



**HAL**  
open science

## Anomaly detection for the individual analysis of brain PET images

Ninon Burgos, Jorge M. Cardoso, Jorge Samper-González, Marie-Odile Habert, Stanley Durrleman, Sébastien Ourselin, Olivier Colliot

► **To cite this version:**

Ninon Burgos, Jorge M. Cardoso, Jorge Samper-González, Marie-Odile Habert, Stanley Durrleman, et al.. Anomaly detection for the individual analysis of brain PET images. *Journal of Medical Imaging*, 2021, 8 (02), pp.024003. 10.1117/1.JMI.8.2.024003 . hal-03193306

**HAL Id: hal-03193306**

**<https://inria.hal.science/hal-03193306v1>**

Submitted on 8 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Anomaly detection for the individual analysis of brain PET images

Ninon Burgos<sup>a,b,c,d,e,f,\*</sup>, M. Jorge Cardoso<sup>g</sup>, Jorge Samper-González<sup>a,b,c,d,e,f</sup>,  
Marie-Odile Habert<sup>h,i,j</sup>, Stanley Durrleman<sup>f,a,b,c,d,e</sup>, Sébastien Ourselin<sup>g</sup>,  
Olivier Colliot<sup>a,b,c,d,e,f</sup>, for the Alzheimer's Disease Neuroimaging Initiative<sup>†</sup>, and the  
Frontotemporal Lobar Degeneration Neuroimaging Initiative<sup>‡</sup>

<sup>a</sup>Paris Brain Institute, Hôpital Pitié-Salpêtrière, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>b</sup>Inserm, U 1127, Hôpital Pitié-Salpêtrière, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>c</sup>CNRS, UMR 7225, Hôpital Pitié-Salpêtrière, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>d</sup>Sorbonne Université, Hôpital Pitié-Salpêtrière, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>e</sup>AP-HP, Hôpital Pitié-Salpêtrière, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>f</sup>Inria, Aramis project-team, Hôpital Pitié-Salpêtrière, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>g</sup>Department of Imaging and Biomedical Engineering, King's College London, Strand, London WC2R 2LS, UK

<sup>h</sup>AP-HP, Pitié-Salpêtrière Hospital, Department of Nuclear Medicine, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>i</sup>Laboratoire d'Imagerie Biomédicale, Sorbonne Université, Inserm U 1146, CNRS UMR 7371, Hôpital Pitié-Salpêtrière, 47 boulevard de l'hôpital, 75013 Paris, France

<sup>j</sup>Centre Acquisition et Traitement des Images, Hôpital Pitié-Salpêtrière, 47 boulevard de l'hôpital, 75013 Paris, France

## Abstract.

**Purpose:** In clinical practice, positron emission tomography (PET) images are mostly analysed visually, but the sensitivity and specificity of this approach greatly depends on the observer's experience. Quantitative analysis of PET images would alleviate this problem by helping define an objective limit between normal and pathological findings. We present an anomaly detection framework for the individual analysis of PET images.

**Approach:** We created subject-specific abnormality maps that summarise the pathology's topographical distribution in the brain by comparing the subject's PET image to a model of healthy PET appearance that is specific to the subject under investigation. This model was generated from demographically and morphologically-matched PET scans from a control dataset.

**Results:** We generated abnormality maps for healthy controls, patients at different stages of Alzheimer's disease and with different frontotemporal dementia syndromes. We showed that no anomalies were detected for the healthy controls and that the anomalies detected from the patients with dementia coincided with the regions where abnormal uptake was expected. We also validated the proposed framework using the abnormality maps as inputs of a classifier and obtained higher classification accuracies than when using the PET images themselves as inputs.

**Conclusions:** The proposed method was able to automatically locate and characterise the areas characteristic of dementia from PET images. The abnormality maps are expected to i) help clinicians in their diagnosis by highlighting, in a data-driven fashion, the pathological areas, and ii) improve the interpretability of subsequent analyses, such as computer-aided diagnosis or spatio-temporal modelling.

**Keywords:** Anomaly detection, Dementia, Image synthesis, Positron emission tomography.

\*Ninon Burgos, [ninon.burgos@inria.fr](mailto:ninon.burgos@inria.fr)

†Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://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).

‡Data used in preparation of this article were obtained from the Frontotemporal Lobar Degeneration Neuroimaging Initiative (FTLDNI) database (<http://4rtmi-ftldni.ini.usc.edu/>). The investigators at NIFD/FTLDNI contributed to the design and implementation of FTLDNI and/or provided data, but did not participate in analysis or writing of this report (unless otherwise listed).

## 1 Introduction

Neurological diseases such as epilepsy<sup>1</sup> or dementia<sup>2,3</sup> show heterogeneous patterns of anomalies on neuroimages, for example positron emission tomography (PET) images. These specific patterns of anomaly are important to distinguish between different syndromes and establish an accurate diagnosis.<sup>4-9</sup> In clinical practice, PET images are mostly analysed visually. The sensitivity and specificity of this approach greatly depends on the observer's experience and is not in favour of centres where advanced expertise in image reading is unavailable.<sup>10</sup> Quantitative analysis of PET images would alleviate this problem by helping define an objective limit between normal and pathological findings.

PET uptake can be quantitatively evaluated either regionally or on a voxel-by-voxel basis. In regional analysis, the regional uptake is compared with the regional uptake expected in a normal control population. This analysis usually requires prior knowledge to select the appropriate atlas and relevant discriminant regions, which should be adapted to a specific pathology, limiting its use.<sup>11</sup>

In voxel-wise analysis, a subject's PET image is usually aligned to a standardised group space to compare the uptake of the spatially normalised scan to a distribution obtained from normal control scans, on a voxel-by-voxel basis. The approach implemented in Neurostat<sup>12,13</sup> consists of registering the PET image of the subject under investigation to a standard space and comparing it to a population of controls by means of a Z-score. The Z-score map is then projected onto different surfaces resulting in three-dimensional stereotactic surface projections that are used for image interpretation. Other software tools implementing a similar technique have been used for the analysis of PET data, such as NeuroGam (GE Healthcare, Waukesha, WI, USA).<sup>14</sup> Signorini et al.<sup>11</sup> used the general linear model implemented in the Statistical Parametric Mapping (SPM) software package<sup>15</sup> to compare a subject PET image to a population of controls. The t statistic corresponds to the difference between the mean uptake of the control group and the uptake of the subject being analysed, divided by the error estimated with the control group data (after correction for age and global metabolism). A similar approach was also used in 16, 10 and 17. Exploratory in nature, voxel-wise techniques require less prior information than regional analysis, but their sensitivity is limited by inter-subject variability in non-pathological tracer uptake, making pathological effects harder to detect.<sup>11</sup> The fact that the images have to be registered to a standard space can also decrease the sensitivity as the non-linear registration step may conceal subtle anomalies.

In this paper, we propose a framework for the individual analysis of PET data that consists of creating a subject-specific model of healthy PET appearance and comparing the patient's PET image to the model via a regularised Z-score. The resulting voxel-wise Z-score map can be interpreted as an abnormality map, as it statistically evaluates the localised deviation of the subject-specific uptake with respect to the healthy uptake distribution. The abnormality maps are meant to help clinicians identify more easily pathological areas and also improve the interpretability of subsequent computer-aided analyses. We validate the proposed framework first by generating abnormality maps for healthy control subjects to ensure that no erroneous abnormalities are detected. We then generate abnormality maps for subjects at different stages of Alzheimer's disease (AD) and use them as features to feed a classifier. This is to ensure that the proposed method is able to extract for each individual the signal characteristic of abnormality. Finally, we apply the proposed framework to a dataset of subjects with different types of frontotemporal dementia (FTD) syndromes. Preliminary work has been presented at conferences.<sup>18,19</sup> Contributions specific to this

paper include a novel regularised Z-score, experiments to optimise the parameters, an extended validation using an AD cohort, and an application of the framework to other types of dementia.

