume of the bilateral putamen was found in the group with RE. This follows along similar lines as the study by Qin et al. (2004). Therefore, in this work we will analyze the shape of the left and right putamen in a classification problem. In other words,

☆ The data (putamen surfaces) and code in MATLAB and free software R are available at http://www3.uji.es/~epifanio/RESEARCH/pufda.zip.

* Corresponding author.
E-mail address: venturan@uji.es (N. Ventura-Campos).

https://doi.org/10.1016/j.neuroimage.2020.117209
Received 15 May 2020; Accepted 24 July 2020
Available online 7 August 2020

we will analyze the morphological changes in the left and right putamen between RE and non-RE groups.

1.2. Shape modeling

In neuroimaging studies, shapes have been modeled using different approaches. Some of them are non-parametric, such as medial representation, where the structure is represented by a skeleton (Styner et al., 2003); the distance map approach (Golland et al., 2001); deformation fields (Joshi et al., 1997) and the landmark approach (Park et al., 2008; Shen et al., 2012). However, we opt for a parametric approach: the use of spherical harmonic representation (SPHARM), which has been successfully applied to model several subcortical structures (Chung et al., 2007, 2010; Gerardin et al., 2009; Gerig et al., 2001; Gu et al., 2004; Shen et al., 2004, 2009). Furthermore, previous studies (Styner et al., 2004) have found good concordance of results based on SPHARM and M-rep shape analysis. In SPHARM, we consider the basis functions of spherical harmonics or its weighted version (the weighted spherical harmonic representation), then a set of coefficients weighting the basis functions parametrizes each surface. As a consequence, SPHARM is a way of smoothing functional data.

Functional data analysis (FDA) is the statistical branch that studies functional observations, i.e. when a whole function is a datum. Although typical functional data comprises uni-dimensional functions with only one argument, usually time, in our case we work with trivariate functions with two arguments (angles), which represent spatial locations. Functional data are recorded discretely, but a continuous function lies behind these data. The discrete observations are converted into a true functional form by approximating (smoothing) each function by a weighted sum (a linear combination) of known basis functions. In our case, each surface is initially described by a set of points belonging to the surface, and then it is converted into a functional datum by smoothing it using spherical harmonics.

FDA shares the same objectives as any other branch of statistics. An excellent overview can be found in Ramsay and Silverman (2005), while a non-parametric point of view of FDA is given in Ferraty and Vieu (2006). A recent review of FDA methods can be found in Wang et al. (2016), although it is centered on univariate functional data. As regards applications, Ullah and Finch (2013) review different applications in different fields, and Sørensen et al. (2013) review FDA with medical applications. To the best of our knowledge, in brain imaging studies, FDA has been applied to the analysis of neuroimaging signal, time courses (Lazar, 2008; Tian, 2010; Viviani et al., 2005), but not to the analysis of brain structures, with some exceptions (Epifanio and Ventura-Campos, 2014; Lila et al., 2016). Note that although SPHARM has been used extensively in neuroscience literature, the majority of these works have not been used in the context of FDA and, therefore, functional data techniques have not been exploited in this field.

1.3. Our contributions

We proposed to use FDA for analyzing the shape of 3D brain structures for the first time in Epifanio and Ventura-Campos (2014). In Epifanio and Ventura-Campos (2014) the hippocampus surfaces, for the study of Alzheimer's disease, were described by multivariate (three) functions with two arguments. We extended principal component analysis (PCA) to deal with trivariate functional data with two arguments. Functional independent component analysis (FICA) was also discussed in Epifanio and Ventura-Campos (2014).

Here, as in Epifanio and Ventura-Campos (2014), we deal with a classification problem. The novelty of our contribution is twofold. On the one hand, in this work our proposal will improve on the method presented in Epifanio and Ventura-Campos (2014). In fact, it will improve on the results of other well-known methodologies, both from the numerical performance point of view and from the interpretative and visual point of view. On the other hand, the new methodology will be applied to the study of RE using the putamen, which is a novel classification problem in neuroeducation.

High dimensionality is one of the greatest difficulties in this kind of classification problems. Unless we consider simple features such as volumetry, the number of variables used to describe the anatomical shape of brain structures is always much larger than the number of subjects (observations). This is the case when using SPHARM coefficients as features in the classification problem. To overcome this problem, we can consider two approaches. In the first approach, one can either select or extract a small subset of relevant features, as in Gerardin et al. (2009), where the most discriminative features are selected using a bagging strategy for subsequent classification with a support vector machine (SVM) classifier, or as in Clemmensen et al. (2011), where sparse discriminant analysis is proposed by extending linear discriminant analysis (LDA) to the high-dimensional setting. Instead of selecting features, in the second approach, one can reduce the dimension, i.e. one can use all the features to construct new components which summarize the original variables. This is the case of Boulesteix (2004) or Epifanio and Ventura-Campos (2014). The method in Boulesteix (2004) consists of Partial Least Squares (PLS) dimension reduction and linear discriminant analysis applied to the PLS components. Although it was originally developed in the context of classification with high-dimensional microarray data, it can be used for classification of any high-dimensional data. In Epifanio and Ventura-Campos (2014), linear discriminant analysis is applied to the new components obtained by dimension reduction techniques for functional data.

Our proposal consists of combining both approaches. The idea is simple, but effective. First we reduce the dimension by using trivariate functional principal component analysis (FPCA) with two arguments or other techniques, such as FICA, as described in Epifanio and Ventura-Campos (2014), but we improve on this and go one step further. We then select the most discriminative components (principal components, PCs or independent components depending on whether we use FPCA or FICA) by stepwise variable selection (SVS) (Weihs et al., 2005) for LDA classification. Note that although the dimension has been reduced in the first step, the number of predictor variables may continue to be high, because the number of subjects is usually small in this kind the classification problems. Note that in the classical multivariate case we can use LDA after PCA, assuming the covariance matrix is the same for all groups (Jolliffe, 2002). Nevertheless, the separation between groups does not necessarily have to occur in the PCs with highest variance, but can occur in the last PCs, those with low variance (Epifanio and Ventura-Campos, 2011).

On the other hand, besides the method's predictive accuracy, human interpretability is one of the most valuable characteristics in a classifier. So the point is to propose methods that are interpretable, rather than trying to explain black box machine learning models (Rudin, 2019). Our proposal is based on the application of LDA in the final step. Despite the simplicity of LDA, its performance is typically almost as good as that of more complicated methods (Hand, 2006; Pierola et al., 2016). Furthermore, LDA provides low-dimensional projections of the data onto the most discriminative directions. We have taken advantage of this and we have defined a discriminative function for 3D shapes. From the interpretation point of view, this is very useful because it reveals the exact locations and directions of the main differences between the groups. In other words, we can visualize how the brain structure changes between the groups. Note that our proposed visual representations are not significance maps, which are common visualization tools in neuroscience. In summary, we propose a local shape methodology that can spatially localize shape changes.

The outline of the paper is as follows. Section 2 describes our data and the proposed methodology. Section 3 includes the numeric results obtained by applying this methodology to our database, together with a comparison with other well-known methods. This section also includes the visualization and graphical interpretation of the results using our procedure. In Section 4 conclusions are given.

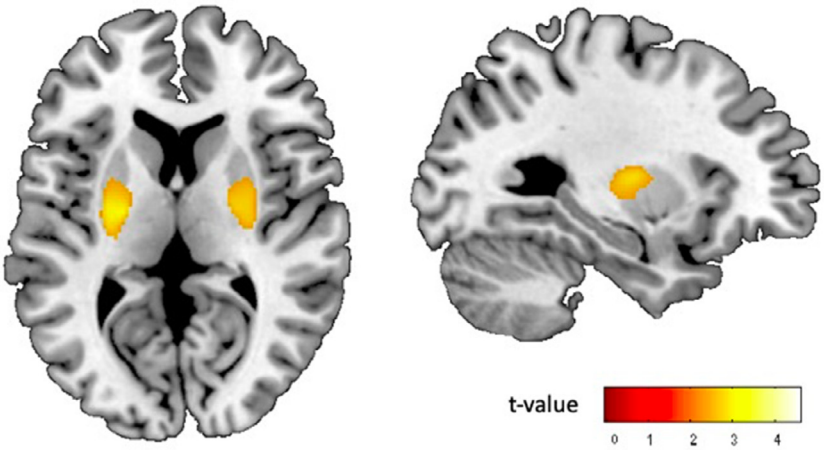

Fig. 1. Neuroimaging results of the two-sample *t*-test performed between groups. It represents the contrast: RE group vs. non-RE group of our study ($p < 0.05$ FWE cluster-corrected using a threshold of $p < 0.005$ at the uncorrected voxel level and a cluster size higher or equal to 813 voxels). The MNI coordinates of the left putamen were x=-23, y=-12, z=0 with Z-value=3.96 (k=938), and the MNI coordinates of the right putamen were x=26, y= -8, z=3 with Z-value= 3.41 (k=813).

2. Materials and methods

The data (putamen surfaces) (Ferrando et al., 2020) and code in MATLAB and free software R (R Core Team, 2020) are available at http://www3.uji.es/~epifanio/RESEARCH/pufda.zip for reproducibility purposes.

