



# Functional Brain Imaging Synthesis Based on Image Decomposition and Kernel Modeling: Application to Neurodegenerative Diseases

## OPEN ACCESS

### Edited by:

Arjen van Ooyen,  
VU University Amsterdam,  
Netherlands

### Reviewed by:

Frithjof Kruggel,  
University of California, Irvine,  
United States  
Arun Bokde,  
Trinity College, Ireland

### \*Correspondence:

Francisco J. Martinez-Murcia  
fjesusmartinez@ugr.es

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

**Received:** 27 March 2017

**Accepted:** 02 November 2017

**Published:** 14 November 2017

### Citation:

Martnez-Murcia FJ, Górriz JM, Ramírez J, Illán IA, Segovia F, Castillo-Barnes D and Salas-Gonzalez D for the Alzheimer's Disease Neuroimaging Initiative (2017) Functional Brain Imaging Synthesis Based on Image Decomposition and Kernel Modeling: Application to Neurodegenerative Diseases. *Front. Neuroinform.* 11:65. doi: 10.3389/fninf.2017.00065

Francisco J. Martinez-Murcia<sup>1\*</sup>, Juan M. Górriz<sup>1</sup>, Javier Ramírez<sup>1</sup>, Ignacio A. Illán<sup>2</sup>, Fermín Segovia<sup>1</sup>, Diego Castillo-Barnes<sup>1</sup>, and Diego Salas-Gonzalez<sup>1</sup> for the Alzheimer's Disease Neuroimaging Initiative<sup>†</sup>

<sup>1</sup> Signal Processing and Biomedical Application, Department of Signal Theory, Networking and Communication, University of Granada, Granada, Spain, <sup>2</sup> Department of Scientific Computing, Florida State University, Tallahassee, FL, United States

The rise of neuroimaging in research and clinical practice, together with the development of new machine learning techniques has strongly encouraged the Computer Aided Diagnosis (CAD) of different diseases and disorders. However, these algorithms are often tested in proprietary datasets to which the access is limited and, therefore, a direct comparison between CAD procedures is not possible. Furthermore, the sample size is often small for developing accurate machine learning methods. Multi-center initiatives are currently a very useful, although limited, tool in the recruitment of large populations and standardization of CAD evaluation. Conversely, we propose a brain image synthesis procedure intended to generate a new image set that share characteristics with an original one. Our system focuses on nuclear imaging modalities such as PET or SPECT brain images. We analyze the dataset by applying PCA to the original dataset, and then model the distribution of samples in the projected eigenbrain space using a Probability Density Function (PDF) estimator. Once the model has been built, we can generate new coordinates on the eigenbrain space belonging to the same class, which can be then projected back to the image space. The system has been evaluated on different functional neuroimaging datasets assessing the: resemblance of the synthetic images with the original ones, the differences between them, their generalization ability and the independence of the synthetic dataset with respect to the original. The synthetic images maintain the differences between groups found at the original dataset, with no significant differences when comparing them to real-world samples. Furthermore, they featured a similar performance and generalization capability to that of the original dataset. These results prove that these images are suitable for standardizing the evaluation of CAD pipelines, and providing data augmentation in machine learning systems -e.g. in deep learning-, or even to train future professionals at medical school.

**Keywords:** Alzheimer's Disease (AD), Parkinson's Disease (PD), Neuroimaging, Synthesis, density estimation, data augmentation, validation, evaluation

## 1. INTRODUCTION

With the rise of neuroimaging in research and practice and the development of the machine learning paradigm, there has been an exponential trend in computer-aided methodologies (Frisoni et al., 2010; Martinez-Murcia et al., 2016; Rathore et al., 2017). Many Computer Aided Diagnosis (CAD) systems are being developed with application to structural and functional imaging in different diseases and disorders, such as Alzheimer's Disease (AD) (Stoeckel et al., 2004; Illán et al., 2011; Khedher et al., 2015) or Parkinson's Disease (PD) (Eckert and Edwards, 2007; Spetsieris et al., 2009; Segovia et al., 2016). However, these algorithms are usually tested in proprietary datasets to which the access is limited. This causes a series of problems in the evaluation of these systems, since a direct performance comparison is not enough to ensure validity. Furthermore, the false discovery rate (type I error) is often high in these studies due to a small sample size, significantly affecting reproducibility (Raudys and Jain, 1991; Poldrack et al., 2017).

A useful solution to perform direct comparison between different CADs is the use of multi-center datasets, and it has already been used in several challenges, such as the CAD Dementia challenge (Bron et al., 2015) or a recent Mild Cognitive Impairment (MCI) prediction challenge from Magnetic Resonance Imaging (MRI) at Kaggle (Sarica et al., 2016). Multi-center initiatives, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Weiner et al., 2012), the Autism Brain Imaging Data Exchange (ABIDE) (Di Martino et al., 2014) or the Parkinson's Progression Markers Initiative (PPMI) (Marek et al., 2011) provide large image samples that reduce the type I error and ease a direct comparison between systems. However, this approach poses some fundamental limitations. First of all, although open data is gaining support in the community (Poldrack and Gorgolewski, 2014), access to these large datasets requires the approval of principal investigators or their teams, a common problem in neuroscience (Ferguson et al., 2014). Secondly, focusing on one, static dataset such as the aforementioned ADNI or PPMI might increase overfitting, reducing their generalization ability. And finally, inhomogeneities in scanner, population, or techniques might cause the apparition of false positives that are not related to the signal (Pearlson, 2009; Van Horn and Toga, 2009).

For its part, synthetic datasets have been widely used in the pattern recognition community in different ways. Synthetic images have all the information about how they have been generated, and so can be used as a ground truth for automatic systems (Black et al., 2003; Ros et al., 2016; Varol et al., 2017). They can also significantly increase the sample size in a procedure commonly known as data augmentation, via analysis-synthesis (Cui et al., 2004) or performing deformations (Krizhevsky et al., 2012). This is a key feature that allowed a faster development of deep learning approaches that need huge amounts of data to build models.

In neuroimaging, some of the previous approaches have already percolated research practices. Recent machine learning challenges such as the aforementioned Kaggle MCI-MRI (Sarica et al., 2016), included synthesized morphological data to prevent

overfitting and *ad-hoc* model training. There are initiatives to build ground-truth phantoms for developing and studying new scanner technologies (Jan et al., 2004; Segars et al., 2010; Stute et al., 2011), to study generative models of functional activation in fMRI (Yarkoni et al., 2011; Erhardt et al., 2012) or to evaluate segmentation procedures such as the BrainWeb initiative (Kwan et al., 1999). However, brain image synthesis algorithms are rarely used for standardization or data augmentation, with few examples of spatial transformations and deformations in MRI (Zhu et al., 1994; Xue et al., 2006) or SPECT-DaTSCAN (Ronneberger et al., 2015; Martinez-Murcia et al., 2017).

In this work, we provide a novel brain synthesis technique (see **Figure 1**) to address the data augmentation and the ground truth problems in neuroimaging by generating a new set of images that share characteristics with a known dataset. Our system performs an analysis of that existing database and extracts a common orthogonal "eigenbrain" basis of a multidimensional space where the different subjects are represented as a coordinate vector, using Principal Component Analysis (PCA), as in many neuroimaging analysis and synthesis papers (Zhu et al., 1994; Markiewicz et al., 2009; Illán et al., 2011; Khedher et al., 2015). In contrast to them, our methodology creates per-class/modality models of the statistical distribution of the coordinates in this space. Together with a random sampling based on the Cumulative Density Function (CDF), it allows us to draw new uncorrelated coordinates from each model, thus setting the ground truth. This is of special importance where class clusters overlap due to the progression of the disease (e.g., MCI and AD), providing samples for standardizing the evaluation of CAD systems.

We have tested this generation approach against two widely used datasets: a Positron Emission Tomography (PET) dataset from the ADNI initiative, and a Single Photon Emission Computed Tomography (SPECT) dataset using the drug DaTSCAN, from the Parkinson's Progression Markers Initiative (PPMI). Our purpose is to assess that the intrinsic properties of each class in these modalities is kept, while providing uncorrelated images that can effectively predict real-world samples, thus reducing bias in the evaluation of new CAD methods. Additionally, the educational use of synthetic images to train future professionals is an inviting possibility yet to explore.

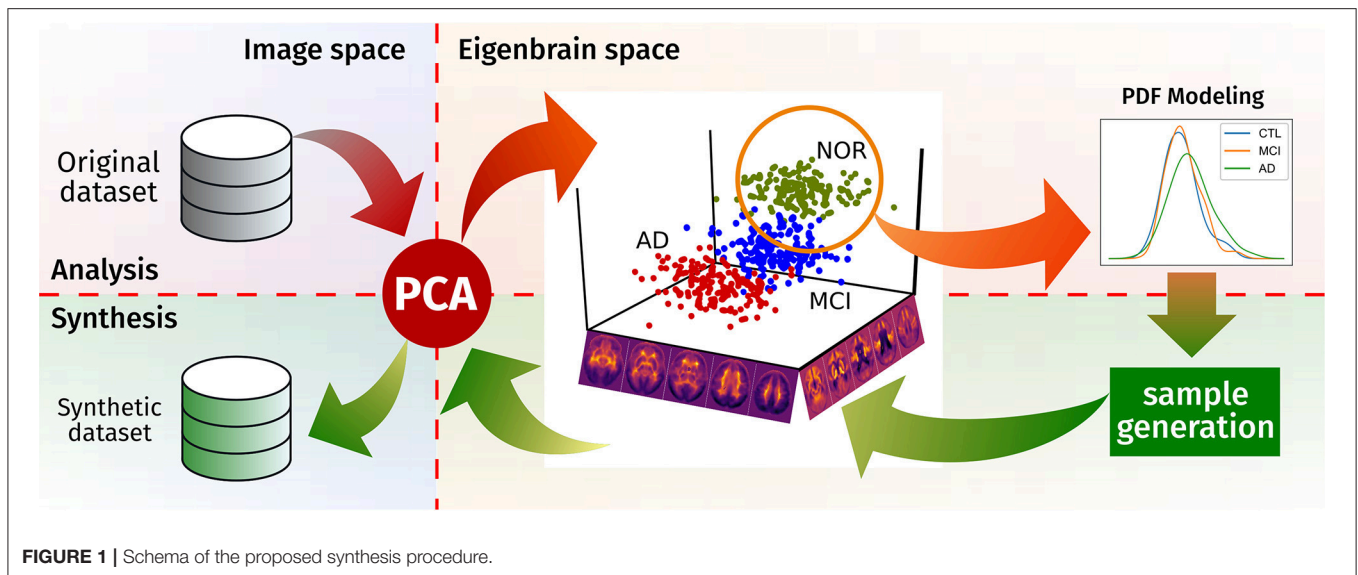
## 2. METHODOLOGY

### 2.1. Databases and Preprocessing

We tested our synthesis methodology on two large datasets comprising AD and PD:

#### 2.1.1. Alzheimer's Disease Neuroimaging Initiative (ADNI)

The first dataset used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) ([adni.loni.usc.edu](http://adni.loni.usc.edu)). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of Mild Cognitive Impairment (MCI)



and Alzheimer's Disease (AD). For up-to-date information on the data used, the recruitment process and the image properties, see [www.adni-info.org](http://www.adni-info.org) and Weiner et al. (2012).