## 2 Methods

The proposed anomaly detection framework consists of selecting, in a control dataset, the subjects that are the most similar to the subject being analysed in terms of demographic characteristics and morphology, creating a subject-specific model of healthy PET uptake from the selected controls and the target subject's anatomical magnetic resonance (MR) image, and using the resulting model to create a subject-specific abnormality map. The steps necessary to generate a subject-specific abnormality map are summarised in Appendix A.

### 2.1 Database of control atlases

The approach relied on a database of atlases, each composed of a pair of co-registered anatomical MR and PET images. The atlases were globally aligned in a common space. This was performed by mapping to a common coordinate frame the MR images from all the atlases via an affine group-wise registration.<sup>20</sup> The transformations were then applied to the MR and PET images by updating their image coordinate system (without resampling).

### 2.2 Atlas pre-selection

An atlas pre-selection step was performed to discard the control atlases too dissimilar to the target and thus limit the computational time while maintaining a high synthesis accuracy. Two strategies were explored: one relying on the demographic characteristics, the other on the anatomical images themselves.

#### 2.2.1 Demographic-based pre-selection

Both age and sex have been shown to influence brain metabolism, even though it is not clear whether this can be explained by the underlying morphology.<sup>21–25</sup> To limit this influence, the demographic-based pre-selection first consisted of selecting the control atlases of the same gender as the target. The atlases closest in terms of age to the target were then picked.

#### 2.2.2 Image-based pre-selection

The image-based strategy preselects the atlases according to their global morphological similarity to the target, as assessed by a global similarity measure, the normalised cross-correlation (NCC). The anatomical MR image of a randomly-chosen reference atlas was affinely registered to the MR image of the target subject. Because all the control atlases were pre-aligned with each other, the resulting affine transformation was applied to the anatomical MR image of each control atlas and the NCC was computed between each resampled control atlas and the target subject. The control atlases with the highest NCC were selected.

### 2.3 Inter-subject registration and atlas local selection

The anatomical MR images of the pre-selected atlases were non-rigidly registered to the target subject's MR image,<sup>26</sup> and the PET images of the atlases, pre-aligned with the MR images, were mapped using the same transformation to the target subject. This inter-subject coordinate mapping was obtained using a symmetric global registration followed by a cubic B-spline parametrised non-rigid registration as implemented in NiftyReg.<sup>26</sup> The normalised mutual information was used as similarity measure. The non-rigid registrations were performed with a pyramidal approach with three levels. The finer lattice of control points had a spacing of 5 mm along each axis.

Once non-rigidly aligned with the target subject, the atlases closest in terms of morphology to the target subject were identified. This morphological similarity was assessed at the voxel level using a local image similarity measure, the structural image similarity (SSIM).<sup>27</sup> The SSIM between two anatomical MR images  $I^{\text{MRI}}$  and  $J^{\text{MRI}}$  at voxel  $x$  is given by

$$\text{SSIM}(I^{\text{MRI}}(x), J^{\text{MRI}}(x)) = \frac{2\mu_{I^{\text{MRI}}}(x)\mu_{J^{\text{MRI}}}(x) + C_1}{\mu_{I^{\text{MRI}}}^2(x) + \mu_{J^{\text{MRI}}}^2(x) + C_1} \times \frac{2\sigma_{I^{\text{MRI}}, J^{\text{MRI}}}(x) + C_2}{\sigma_{I^{\text{MRI}}}^2(x) + \sigma_{J^{\text{MRI}}}^2(x) + C_2}. \quad (1)$$

$C_1$  and  $C_2$  are two constants used to improve the stability of the structural similarity that depend on the range of the voxel values<sup>27</sup> (formulas given in Appendix A.4). The means and standard deviations were calculated using a Gaussian kernel  $G_{\sigma_G}$  with standard deviation  $\sigma_G$  through density normalised convolution<sup>28,29</sup>

$$\begin{aligned} \mu_{I^{\text{MRI}}}(x) &= \frac{[G_{\sigma_G} * I^{\text{MRI}}](x)}{[G_{\sigma_G} * \Omega](x)}, & \sigma_{I^{\text{MRI}}}^2(x) &= \mu_{I^{\text{MRI}^2}}(x) - \mu_{I^{\text{MRI}}}(x)^2, \\ \sigma_{I^{\text{MRI}}, J^{\text{MRI}}}(x) &= \mu_{I^{\text{MRI}} \cdot J^{\text{MRI}}}(x) - \mu_{I^{\text{MRI}}}(x) \cdot \mu_{J^{\text{MRI}}}(x), \end{aligned}$$

where  $*$  denotes the convolution operator.  $G_{\sigma_G} * \Omega$  represents a density normalisation term that compensates for areas with missing information where  $\Omega$  is a density function equal to 1 where the fields of view overlap and 0 otherwise. An illustration of mismatches between fields of view is available in 30 (Fig. 6.1).

The local selection was performed via a weighted scheme. The weights reflect the contribution of each control atlas to the model. They were obtained at each voxel  $x$  by ranking the SSIM across the  $N$  control atlases globally pre-selected and by applying an exponential decay function:

$$w_n(x) = e^{-\beta r_n(x)}, \quad (2)$$

where  $r_n(x)$  denotes the rank of the  $n^{\text{th}}$  control atlas. Using the rank instead of the SSIM value means that the sum and separation of the weights for different voxels are the same at each voxel location, leading to more stable results.<sup>31</sup> For each voxel, the atlases contributing the most to the model are the ones with the highest morphological similarity to the target subject.

### 2.4 Subject-specific models of healthy PET appearance

To generate the subject-specific model, which is composed of two elements: a spatially-varying weighted average and a spatially-varying weighted standard deviation, the control atlases locally selected were fused based on their morphological similarity to the target subject. For each of the

$N$  pre-selected atlases in the control dataset, let the  $n^{\text{th}}$  mapped PET image be denoted by  $J_n^{\text{PET}}$ . The two subject-specific model elements ( $I_\mu^{\text{PET}}, I_\sigma^{\text{PET}}$ ) were computed as follows:

$$I_\mu^{\text{PET}}(x) = \frac{\sum_{n=1}^N w_n(x) \cdot J_n^{\text{PET}}(x)}{\sum_{n=1}^N w_n(x)},$$

$$I_\sigma^{\text{PET}}(x) = \sqrt{\frac{N_w}{N_w - 1} \frac{\sum_{n=1}^N w_n(x) \cdot (J_n^{\text{PET}}(x) - I_\mu^{\text{PET}}(x))^2}{\sum_{n=1}^N w_n(x)}}$$
(3)

where  $N_w$  is the number of non-zero weights.