2.1. Processing of structural magnetic resonance imaging (smri) brain scans

Thirty-three participants (20 females) with ages ranging between 18 and 26 years (mean age: 22.03, SD: 2.36) were analyzed. All participants were students of Universitat Jaume I. Before participating, they signed a written consent form. All experimental procedures followed the guidelines of the research ethics committee at Universitat Jaume I. The exclusion criteria were trauma with loss of consciousness for more than one hour, typical contraindications to MRI, such as metal implants, and the presence of medical or neurological illness.

Apart from having an MRI scan, the participants carried out a behavioral task. The subjects answered 16 questions, where a mathematical equation had to be built for each of the statements presented. We used an application similar to González-Calero et al. (2015). See details in Ferrando (2019). We established two groups, those who committed REs vs. non-RE, according to the number of errors they made. One group comprised those who failed more than 40% of the equations (RE group: 15 subjects, 4 males, mean age: 21.466, SD: 2.1), and the second group was formed by those who answered 100% of the questions correctly (non-RE group: 18 subjects, 9 male, mean age: 22.5; SD: 2.53).

The sMRI scans were acquired using two scanners. A 3 Tesla Philips scanner and 1.5 Tesla Siemens Symphony scanner (Erlangen, Germany). High-resolution T1-weighted, TR = 8.4 ms, TE = 3.8 ms, matrix size = $320 \times 320 \times 250$ and voxel size = $0.75 \times 0.5 \times 0.8$ mm was used with the Philips scanner. However, high-resolution T1-weighted, TR = 2200 ms, TE = 3 ms, flip angle = 90°, matrix size = $256 \times 256 \times 160$ and voxel size = $1 \times 1 \times 1$mm was used with the Siemens Symphony scanner. The scanner acquisitions covered the entire brain and were performed in parallel to the anterior commissure-posterior commissure plane (AC-PC).

The Voxel Based Morphometry (VBM) analysis was conducted with SPM12 (SPM12 (v7219), Wellcome Trust Centre for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm/software/spm12). We performed the preprocessing steps using the CAT12 toolbox with the default setting (CAT12.5, http://dbm.neuro.uni-jena.de/cat/). We used the GM images to identify the differences in volume between the groups. Having segmented, modulated, and normalized these images, they were smoothed using 8-mm full-width-half-maximum Gaussian smoothing and then fed into a two-sample *t*-test analysis in SPM12. The VBM results showed a greater bilateral posterior putamen volume in the group with RE in comparison to the non-RE group (see Fig. 1). Following these results, we consider the left and right putamen as the Region of Interests (ROIs) to our classification study, since they were the most significant structures. So, the next step was to extract the left and right putamen for each participant. To obtain these ROIs, we segmented each putamen by using the *imcalc* toolbox of SPM12 and performing an intersection between the GM image of each participant and the ROI of the putamen of the AAL atlas. Finally, the slices of each putamen were put together using the *isosurface* function in MATLAB, which returns the faces and vertices of the triangle mesh.

2.2. Surface parametrization

We need three functions with two angular parameters to represent the putamen surfaces: $x(\theta, \phi)$, $y(\theta, \phi)$, $z(\theta, \phi)$ (more details in Shen et al. (2009)). In fact, the surface of these closed 3D objects is mapped onto a unit sphere under a one-to-one. There are several well-known surface flattening techniques that provide this bijective mapping: area preserving mapping (Brechbühler et al., 1995; Shen et al., 2004), conformal mapping (Gu et al., 2004), the deformable surface algorithm (Macdonald et al., 2000) or semi-isometric mapping (Timsari and Leahy, 2000), for instance. Nevertheless, the implementation of these flattening methods is not trivial and, especially, their computational intensity leads us to use the method proposed by Chung et al. (2010) as a better alternative for objects that are almost convex or star-shaped. This method considers the equilibrium state of heat diffusion by tracing the geodesic path of heat equilibrium state from a heat source (the putamen) to a heat sink (the unit sphere). This flattening technique is numerically simpler than the previous methods, since the solution of an isotropic heat equation in a 3D image is computationally trivial and it does not require either to optimize a target function (for more details see Chung et al. (2010)). In any case, any flattening method could be used without altering the subsequent analysis.

Once the surface is projected into the sphere, the angles will act as coordinates for the surfaces of the putamen. Fig. 2 shows an illustration of the surface flattening process for a left putamen with this procedure and the surface parametrization using the angles (θ, φ). The north pole of a unitary sphere coincides with the point $\theta = 0$.

2.3. Representing putamens using SPHARM

Although each putamen could be described by a set of points or discrete observations of its surface, each putamen is truly a smooth surface, a function whose two arguments are angles. Smoothing allows us to convert the discrete data to functions and to perform FDA. This smoothing is carried out using a basis system, SPHARM, and considering the coefficients of each putamen in this basis function expansion. As said before, we have chosen SPHARM because it has previously been applied

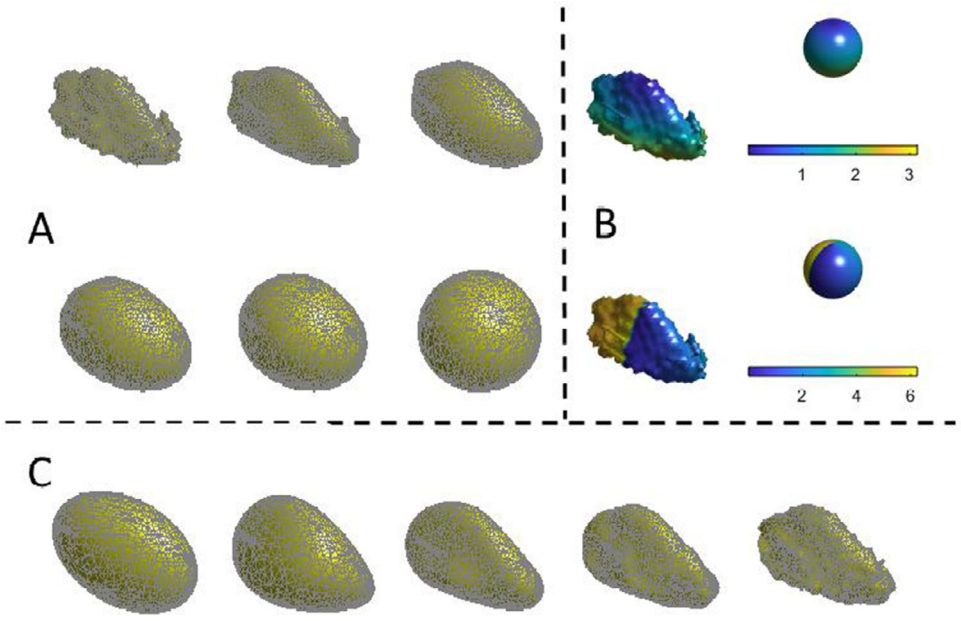

Fig. 2. A left putamen example: (A): The surface flattening process from the original putamen (top left) to the sphere (bottom right). The same level sets as in (Chung et al., 2010) are used (1.0, 0.6, 0.2, -0.2, -0.6, -1.0). (B): The spherical angles are projected on the putamen surface and the unitary sphere, for θ (first row) and φ (second row). (C): The putamen representation using a different number of spherical harmonics: $L = 1, 2, 5, 11, 30$ (from left to right). This will be explained below. The head of the putamen is on the left, while the tail is on the right. This orientation is considered the reference throughout the paper.

successfully to model several subcortical structures. Furthermore, its orthogonality will simplify calculations. Nevertheless, other less common bases, such as spherical splines (Alfeld et al., 1996; He et al., 2005), the weighted Fourier series (Chung et al., 2007) or spherical wavelets (Nain et al., 2007; Yu et al., 2007), could also be used.

We have used the real spherical harmonics as in Chung et al. (2007, 2010), although we could have used spherical harmonics of complex value as in Gerig et al. (2001) or Shen et al. (2004). We prefer to set up a real-valued model because we only need real-valued spherical functions in our application, as in most applications.

A real basis of spherical harmonics, where l is the degree and m is the order, is given by:

$$Y_{lm}(\theta,\varphi) = \begin{cases} \sqrt{2}N_{(l,m)}\cos(m\varphi)P_l^m(cos\theta) & \text{if } m > 0 \\ N_{(l,0)}P_l^0(cos\theta) & \text{if } m = 0 \\ \sqrt{2}N_{(l,|m|)}\sin(|m|\varphi)P_l^{|m|}(cos\theta) & \text{if } m < 0 \end{cases} \quad (1)$$

where P_l^m is the associated Legendre polynomial of order m defined over the range $[-1, 1]$:

$$P_l^m(x) = \frac{(-1)^m}{2^l l!}(1-x^2)^{m/2}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l$$

and $N_{(l,m)} = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}$.

Let S^2 be the unit sphere in $\mathbb{R}^3$, and f and $g \in L^2(S^2)$. The inner product is given by:

$$< f, g >= \int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi} f(\theta,\varphi)g(\theta,\varphi)d\Omega = \int_{S^2} f(\theta,\varphi)g(\theta,\varphi)d\Omega = \int_{S^2} f g d\Omega \quad (2)$$

where $d\Omega = \sin(\theta)d\varphi d\theta$. The spherical harmonics satisfy the orthonormal condition with respect to the inner product:

$$\int_{S^2} Y_{lm}Y_{l'm'}d\Omega = \delta_{ll'}\delta_{mm'}$$

where δ_{ij} is the Kronecker delta.