For this work, we focus on the  $^{18}\text{F}$ -FDG PET images available. This radiopharmaceutical is a glucose analog, and its distribution of the brain can be used to trace glucose metabolism, and by extension, brain function. The ADNI PET subset here contains  $N = 403$  images: 95 images from individuals affected by AD, 207 images from MCI subjects and 101 images from Controls (NOR).

### 2.1.2. Parkinson's Progression Markers Initiative (PPMI)

Another large multi-center dataset used in the preparation of this article was obtained from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)). For up-to-date information on the study, visit [www.ppmi-info.org](http://www.ppmi-info.org) and (The Parkinson Progression Marker Initiative, 2010; Marek et al., 2011). The dataset contains SPECT images obtained using the radiopharmaceutical ioflupane, also known as its trade name DaTSCAN, which binds to the dopamine transporters at the striatum. This allows to quantify the dopaminergic deficit associated to Parkinson's Disease (PD). In this study we use 269 DaTSCAN images belonging to 111 NOR and 158 subjects affected by PD.

### 2.1.3. Database Preprocessing

The aforementioned datasets have been preprocessed to account for spatial and intensity differences. A spatial normalization, also known as registration, has been applied to ensure that the same MNI coordinate corresponds to the same spatial position inside the brain. We have used the SPM8 software (Friston et al., 2007) to perform this task, using either the included PET template (for the ADNI dataset) and a custom template (Salas-Gonzalez et al., 2015) for the PPMI and VDLN.

Later, intensity normalization was applied to ensure that a direct comparison of the image function encoded in the voxel intensities (dopamine transporters density in DaTSCAN and glucose metabolism in FDG-PET) is possible. This corrects the individual differences (e.g., drug uptake, exposition time, etc.) that affect these values, and is also key since the subsequent analysis needs directly comparable values to quantify variance. We have applied a normalization to the maximum strategy (Saxena et al., 1998) in the form  $\mathbf{I}' = \mathbf{I}/I_n$ , where the image intensity  $\mathbf{I}$  is divided by  $I_n$ , the average of the top 3% intensities. This normalization was proven very useful in Illán et al. (2011, 2012) and Martínez-Murcia et al. (2012, 2014).

## 2.2. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is used to establish a common reference, or basis, to generate new images. It is very extended in analysis and feature extraction in neuroimaging (Markiewicz et al., 2009; Illán et al., 2011; Khedher et al., 2015), and also was used with the same purpose of obtaining a common neuroimaging reference in Zhu et al. (1994).

Intuitively, PCA defines a new space where the first spatial direction is defined so that it explains the maximum variance in the data. The subsequent directions will try to explain the remaining variance in decreasing order. All these directions, or components, are meant to be uncorrelated. This way, the maximum information about the data is contained in the first components, and the remaining can be considered noise.

Mathematically (Brown, 2009), PCA works as an orthogonal transformation that maps a correlated set of observations  $\mathbf{X}$  (in this work, our set of zero-rated images, of size  $K \times N$  containing  $K$  images of length  $N$ ) into a set of uncorrelated data  $\mathbf{S}$  that contains  $K$  vectors in a  $M$ -dimensional space, defined by  $\mathbf{W}$ , a vector whose  $M$  columns contains the basis of the new space. That way, the mapping is obtained by:

$$\mathbf{S} = \mathbf{X}\mathbf{W} \quad (1)$$

where the columns of  $\mathbf{W}$  contain the eigen-values of  $\mathbf{X}^T\mathbf{X}$ , the empirical covariance matrix of  $\mathbf{X}$ .

This is done, ideally, by obtaining the eigen-value decomposition of the empirical covariance matrix of the data. A very extended and fast way of computing  $\mathbf{W}$ , the matrix of eigenvectors, also known as “eigenbrains” in the neuroscience literature (Illán et al., 2011), is via the Singular Value Decomposition (SVD) of  $\mathbf{X}$ :

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* \quad (2)$$

where  $\mathbf{U}$  is a  $K \times K$  orthogonal matrix,  $\mathbf{\Sigma}$  is a  $K \times M$  non-negative real diagonal matrix, and the  $M \times M$  unitary matrix  $\mathbf{V}^*$  denotes the conjugate transpose of the  $M \times M$  unitary matrix  $\mathbf{V}$ . Using this decomposition, we can rewrite Equation (1) as:

$$\mathbf{S} = \mathbf{X}\mathbf{W} = \mathbf{U}\mathbf{\Sigma} \quad (3)$$

Given that the conjugate transpose of a unitary matrix is its inverse, the matrix  $\mathbf{W}$  is equivalent to  $\mathbf{V}$ . A truncated version of the decomposition can also be performed, by retaining the first  $L$  components (ranked by their eigen-values):

$$\mathbf{S}_L = \mathbf{X}\mathbf{W}_L \quad (4)$$

where  $\mathbf{S}_L$  is the truncated estimate of size  $K \times L$ , and  $\mathbf{W}_L$  contains only the  $L$  first columns of  $\mathbf{W}$ . Since PCA does not account for random noise in its model, the noise is included as different components. Therefore, a choice for  $L$  can eliminate random noise and increase the signal to noise ratio of our model. In this work, when not stated, we will use the first  $L = K$  components (where  $K$  is the number of samples in the dataset).

## 2.3. Density Estimation

Once the images have been projected to the eigenbrain space, we want to generate new samples in this new space, in order to synthesize new images. To do so, we assume that the coordinates of the subjects of a certain class in the eigenbrain space are different realizations of a random process with a given Probability Density Function (PDF).

In order to estimate the PDF of the process that generates the coordinates of each class, we use two different density estimation procedures: a multivariate approach under the assumption of a normal distribution and a per-component estimation of density using the empirical Kernel Density Estimation.

### 2.3.1. Multivariate Normal Distribution (MVN)

The Multivariate Normal Distribution (MVN, also known as Multivariate Gaussian Distribution) is a generalization of the random normal distribution to  $n$  dimensions.

Let us note  $\mathbf{S}^c$  the matrix containing only the coordinates in the eigenbrain space of the  $K_c$  individuals belonging to class  $c$ . We can estimate its multivariate PDF  $\hat{f}_{mvn}^c(\mathbf{x})$  by computing the class mean  $\boldsymbol{\mu}^c$  and its class covariance matrix  $\boldsymbol{\Sigma}^c$ . The estimation of  $\boldsymbol{\Sigma}^c$  is performed via shrinkage, which consists in reducing the ratio between the smallest and the largest eigen-value of the empirical covariance matrix using a shrinkage parameter  $\alpha$ . In this work we used the method proposed in

Ledoit and Wolf (2004) to estimate an optimum  $\alpha$  that minimizes the Mean Squared Error between the estimated and the real covariance matrix. The multivariate PDF for class  $c$  would be:

$$\hat{f}_{mvn}^c(\mathbf{x}) = \frac{1}{(2\pi)^{K_c/2} |\boldsymbol{\Sigma}^c|^{1/2}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu}^c)^T \boldsymbol{\Sigma}^c^{-1} (\mathbf{x} - \boldsymbol{\mu}^c)}{2}\right) \quad (5)$$

### 2.3.2. Kernel Density Estimation (KDE)

Kernel Density Estimation (KDE) is an increasingly used method to estimate the PDF of a set of data (Botev et al., 2010; Simonoff, 2012). In this case, we perform a per-component estimation of the PDF. This approach disregards conditional probabilities between components and uses each component’s modeling as independent. While this is not theoretically accurate, the components extracted in PCA are uncorrelated by definition, and in practice the conditional terms are very small.

On the other hand, the per-component KDE is less prone to overfitting by disregarding these constraints. Additionally, the KDE can empirically account for heavy-tailed distributions that are sometimes more common in pathological models (Salas-Gonzalez et al., 2012), which might make this model more suitable.

An estimation of the PDF of the  $l$ th coordinate of class  $c$  is defined using KDE as:

$$\hat{f}_{kde}^{l,c}(x) = \frac{1}{K_c} \sum_{i=1}^{K_c} G_h(x - \mathbf{S}_l^{i,c}) = \frac{1}{K_c h} \sum_{i=1}^{K_c} G\left(\frac{x - \mathbf{S}_l^{i,c}}{h}\right), \quad (6)$$

where  $i = 1, \dots, K_c$ , with  $K_c$  the number of subjects belonging to class  $c$ ,  $h$  is the bandwidth and  $\mathbf{S}_l^{i,c}$  contains the  $l$ th coordinate of the  $i$ th subject of class  $c$ . The kernel  $G(x)$  is a function of  $\mathbb{R}^n$  that must define a probability:

$$\int \dots \int_{\mathbb{R}^n} G(\mathbf{x}) d\mathbf{x} = 1 \quad (7)$$

It also must be centered:

$$\int \dots \int_{\mathbb{R}^n} \mathbf{x}G(\mathbf{x}) d\mathbf{x} = \mathbf{0} \quad (8)$$

and its covariance matrix must be close to identity:

$$\forall \mathbf{u} \in \mathbb{R}^n, \|\mathbf{u}\| = 1 \quad \int_{\mathbb{R}} t^2 G(t\mathbf{u}) dt \approx 1 \quad (9)$$

In this work, we use a gaussian kernel  $G(x) = 1/(2\pi) \exp(-\frac{1}{2}x^2)$  for the estimate. Estimation of the bandwidth  $h$  is performed using the diffusion approximation proposed by Botev et al. (2010).

## 2.4. Brain Image Synthesis

After estimating the empirical PDF of the coordinates, we aim to generate a new set of coordinates for class  $c$   $\hat{\mathbf{S}}^c$  that match the distribution of the originals. To do so, we compute the

Cumulative Distribution Function (CDF) from the PDF that we estimated previously as:

$$F(x) = \int_{-\infty}^x f(t)dt \quad (10)$$

Afterwards, we can use a random number generator to provide uniformly distributed random numbers in the interval  $[0, 1]$ . These numbers are in the range of the CDF (from 0 to 1), and therefore we can consider them as  $F(x)$ , from which we could obtain the value  $x$ . In practice, we perform a numerical approximation to the problem, in which we calculate the full CDF in a wide range of  $x$ , and then interpolate the value of  $x$  using the generated  $F(x)$  as query point.