### 2.5 Subject-specific abnormality maps

To compare the target subject's PET image ( $I^{\text{PET}}$ ) to the subject-specific model, in our preliminary work<sup>18</sup> a Z-score was computed for each voxel of the image. However, we observed that this leads to the generation of high frequency signals in certain areas due to the standard deviation approaching zero. To avoid this problem, we define a regularised Z-score

$$\tilde{Z}(x) = \frac{I^{\text{PET}}(x) - I_\mu^{\text{PET}}(x)}{\hat{I}_\sigma^{\text{PET}}(x)}$$
(4)

with

$$\hat{I}_\sigma^{\text{PET}}(x) = \begin{cases} P_k(I_\sigma^{\text{PET}}) & \text{if } I_\sigma^{\text{PET}}(x) < P_k(I_\sigma^{\text{PET}}) \\ I_\sigma^{\text{PET}}(x) & \text{otherwise} \end{cases},$$
(5)

where  $P_k(I_\sigma^{\text{PET}})$  is the  $k^{\text{th}}$  percentile of  $I_\sigma^{\text{PET}}$ , computed only from brain voxels using a sorting-based algorithm.

The voxel-wise regularised Z-score map can be interpreted as an abnormality map, as it statistically evaluates the localised deviation of the subject-specific uptake with respect to the healthy uptake distribution.

## 3 Validation on Alzheimer's disease

### 3.1 Materials: Alzheimer's disease cohort

Part of the data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative database (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 early AD. For up-to-date information, see [www.adni-info.org](http://www.adni-info.org).

We selected 515 participants from the ADNI2 database who had T1-weighted (T1w) MRI,  $^{18}\text{F}$ -FDG PET, and Florbetapir ( $^{18}\text{F}$ -AV45) PET images at baseline and were diagnosed as cognitively normal (CN), early MCI (EMCI), late MCI (LMCI) or AD. The diagnosis relies on three main criteria: the absence or presence of memory complaints, the Mini-Mental State Examination (MMSE) score and the Clinical Dementia Rating (CDR) score.<sup>32</sup> 23 CN subjects with an amyloid negative status were extracted from the main dataset for parameter optimisation purposes. This set

**Table 1** Summary of the ADNI participants' demographics, clinical scores and amyloid status.

	N	Age	Gender	MMSE	CDR	Amyloid status
CN <sub>opti</sub>	23	74.6 ± 5.8 [65.4; 89.0]	9 F / 14 M	28.4 ± 1.6 [24; 30]	0: 23	23 - / 0 +
CN	131	73.6 ± 6.4 [56.2; 85.6]	67 F / 64 M	29.1 ± 1.1 [25; 30]	0: 131	89 - / 42 +
EMCI	142	70.9 ± 7.0 [55.5; 88.6]	65 F / 77 M	28.4 ± 1.6 [24; 30]	0: 1; 0.5: 141	74 - / 68 +
LMCI	120	72.0 ± 7.9 [55.0; 91.4]	62 F / 58 M	27.7 ± 1.8 [24; 30]	0.5: 119; 1:1	42 - / 78 +
AD	99	74.4 ± 8.0 [55.9; 88.5]	44 F / 55 M	22.9 ± 2.2 [19; 26]	0.5: 46; 1: 52; 2: 1	12 - / 87 +

Abbreviations: CN: cognitively normal; EMCI: early mild cognitive impairment; LMCI: late mild cognitive impairment; AD: Alzheimer's disease; MMSE: Mini-Mental State Examination; CDR: global Clinical Dementia Rating.

is referred to as the CN<sub>opti</sub> dataset in the following. Table 1 summarises the demographic characteristics, clinical scores, and amyloid status of the participants. Subjects were categorised as amyloid positive ( $A\beta^+$ ) or negative ( $A\beta^-$ ) based on a cortical mean cutoff of 1.11 on Florbetapir PET.<sup>33</sup> The database of control atlases used in this paper is composed of the CN participants amyloid negative ( $n = 89$ ).

### 3.1.1 Image acquisition

The acquisition protocols of the 3D T1w images can be found in 34. The images were downloaded after correction of image geometry distortion due to gradient non-linearity (gradwarp) and correction of the image intensity non-uniformity.<sup>34</sup> The ADNI2 FDG PET protocol consisted of a dynamic acquisition of six five-minute frames, 30 to 60 minutes post-injection, and the Florbetapir PET protocol consisted of a dynamic acquisition of four five-minute frames from 50 to 70 minutes post-injection.<sup>35</sup> For both tracers, images were downloaded after several stages of preprocessing: frame averaging, spatial alignment, interpolation to a standard voxel size, and smoothing to a common resolution of 8 mm full width at half maximum.

### 3.1.2 Image preprocessing

For each subject, the T1w MR image was corrected for intensity non-uniformity following a non-parametric intensity non-uniformity normalisation method<sup>36</sup> and was mapped to the PET images using a rigid transformation. The T1w MR images, resampled to the PET voxel grid of  $1.5 \times 1.5 \times 1.5 \text{ mm}^3$ , were then parcellated into 143 different regions using a multi-atlas propagation and fusion algorithm implemented in NiftySeg.<sup>37</sup> The PET images were intensity-normalised using the average uptake in reference regions that were extracted from the parcellated T1w MR images. The pons was used for the FDG PET images<sup>38</sup> and the whole cerebellum for the Florbetapir PET images.<sup>39</sup> The MR and PET images of all the subjects were globally aligned in the common space of the control atlases (Sec. 2.1), via an affine groupwise registration.<sup>20</sup> This step is not necessary to generate abnormality maps but facilitates the subsequent group-space analyses.

## 3.2 Parameter optimisation

Two parameters were optimised using a leave-one-out cross validation on the 23 subjects from the CN<sub>opti</sub> subset: the standard deviation of the Gaussian kernels used in Eq. (1) ( $\sigma_G$ , expressed

in voxels), which controls the size of the neighbourhood where the local similarity measure is computed:

$$\sigma_G = [ 1 \ 3 \ 5 \ 7 \ 9 ] ,$$

and  $\beta$  from Eq. (2) whose value influences the repartition of the weights:

$$\beta = [ 1 \ 0.5 \ 0.25 \ 0.1 \ 0.01 ] .$$

The weighted sum tends to the mean when  $\beta$  is small.

For each tracer, a pseudo-healthy PET image ( $I_\mu$  in Eq. (3)) was generated using the proposed method from the subject's T1w MR image. This pseudo-healthy PET image was then compared to the real PET image ( $I$ ) using the relative mean absolute error, defined as

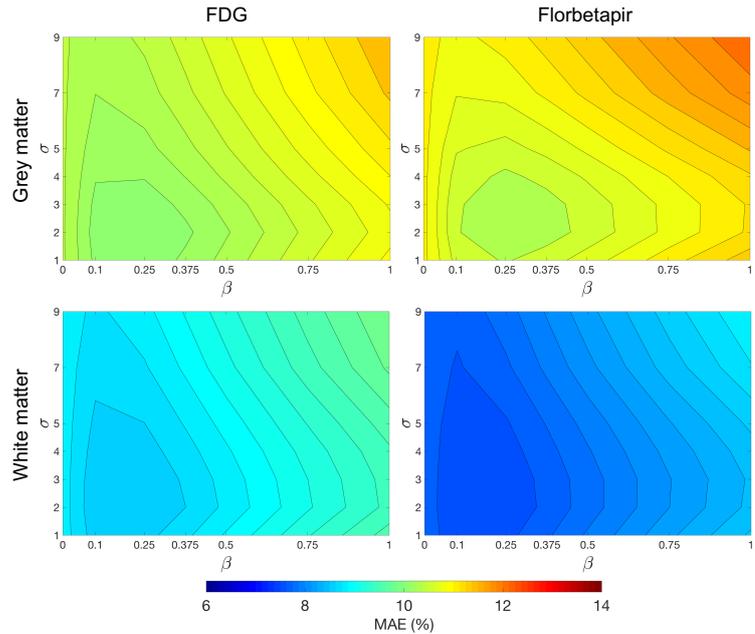
$$\text{rMAE} = 100 * \frac{\sum_x |I_\mu(x) - I(x)|}{\sum_x I(x)} . \quad (6)$$

The rMAE was computed in the grey matter (GM) and white matter (WM) regions for both tracers. Values averaged over all subjects are shown in Fig. 1. For both tracers, the lowest MAE were obtained with  $\sigma_G = 2$  or  $3$  and  $\beta = 0.25$ . We chose  $\sigma_G = 3$  to favour smoother and less patchy images. The selected parameters are thus  $\sigma_G = 3$  and  $\beta = 0.25$ , giving an average rMAE of 10.1% (GM) and 8.6% (WM) for FDG PET and 10.4% (GM) and 7.6% (WM) for Florbetapir PET.