Three functions represent each putamen in terms of the spherical harmonics, where L determines the smoothing degree:

- $x(\theta,\varphi) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^x Y_{lm}(\theta,\varphi)$
- $y(\theta,\varphi) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^y Y_{lm}(\theta,\varphi)$
- $z(\theta,\varphi) = \sum_{l=0}^{L}\sum_{m=-l}^{l} c_{lm}^z Y_{lm}(\theta,\varphi)$

This can be expressed as a vector-valued function:

$$F(\theta,\varphi) = (x(\theta,\varphi), y(\theta,\varphi), z(\theta,\varphi))' = \sum_{l=0}^{L}\sum_{m=-l}^{l}\mathbf{c}_{lm}Y_{lm}(\theta,\varphi) \quad (3)$$

where $\mathbf{c}_{lm} = (c_{lm}^x, c_{lm}^y, c_{lm}^z)'$. The coefficients can be estimated by least squares, since we know the values of each function in a sample of points, $\{(\theta_i,\varphi_i)\}_{i=1}^n$. For $x(\theta,\varphi)$ (and similarly for $y(\theta,\varphi)$ and $z(\theta,\varphi)$), $\mathbf{x} = \{x(\theta_i,\varphi_i)\}_{i=1}^n$ is the vector of observations, $\mathbf{Y} = \{Y_{lm}(\theta_i,\varphi_i)\}_{i=1}^n$ is the matrix of basis function values at the observation points and $\mathbf{c}^x$ is the vector with the coefficients c_{lm}^x, which can be least square estimated by $\mathbf{c}^x = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{x}$ or by the iterative residual fitting algorithm (Chung et al., 2007).

Fig. 2 (C) shows a putamen example represented by SPHARM using different L values. For small L values, the surface is highly smoothed and many details are missing, but for high L values, such as $L = 30$, the surface is quite noisy, since noise is also fitted. The value $L = 11$ represents a trade-off between both positions, and it has been chosen for left putamens following the strategy suggested by Ramsay and Silverman (2005, Section 4.6.2) to determine the number of basis. This strategy consists of computing the unbiased estimate of the residual variance and selecting the number of basis that makes this variance decrease substantially (Millán-Roures et al., 2018). Specifically, we compute the residuals as the square Euclidean norm of the difference between the points in smoothed surfaces and the original surface for every spherical mesh vertex. These values are added for all the vertices and divided by the number of vertices minus the number of basis, which is $(L+1)^2$ for the degree value L). Then, we compute the pooled variance for all the individuals for L values from 2 to 23 and the minimum variance estimate was attained at $L = 11$ for left putamens and $L = 12$ for right putamens.

No alignment is necessary, since each putamen was translated to the same point in such a way that its centroid coincided with that point, i.e. location was removed previously. No rotation is needed, since all the putamens had the same orientation. Scaling is not needed either, i.e. we want to keep size since the volume is used as discriminant feature in many classification problems. If size had to be removed, then we could divide each putamen by the size of the centroid at the beginning, as in Epifanio and Ventura-Campos (2011). Note that no registration is necessary, as happened in Chung et al. (2010), because the coordinates $((\theta, \varphi))$ on the surfaces are corresponding pairs, therefore the coefficients match each other.

For this and the previous Section 2.2, the following packages are very helpful: the SurfStat package (http://www.math.mcgill.ca/keith/surfstat) and its extension (http://www.stat.wisc.edu/~mchung/research/amygdala/) (Chung et al., 2010).

2.4. Functional principal component analysis

Before introducing FPCA, let us remember how PCA works in the standard multivariate case. Let $\mathbf{X}$ be the centered data (the mean has been subtracted) matrix with N rows. N indicates the number of individuals. Let $\mathbf{V}$ be the sample variance-covariance matrix, $\mathbf{V} = (N-1)^{-1}\mathbf{X}'\mathbf{X}$, where $\mathbf{X}'$ indicates the transposition of $\mathbf{X}$. The solution of the following eigenequation, where ρ is an eigenvalue and ξ is an eigenvector of $\mathbf{V}$, provides the PCA solution:

$$\mathbf{V}\boldsymbol{\xi} = \rho\boldsymbol{\xi}, \tag{4}$$

The PC scores for the k-th PC are computed by $s_i^k = \sum_j x_{ij}\xi_j^k$.

In FPCA, PCs are not vectors, but functions, and summations change into integrations. Let us begin by recalling FPCA for the functional univariate case with one scalar argument t. Let $\{x_1(t), \ldots, x_N(t)\}$ be the set of observed functions. The mean function is defined by $\bar{x}(t) = N^{-1}\sum_{i=1}^N x_i(t)$, while the variance-covariance function $v(s,t)$ is defined by $v(s,t) = (N-1)^{-1}\sum_{i=1}^N x_i(s)x_i(t)$, once the data have been centered. The functional counterpart of Eq. 4 (see details in Ramsay and Silverman (2005, Chapter 8)) is:

$$\int v(s,t)\xi(t)dt = \rho\xi(s), \tag{5}$$

where ρ is still an eigenvalue and $\xi(s)$ is not an eigenvector, but an eigenfunction. The score for the kth PC for the ith subject is now calculated by using the inner product for functions: $s_{ik} = \int x_i(s)\xi_k(s)ds$.

To solve 5 there are different alternatives (see Ramsay and Silverman (2005, Sec. 8.4.2) for a review). One of them consists of considering the coefficients in a basis functions. In fact, if the basis is orthonormal, FPCA reduces to the classical multivariate PCA of the coefficient matrix, which reduces the computational cost. The functions $\xi_k(t)$ satisfy the orthonormality constraint, as in the multivariate case.

The maximum number of possible functional PCs is limited by $N-1$, although if the number of basis functions M is less than N, then the maximum would be M. Let K be the number of functional PCs considered, then $x_i(t)$ is described by $\sum_{k=1}^K s_{ik}\xi_k(t)$.

2.4.1. FPCA With multiple functions and multiple arguments

Let $\{F_i(\theta,\varphi)\}_{i=1}^N = ((x_i(\theta,\varphi), y_i(\theta,\varphi), z_i(\theta,\varphi))$ be the set of vector-valued functions with two arguments. Each of them represents the putamen of a subject. As previously, we can calculate pointwisely three mean functions, $\bar{x}(\theta,\varphi)$, $\bar{y}(\theta,\varphi)$ and $\bar{z}(\theta,\varphi)$, three covariance functions $v_{XX}((\theta,\varphi),(\vartheta,\phi))$, $v_{YY}((\theta,\varphi),(\vartheta,\phi))$, $v_{ZZ}((\theta,\varphi),(\vartheta,\phi))$, and cross-covariance functions. For example, the cross-covariance function of the centered data for the combination XY is $v_{XY}((\vartheta,\phi),(\theta,\varphi)) = (N-1)^{-1}\sum_{i=1}^N x_i(\vartheta,\phi)y_i(\theta,\varphi)$, it can be computed analogously for the combination XZ and YZ.

The addition of the inner products of the three components (as defined in 2) yields an inner product on the space of vector-valued functions:

$$< F_1, F_2 >=< x_1, x_2 > + < y_1, y_2 > + < z_1, z_2 > . \tag{6}$$

The PC score for the k-th PC is calculated by $s_i^k =< F_i, \xi^k >= \int_{S^2} x_i\xi_X^k d\Omega + \int_{S^2} y_i\xi_Y^k d\Omega + \int_{S^2} z_i\xi_Z^k d\Omega$, where PCs are now a three-vector $\xi=(\xi_X, \xi_Y, \xi_Z)$ of functions, which are solutions of the following eigenequation system $V\xi = \rho\xi$, which is expressed as

$$\begin{aligned}
&\int_{S^2} v_{XX}((\vartheta,\phi),(\theta,\varphi))\xi_X(\theta,\varphi)d\Omega + \int_{S^2} v_{XY}((\vartheta,\phi),(\theta,\varphi))\xi_Y(\theta,\varphi)d\Omega \\
&\quad + \int_{S^2} v_{XZ}((\vartheta,\phi),(\theta,\varphi))\xi_Z(\theta,\varphi)d\Omega = \rho\xi_X(\vartheta,\phi) \\
&\int_{S^2} v_{YX}((\vartheta,\phi),(\theta,\varphi))\xi_X(\theta,\varphi)d\Omega + \int_{S^2} v_{YY}((\vartheta,\phi),(\theta,\varphi))\xi_Y(\theta,\varphi)d\Omega \\
&\quad + \int_{S^2} v_{YZ}((\vartheta,\phi),(\theta,\varphi))\xi_Z(\theta,\varphi)d\Omega = \rho\xi_Y(\vartheta,\phi) \\
&\int_{S^2} v_{ZX}((\vartheta,\phi),(\theta,\varphi))\xi_X(\theta,\varphi)d\Omega + \int_{S^2} v_{ZY}((\vartheta,\phi),(\theta,\varphi))\xi_Y(\theta,\varphi)d\Omega \\
&\quad + \int_{S^2} v_{ZZ}((\vartheta,\phi),(\theta,\varphi))\xi_Z(\theta,\varphi)d\Omega = \rho\xi_Z(\vartheta,\phi).
\end{aligned} \tag{7}$$

As mentioned before, we consider the basis function expansion of the vector-valued functions to solve the eigenequation system. Each F_i is described by the following vector of basis coefficients $\mathbf{c}^i = (\{c_{ilm}^x\}, \{c_{ilm}^y\}, \{c_{ilm}^z\})$, with $l = 0, \ldots, L$ and $m = -l$ to l, and a matrix $\mathbf{C}$ with N rows (one per subject) is built by stacking those vectors. We only need to compute the PCA of the $N \times 3M$ matrix $\mathbf{C}$ since spherical harmonics are orthonormal (M is 144 for $L = 11$, whereas M is 169 for $L = 12$). Once PCs are calculated, the parts corresponding to each coordinate are separated (see Ramsay and Silverman (2005, Sec. 8.5.1) for details in the case of bivariate FPCA with one argument).