This procedure is repeated for all coordinates  $i = 1, \dots, K$  as many times as the number of subjects of class  $c$  that we want to synthesize. Then, the new set of images can be reconstructed using the eigenbrain basis  $\mathbf{W}$  and the new matrix of scores  $\widehat{\mathbf{S}}^c$ .

$$\widehat{\mathbf{X}}^c = \widehat{\mathbf{S}}^c \mathbf{W}^{-1} \quad (11)$$

The synthesis procedure defined here is available via the **brainSimulator** python package at <https://github.com/SiPBA/brainSimulator>, under the GPL-3+ license.

## 2.5. Validation

Validation of a synthetic dataset is still a matter of discussion. It depends heavily on the specific application of the synthesis. For example, for synthetic phantoms (Jan et al., 2004; Segars et al., 2010; Stute et al., 2011) and automatic segmentation methods (Ma et al., 1993; Kwan et al., 1999), visual inspection was used. Additionally, some studies performed measures of accuracy of segmentation, but that is not our case. Other studies applied spatial deformations for data augmentation (Zhu et al., 1994; Xue et al., 2006; Ronneberger et al., 2015; Martinez-Murcia et al., 2017), mostly without validating the deformed images.

The main purpose of this work is to generate images that could have been drawn from the same population of a given dataset, sharing relevant characteristics and, at the same time, being independent from the existing samples. Assessing this is not trivial. Therefore, we use two kinds of analyses:

- The well-known **Statistical Parametric Mapping** analysis (SPM) (Friston et al., 2007), to obtain a visual identification of the differences and similarities between classes and datasets. In particular, mass-univariate two-sample  $t$ -maps were obtained using the SPM12 software, using Family-wise error (FWE) correction  $t$ -threshold for a  $p < 0.05$ , with no masking applied.
- A classification analysis using **Voxel as Features** (VAF) (Stoeckel et al., 2004), which yields classification performance on different experimental setups. We cross-validate (10-fold) a Support Vector Machine Classifier (SVC) with linear kernel where the voxel intensities are used as features. To estimate the SVM regularization parameter ( $C$ ), we perform a grid search in an inner 5-fold cross-validation loop on the training set. The following performance values are provided: accuracy (acc), sensitivity (sens), specificity (spec), and their standard deviations (SD).

These analyses will be applied to evaluate the:

- **E1: Similarity and generalization ability** of the synthetic datasets via three different approaches:
  - **E1.1.** Assessing the **differences between classes** in both the original and the synthetic images using SPM and classification analysis. Similar performance results and SPM significant areas is expected if the original datasets are accurately modeled. We also study here the dependence of the synthetic images on the model parameters  $L$  (number of components) and  $N$  (number of generated subjects).
  - **E1.2.** Assessing the **differences between original and synthetic datasets**, using both SPM and classification analysis. If we assume that the synthetic datasets are a new sampling of the population from which the original datasets were drawn, there should be no significant differences between them.
  - **E1.3.** Evaluating the **predictive power of the synthetic images** on the original datasets. In this case, we use the synthetic images as a training set and test de trained model on real-world samples.
- **E2: Independence of the synthetic images.** To assess how much the synthesis procedure depends on the original dataset we choose a strategy based on the resubstitution error (Neto and Dougherty, 2015). Resubstitution (using the training set as test set) estimates the performance loss in the training. We will obtain the resubstitution error of the original training set and a synthetic dataset derived from that original training set. If this performance loss is similar to the performance loss in the original test set (E1.1), this means that we can consider that our synthesis algorithm produces images independent from the original set.

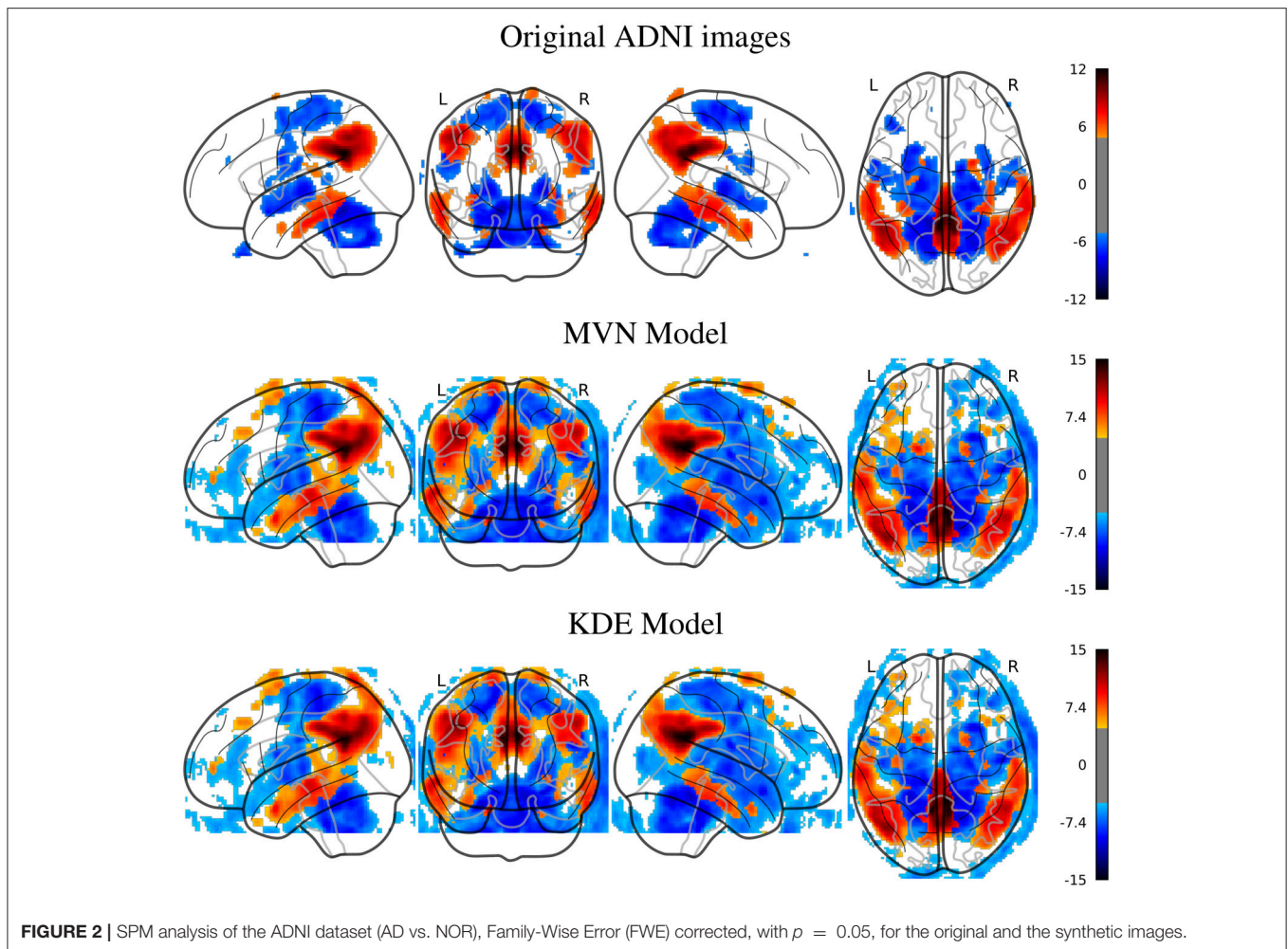
## 3. RESULTS

### 3.1. E1: Similarity and Generalization Ability of the Synthetic Datasets

#### 3.1.1. E1.1: Differences between Classes

The first approach we used to verify whether the synthetic images are similar to their original counterparts is to quantify and localize differences between classes in both the original and the synthetic datasets. We assess this similarity using both a map of the statistical differences between classes in all these datasets (SPM) and the classification performance when using the original and the synthetic images.

In **Figure 2**, we can look at the differences ( $t$ -maps, FWE corrected,  $p < 0.05$ ) between AD and NOR images with the original, and two synthesized datasets with 200 samples per class, using either MVN or KDE modeling. In these maps, the differences are located in similar places in both synthesized databases, that are as well, although less intense, represented in the SPM analysis of the original dataset. The aforementioned regions such as the precuneus, angular, mid-temporal lobe, hippocampus, amygdala, among others, are represented in both the original and synthetic datasets, with a special mention of the



cingulum, which also is the main difference in the MCI vs. NOR scenario (see Supplementary Material).

In the PPMI dataset, the SPM analysis revealed (**Figure 3**) that the main differences are located in the posterior part of the striatum, specifically at the posterior part of the putamen and globus pallidus. This behavior is consistent in both original and synthesized images, although with more statistical significance in the case of the MVN model, and a more homogeneous distribution of the negative differences under the KDE model.

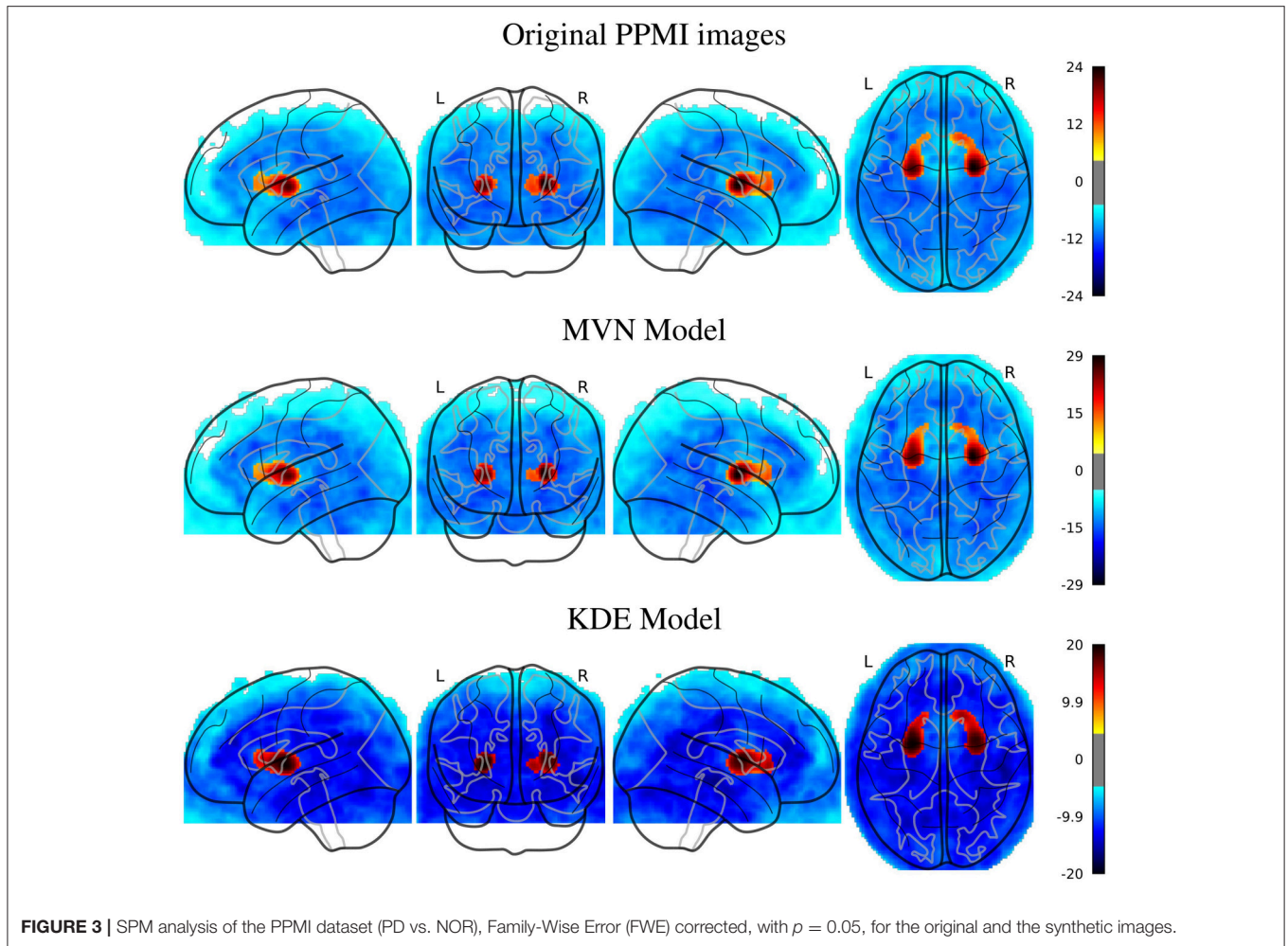
The choice of these two datasets was not random. They represent two extremes of the nuclear imaging spectrum: a very specific radiotracer (DaTSCAN) in two distinguishable states of a disease (PPMI) and a general radiopharmaceutical (FDG PET) in a progressive neurodegenerative disorder with very overlapping classes (ADNI). The separation between AD and NOR should be, therefore, higher than between the intermediate MCI and the two better-defined classes. However, even the AD and NOR classes overlap in the original dataset, perhaps due to a lower specificity of the biomarker (glucose metabolism) and noise in the diagnostic tests used to label the patients (Chapman et al., 2016).