The percentile used in Eq. (5) to regularise the Z-score was determined experimentally as a compromise between reducing the number of outliers and preserving enough standard deviation information. It was set to  $k = 10$ .

### 3.3 Validation scheme

The validation was performed in two steps. First, the pseudo-healthy PET images of the CN subjects amyloid negative were compared with their real PET images. The abnormality maps were then used as features to perform individual classification.



**Fig 1** Average rMAE between the real and pseudo-healthy PET images in the grey matter (GM) and white matter (WM) regions for varying values of  $\sigma_G$  and  $\beta$ . The selected parameters are  $\sigma_G = 3$  and  $\beta = 0.25$ , giving an average rMAE of 10.1% (GM) and 8.6% (WM) for FDG PET and 10.4% (GM) and 7.6% (WM) for Florbetapir PET.

### 3.3.1 Synthesis accuracy

The first step of the validation consisted of comparing the pseudo-healthy PET images of the CN subjects amyloid negative to their real PET images. As these subjects should not present abnormalities, the pseudo-healthy and real PET images should be as similar as possible. This similarity was assessed using the rMAE.

### 3.3.2 Individual classification

To assess the ability of the abnormality maps to extract relevant information from PET data, the abnormality maps were used as features to feed a linear SVM classifier.

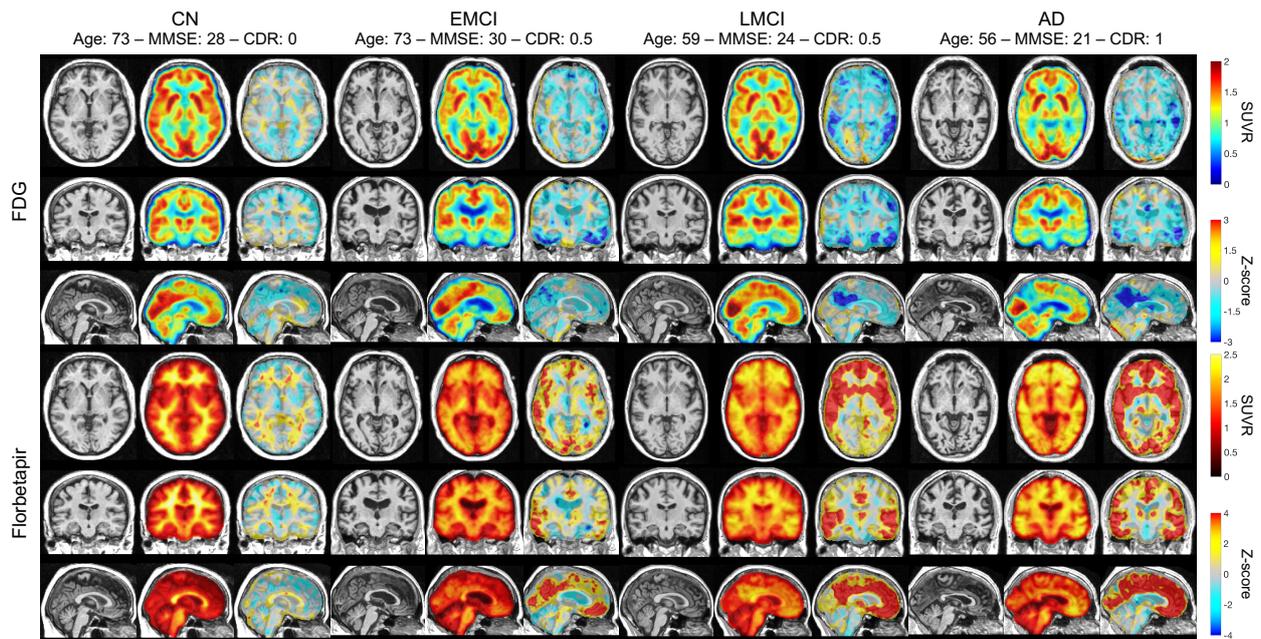
**Non-linear alignment to group space** A way to compare the abnormality maps, each generated in the subject's native space, across all the subjects, is to align them with each other. As the T1 images from all the subjects were already mapped to a common coordinate frame via an affine groupwise registration, the T1 images were subsequently non-rigidly registered to the group-space. The same transformations were then applied to the abnormality maps.

**Linear SVM classifier** We chose a linear SVM to classify the abnormality maps. A linear kernel was calculated using the inner product for each pair of abnormality maps available in the dataset (using all the brain voxels). The cross-validation (CV) procedure included two nested loops: an outer loop evaluating the classification performance and an inner loop used to optimise the hyperparameter  $C$  that regularises the SVM. For the outer loop, we used 250 stratified shuffle splits with a test size of 30%. Note that, for a same task, the splits were kept the same between the different types of features tested. We used an inner  $k$ -fold with  $k = 10$ . This individual classification was performed with tools implemented in Clinica<sup>40</sup> that rely on scikit-learn.<sup>41</sup>

**Classification tasks** The experiments consisted of different tasks with varying degrees of difficulty:

- differentiating cognitively normal subjects from subjects with a disease, i.e. CN  $A\beta^-$  vs AD  $A\beta^+$ , CN  $A\beta^-$  vs LMCI  $A\beta^+$ , CN  $A\beta^-$  vs EMCI  $A\beta^+$  (using FDG PET);
- differentiating between subjects at the beginning and at the end of the early stage of the disease, i.e. EMCI  $A\beta^+$  vs LMCI  $A\beta^+$  (using FDG PET);
- differentiating between amyloid negative and amyloid positive subjects, i.e.  $A\beta^-$  vs  $A\beta^+$  (using Florbetapir PET).

For the first two experiments, 322 subjects (89 CN  $A\beta^-$ , 68 EMCI  $A\beta^+$ , 78 LMCI  $A\beta^+$  and 87 AD  $A\beta^+$ ) were considered, while for the last experiment 492 subjects (217  $A\beta^-$  and 275  $A\beta^+$ ) were analysed.



**Fig 2** Examples of FDG (top) and Florbetapir (bottom) abnormality maps obtained for a cognitively normal subject (CN) and patients with early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI) and Alzheimer’s disease (AD). For each subject, the T1w MRI, PET and abnormality map (demographic-based pre-selection) are displayed. A highly negative Z-score (blue) indicates a reduced PET uptake in the subject relative to the controls while a highly positive Z-score (red) indicates an increased PET uptake in the subject relative to the controls.

**Comparison with standard approaches** To set the results in perspective, the subjects’ PET image itself and two standard Z-maps were also used as features and fed to the classifier. For each subject, a first Z-map was obtained by comparing the subject’s PET image in the group space to the mean and standard deviation computed from the 89 subjects of the control dataset, also in the group space. A second Z-map was obtained by first pre-selecting the control subjects using demographic characteristics, as in Sec. 2.2.1, and comparing the subject’s PET image in the group space to the mean and standard deviation computed from the pre-selected control subjects, also in the group space.