As in the multivariate case, the proportion of variance explained by each eigenfunction was given by each eigenvalue ρ divided by the sum of all eigenvalues. In addition, for the j-th principal component $\xi^j = (\xi_X^j, \xi_Y^j, \xi_Z^j)$, we can compute the variation accounted for each coordinate by $< \xi_X^j, \xi_X^j >$, $< \xi_Y^j, \xi_Y^j >$ and $< \xi_Z^j, \xi_Z^j >$ respectively, since their sum is one by definition.

2.5. Functional linear discriminant analysis (FLDA)

In order to obtain meaningful results of LDA with functions, some kind of regularization or filtering is necessary (naively, we could apply the linear discriminant method to the high-dimensional vectors, but this approach does not give meaningful results, see (Ramsay and Silverman, 2002, ch.8) for details about the explanation for this). A common regularization approach consists of carrying out LDA on the first PCs (or other types of dimension reduction techniques), i.e. all the PCs up to a certain number are considered in LDA. This idea has been used in the functional univariate case with one argument (Epifanio, 2008; Hall et al., 2001), with multivariate functions with one argument (Ramsay and Silverman, 2002, ch.8), and with two arguments (Epifanio and Ventura-Campos, 2014).

However, as explained in Section 1.3, this dimension reduction may not be enough, since we are dealing with small sizes. Furthermore, PCA is an unsupervised statistical learning technique, and its application does not ensure that the separation between classes occurs in the first few components, but it can occur in the last PCs (Jolliffe, 2002). On the one hand, if these last PCs were not considered, the accuracy of classification would be affected. On the other hand, if we consider a very high number of components, i.e. a very high number of predictors with few observations, we return to a high-dimensionality problem again. This is the reason why we propose to consider only the most discriminative PCs by SVS for LDA classification.

2.6. Stepwise variable selection for classification

There are many methods for variable section for classification. We consider a stepwise forward variable selection based on Wilks' lambda criterion. The method is implemented in the *greedy.wilks* function of the R package klaR (Weihs et al., 2005). The variable which separates the groups most constitutes the initial model. Then, more variables are added to the model depending on Wilks' lambda criterion: we add the variable that minimizes Wilks' lambda of the model, including the variable if its p-value still shows statistical significance.

Another variable selection method could also be considered, such as the method implemented in the *stepclass* function of the R package klaR (Weihs et al., 2005). However, the selection of variables in this kind of method is based on optimizing a performance measure, such as accuracy, which is estimated by cross-validation (CV). As the sample size is very small in our application, the results of this kind of method are unstable due to the cross-validation step. In other words, the variables selected can change a lot depending on how the data are split. This is why we opt for a deterministic method like *greedy.wilks*.

In summary, our proposal, which is referred to as FPCA-SVS-LDA, consists of SPHARM representing putamen, applying FPCA, selecting the scores of the PCs by SVS using the *greedy.wilks* method and carrying out LDA on this selection.

2.6.1. Linear discriminant function

As in the multivariate case (Mardia et al., 1979, Sect. 11.5), we can define the linear discriminant function in the functional case. The linear discriminant vector function $\lambda^j(\theta, \varphi) = (\lambda_X^j(\theta,\varphi), \lambda_Y^j(\theta,\varphi), \lambda_Z^j(\theta,\varphi))$ is the functional counterpart of the linear discriminant vector or canonical variate in the multivariate case. Therefore, the score or discriminant value of F_i can be obtained by $d_i^j = < F_i, \lambda^j >= \int_{S^2} x_i \lambda_X^j d\Omega + \int_{S^2} y_i \lambda_Y^j d\Omega + \int_{S^2} z_i \lambda_Z^j d\Omega$.

Let us express both functions in the orthonormal base defined by the PCs, then $d_i^j = < F_i, \lambda^j > = < \sum_{k=1}^K s_i^k \xi^k, \sum_{k=1}^K l_k^j \xi^k >$. Due to the orthonormality, $d_i^j = < \mathbf{s}_i, \mathbf{l}^j >$, i.e. the vector $\mathbf{l}^j$ is the j-th canonical variate for the $N \times K$ matrix $\mathbf{S}$ with the scores for the N individuals, where each row in $\mathbf{S}$ is formed by $\mathbf{s}_i$, which has the K scores for the individual i.

In summary, if there are Q groups, each of them with size N_i ($\sum_{i=1}^Q N_i = N$), LDA is applied to $N \times K$ matrix $\mathbf{S}$ of PC scores. The $K \times r$ matrix $\mathbf{L}$, where $r = \min\{K, Q-1\}$ is the number of discriminant functions, contains the linear discriminant vectors $\mathbf{l}^j$, while the $N \times r$ matrix $\mathbf{D} = \mathbf{SL}$ contains the discriminant values, and the linear discriminant function $\lambda^j(\theta, \varphi)$ ($j = 1, \ldots, r$) is $\sum_{k=1}^K l_k^j \xi^k$, where l_k^j is the element (k, j) of the matrix $\mathbf{L}$.

2.7. Visualization of the results

The effect of each functional PC (FPC) or linear discriminant function can be displayed by adding a suitable multiple, which can be positive or negative, of that function to the mean function (mean putamen). This approach is common in the literature on shape analysis (Dryden and Mardia, 1998) and FDA (Ramsay and Silverman, 2002). We can plot a vector map, where vectors are drawn from the mean putamen to the surface formed by the mean plus the multiple of the function in question. Or we can also color the mean putamen using the magnitude (norm) of those vectors. Furthermore, the PCA scores and the discriminant values can be also displayed.

2.8. Functional independent component analysis

Let us introduce the methodology when the FPCA step is exchanged for another dimension reduction technique: FICA.

Let us remember ICA for the multivariate case. The data matrix $\mathbf{X}$ is expressed as a linear combination of non-Gaussian (independent) components: $\mathbf{X} = \mathbf{SA}$, where columns of $\mathbf{S}$ contain the independent components and $\mathbf{A}$ is a linear mixing matrix. ICA seeks to "un-mix" the data by estimating an un-mixing matrix $\mathbf{W}$ such that $\mathbf{XW} = \mathbf{S}$. Under this assumption, the "signals" in $\mathbf{X}$ will be "more Gaussian" than the source components in $\mathbf{S}$ due to the Central Limit Theorem. Therefore, the objective is to find an un-mixing matrix $\mathbf{W}$ that maximizes the non-gaussianity of the sources.

For the functional univariate case, let $x_1(t), \ldots, x_N(t)$ be N linear mixtures of K independent components $s_j(t)$: $x_i(t) = \sum_{j=1}^K a_{ij} s_j(t)$, for all i. Assume the following basis expression for each function: $x_i(t) = \sum_{m=1}^M b_{im} G_m(t)$.

Let x be a vector-valued function with components $x_1, \ldots, x_N$ and G the vector-valued function with components $G_1, \ldots, G_M$, then the simultaneous expansion of all N functions can be expressed by $x = \mathbf{B}G$, where $\mathbf{B}$ is the coefficient matrix, with size $N \times M$. ICA can be performed on $\mathbf{B}'$, thus $\mathbf{B}' = \mathbf{S_b A_b}$ and $x = \mathbf{B}G = \mathbf{A_b' S_b'} G$. In other words, the data x are generated by a process of mixing the K components $I = \mathbf{S_b'} G$ (the independent components are the rows of $\mathbf{S_b'}$).

For any function $\tilde{x}(t)$ that is not contained in the initial data set, its expansion in terms of those ICA components is $\tilde{x}(t) = \sum_{j=1}^K \tilde{a}_j I_j(t)$, where $I_j(t)$ is the j-th component of I. If I and G are estimated in p points ($\{t_k; k = 1, \ldots, p\}$), the $p \times K$ matrix $\mathbf{I}$ and the $p \times M$ matrix $\mathbf{G}$ can be defined, as well as $\mathbf{I} = \mathbf{GS_b}$. Then, we can compute the K-vector $\tilde{\mathbf{a}}$ with the coefficients $\tilde{a}_j$ by least squares fitting (Ramsay and Silverman, 2005): $\tilde{\mathbf{a}} = (\mathbf{I'I})^{-1}\mathbf{I'}\tilde{\mathbf{x}}$, with $\tilde{\mathbf{x}} = \{\tilde{x}(t_k)\}_{k=1}^p$, i. e. $\tilde{\mathbf{a}} = (\mathbf{S_b' G' G S_b})^{-1}\mathbf{S_b' G'}\tilde{\mathbf{x}}$. Similarly, we can estimate the M-vector with the coefficients $\tilde{b}_m$ by $\tilde{\mathbf{b}} = (\mathbf{G'G})^{-1}\mathbf{G'}\tilde{\mathbf{x}}$, where G is the basis and $\tilde{x}(t) = \sum_{m=1}^M \tilde{b}_m G_m(t)$. In the case where the basis G is orthonormal, i.e. $\mathbf{G'G}$ is the identity matrix, then

$$\tilde{\mathbf{a}} = (\mathbf{S_b' G' G S_b})^{-1}\mathbf{S_b' G'}\tilde{\mathbf{x}} = (\mathbf{S_b' S_b})^{-1}\mathbf{S_b'}\tilde{\mathbf{b}}. \tag{8}$$

The same discussion can be adopted when the functions have more than one argument. In the case of multivariate functional data, the coefficients for each function can be concatenated into a single long vector, as done in Section 2.4.1 with multivariate FPCA. In this case, $\tilde{\mathbf{b}}$ would be $(\mathbf{c}^i)'$.