A more descriptive analysis can be performed via the VAF performance of the original and synthetic datasets. These results

can be used to quantify the impact of some model parameters (for example, the number of subjects generated  $N$  or the number of components  $L$  used in the model) on the differences between classes. To do so, we first establish a baseline: the VAF performance of the original dataset, that can be checked at **Table 1**.

Now, two of the model parameters,  $L$  and  $N$  have been varied to explore their impact on the synthetic images. First, we vary  $L$ , the number of components used to synthesize images. It is logic to assume that a small number of components will be insufficient to acknowledge all variance in the original dataset, therefore producing low-quality images. On the other hand, a large number of components could lead to overfitting, especially in the MVN model, which takes into account not only the component scores' distribution, but also the relations between components.

This evolution is assessed in **Figure 4**, compared to the VAF performance of the original dataset (black dashed line) and standard deviation (shadowed area). There, we can check that our assumptions were accurate. Small  $L$ s lead to inaccurate models where the VAF performance is low. On the other hand, larger  $L$ s have a stronger impact on the MVN modeling, while the KDE performance remains almost unaltered after a certain  $L$ -value. This behavior might be an indication that the MVN approach



**FIGURE 3** | SPM analysis of the PPMI dataset (PD vs. NOR), Family-Wise Error (FWE) corrected, with  $p = 0.05$ , for the original and the synthetic images.

**TABLE 1** | Original VAF performance of the two datasets, including MCI scenario in ADNI.

Database	Scenario	acc [SD]	sens [SD]	spec [SD]
PET ADNI	AD vs. NOR	0.882 [0.012]	0.865 [0.091]	0.901 [0.118]
	MCI vs. NOR	0.698 [0.042]	0.791 [0.064]	0.504 [0.179]
	MCI vs. AD	0.702 [0.117]	0.444 [0.219]	0.822 [0.258]
DAT PPMI	PD vs. NOR	0.923 [0.057]	0.929 [0.090]	0.918 [0.088]

converges faster to an optimum modeling (approximate error 0.15 in ADNI and 0.08 in PPMI), and afterwards starts to produce more concentrated data clusters in the eigenbrain space due to the restrictions imposed by its  $L$ -dimensional nature. The larger the number of components  $L$  is, the more concentrated the data clusters might be, a possible indication of overfitting. We can also observe that the MVN surpasses the original VAF performance with a smaller  $L$ , which points to the fact that the variability in DaTSCAN imaging is far smaller than in PET, and therefore, less components are needed for an accurate model.

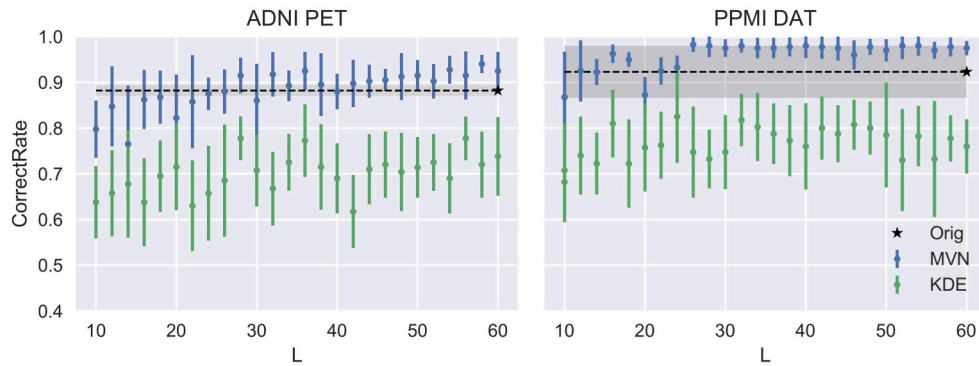
We can assume that the optimum model uses the  $L$  that produces the more detailed images while maintaining a similar

performance to the original dataset. In this work we propose to choose the highest  $L$  whose average performance remains within 1 standard deviation of the original dataset performance. In our case, it would be  $L = 40$  for the FDG-PET and  $L = 24$  for DaTSCAN. Now, by fixing these two  $L$ s, we will analyze how the VAF performance changes when increasing the number of synthetic images. We may assume that increasing  $N$  would reduce progressively the standard deviation of the performance estimates, converging toward the original VAF performance.

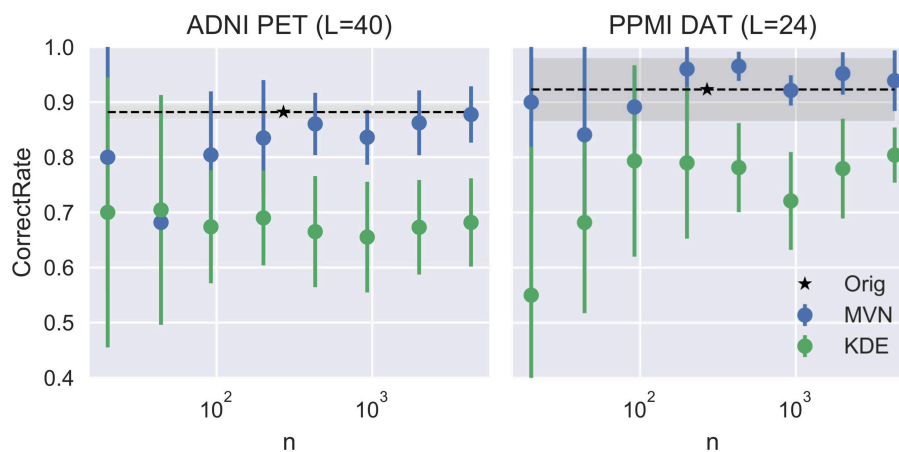
Figure 5 shows that, while the VAF performance of the MVN modeling increases and reduces its variance, eventually reaching the original performance, the KDE estimation hardly varies after a decent sample size (100 subjects, 50 subjects per class) has been reached.

### 3.2. E1.2: Differences between Original and Synthetic Images

It is now important to ensure that the synthetic images statistically belong to the same population from which the original images were drawn. To do so, we perform a SPM analysis (massive univariate  $t$ -test with FWE corrected with  $p = 0.05$ ) that compares the original datasets to:



**FIGURE 4** | Evolution of the accuracy when varying the number of components  $L$ .



**FIGURE 5** | Evolution of the accuracy when varying the number of synthetic images  $N$ .

- KDE-synthetic (with maximum  $L$ ) datasets.
- KDE-synthetic (with  $L = 40/24$  for ADNI/PPMI) datasets.
- MVN-synthetic (with maximum  $L$ ) datasets.
- MVN-synthetic (with  $L = 40/24$  for ADNI/PPMI) datasets.

Therefore, eight analyses were performed (four per dataset). None of these analyses yielded significant ( $p < 0.05$ , FWE) differences between the two populations, regardless of the model type (MVN or KDE) or database. Therefore, the group differences between original and synthetic images were relatively sparse and of small effect sizes. This might indicate that both the original and the synthetic images belong to the same population.

### 3.3. E1.3: Generalization Ability of the Synthetic Images

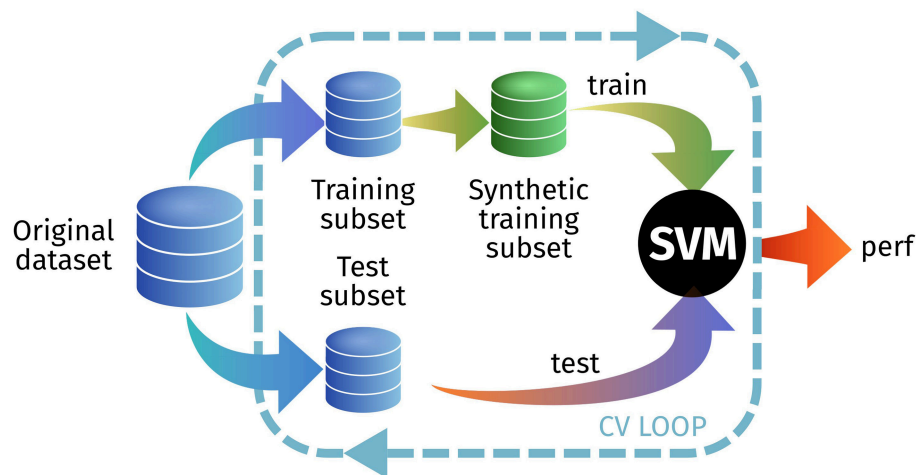
In the first experiment, we test how a model trained with synthetic images is able to predict real-world images. We do this by generating a synthetic training set from the original training set within the cross-validation loop, as seen in **Figure 6**. We tested this approach on the PET ADNI and the DAT PPMI

datasets, using all PDF estimators and different  $L$ -values. In each cross-validation iteration we synthesize a new set with 200 samples per class. Results are shown at **Table 2**.