### 3.4 Results

Abnormality maps were generated for each of the 492 ADNI2 participants selected, for both the FDG and Florbetapir PET images. Note that for the CN  $\beta^-$  subjects (forming the control dataset), a leave-one-out strategy was used, i.e. the images of the CN subject being processed were excluded from the control database. For the demographic-based pre-selection, the 30 control atlases of the same gender and closest in terms of age to the target subject were selected. For the image-based pre-selection, the 30 control atlases with the highest NCC were selected. The number of pre-selected atlases was chosen as a compromise between having a representative database of controls and computational time.

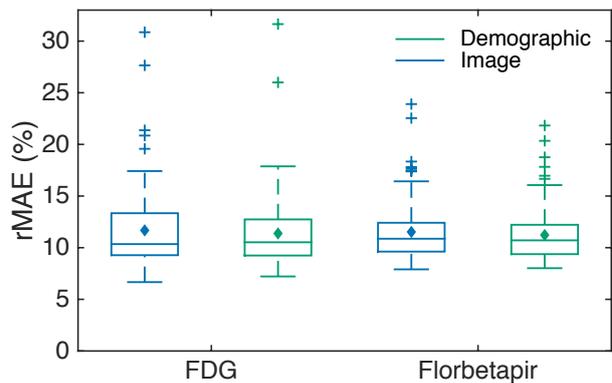
Examples of abnormality maps are displayed in Fig. 2 for a CN, an early MCI, a late MCI and an AD subject. We observe that, as expected, no specific signal is being detected for the CN subject, for both the FDG and Florbetapir tracers. For the EMCI and LMCI subjects, abnormal glucose

uptake is detected mainly in the precuneus and in the medial-lateral temporal lobe, and abnormal amyloid deposition is detected in the frontal, parietal, temporal and cingulate cortices, which is consistent with previous observations.<sup>42</sup> Finally, for the AD subject, abnormal glucose uptake is also detected mainly in the precuneus and in the medial-lateral temporal lobe, and abnormal amyloid deposition is detected in all the cortex, which is typical of AD.<sup>43</sup>

### 3.4.1 Synthesis accuracy

The rMAE results obtained when comparing the pseudo-healthy PET images of the CN subjects amyloid negative with their real PET images are displayed in Fig. 3.

The pre-selection strategy has no significant impact ( $p > 0.05$ , paired t-test) on the synthesis accuracy: for the FDG tracer, the average rMAE is of 11.6% for the image-based and 11.4% for the demographic-based pre-selection; and for the Florbetapir tracer, the average rMAE is of 11.5% for the image-based and 11.3% for the demographic-based pre-selection. These results are consistent with the ones obtained for the optimisation dataset (rMAE in the brain of 11.2% for FDG and 10.9% for Florbetapir).



**Fig 3** Boxplots displaying the mean (diamond), median, lower and upper quartiles, and minimum and maximum of the rMAE between the real and pseudo-healthy PET images for the CN amyloid negative subjects.

### 3.4.2 Individual classification

The balanced accuracies obtained for the different features and tasks are displayed in Table 2. No matter the task, the abnormality maps obtained with the demographic-based pre-selection lead to similar (Florbetapir PET tasks) or higher (FDG PET tasks) balanced accuracies than the abnormality maps obtained with the image-based pre-selection. The effect of the demographic-based pre-selection is also visible when comparing the Z-maps computed from the whole control dataset or from pre-selected control subjects: the accuracy is usually equivalent or higher in the later case. The classification performance is generally higher when the Z-maps and abnormality maps are used as features rather than the PET images themselves. For the majority of the FDG tasks, the balanced accuracy obtained for the proposed method with demographic-based pre-selection is slightly higher than the balanced accuracy obtained using the Z-maps with demographic-based pre-selection as features. The benefits of the abnormality maps seem to slightly increase with the difficulty of the task: 62.2% vs 64.7% for the CN vs EMCI task and 60.6% vs 61.1% for the EMCI vs LMCI task.

When analysing Florbetapir data, using the PET images themselves, the state-of-the-art Z-maps or the proposed abnormality maps leads to similar, highly accurate, classification results. These highly accurate results were expected, but are here confirmed, as differentiating amyloid negative from amyloid positive subjects based on features extracted from Florbetapir data is a quite trivial task.

**Table 2** Balanced accuracy obtained when using PET images, state-of-the-art Z-maps, and the subject-specific abnormality maps (both with image-based and demographic-based pre-selection) as features of the linear SVM classification algorithm. The average  $\pm$  SD balanced accuracy, obtained over 250 repeats, is expressed in percentages. For each task, the highest balanced accuracy is highlighted with bold font.

Tracer	Task	PET	Z-map (No pre-select.)	Z-map (Demographic pre-select.)	Abn. map (Image pre-select.)	Abn. map (Demographic pre-select.)
FDG	CN $A\beta^-$ vs AD $A\beta^+$	89.3 $\pm$ 4.3	90.8 $\pm$ 3.9	90.9 $\pm$ 4.1	91.4 $\pm$ 3.9	<b>92.4 <math>\pm</math> 3.7</b>
	CN $A\beta^-$ vs LMCI $A\beta^+$	73.4 $\pm$ 6.1	75.3 $\pm$ 5.2	<b>77.2 <math>\pm</math> 5.3</b>	74.9 $\pm$ 5.2	76.7 $\pm$ 5.3
	CN $A\beta^-$ vs EMCI $A\beta^+$	58.3 $\pm$ 6.7	62.6 $\pm$ 6.2	62.2 $\pm$ 5.8	60.0 $\pm$ 6.2	<b>64.7 <math>\pm</math> 5.8</b>
	EMCI $A\beta^+$ vs LMCI $A\beta^+$	58.3 $\pm$ 6.0	59.1 $\pm$ 6.2	60.6 $\pm$ 6.1	56.3 $\pm$ 6.4	<b>61.1 <math>\pm</math> 6.0</b>
Florbetapir	$A\beta^-$ vs $A\beta^+$	95.0 $\pm$ 1.5	<b>96.3 <math>\pm</math> 1.3</b>	96.2 $\pm$ 1.3	96.1 $\pm$ 1.3	96.1 $\pm$ 1.4

Overall, the classification results obtained for the abnormality maps confirms their ability to detect meaningful signal from both FDG and Florbetapir PET images.