In order to reduce noise and prevent overlearning (Hyvärinen et al., 2001, Section 13.2), data dimension reduction by PCA should be carried out prior to the application of the ICA algorithm (see Hyvärinen et al. (2000, Section 5) for details). Therefore, PCA is computed first with a concrete number of components, then the same number of independent components as the PCA reduced dimension are estimated.

As previously, let Q be the number of groups, with size N_i ($\sum_{i=1}^Q N_i = N$). For computing the linear discriminant vector function $\lambda^j(\theta, \varphi)$, LDA is applied to the $K \times N$ matrix $\mathbf{A}$ with the coefficients of the K ICA components. Then, we obtain a $K \times r$ matrix $\mathbf{L}$ ($r = \min\{K, Q-1\}$ is the number of discriminant functions), which yields the $r \times N$ matrix $\mathbf{D}$ of discriminant values ($\mathbf{D} = \mathbf{L'A}$). According to Eq. 8, $\mathbf{A} = (\mathbf{S_b' S_b})^{-1}\mathbf{S_b' C'}$, where $\mathbf{S_b}$ is the $3M \times K$ matrix that contains the independent components of $\mathbf{C'}$, the $N \times 3M$ matrix with the SPHARM coefficients. As $\mathbf{D} = \mathbf{\Lambda C'}$, where $\mathbf{\Lambda}$ is the $r \times 3M$ matrix with the SPHARM coefficients of the r functions $\lambda^j(\theta, \varphi)$ ($j = 1, \ldots, r$) (λ^j is the j-th row), then $\mathbf{\Lambda} = \mathbf{L'}(\mathbf{S_b' S_b})^{-1}\mathbf{S_b'}$.

In summary, when FPCA is exchanged for FICA in our methodology, while keeping the remaining steps the same, we call this procedure FICA-SVS-LDA.

3. Results and discussion

FPCA-SVS-LDA and FICA-SVS-LDA are applied to our data set. Due to the small sample size of our data set, the findings cannot be conclusive from the neuroeducational point of view. However, the proposed methodology can be applied without modification to a larger data set. The first two FPCs account for 46% of the variance for the left putamens (62% for the right putamens, respectively), while the first 19 FPCs explain 95% of the variation (the first 18 FPCs for the right putamens, respectively). Note that the 1st FPC is not selected by our procedure until step 17 for the left putamens, which demonstrates again that the first PCs are not necessarily the most discriminative ones, as explained in Sect. 1.3. The most discriminating FPCs for the left putamens are the 20th and 11th. Fig. 3 displays the scores for these two components. Blue stars and red circles represent the RE and non-RE subjects, respectively. The separation between both groups can be perceived visually. Fig. 3 also displays the discriminant values for the left putamens. As we only have two classes, only one discriminant function can be defined. In this case, the whole training set is classified correctly.

Fig. 4 shows the effect of the first three FPCs on the mean left putamen by representing their magnitudes with two standard deviations.

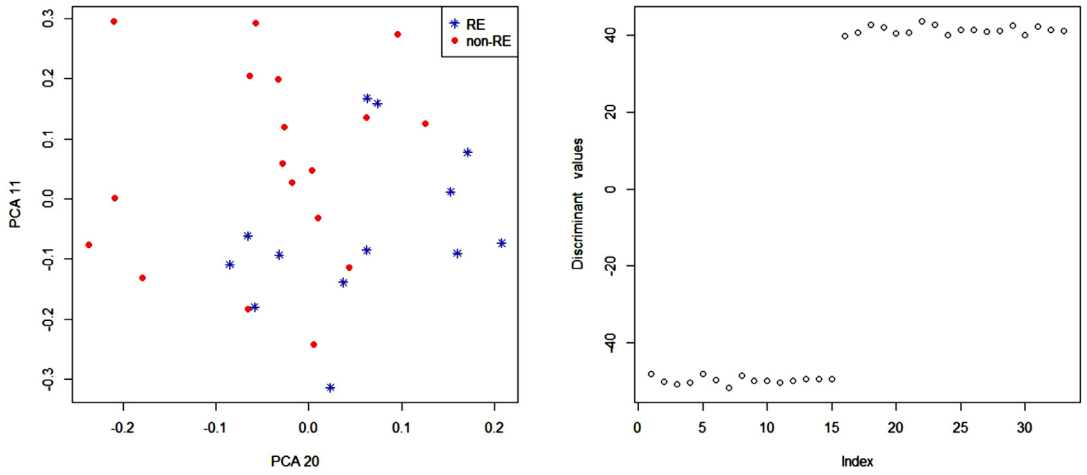

Fig. 3. Left side: Score plot of 20th and 11th PCs for the left putamens. The legend indicates the groups. Right side: Index versus Discriminant values. Note that the first 15 subjects correspond to the RE group, while the remaining individuals belong to the non-RE group.

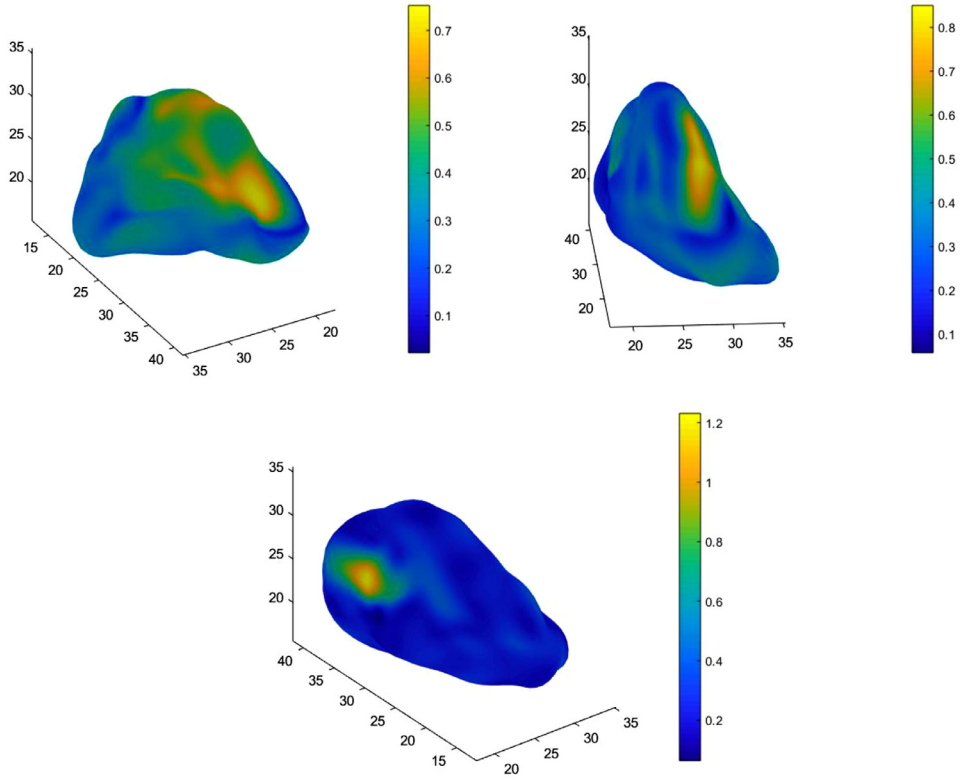

Fig. 4. The effect of first FPCs on the mean shape (from left to right and from top to bottom) for the left putamens. The viewpoint of the first image is defined by azimuth of 150° and elevation of 31°. The viewpoint of the second image is defined by azimuth of -4° and elevation of 20°, while the third one has the same orientation as the reference.

The viewpoints have been selected in order to show the effect of each component better. As the code is available, figures can be reproduced and rotated. The first FPC shows a more global effect than the second and third FPCs, which have a more localized effect. In other words, for the first FPC the effect is distributed throughout a large zone of the putamen; while the effect of the other two components is concentrated on a specific part of the putamen. Fig. 5 displays the linear discriminant function on the mean left putamen by FPCA-SVS-LDA. The directions in which the discriminant score increases fastest are shown by the arrows. The norm of these arrows is displayed by color. The differences between both groups are located in the yellow/orange zones of the putamen. Analogously, Fig. 6 displays the linear discriminant function on the mean right putamen by FICA-SVS-LDA. It seems that the differences are more localized in small zones in the right putamen than in the left putamen, where the differences are more spread out. This also occurred in Sandman et al. (2014) in another study on neuroeducation.