The prediction accuracy, sensitivity and specificity are higher when using the MVN estimator than with the KDE in both ADNI and PPMI. The MVN modeling performance is also closer to the original baseline performance (see previous section), especially when using the suggested  $L = 40$  for AD and  $L = 24$  for PD. The KDE performance, however, degrades significantly when reducing the number of components.

Regarding the PET ADNI dataset, we tested three different scenarios: AD vs. NOR, MCI vs. NOR, and MCI vs. AD. The higher performance is obtained in the AD vs. NOR, but when including MCI subjects, the results vary. In the case of the MVN estimator, the MCI vs. NOR scenario performs slightly better than the MCI vs. AD. On the other hand, when using the KDE, MCI vs. AD obtains better results than MCI vs. NOR. However, the big difference among them is that, whereas the MVN estimator achieves similar performance to the baseline, the KDE-synthetic images lead to smaller predictive power.





**FIGURE 6** | Outline of the experimental setup for E1.3 to test the generalization ability of the synthetic images.

**TABLE 2** | Performance of E1.3, demonstrating the predictive ability of the synthetic images over the real dataset.

Database	Est.	$L$	Scenario	acc [SD]	sens [SD]	spec [SD]	
PET ADNI	MVN	403	AD vs. NOR	0.852 [0.078]	0.804 [0.151]	0.900 [0.137]	
			MCI vs. NOR	0.688 [0.067]	0.743 [0.108]	0.572 [0.169]	
			MCI vs. AD	0.675 [0.121]	0.468 [0.192]	0.774 [0.224]	
		40	AD vs. NOR	0.790 [0.084]	0.813 [0.165]	0.770 [0.207]	
			MCI vs. NOR	0.496 [0.139]	0.374 [0.359]	0.750 [0.392]	
			MCI vs. AD	0.622 [0.152]	0.563 [0.298]	0.651 [0.320]	
	KDE	403	AD vs. NOR	0.770 [0.114]	0.747 [0.153]	0.790 [0.169]	
			MCI vs. NOR	0.642 [0.044]	0.726 [0.098]	0.472 [0.188]	
		40	MCI vs. AD	0.672 [0.081]	0.484 [0.142]	0.760 [0.187]	
			AD vs. NOR	0.622 [0.114]	0.784 [0.290]	0.459 [0.394]	
DAT PPMI	MVN	268	PD vs. NOR	0.948 [0.041]	0.948 [0.055]	0.946 [0.064]	
			PD vs. NOR	0.925 [0.047]	0.910 [0.079]	0.945 [0.083]	
		24	PD vs. NOR	0.914 [0.082]	0.916 [0.113]	0.909 [0.121]	
	KDE		268	PD vs. NOR	0.914 [0.082]	0.916 [0.113]	0.909 [0.121]
			24	PD vs. NOR	0.833 [0.146]	0.834 [0.292]	0.819 [0.248]

### 3.4. E2: Dependence on the Original Images

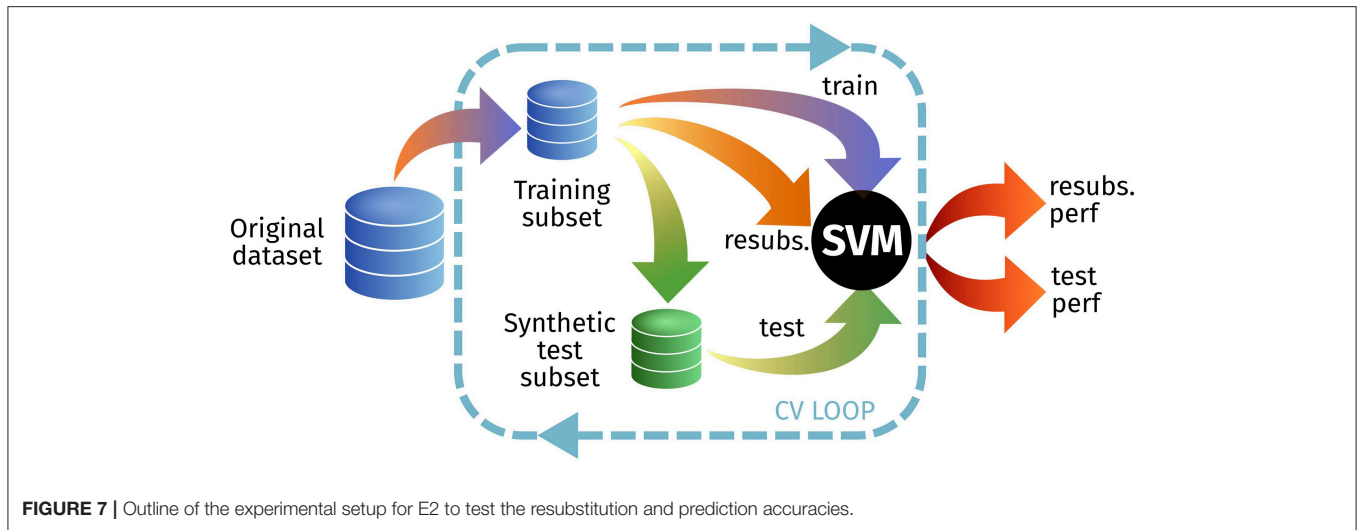
In experiment 2, we will assess the dependence of the synthetic images on the original datasets, using the resubstitution error (Neto and Dougherty, 2015). A classifier trained and tested on the same set (resubstitution) usually has a low error rate (resubstitution error or training error). It is generally assumed that the generalization error is the test error, or the error achieved by a test set, different from the training set.

In this experiment, we first estimate the resubstitution accuracy ( $1 - \text{resubstitution error}$ ) testing on the original training set, and then use that same set to synthesize (at different  $L$ s) a new test set (Figure 7). Depending on the performance loss, that could imply realistic images—different from the training set—or

images that are almost identical to the training set—overfitting. The results for this experiment are shown at Table 3.

When using all available components ( $L = 402$  for ADNI and  $L = 268$  for PPMI) we observe that the performance loss of the MVN estimator is almost null, probably due to the aforementioned overfitting. Conversely, the KDE model produces more different images, increasing the prediction error up to a 0.2 under all scenarios in ADNI and 0.15 in PPMI.

However, this situation changes completely when using the optimal  $L$  in the MVN synthesis. In this case, the performance loss is higher in all cases, demonstrating that the images are different from the original training set, and less concentrated than when using higher  $L$ s. Moreover, whereas the KDE achieved similar performance under all scenarios (with all  $L$ s), the MVN



**TABLE 3 |** Performance of E2, showing the resubstitution error (resubs.) and the test error of synthetic test sets with different  $L$ s.

Database	Scenario	Test (model, $L$ )	acc [SD]	sens [SD]	spec [SD]
PET ADNI	AD vs. NOR	resubs.	1.000 [0.000]	1.000 [0.000]	1.000 [0.000]
		MVN 402	0.994 [0.003]	0.988 [0.006]	1.000 [0.007]
		KDE 402	0.811 [0.014]	0.791 [0.021]	0.832 [0.028]
		MVN 40	0.906 [0.014]	0.882 [0.020]	0.930 [0.034]
	MCI vs. NOR	resubs.	0.999 [0.001]	0.998 [0.002]	1.000 [0.001]
		MVN 402	1.000 [0.000]	1.000 [0.000]	1.000 [0.000]
		KDE 402	0.823 [0.018]	0.843 [0.024]	0.801 [0.033]
		MVN 40	0.722 [0.021]	0.863 [0.024]	0.580 [0.143]
	MCI vs. AD	resubs.	0.692 [0.015]	0.829 [0.023]	0.556 [0.139]
		MVN 402	1.000 [0.000]	1.000 [0.000]	1.000 [0.000]
		KDE 402	0.820 [0.014]	0.799 [0.022]	0.841 [0.030]
		MVN 40	0.775 [0.025]	0.615 [0.047]	0.934 [0.163]
DAT PPMI	PD vs. NOR	resubs.	1.000 [0.000]	1.000 [0.000]	1.000 [0.000]
		MVN 268	0.998 [0.002]	0.997 [0.003]	1.000 [0.003]
		KDE 268	0.848 [0.013]	0.809 [0.021]	0.887 [0.043]
		MVN 24	0.966 [0.012]	0.961 [0.018]	0.971 [0.015]
		KDE 24	0.854 [0.024]	0.839 [0.035]	0.869 [0.035]

with optimal  $L$  replicates the baseline behavior found at **Table 1**; that is, a much higher discrimination ability in AD vs. NOR than when comparing to the prodromal state MCI. This is even true under the two MCI scenarios: the accuracy is higher under the MCI vs. AD scenario (0.775) than in the MCI vs. NOR (0.722), the same that occurred in the original dataset (0.702 and 0.698, respectively).

## 4. DISCUSSION

In this work, we propose an brain image synthesis algorithm that analyses a dataset, extracts its most relevant characteristics

and then generates new images that share the same properties. The algorithm is based on PCA, which defines a new common space for each dataset (the “eigenbrain” space), in which the individual images are represented as points. There, it models the distribution of these points, using them to construct a generative model in the eigenbrain space. After generating new samples in the eigenbrain space, these can be inversely transform to the image space, producing new samples of the same population (see **Figure 1** for a graphical representation of the procedure).

The use of PCA for feature extraction neuroimaging in components is widely documented (Zhu et al., 1994; Markiewicz et al., 2009; Illán et al., 2011; Khedher et al., 2015;

Martinez-Murcia et al., 2016). The PCA components model different orthogonal sources of variance in the original data. In AD, these components, or eigenbrains, represent features that have been associated with the progression of the disease. The contribution of each eigenbrain (or the coordinates of each subject in the eigenbrain space) has been proven to effectively model the advance of the disease in many works. PCA has also been used to build up a reference of the sources of variance in a generative model in Zhu et al. (1994), using a different simulation strategy. Therefore, it was the optimal tool, well-known and with proven utility, to be used in this work.

In **Figure 8** we show the first four eigenbrains (zero-centered) for the ADNI-PET dataset. The positive/negative contributions of each eigenbrain are shown in blue/red color, respectively. Each component accounts for different variability terms in the original dataset, for example, a negative contribution at the cingulate gyri, precuneus, and areas around the thalamus in Component 0 (Stoeckel et al., 2004; Illán et al., 2011), contrast between the anterior and posterior part of the brain in Component 1 and so on. Component 2 accounts for uptake differences at the angular gyrus and precuneus that have been linked to AD (see SPM analysis at section 3.1.1), also present in Component 3. These features prove that the computed eigenbrains are representative of independent structures and activities that together can positively influence the synthesis of new brain images via a correct parametrization and estimation of the component scores.