## 4 Application to frontotemporal dementia

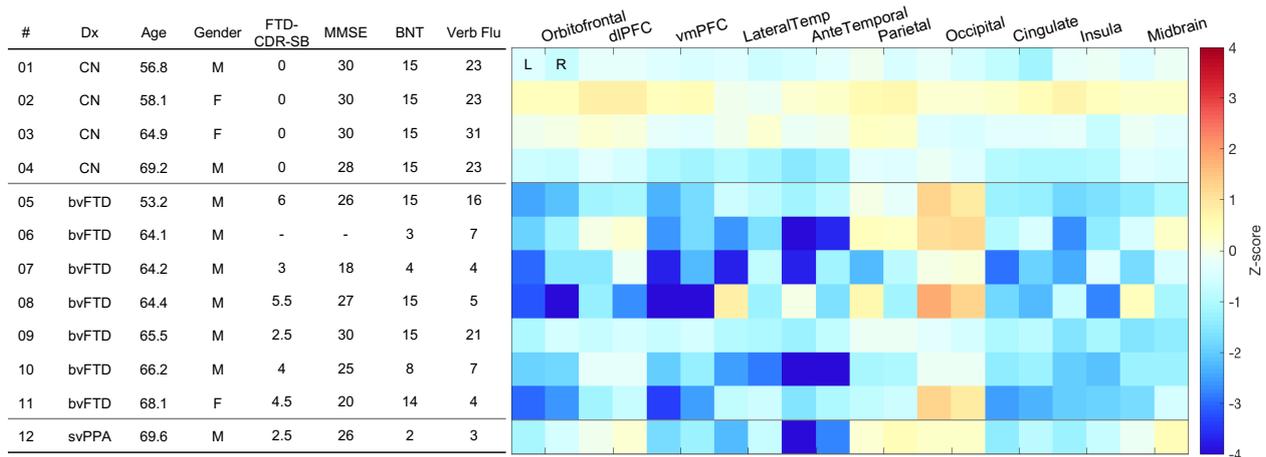
### 4.1 Materials: frontotemporal dementia cohort

Data for the FTD cohort were obtained from NIFD, which uses the ADNI platform to make available data from the frontotemporal lobar degeneration neuroimaging initiative (FTLDNI). FTLDNI (<http://memory.ucsf.edu/research/studies/nifd>) is founded through the National Institute of Aging, and started in 2010. The primary goals of FTLDNI are to identify neuroimaging modalities and methods of analysis for tracking frontotemporal lobar degeneration and to assess the value of imaging versus other biomarkers in diagnostic roles. The Principal Investigator of FTLDNI is Dr. Howard Rosen, MD at the University of California, San Francisco. The data is the result of collaborative efforts at three sites in North America.

We focused on the 12 participants who had T1w MRI and  $^{18}\text{F}$ -FDG PET images with an iterative reconstruction at baseline. Four subjects were cognitively normal (CN), seven diagnosed as behavioural variant FTD (bvFTD) and one as semantic variant primary progressive aphasia (svPPA). The neuropsychological battery included functional measures, such as the FTD-specific Clinical Dementia Rating (CDR), and cognitive measures, such as the Mini-Mental State Examination (MMSE) that includes tests of orientation, attention, memory, language and visuospatial ability, the Boston Naming Test (BNT) that assesses word-finding ability, and a verbal fluency test assessing the ability to retrieve specific information within the animal category.<sup>44,45</sup> The subjects' demographic characteristics and clinical scores are displayed in Fig. 4.

#### 4.1.1 Image acquisition

Both the MRI and PET images were acquired at the Mayo Clinic, Rochester. 3D T1w MR images were acquired on a GE Discovery MR750 or GE Signa HDxt 3T scanner using the following sequence parameters: TR  $\approx$  7 ms, TE  $\approx$  3 ms, TI = 900 ms, flip angle = 8°, slice thickness = 1.2 mm, in-plane resolution = 1.0  $\times$  1.0 mm<sup>2</sup>, matrix = 256  $\times$  256  $\times$  166. The images were downloaded after correction of image geometry distortion due to gradient non-linearity (gradwarp) and correction of the image intensity non-uniformity. The FDG PET images were acquired on a GE Discovery RX PET/CT scanner following a protocol that consisted of a dynamic acquisition of



**Fig 4** Left: NIFD participants' demographics and clinical scores. Right: Average Z-scores obtained with the proposed method for each subject and ROI. The subjects were sorted by diagnosis and age. A highly negative Z-score (blue) indicates a reduced FDG uptake in the subject relative to the controls while a highly positive Z-score (red) indicates an increased FDG uptake in the subject relative to the controls. Abbreviations: CN: cognitively normal; bvFTD: behavioural variant frontotemporal dementia (FTD); svPPA: semantic variant primary progressive aphasia; MMSE: Mini-Mental State Examination; FTD-CDR-SB: FTD-specific Clinical Dementia Rating sum of boxes; BNT: Boston Naming Test; Verb Flu: verbal fluency; dIPFC: dorsolateral prefrontal region; vmPFC: ventromedial prefrontal region; LateralTemp: lateral temporal region; AnteTemporal: anterior temporal region; L: left; R: right.

six five-minute frames, 30 to 60 minutes post-injection. PET images were reconstructed using a 3D iterative method and displayed in a  $256 \times 256 \times 47$  matrix (voxel size  $1.17 \times 1.17 \times 3.27 \text{ mm}^3$ ).

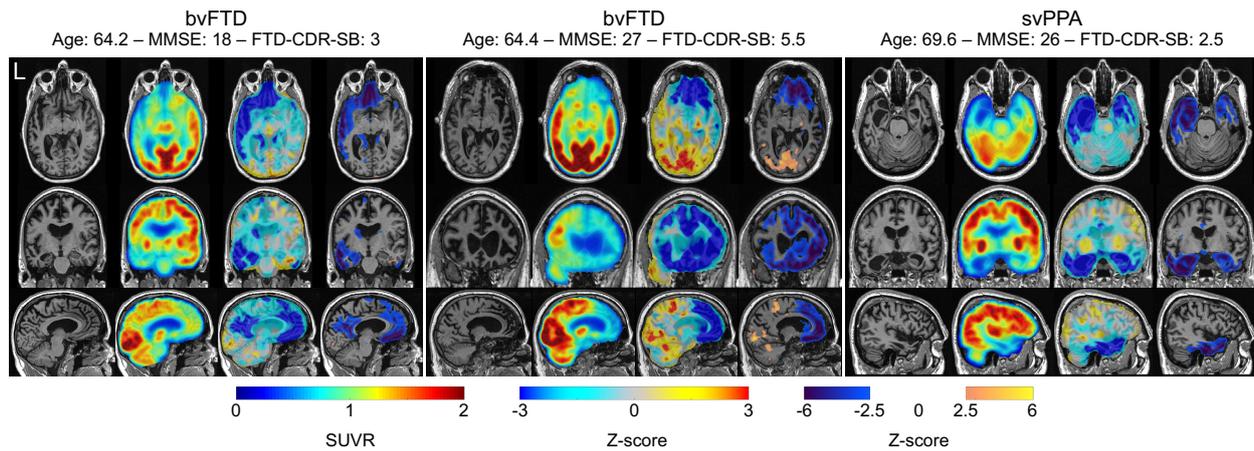
#### 4.1.2 Image preprocessing

The T1w MR images were corrected for intensity non-uniformity<sup>36</sup> and parcellated into 143 different regions.<sup>37</sup> For each PET acquisition, each frame was rigidly registered to the first frame and the co-registered frames were then averaged. The averaged PET image was mapped to the T1w MR image using a rigid transformation. The PET image, resampled to the T1w MRI, was smoothed using a Gaussian kernel with a standard deviation of two voxels (to obtain images with a resolution similar to that of the ADNI database) and intensity-normalised using the average uptake in the pons, which was extracted from the parcellated T1w MR image.

#### 4.2 Results

Abnormality maps were generated for the 12 subjects of the FTD cohort using the CN subjects amyloid negative from the ADNI cohort as atlases. Demographic-based pre-selection was performed by selecting the 30 control atlases of the same gender and closest in terms of age to the target subject.