In order to assess the performance of FPCA-SVS-LDA, we estimate it by leave-one-out (LOO) cross-validation. In each trial, one individual is left out, while FPCA-SVS-LDA is applied to the remaining individuals, which constitute the training set of that trial. Then, the FPCA scores for the individual that was left out, which is the test set, are computed and used to predict its class. This procedure is repeated for each individual of the data set. So, finally the performance estimates by LOO are obtained and shown in Table 1 for the left putamen and in Table 2 for the right putamen, together with the LOO performance of FICA-SVS-LDA and other methods explained in Section 3.1. Note that in each trial the FPCs are different, since the training sets are different, and the number of selected FPCs for classification varies for each trial. This is why the

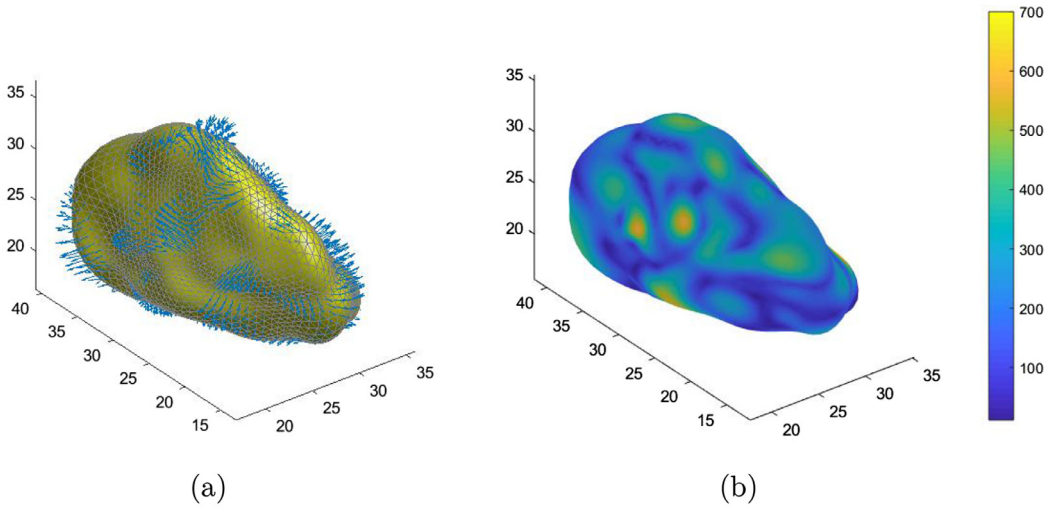

Fig. 5. Representation of the linear discriminating function by FPCA-SVS-LDA, with a vector map (left side) and magnitude map (right side) for the left putamens. The images have the same orientation as the reference (see Fig. 2).

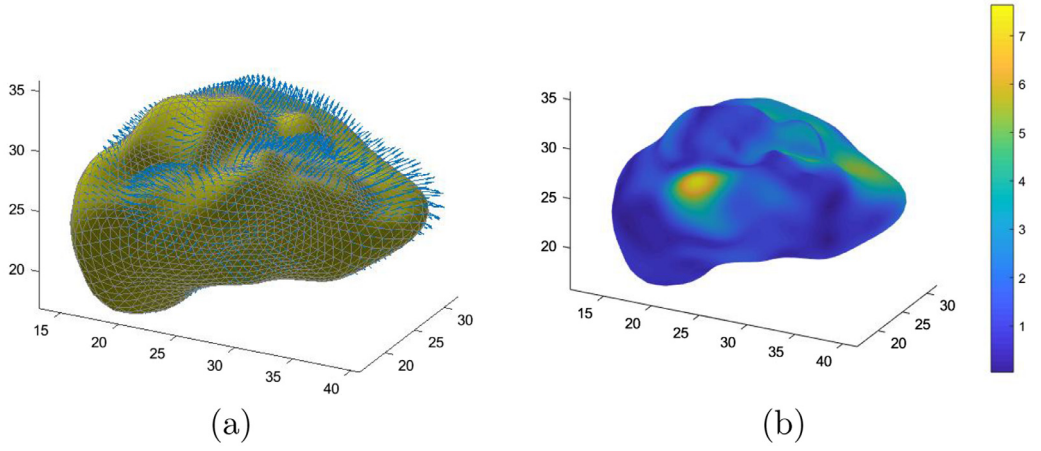

Fig. 6. Representation of the linear discriminating function by FICA-SVS-LDA, with a vector map (left side) and magnitude map (right side) for the right putamens. The viewpoint of the images are defined by azimuth of 63° and elevation of -23°.

Table 1
Left putamen. LOO performance for different methods: accuracy, recall or sensitivity, specificity, precision or positive predictive value, and negative predictive value (NPV), assuming the RE class as the positive class. The maximum value in each column appears in bold.

Method	No. features	Accuracy	Recall	Specificity	Precision	NPV
FPCA-SVS-LDA	14.12	**0.6364**	**0.5333**	**0.7222**	**0.6154**	**0.6500**
FICA-SVS-LDA	1.94	0.5758	0.4667	0.6667	0.5385	0.6
Volume	1	0.4545	0.2	0.6667	0.3333	0.5
SVM	33	0.5152	0.4667	0.5556	0.4667	0.5556
SDA	20	0.4848	0.4	0.5556	0.4286	0.5263
PLS-LDA	10.91	0.4242	0.4	0.4444	0.3750	0.4706
FPCA-LDA	23	0.4848	0.4667	0.5	0.4375	0.5294
FICA-LDA	12	0.4848	0.4667	0.5	0.4375	0.5294

mean of the number of selected FPCs in each trial is shown in the 'No. features' column of those tables.

3.1. Comparison with other methods

We apply different classification methodologies in order to compare the results. The first and simplest one is based on putamen volumetry. The putamen volume is estimated by the sum of the slice areas, i.e. the number of pixels that belong to each slice of the putamen. We perform LDA with this data using LOO cross-validation, and the results are shown in the 'Volume' row of Table 1 and Table 2.

The second methodology is that used in Gerardin et al. (2009). As mentioned previously, an SVM is used to classify the SPHARM coefficients. These are selected with a bagging strategy, where *t*-tests are used for finding the coefficients that best separate the classes. The number of coefficients used is selected by double or nested leave-one-out cross-validation. The results are shown in the 'SVM' row of Table 1 and Table 2.

Table 2
Right putamen. LOO performance for different methods: accuracy, recall or sensitivity, specificity, precision or positive predictive value, and negative predictive value (NPV), assuming the RE class as the positive class. The maximum value in each column appears in bold.

Method	No. features	Accuracy	Recall	Specificity	Precision	NPV
FPCA-SVS-LDA	13.73	0.6364	0.6000	0.6667	0.6000	0.6667
FICA-SVS-LDA	5.42	**0.7273**	**0.667**	**0.7778**	**0.7143**	**0.7368**
Volume	1	0.4242	0	**0.7778**	0	0.4828
SVM	27	0.6364	0.6	0.6667	0.6	0.6667
SDA	20	0.4848	0.4	0.5556	0.4286	0.5263
PLS-LDA	10	0.4848	0.4	0.5556	0.4286	0.5263
FPCA-LDA	7	0.6061	0.4667	0.7222	0.5833	0.6190
FICA-LDA	8	0.6970	0.6	**0.7778**	0.6923	0.7

The third methodology is sparse discriminant analysis (SDA), proposed by Clemmensen et al. (2011). We apply SDA (Sjöstrand et al., 2018) to the SPHARM coefficients. The number of variables is selected by nested leave-one-out cross-validation. The results are shown in the 'SDA' row of Table 1 and Table 2.

The fourth methodology is the method proposed by Boulesteix (2004) and it is implemented in the R package *plsgenomics* (Boulesteix et al., 2018). The choice of the number of latent components is performed by the cross-validation method proposed by Boulesteix (2004). The results are shown in the 'PLS-LDA' row of Table 1 and Table 2.

The fifth and sixth methods are the procedures proposed by Epifanio and Ventura-Campos (2014), where LDA is applied to the coefficients of FPCA and FICA, but without selection of components. The number of components is selected by nested leave-one-out cross-validation. The results are shown in the 'FPCA-LDA' and 'FICA-LDA' rows of Table 1 and Table 2.

For the left putamen, the method that obtains the best results for all the performance measures is our proposal FPCA-SVS-LDA, while the second one is our other proposal, FICA-SVS-LDA. The third best method yields worse results, particularly in terms of accuracy, despite using a high number of features. The importance of considering selection of variables after the dimension reduction step is revealed. Note the great improvement in the measures when SVS is performed. For example, the accuracy goes from 48.48% for FPCA-LDA or FICA-LDA to 63.64% for FPCA-SVS-LDA and 57.58% for FICA-SVS-LDA. It is clear that using a variable selection step after the dimension reduction step has been a success. Note that in this comparison, all the methods except that based on volume, are local shape methods based on SPHARM and the same pre-processing steps have been carried out for all of them. In this way, we have compared the different methods once the SPHARM representation is available.

For the right putamen, the method that obtains the best results for all the performance measures is our proposal FICA-SVS-LDA, which returns better results than those for the left putamen. The second best method in terms of accuracy is FICA-LDA, while the third best are FPCA-SVS-LDA and SVM. As happened with the left putamen, using a variable selection step after the dimension reduction step has improved the results. The accuracy obtained with FICA-SVS-LDA and the right putamen is higher (0.7273) than that obtained with the left putamen.