The distribution of the samples in the eigenbrain space was modeled using two estimation methods: a Multivariate Normal

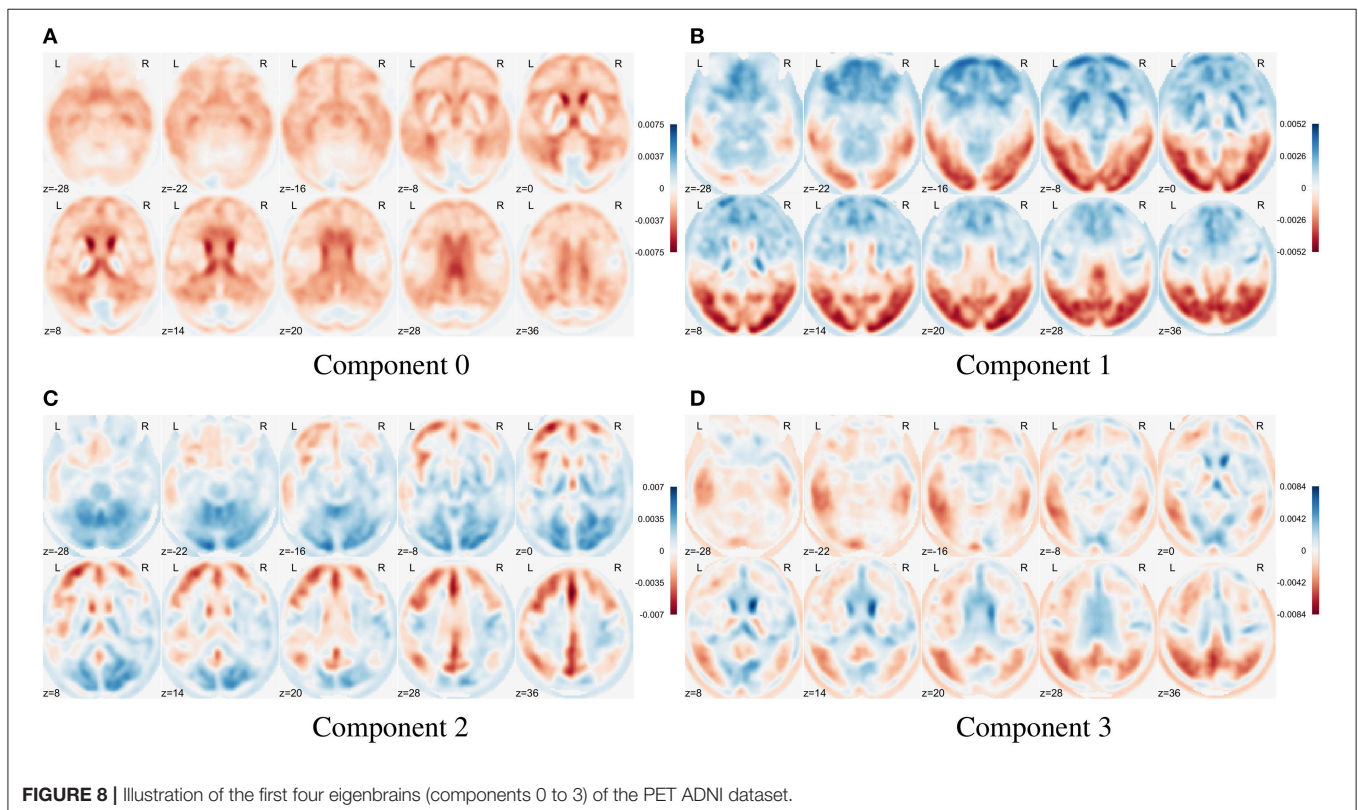
Distribution (MVN) and the Kernel Density Estimation (KDE) via diffusion. An accurate estimation of the distribution of the scores belonging to a certain class is paramount to obtain a reliable brain image synthesis.

In **Figure 9** we compare the two PDF estimation methods in two relevant eigenbrains: 2 and 10 (only AD and NOR groups are shown for simplicity), setting the original classes histogram as reference. Note that, since the MVN is multivariate in nature, we have projected the 2nd and 10th components to a single component for an estimated model of  $L = K$  components to obtain a clearer visualization.

Two major features are shown in these figures: class separability and quality of the modeling. As it can be seen, a large proportion of the variability contained in component 2 correspond to class differences (see also **Figure 8** and the positive weights of AD-related brain regions). These class differences are higher in the MVN model with  $L = 403$  (more concentrated classes) than in the KDE model. Furthermore, the KDE model adapts better to the empirical distribution, as it can be seen in component 10.

On the other hand, the scores distribution may not contain class differences, as in component 10. However, in this case, the distribution of scores in both classes is asymmetric, with longer tails and less gaussian than component 2. So, it is easier to assume that the KDE modeling will also perform better in this case.

Nevertheless, we cannot assume that the PCA components are statistically independent, and therefore, the KDE per-component model is deprecating substantial information. The multivariate



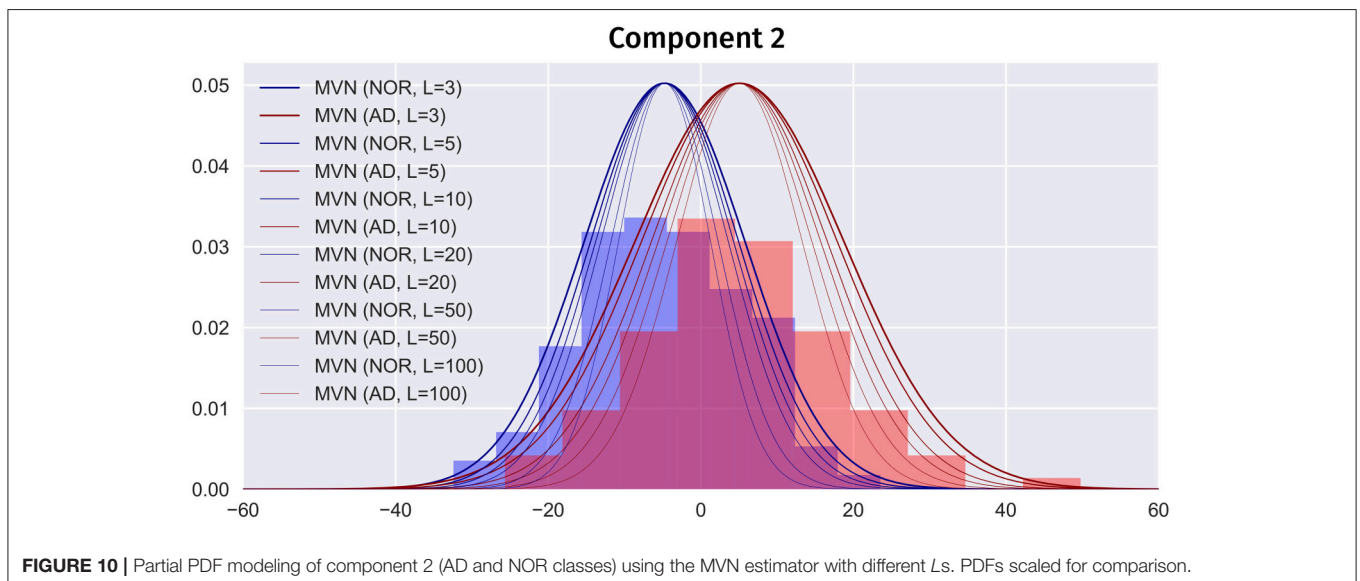
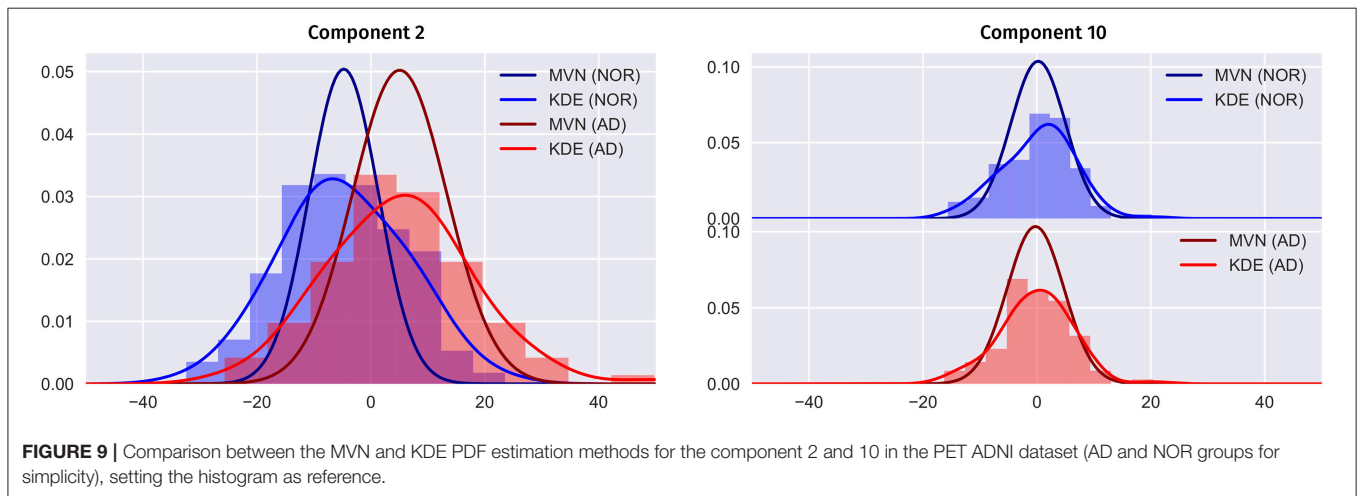
nature of the MVN does consider these possible dependences, which are then included in the model. This makes it more accurate but at the same time more prone to overfitting.

Overfitting in the MVN model means that, when  $L \rightarrow \infty$ , the model would only produce the average image of each class. So, choosing an optimum  $L$  has a strong impact on the simulated images. A higher  $L$  would contain more high frequency information, but it also overfits the model, producing images more similar to the average. On the other hand, small  $L$ s will lead to more overlapped classes, more similar to the real world, but less detail. **Figure 10** shows how increasing  $L$  affects the class overlapping, reducing the variability and eventually leading to images that are very close to each other. In this work, we carefully chose  $L = 40$  for ADNI and  $L = 24$  for PPMI by selecting the highest number of components that obtained similar performance to the original dataset. A significantly higher performance might be considered overfitting. Other approaches to select an optimum  $L$  such as the Variance of Reconstruction

Error (VRE) proposed in Mnassri et al. (2010) or the modified Bayesian model selection criterion of Kazianka and Pilz (2016) might be considered in the future.

A visual analysis of the synthetic images in **Figures 11, 12** reveals that the synthetic images preserve similar characteristics of the original datasets. For example, it is easy to appreciate differences in glucose metabolism in **Figure 11** typically associated with AD, such as a smaller activity at the temporal lobe or a less homogeneous distribution of the radiopharmaceutical in the gray matter (Stoeckel et al., 2004; Illán et al., 2011). In the synthetic DaTSCAN images (**Figure 12**) differences in shape and intensity of the striatum, and bilateral differences (Lozano et al., 2010; Towey et al., 2011; Illán et al., 2012; Martínez-Murcia et al., 2014) can also be noticed, using both MVN and KDE modeling, although variability of the PD class is perhaps better modeled in KDE. This is again a probable case of overfitting.