Abnormality maps obtained for two bvFTD subjects and the svPPA subject are displayed in Fig. 5. This figure highlights the fact that bvFTD is a very heterogeneous syndrome: neurodegeneration can affect the frontal lobe (Fig. 5, centre), or also the temporal and parietal lobes (Fig. 5, left).<sup>46,47</sup> The svPPA subject shows typical asymmetric hypometabolism affecting mainly the temporal pole, entorhinal area and hippocampus.<sup>48</sup>



**Fig 5** Examples of abnormality maps obtained for two subjects with bvFTD and a subject with svPPA. For each subject, the T1w MRI, PET and abnormality map, both with and without thresholding, are displayed. The threshold is set at 2.5. A highly negative Z-score (blue) indicates a reduced FDG uptake in the subject relative to the controls while a highly positive Z-score (red) indicates an increased FDG uptake in the subject relative to the controls. The first bvFTD subject corresponds to subject #07 in Fig. 4, and the second to subject #08.

Comparisons were restricted to ten clinically relevant regions for the sake of brevity. These regions were either selected to represent the areas where abnormal uptake, compared with controls, is expected for bvFTD:

- the orbitofrontal region, comprising the anterior, posterior, medial and lateral orbital gyri;<sup>49–52</sup>
- the dorsolateral prefrontal region, comprising the inferior, middle and superior frontal gyri;<sup>49–53</sup>
- the ventromedial prefrontal region, comprising the gyrus rectus, medial frontal cortex, subcallosal area, superior frontal gyrus medial segment;<sup>50–53</sup>
- the lateral temporal region, comprising the inferior, middle and superior temporal gyri;<sup>51–53</sup>
- the parietal region, comprising the precuneus and the supramarginal and angular gyri;<sup>49–52</sup>
- the cingulate;<sup>50–53</sup>
- the insula;<sup>50–52</sup>
- the midbrain region, comprising the caudate, pallidum, putamen and thalamus;<sup>49–51, 53</sup>

or svPPA:

- the anterior temporal region, comprising the hippocampus, amygdala, temporal pole and entorhinal area;<sup>48</sup>

or to act as a neutral region where no hypometabolism is expected:

- the occipital region, comprising the inferior, middle and superior occipital gyri, and occipital fusiform gyrus.

Each hemisphere was analysed separately to account for left/right asymmetry.

Fig. 4 displays average Z-scores obtained with the proposed method for each ROI. We note that the controls have approximately zero Z-scores in all regions while highly negative Z-scores are observed for bvFTD and svPPA subjects in regions where a reduced uptake is expected when compared with healthy controls. The svPPA subject shows typical asymmetric hypometabolism affecting the anterior temporal region.<sup>48</sup> This figure again highlights the heterogeneity of bvFTD: the areas affected and the degree of hypometabolism vary across subjects. Several subjects present hypometabolism only in the frontal lobe (e.g. subject #08), while for others hypometabolism is

present in both the frontal and temporal lobes (e.g. subject #07). The parietal region can also be affected, even though not as much as the frontal and temporal regions (subject #07). On the contrary, we observe for some patients a hypermetabolism in the occipital region (e.g. subject #08), which has already been described.<sup>53</sup> The subjects may or may not present left/right observed asymmetries. The degree of hypometabolism appears consistent with the neuropsychological scores. For example, the low degree of anomaly observed for subject #09 can be explained by his high MMSE, BNT and verbal fluency scores, and low FTD-CDR score, compared with the other bvFTD subjects. Subjects with the highest degree of anomaly in the left anterior temporal region are the ones with the lowest verbal fluency score, which is in accordance with the fact that this region has been shown to be critical for semantic abilities.<sup>54</sup>

## 5 Discussion

We presented a method for the individual analysis of PET data providing voxel-wise statistics of normality/abnormality. The method consists of creating a subject-specific model of healthy PET appearance and comparing the patient's PET image to the model via a regularised Z-score. We validated the proposed method by generating abnormality maps for healthy controls and subjects at different stages of Alzheimer's disease, and we applied the framework to the analysis of frontotemporal dementia.

We first ensured that the method was able to generate accurate subject-specific models of healthy PET appearance by applying it to normal control subjects and showed that no anomalies were detected for this population (Fig. 2, 3 and 4). We then applied it to subjects at different stages of Alzheimer's disease and to subjects with two subtypes of frontotemporal dementia. In both cases, we observed that the abnormality maps obtained with the proposed method coincided with the regions where uptake abnormalities were expected (Fig. 2, 4 and 5).

The proposed framework was also validated using the abnormality maps as inputs of a classifier. For comparison, we also used the PET images themselves and standard Z-maps as features. The different approaches produced comparable accuracies for all the tasks tested: differentiating CN from early MCI, late MCI and AD, differentiating early from late MCI, and differentiating amyloid negative subjects from amyloid positive subjects. More accurate results were systematically obtained with the proposed method than with the PET images themselves. These results demonstrated that the proposed approach was able to automatically locate and characterise the areas characteristic of dementia. The classification results also highlighted the importance of selecting control subjects matching the demographic characteristics of the subject under investigation. This may be due to the fact that age and sex influence brain metabolism, and that these differences are not captured when comparing structural MR images. Note that the objective of the proposed framework was not to generate features leading to a higher classification accuracy, but rather to provide feature maps easily interpretable.

The subject-specific model is composed of two elements: a spatially-varying weighted average and a spatially-varying weighted standard deviation. The former could be obtained using other image synthesis approaches, such as deep learning methods,<sup>55-60</sup> which have been shown to be more accurate than atlas-based methods.<sup>61-63</sup> However, they would not generate the later in a straightforward manner. The standard deviation is an important component of the model as it decreases the degree of the anomalies detected in areas where the uncertainty is high, reducing the amount of false positives.

Others have used image synthesis for anomaly detection. Ye et al. proposed a patch-based modality propagation technique to synthesise pseudo-healthy T2-weighted from T1w MR images for tumour detection.<sup>64</sup> A patch-based approach has also been developed in 65 to synthesise pseudo-normal chest radiographs. In 66, a regression-based method has been proposed to synthesise pseudo-healthy FLAIR from T1w images for white matter lesion segmentation. Yang et al.<sup>67</sup> synthesised pseudo-normal images from images with lesions to improve registration. They used a variational auto-encoder to learn the brain appearance from the normal areas of the pathological images only and estimate the reconstruction uncertainty of the predicted pseudo-normal image. Various auto-encoder models have been explored in 68–74 for brain lesion detection. Adversarial learning has been used as well to learn mappings between abnormal and normal tissues, and detect tumours on brain MR images<sup>60,75,76</sup> or fluid on optical coherence tomography images of the retina.<sup>77</sup> Generative adversarial networks have also been used to generate pseudo-healthy PET images from T1w MR images to detect hypometabolic lesions in the context of epilepsy.<sup>78</sup> Except for 67, these approaches do not estimate the synthesis uncertainty, which is important to assess the significance of the anomalies detected. Future work will consist in comparing the proposed method with several of these deep learning-based approaches.

A limitation of the proposed strategy is the computational cost. The bottleneck is the inter-subject registration, which takes on average 18 min per atlas when run on two standard CPUs for a reference image of size  $256 \times 256 \times 166$ . When the inter-subject registrations are run in parallel for all the atlases, an abnormality map can be generated in less than 25 min. The inter-subject registration step could be accelerated using deep learning-based approaches that can perform pairwise registration in less than a minute.<sup>79–81</sup>

## 6 Conclusion

We proposed an anomaly detection framework for the individual analysis of PET images. Subject-specific abnormality maps were obtained by comparing the subject's PET image to a model of healthy PET appearance that is specific to the subject under investigation. This model was generated from demographically and morphologically-matched PET scans from a control dataset. The abnormality maps could have two complementary uses. They could i) help clinicians in their diagnosis by highlighting, in a data-driven fashion, the pathological areas obtained from the individual PET data, and ii) improve the interpretability of subsequent analyses, such as computer-aided diagnosis or spatio-temporal modelling.