3.1.1. Multivariate linear model

Although discriminant analysis and testing of mean group difference are different problems, we apply the methodology in Chung et al. (2010) for emphasizing the usefulness of the linear discriminant function applied to brain structures as in Figs. 5 and 6, and for differentiating it from the significance maps of group differences that are commonly used in the neuroimaging literature together with classification results (Gerardin et al., 2009). Multivariate linear modeling (Taylor and Worsley, 2008) is carried out on SPHARM, and the effect of the group variable on the model is tested. Fig. 7 shows the *F*-statistic

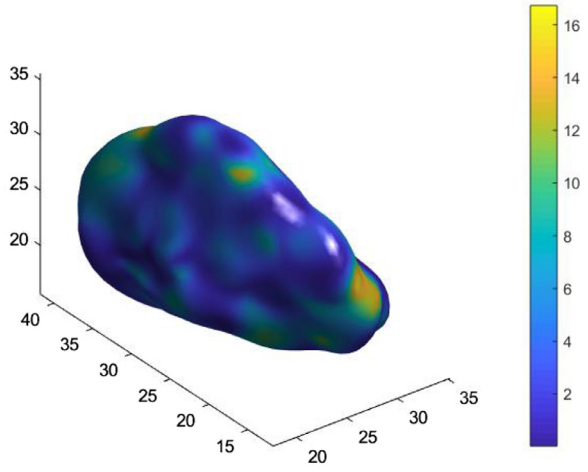

Fig. 7. Left putamen. A map view of the F-statistic of the difference in shape between the subjects in the RE group and the subjects in the non-RE group (the random field-based threshold corresponding to level α=0.05 is 33.93 and for level α=0.1 is 30.26, while the maximum *F*-statistic value is 16.73). The image has the same orientation as the reference (see Fig. 2).

value on the mean left putamen. No statistically significant differences in shape with α=0.05 are found when we test for group differences at each vertex of the putamen surface. However, we can check that discriminant analysis is worthwhile in this problem. In order to check that discrimination is worthwhile in this problem, we apply multivariate analysis of variance (MANOVA) to the PC scores selected to test the hypothesis of equality of group means, as explained in Mardia et al. (1979, Sect. 11.4). In this problem, the $p - value$ is 4-e8, which is small enough to reject the null hypothesis of equality of group means, and discriminant analysis is therefore worthwhile. We have therefore shown the superiority of our approach versus the use of the methodology in Chung et al. (2010) in this problem.

3.1.2. Laplace-Beltrami spectra

We compare our proposal with a methodology for global shape comparison based on the Laplace-Beltrami eigenvalues, as described by Reuter et al. (2006, 2009), Wachinger et al. (2015). The spectra for all the putamens are computed. To compute the eigenfunctions of the Laplace-Beltrami operator, we use the finite element method (FEM) described by Chung and Taylor (2004). Then, we consider a nearest neighbor algorithm (1-NN), with two distances for the eigenvalues that describe each shape, as explained by Wachinger et al. (2015). On the one hand, the Euclidean distance with a linear re-weighting of eigenvalues is considered. We call this procedure LB-WE. On the other hand, the Mahalanobis distance is considered. We call this procedure LB-M. We assess the performance by LOO cross-validation. The number of eigenvalues

used is selected by double or nested leave-one-out cross-validation. For the left putamen, the accuracy is 0.5758 (with 5 eigenvalues) for LB-WE and 0.7272 (with 27 eigenvalues) for LB-M, while for the right putamen, the accuracy is 0.4242 (for any number of eigenvalues considered) for LB-WE and 0.7272 (with 12 eigenvalues) for LB-M. Therefore, the performance of our proposal is better than that of LB-WE. For LB-M, the performance for the right putamen is equal to that of our proposal, but better than our proposal for the left putamen. In any case, the best classification rate is attained for the right putamen with both LB-M and FICA-SVS-LDA. However, LB-WE or LB-M are black box methodologies, since they do not explain why or how these differences occur; they are lacking in human interpretability, unlike our proposal. As discussed previously, it is desirable to have information that provides qualitative understanding (Hastie et al., 2009).

3.2. Limitations

We have seen the advantages of FPCA-SVS-LDA, in terms of both performance and interpretability, especially in comparison with other methodologies. As regards the limitations of our methodology, they may result from the limitations of LDA (Clemmensen et al., 2011) and the variable selection step. On the one hand, a situation where LDA can fail is when the groups cannot be separated by linear boundaries. Then SVS could still be used, but instead of LDA, quadratic discriminant analysis could perhaps be used. A difficult situation occurs when we have unbalanced groups, and the sample size of one of them is very small. In the most extreme case, a single observation per group is insufficient to compute LDA. Furthermore, the confidence in SVS decisions based on very few samples could decrease. On the other hand, the SVS used is based on Wilks' lambda criterion, which may not be the best option for non-Gaussian distributions.

Our methodology is based on local shape analysis rather than global shape analysis. In local shape analysis approaches, one-to-one correspondences between surfaces need to be established. This could be seen as a limitation, since global shape analysis approaches may need fewer pre-processing steps. However, our methodology yields spatially localized results that are straightforward to interpret, unlike global shape analysis approaches. Global shape analysis approaches can be seen as black boxes that do not explain their predictions in a way that humans can understand them easily. Nowadays, it is therefore preferable to use models that are inherently interpretable (Rudin, 2019).

4. Conclusions

We have proposed a methodology based on the use of SPHARM representation of brain structures for classification. The procedure is an improvement on that proposed by Epifanio and Ventura-Campos (2014). We have shown that our proposal not only performs well in terms of predictive power, but also yields interpretable classification in the high-dimensional setting. Furthermore, it has been applied to a novel classification problem in neuroeducation.

Although the procedure has been applied to a binary classification problem, it can be used in multiclass classification problems. In that case, more than one discriminant function can be obtained, as in classical LDA.

As future work, from the practical point of view, the proposed methodology can be applied to any classification problem in neuroscience where the anatomical structures can be expressed with SPHARM coefficients. From the theoretical point of view, other variable selection methods could be studied. Furthermore, we could extend the methodology to the problem of ordinal classification, i.e. when groups (categories or classes) are ordered. In addition, we could extend the methodology to combine not only functional data with SPHARM coefficients, but also multivariate features, such as variables related to education. In other words, we have to define FPCA for hybrid data with vector and multivariate functions, similarly to what Ramsay and Silverman (2005, Chapter 10) did for univariate functions. Finally, the FDA approach could be used not only for classification but also in other problems where the data are 3D brain structures.

Funding

This work has been partially supported by the following grants: DPI2017-87333-R from the Spanish Ministry of Science, Innovation and Universities (AEI/FEDER, EU) and UJI-A2017-8 and UJI-B2017-13 from Universitat Jaume I.

Declaration of Competing Interest

The authors have declared that no competing interests exist.