When analyzing the particularities of the MVN and KDE modeling under a classification analysis, we have already



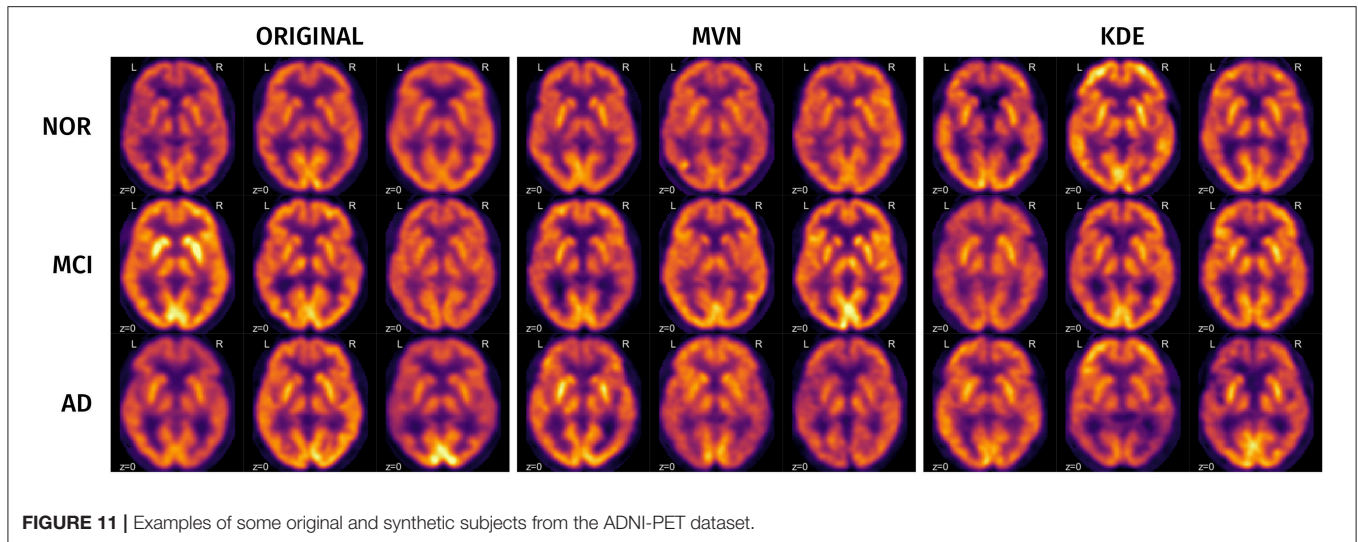


FIGURE 11 | Examples of some original and synthetic subjects from the ADNI-PET dataset.

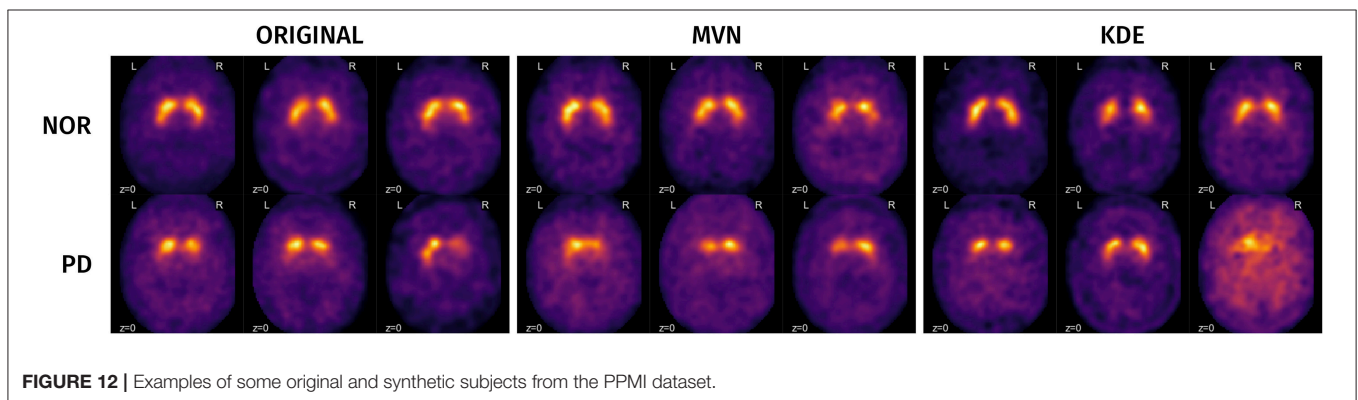


FIGURE 12 | Examples of some original and synthetic subjects from the PPMI dataset.

demonstrated that MVN tends heavily to overfitting when the number of components is high ( $L = K$ ), but it considers all conditional probabilities in the estimation. For its part, the KDE modeling is more robust to overfitting, mainly due to its univariate nature (per-component modeling), keeping a relevant prediction ability over real-world images, at the same time that its dependence on the original dataset is reduced. However, when an optimum  $L$  is chosen, the MVN model resembles more the original dataset performance in both the generalization ability (E1.3) and the dependence of the synthetic images on the original dataset (E2).

Each PDF estimation method has advantages and disadvantages. The KDE modeling works “out of the box,” producing more heterogeneous classes and images. On the other hand, the MVN modeling requires more fine-tuning of the  $L$  parameter, but it preserves conditional probabilities that may add relevant information to the synthetic images. New multidimensional PDF models, such as a multivariate KDE estimation or alpha-stable (heavy-tailed) modeling of the non-gaussian components, could also be used to preserve these conditional probabilities and improve each component modeling.

Still, evaluation and data availability are currently major bottlenecks for assessing validity and comparing machine learning approaches in neuroimaging, especially with the increasingly popular deep learning approaches. Testing on large samples of dynamically generated images that belong to the same population as the original (as in E2) can produce more realistic performance estimates, leading to a standardization of the evaluation of CAD systems that provides an idea of their generalizing capacity.

Our methodology provides functional brain images that could be drawn from the same population as the original dataset; images that can effectively predict real world samples at the same time that they remain independent from the database used in the modeling. Compared to other widely used data augmentation procedures, such as affine or elastic deformations, it is a more advanced paradigm that can simulate functional patterns associated with a given disease, which could increase the generalization ability of our models. This application is the main purpose of this paper, but apart from this there exist many application possibilities yet to be explored, e.g., using the synthetic images in clinical training of future professionals, or in standardized automatic evaluation procedures.

## 5. CONCLUSIONS

In this work, we have proposed a novel brain synthesis algorithm that could be used, among other things, in standardizing evaluation of CAD systems or as a data augmentation procedure. The algorithm consists of an analysis of an existing dataset using Principal Component Analysis, building a space defined by these principal components, or eigenbrains. In this space, we have modeled the spatial distribution of each class, a model from which we can derive new coordinates in the eigenbrain space that can be projected back into the original image space.

We have tested the algorithm in two well-known databases: one FDG-PET database from the Alzheimer's Disease Neuroimaging Initiative (ADNI), and a DaTSCAN SPECT database from the Parkinson's Progression Markers Initiative (PPMI). A visual analysis of the synthetic images revealed that they visually resemble the originals, sharing functional patterns that have been associated with Alzheimer's Disease and Parkinson's Disease in the literature. A Statistical Parametric Mapping analysis revealed similar regions in both the original and the synthetic datasets when studying the significant differences between classes.

We tested different features of the synthetic datasets under three experiments, aimed to prove their ability on detecting real-world image patterns and quantifying their dependence on the original dataset and the number of components used in the modeling. When comparing the two PDF estimation procedures, we found that the Multivariate Normal distribution (MVN) was more accurate, but also more affected by overfitting, whereas the synthesis using Kernel Density Estimation (KDE) produced more overlapped classes at any number of components considered, probably due to missing information about conditional probabilities. Our system, regardless of the PDF estimation, proved to be a useful tool for generating synthetic images that could be used for data augmentation, standardization of CAD system evaluation and even educational purposes.

## AUTHOR CONTRIBUTIONS

Conception or design of the work: FM-M, II, JG, and JR. Data collection: II, JR, JG, and DS-G. Data analysis and interpretation: FM, JG, and JR. Drafting of the article: FM-M, JG, and JR. Critical revision of the article: FS, JG, JR, II, FS, DC-B, and DS-G. Major revision of the article: FM-M, JR, and JG.

## FUNDING

This work was partly supported by the MINECO/ FEDER under the TEC2015-64718-R project and the Consejería de Economía,

Innovación, Ciencia y Empleo (Junta de Andalucía, Spain) under the Excellence Project P11-TIC- 7103.

## ACKNOWLEDGMENTS

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

PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including AbbVie; Avid Radiopharmaceuticals; Biogen; Bristol-Myers Squibb Company; Covance; GE Healthcare; Genentech, Inc.; GlaxoSmithKline; Eli Lilly & Co.; Lundbeck; Merck; Meso Scale Discovery; Pfizer Inc.; Piramal Imaging; Roche CNS group; Sanofi Genzyme; Servier; Teva; UCB; and Golub Capital.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fninf.2017.00065/full#supplementary-material>

Supplementary figures of the SPM analysis of all scenarios and datasets, including the original montages of SPM12 and the glass brain visualization.

## REFERENCES

Black, J., Ellis, T., and Rosin, P. (2003). "A novel method for video tracking performance evaluation," in *Proceedings of the IEEE International Workshop on*

*Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS 03)* (Nice), 125–132.