References

Alfeld, P., Neamtu, M., Schumaker, L.L., 1996. Fitting scattered data on sphere-like surfaces using spherical splines. J. Comput. Appl. Math. 73 (1–2), 5–43.
Anderson, J.R., Betts, S., Ferris, J., Fincham, J., 2012. Tracking children's mental states while solving algebra equations. Hum. Brain Mapp. 33 (11), 2650–2665.
Boulesteix, A.L., 2004. PLS Dimension reduction for classification with microarray data. Stat. Appl. Genet. Mol. Biol. 3 (1), 1–30.
Boulesteix, A.-L., Durif, G., Lambert-Lacroix, S., Peyre, J., Strimmer, K., 2018. plsgenomics: PLS analyses for genomics. R package version 1.5-2. https://CRAN.R-project.org/package=plsgenomics.
Brechbühler, C., Gerig, G., Kübler, O., 1995. Parametrization of closed surfaces for 3-d shape description. Comput. Vision Image Understanding 61 (2), 154–170.
Chung, M.K., Dalton, K.M., Shen, L., Evans, A.C., Davidson, R.J., 2007. Weighted Fourier series representation and its application to quantifying the amount of gray matter. IEEE Trans. Med. Imaging 26 (4), 566–581.
Chung, M.K., Taylor, J., 2004. Diffusion smoothing on brain surface via finite element method. In: 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro, pp. 432–435. Vol. 1
Chung, M.K., Worsley, K.J., Nacewicz, B.M., Dalton, K.M., Davidson, R.J., 2010. General multivariate linear modeling of surface shapes using surfstat. Neuroimage 53 (2), 491–505.
Clement, J., 1982. Algebra word problem solutions: thought processes underlying a common misconception. J. Res. Math. Educ. 13 (1), 16–30.
Clement, J., Lochhead, J., Monk, G.S., 1981. Translation difficulties in learning mathematics. The American Mathematical Monthly 88, 286. 04
Clement, J., Lochhead, J., Soloway, E., 1980. Positive Effects of Computer Programming on Students Understanding of Variables and Equations. In: Proceedings of the ACM, pp. 467–474.
Clemmensen, L., Hastie, T., Witten, D.B.E., 2011. Sparse discriminant analysis. Technometrics 53 (4), 406–413.
Cooper, M., 1986. The dependence of multiplicative reversal on equation format. Journal of Mathematical Behaviour 5 (2), 115–120.
Dryden, I.L., Mardia, K.V., 1998. Statistical shape analysis. Wiley, Chichester.
Epifanio, I., 2008. Shape descriptors for classification of functional data. Technometrics 50 (3), 284–294.
Epifanio, I., Ventura-Campos, N., 2011. Functional data analysis in shape analysis. Computational Statistics & Data Analysis 55 (9), 2758–2773.
Epifanio, I., Ventura-Campos, N., 2014. Hippocampal shape analysis in alzheimer's disease using functional data analysis. Stat. Med. 33 (5), 867–880.
Ferrando, L., 2019. Classificació d'estructures cerebrals en 3D amb anàlisi de dades funcionals. Universitat Jaume I. Aplicació a problemes verbals amb errors d'inversió. master's thesis.
Ferrando, L., Ventura-Campos, N., Epifanio, I., 2020. A neuroimaging data set on problem solving in the case of the reversal error: Putamen data. Data in Brief (submitted).
Ferraty, F., Vieu, P., 2006. Nonparametric functional data analysis: Theory and practice. Springer.
Gerardin, E., Chetelat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H., Niethammer, M., Dubois, B., Lehericy, S., Garnero, L., Eustache, F., Colliot, O., 2009. Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage 47 (4), 1476–1486.
Gerig, G., Styner, M., Jones, D., Weinberger, D., Lieberman, J., 2001. Shape Analysis of Brain Ventricles Using SPHARM. In: Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'01), pp. 171–178.
Golland, P., Grimson, W.E.L., Shenton, M.E., Kikinis, R., 2001. Deformation Analysis for Shape Based Classification. In: Proceedings of the 17th International Conference on Information Processing in Medical Imaging. IPMI '01, pp. 517–530.
González-Calero, J.A., Arnau, D., Laserna-Belenguer, B., 2015. Influence of additive and multiplicative structure and direction of comparison on the reversal error. Educational Studies in Mathematics 89 (1), 133–147.
Gu, X., Wang, Y., Chan, T.F., Thompson, P.M., Yau, S., 2004. Genus zero surface conformal mapping and its application to brain surface mapping. IEEE Trans. Med. Imaging 23 (8), 949–958.
Hall, P., Poskitt, D., Presnell, B., 2001. A functional data-analytic approach to signal discrimination. Technometrics 43, 1–9.

Hanakawa, T., Honda, M., Okada, T., Fukuyama, H., Shibasaki, H., 2003. Neural correlates underlying mental calculation in abacus experts: a functional magnetic resonance imaging study. Neuroimage 19 (2), 296–307.
Hand, D.J., 2006. Classifier technology and the illusion of progress. Statistical Science 21 (1), 1–14.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. Springer, New York.
He, Y., Li, X., Gu, X., Qin, H., 2005. Brain Image Analysis Using Spherical Splines. In: Proc. of Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 633–644.
Hyvärinen, A., Karhunen, J., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Networks 13, 411–430.
Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent component analysis. Wiley, New York.
Jolliffe, I.T., 2002. Principal component analysis, 2nd edition Springer.
Joshi, S.C., Miller, M.I., Grenander, U., 1997. On the geometry and shape of brain sub--manifolds. Internation Journal of Pattern Recognition and Artificial Intelligence 11 (8), 1317–1343.
Lazar, N., 2008. The statistical analysis of functional MRI data. statistics for biology and health. Springer, New York.
Lila, E., Aston, J.A.D., Sangalli, L.M., 2016. Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. Ann. Appl. Stat. 10 (4), 1854–1879.
Macdonald, D., Kabani, N., Avis, D., Evans, A.C., 2000. Automated 3-d extraction of inner and outer surfaces of cerebral cortex from MRI. Neuroimage 12 (3), 340–356.
Mardia, K., Kent, J., Bibby, J., 1979. Multivariate analysis. probability and mathematical statistics. Academic Press.
Millán-Roures, L., Epifanio, I., Martínez, V., 2018. Detection of anomalies in water networks by functional data analysis. Mathematical Problems in Engineering 2018. (Article ID 5129735), Article ID 5129735
Nain, D., Styner, M., Niethammer, M., Levitt, J.J., Shenton, M., Gerig, G., Tannenbaum, A., 2007. Statistical Shape Analysis of Brain Structures Using Spherical Wavelets. In: Proceedings of the Fourth IEEE International Symposium on Biomedical Imaging, pp. 209–212.
Park, Y., Priebe, C.E., Miller, M.I., Mohan, N.R., Botteron, K.N., 2008. Statistical analysis of twin populations using dissimilarity measurements in hippocampus shape space. Journal of Biomedicine and Biotechnology 2008. (Article ID 694297, doi:10.1155/2008/694297), Article ID 694297
Pierola, A., Epifanio, I., Alemany, S., 2016. An ensemble of ordered logistic regression and random forest for child garment size matching. Computers & Industrial Engineering 101, 455–465.
Qin, Y., Carter, C.S., Silk, E.M., Stenger, V.A., Fissell, K., Goode, A., Anderson, J.R., 2004. The change of the brain activation patterns as children learn algebra equation solving. Proceedings of the National Academy of Sciences of the United States of America 101, 5686–5691. 05
Ramsay, J.O., Silverman, B.W., 2002. Applied functional data analysis. Springer.
Ramsay, J.O., Silverman, B.W., 2005. Functional data analysis, 2nd edition Springer.
Reuter, M., Wolter, F.-E., Peinecke, N., 2006. Laplace-beltrami spectra as “shape-DNA” of surfaces and solids. Comput. Aided Des. 38 (4), 342–366.
Reuter, M., Wolter, F.-E., Shenton, M., Niethammer, M., 2009. Laplace Beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis. Comput.-Aided Des. 41 (10), 739–755.
Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1, 206–215.
R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/
Sandman, C.A., Head, K., Muftuler, L.T., Su, L., Buss, C., Davis, E.P., 2014. Shape of the basal ganglia in preadolescent children is associated with cognitive performance. Neuroimage 99, 93–102.
Shen, L., Ford, J., Makedon, F., Saykin, A., 2004. A surface-based approach for classification of 3D neuroanatomic structures. Intell. Data Anal. 8 (6), 519–542.
Shen, K., Fripp, J., Meriaudeau, F., Chetelat, G., Salvado, O., Bourgeat, P., 2012. Detecting global and local hippocampal shape changes in alzheimer’s disease using statistical shape models. Neuroimage 59 (3), 2155–2166.
Shen, L., Farid, H., McPeek, M., 2009. Modeling 3-dimensional morphological structures using spherical harmonics. Evolution 4 (63), 1003–1016.
Sjöstrand, K., Clemmensen, L., Larsen, R., Einarsson, G., Ersbøll, B., 2018. SpaSM: a MATLAB toolbox for sparse statistical modeling. J. Stat. Softw. 84 (10), 1–37.
Sørensen, H., Goldsmith, J., Sangalli, L.M., 2013. An introduction with medical applications to functional data analysis. Stat. Med. 32 (30), 5222–5240.
Styner, M., Gerig, G., Joshi, S.C., Pizer, S.M., 2003. Automatic and robust computation of 3D medial models incorporating object variability. Int. J. Comput. Vis. 55 (2–3), 107–122.
Styner, M., Lieberman, J.A., Pantazis, D., Gerig, G., 2004. Boundary and medial shape analysis of the hippocampus in schizophrenia. Medical Image Analysis Journal 8 (3), 197–203.
Taylor, J., Worsley, K., 2008. Random fields of multivariate test statistics, with applications to shape analysis. Ann. Stat. 36 (1), 1–27.
Tian, T.S., 2010. Functional data analysis in brain imaging studies. Front. Psychol. 1, 35.
Timsari, B., Leahy, R.M., 2000. Optimization method for creating semi-isometric flat maps of the cerebral cortex. In: Proc. SPIE, Medical Imaging 3979, 698–708.
Ullah, S., Finch, C.F., 2013. Applications of functional data analysis: asystematic review. BMC Med. Res. Methodol. 13 (1), 43.
Viviani, R., Grön, G., Spitzer, M., 2005. Functional principal component analysis of fMRI data. Hum. Brain Mapp. 24 (2), 109–129.
Wachinger, C., Golland, P., Kremen, W., Fischl, B., Reuter, M., 2015. Brainprint: a discriminative characterization of brain morphology. Neuroimage 109, 232–248.
Wang, J.-L., Chiou, J.-M., Müller, H.G., 2016. Functional data analysis. Annu. Rev. Stat. Appl. 3 (1), 257–295.
Weihs, C., Ligges, U., Luebke, K., Raabe, N., 2005. klaR Analyzing German Business Cycles. In: Data Analysis and Decision Support. Springer Berlin, Heidelberg, Berlin, Heidelberg, pp. 335–343.
Wollman, W., 1983. Determining the sources of error in a translation from sentence to equation. J. Res. Math. Educ. 14 (3), 169–181.
Yu, P., Grant, P.E., Qi, Y., Han, X., Ségonne, F., Pienaar, R., Busa, E., Pacheco, J., Makris, N., Buckner, R.L., Golland, P., Fischl, B., 2007. Cortical surface shape analysis based on spherical wavelets. IEEE Trans. Med. Imaging 26 (4), 582–597.