Botev, Z., Grotowski, J., and Kroese, D. (2010). Kernel density estimation via diffusion. *Ann. Stat.* 38, 2916–2957. doi: 10.1214/10-AOS799

- Bron, E. E., Smits, M., Van Der Flier, W. M., Vrenken, H., Barkhof, F., Scheltens, P., et al. (2015). Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the caddementia challenge. *Neuroimage* 111, 562–579. doi: 10.1016/j.neuroimage.2015.01.048
- Brown, J. D. (2009). Principal components analysis and exploratory factor analysis—definitions, differences, and choices definitions, differences, and choices. *Shiken JALT Test. Eval. SIG Newsllett.* 13, 26–30.
- Chapman, K. R., Bing-Canar, H., Alosco, M. L., Steinberg, E. G., Martin, B., Chaisson, C., et al. (2016). Mini mental state examination and logical memory scores for entry into Alzheimer's disease trials. *Alzheimers Res. Ther.* 8:9. doi: 10.1186/s13195-016-0176-z
- Cui, J., Wang, Y., Huang, J., Tan, T., and Sun, Z. (2004). "An iris image synthesis method based on PCA and super-resolution," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, Vol. 4 (Cambridge, UK: IEEE), 471–474.
- Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* 19, 659–667. doi: 10.1038/mp.2013.78
- Eckert, T., and Edwards, C. (2007). The application of network mapping in differential diagnosis of parkinsonian disorders. *Clin. Neurosci. Res.* 6, 359–366. doi: 10.1016/j.cnr.2007.05.001
- Erhardt, E. B., Allen, E. A., Wei, Y., Eichele, T., and Calhoun, V. D. (2012). SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability. *Neuroimage* 59, 4160–4167. doi: 10.1016/j.neuroimage.2011.11.088
- Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E., and Martone, M. E. (2014). Big data from small data: data-sharing in the 'long tail' of neuroscience. *Nat. Neurosci.* 17, 1442–1447. doi: 10.1038/nn.3838
- Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., and Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* 6, 67–77. doi: 10.1038/nrneurol.2009.215
- Friston, K., Ashburner, J., Kiebel, S., Nichols, T., and Penny, W. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. London, UK: Academic Press.
- Illán, I., Górriz, J., Ramírez, J., Segovia, F., Jiménez-Hoyuela, J., and Ortega Lozano, S. (2012). Automatic assistance to Parkinsons disease diagnosis in DaTSCAN SPECT imaging. *Med. Phys.* 39, 5971–5980. doi: 10.1118/1.4742055
- Illán, I. A., Górriz, J. M., Ramírez, J., Salas-Gonzalez, D., López, M. M., Segovia, F., et al. (2011). <sup>18</sup>F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis. *Inform. Sci.* 181, 903–916. doi: 10.1016/j.ins.2010.10.027
- Jan, S., Santin, G., Strul, D., Staelens, S., Assie, K., Autret, D., et al. (2004). Gate: a simulation toolkit for pet and spect. *Phys. Med. Biol.* 49:4543. doi: 10.1088/0031-9155/49/19/007
- Kazianka, H., and Pilz, J. (2016). A corrected criterion for selecting the optimum number of principal components. *Aust. J. Stat.* 38, 135–150. doi: 10.17713/ajs.v38i3.268
- Khedher, L., Ramirez, J., Górriz, J., Brahim, A., and Segovia, F. (2015). Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing* 151, 139–150. doi: 10.1016/j.neucom.2014.09.072
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (Stateline, NV), 1097–1105.
- Kwan, R.-S., Evans, A. C., and Pike, G. B. (1999). MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans. Med. Imaging* 18, 1085–1097. doi: 10.1109/42.816072
- Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* 88, 365–411. doi: 10.1016/S0047-259X(03)00096-4
- Lozano, S. O., Torres, M. M. d. V., Moreno, E. R., Viedma, S. S., Raissouni, T. A., and Jiménez-Hoyuela, J. (2010). [Quantitative evaluation of SPECT with FP-CIT. Importance of the reference area]. *Rev. Espa. Med. Nucl.* 29, 246–250. doi: 10.1016/j.remnu.2010.03.006
- Ma, Y., Kamber, M., and Evans, A. (1993). 3D simulation of pet brain images using segmented mri data and positron tomograph characteristics. *Comput. Med. Imaging Graphics* 17, 365–371. doi: 10.1016/0895-6111(93)90030-Q
- Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., et al. (2011). The Parkinson Progression Marker Initiative (PPMI). *Progr. Neurobiol.* 95, 629–635. doi: 10.1016/j.pneurobio.2011.09.005
- Markiewicz, P., Matthews, J., Declerck, J., and Herholz, K. (2009). Robustness of multivariate image analysis assessed by resampling techniques and applied to FDG-PET scans of patients with Alzheimer's disease. *Neuroimage* 46, 472–485. doi: 10.1016/j.neuroimage.2009.01.020
- Martinez-Murcia, F., Górriz, J., and Ramirez, J. (2016). "Computer aided diagnosis in neuroimaging," in *Computer-aided Technologies - Applications in Engineering and Medicine*, ed R. Udroui (London, UK: InTech), 137–160.
- Martinez-Murcia, F., Górriz, J., Ramirez, J., Moreno-Caballero, M., Gómez-Río, M., Initiative, P. P. M., et al. (2014). Parametrization of textural patterns in 123I-ioflupane imaging for the automatic detection of Parkinsonism. *Med. Phys.* 41:012502. doi: 10.1118/1.4845115
- Martinez-Murcia, F., Górriz, J., Ramirez, J., Puntinet, C., and Salas-González, D. (2012). Computer aided diagnosis tool for Alzheimer's disease based on Mann-Whitney-Wilcoxon U-test. *Exp. Syst. Appl.* 39, 9676–9685. doi: 10.1016/j.eswa.2012.02.153
- Martinez-Murcia, F. J., Ortiz, A., Górriz, J. M., Ramirez, J., Segovia, F., Salas-Gonzalez, D., et al. (2017). "A 3D convolutional neural network approach for the diagnosis of Parkinson's disease," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*, Vol. 10337, *Lecture Notes in Computer Science*, eds J. Ferrández Vicente, J. Álvarez Sánchez, F. de la Paz López, J. Toledo-Moreo, and H. Adeli (A Coruña: Springer), 324–333.
- Mnassri, B., Ouladsine, M., El Adel, E. M., and Ananou, B. (2010). "Selection of the number of principal components based on the fault reconstruction approach applied to a new combined index," in *2010 49th IEEE Conference on Decision and Control (CDC)* (Atlanta, GA: IEEE), 3307–3312.
- Neto, U. M. B., and Dougherty, E. R. (2015). *Error Estimation for Pattern Recognition*. Hoboken, NJ: John Wiley & Sons.
- Pearlson, G. (2009). Multisite collaborations and large databases in psychiatric neuroimaging: advantages, problems, and challenges. *Schizophr. Bull.* 35, 1–2. doi: 10.1093/schbul/sbn166
- Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., et al. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* 18, 115–126. doi: 10.1038/nrn.2016.167
- Poldrack, R. A., and Gorgolewski, K. J. (2014). Making big data open: data sharing in neuroimaging. *Nat. neurosci.* 17, 1510–1517. doi: 10.1038/nn.3818
- Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., and Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage* 155, 530–548. doi: 10.1016/j.neuroimage.2017.03.057
- Raudys, S., and Jain, A. (1991). Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans. Patt. Anal. Mach. Intell.* 13, 252–264. doi: 10.1109/34.75512
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597*.
- Ros, G., Sellart, L., Materzynska, J., Vazquez, D., and Lopez, A. (2016). "The SYNTHIA Dataset: a large collection of synthetic images for semantic segmentation of urban scenes," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: IEEE), 3234–3243.
- Salas-Gonzalez, D., Gorriz, J., Ramirez, J., Martinez, F., Chaves, R., Segovia, F., et al. (2012). "Intensity normalization of FP-cit spect in patients with parkinsonism using the  $\alpha$ -stable distribution," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2012 IEEE (New York, NY: IEEE), 3944–3946.
- Salas-Gonzalez, D., Górriz, J. M., Ramirez, J., Illán, I. A., Padilla, P., Martínez-Murcia, F. J., et al. (2015). Building a FP-CIT SPECT brain template using a posterization approach. *Neuroinformatics* 13, 391–402. doi: 10.1007/s12021-015-9262-9
- Sarica, A., Cerasa, A., Quattrone, A., Calhoun, V., and for the Alzheimer's Disease Neuroimaging Initiative (2016). A machine learning neuroimaging challenge for automated diagnosis of mild cognitive impairment. *J. Neurosci. Methods*. Available online at: <https://www.kaggle.com/c/mci-prediction>
- Saxena, P., Pavel, D. G., Quintana, J. C., and Horwitz, B. (1998). "An automatic threshold-based scaling method for enhancing the usefulness of Tc-HMPAO SPECT in the diagnosis of Alzheimer's disease," in *Medical Image Computing*

- and *Computer-Assisted Intervention - MICCAI, Vol. 1496, Lecture Notes in Computer Science*, (Cambridge MA: Springer), 623–630.
- Segars, W., Sturgeon, G., Mendonca, S., Grimes, J., and Tsui, B. M. (2010). 4D XCAT phantom for multimodality imaging research. *Med. Phys.* 37, 4902–4915. doi: 10.1118/1.3480985
- Segovia, F., Gorriz, J., Ramírez, J., and Salas-Gonzalez, D. (2016). “Multiclass classification of 18 F-DMFP-PET data to assist the diagnosis of Parkinsonism,” in *2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI)* (Trento: IEEE), 1–4.
- Simonoff, J. S. (2012). *Smoothing Methods in Statistics*. New York, NY: Springer Science & Business Media.
- Spetsieris, P. G., Ma, Y., Dhawan, V., and Eidelberg, D. (2009). Differential diagnosis of Parkinsonian syndromes using functional PCA-based imaging features. *Neuroimage* 45, 1241–1252. doi: 10.1016/j.neuroimage.2008.12.063
- Stoeckel, J., Ayache, N., Malandain, G., Koulibaly, P. M., Ebmeier, K. P., and Darcourt, J. (2004). “Automatic classification of SPECT images of Alzheimer’s disease patients and control subjects,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI, Vol. 3217, Lecture Notes in Computer Science* (St. Malo: Springer), 654–662.
- Stute, S., Carlier, T., Cristina, K., Noblet, C., Martineau, A., Hutton, B., et al. (2011). Monte Carlo simulations of clinical PET and SPECT scans: impact of the input data on the simulated images. *Phys. Med. Biol.* 56, 6441–6457. doi: 10.1088/0031-9155/56/19/017
- The Parkinson Progression Marker Initiative (2010). *Imaging Technical Operations Manual, 2nd Edn*. PPMI.
- Towey, D. J., Bain, P. G., and Nijran, K. S. (2011). Automatic classification of 123I-FP-CIT (DaTSCAN) SPECT images. *Nucl. Med. Commun.* 32, 699–707. doi: 10.1097/MNM.0b013e328347cd09
- Van Horn, J. D., and Toga, A. W. (2009). Multi-site neuroimaging trials. *Curr. Opin. Neurol.* 22, 370–378. doi: 10.1097/WCO.0b013e32832d92de
- Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M. J., Laptev, I., et al. (2017). “Learning from synthetic humans,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)* (Honolulu, HI).
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2012). The Alzheimer’s Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimers Dement.* 8(1 Suppl.), S1–S68. doi: 10.1016/j.jalz.2011.09.172
- Xue, Z., Shen, D., Karacali, B., Stern, J., Rottenberg, D., and Davatzikos, C. (2006). Simulating deformations of MR brain images for validation of atlas-based segmentation and registration algorithms. *NeuroImage* 33, 855–866. doi: 10.1016/j.neuroimage.2006.08.007
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670. doi: 10.1038/nmeth.1635
- Zhu, X. P., Hutchinson, C. E., Hawnaur, J. M., Cootes, T. F., Taylor, C. J., and Isherwood, I. (1994). Magnetic resonance image synthesis using a flexible model. *Brit. J. Radiol.* 67, 976–982.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Martinez-Murcia, Gorriz, Ramírez, Illán, Segovia, Castillo-Barnes, and Salas-Gonzalez for the Alzheimer’s Disease Neuroimaging Initiative. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.