## Appendix A: Summary of the workflow

In this appendix we summarise the steps necessary to generate a subject-specific abnormality map from a patient's MR and PET images using a dataset of MR and PET images from control subjects. In particular, we specify the tools and parameters used.

### A.1 Intra-subject registration

To build the control dataset, each control MR image is rigidly registered to the PET image using NiftyReg<sup>1</sup>:

```
reg_aladin -ref <ControlPET> -flo <ControlMRI>
           -aff <rigid_ ControlMRI _to_ ControlPET>
           -res <ControlMRI.to.PET>
           -maxit 10 -rigOnly
```

The target patient's MR image is registered to its PET image in the same way.

### A.2 Atlas pre-selection

Only the controls with the same gender and closest in terms of age to the target patient are pre-selected. Pre-selecting around 30 atlases reduces the computational time while maintaining a good synthesis accuracy.<sup>30</sup>

### A.3 Inter-subject registration

For each pre-selected atlas, the control MR image is non-rigidly registered to the target patient MR image and the same transformation is then applied to the control PET image using NiftyReg:

```
reg_aladin -ref <TargetMRI.to.PET> -flo <ControlMRI.to.PET>
           -aff <affine_ ControlMRI.to.PET _to_ TargetMRI.to.PET>
reg_f3d -ref <TargetMRI.to.PET> -flo <ControlMRI.to.PET>
        -aff <affine_ ControlMRI.to.PET _to_ TargetMRI.to.PET>
        -cpp <cpp_ ControlMRI.to.PET _to_ TargetMRI.to.PET>
        -res <res_ ControlMRI.to.PET _to_ TargetMRI.to.PET>
        -sx 5 --nmi -ln 3 -lp 3 -be 0.01 -maxit 1000
reg_resample -ref <TargetMRI.to.PET> -flo <ControlMRI.to.PET>
            -trans <cpp_ ControlMRI.to.PET _to_ TargetMRI.to.PET>
            -res <res_ ControlPET _to_ TargetMRI.to.PET>
            -pad nan -inter 1
```

### A.4 Atlas local selection

The registered controls morphologically the most similar to the target patient are identified at a local scale. For each voxel, the SSIM is computed between the target subject MR image and each control MR image using Eq. (1) with  $\sigma_G = 3$ ,  $C_1 = \left( \frac{\max(\max(I^{MRI}), \max(J^{MRI}))}{\max(\min(\min(I^{MRI}), \min(J^{MRI})), 1)} \times 0.001 \right)^2$  and  $C_2 = \left( \frac{\max(\max(I^{MRI}), \max(J^{MRI}))}{\max(\min(\min(I^{MRI}), \min(J^{MRI})), 1)} \times 0.003 \right)^2$ .

---

<sup>1</sup><https://github.com/KCL-BMEIS/niftyreg>

#### *A.5 Subject-specific model of healthy PET appearance*

The similarity measure is then ranked across all the control subjects and the ranks are converted to weights using Eq. (2) with  $\beta = 0.25$ . To generate the subject-specific healthy-population mean and standard deviation, the controls' registered PETs are fused using a weighted average and a weighted standard deviation as described in Eq. (3).

#### *A.6 Subject-specific abnormality map*

The subject-specific abnormality map is finally obtained by comparing the target patient's PET image with the subject-specific model using the regularised Z-score proposed in Eq. (4) with the percentile set to  $k = 10$ .

## *Disclosures*

Authors have no conflict of interest regarding this article.

## *Acknowledgments*

The authors would like to thank Anne Bertrand for her advice and support during the development of this work.

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement N° PCOFUND-GA-2013-609102, through the PRESTIGE programme coordinated by Campus France, from the European Research Council (ERC) under grant agreement N° 678304 and from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" programme, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6).

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 NeuroImaging at the University of Southern California.

Data collection and sharing for this project was also funded by the Frontotemporal Lobar Degeneration Neuroimaging Initiative (National Institutes of Health Grant R01 AG032306). The study is coordinated through the University of California, San Francisco, Memory and Aging Center. FTLDNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## References

- 1 X. De Tiege, S. Goldman, S. Laureys, *et al.*, “Regional cerebral glucose metabolism in epilepsies with continuous spikes and waves during sleep,” *Neurology* **63**(5), 853–857 (2004).
- 2 L. Mosconi, W. H. Tsui, K. Herholz, *et al.*, “Multicenter Standardized 18F-FDG PET Diagnosis of Mild Cognitive Impairment, Alzheimer’s Disease, and Other Dementias,” *J. Nucl. Med.* **49**(3), 390–398 (2008).
- 3 C. Cerami, A. Dodich, G. Lettieri, *et al.*, “Different FDG-PET metabolic patterns at single-subject level in the behavioral variant of fronto-temporal dementia,” *Cortex* **83**, 101–112 (2016).
- 4 K. Herholz, “FDG PET and differential diagnosis of dementia,” *Alzheimer Dis. Assoc. Disord.* **9**(1), 6–16 (1995).
- 5 P. K. Panegyres, J. M. Rogers, M. McCarthy, *et al.*, “Fluorodeoxyglucose-Positron Emission Tomography in the differential diagnosis of early-onset dementia: a prospective, community-based study,” *BMC Neurology* **9**(1) (2009).
- 6 W.-D. Heiss and S. Zimmermann-Meinzingen, “PET imaging in the differential diagnosis of vascular dementia,” *J. Neurol. Sci.* **322**(1-2), 268–273 (2012).
- 7 R. Ossenkoppele, N. D. Prins, Y. A. Pijnenburg, *et al.*, “Impact of molecular imaging on the diagnostic process in a memory clinic,” *Alzheimer’s & Dementia* **9**(4), 414–421 (2013).
- 8 L. Rice and S. Bisdas, “The diagnostic value of FDG and amyloid PET in Alzheimer’s disease—A systematic review,” *Eur. J. Radiol.* **94**, 16–24 (2017).
- 9 F. Bouwman, S. Orini, F. Gandolfo, *et al.*, “Diagnostic utility of FDG-PET in the differential diagnosis between different forms of primary progressive aphasia,” *Eur. J. Nucl. Med. Mol. Imaging* **45**(9), 1526–1533 (2018).
- 10 D. Perani, P. A. Della Rosa, C. Cerami, *et al.*, “Validation of an optimized SPM procedure for FDG-PET in dementia diagnosis in a clinical setting,” *NeuroImage: Clinical* **6**, 445–454 (2014).
- 11 M. Signorini, E. Paulesu, K. Friston, *et al.*, “Rapid Assessment of Regional Cerebral Metabolic Abnormalities in Single Subjects with Quantitative and Nonquantitative [18f]FDG PET: A Clinical Validation of Statistical Parametric Mapping,” *NeuroImage* **9**(1), 63–80 (1999).
- 12 S. Minoshima, K. A. Frey, R. A. Koeppe, *et al.*, “A Diagnostic Approach in Alzheimer’s Disease Using Three-Dimensional Stereotactic Surface Projections of Fluorine-18-FDG PET,” *J. Nucl. Med.* , 12 (1995).
- 13 A. Drzezga, T. Grimmer, M. Riemenschneider, *et al.*, “Prediction of Individual Clinical Outcome in MCI by Means of Genetic Assessment and <sup>18</sup>F-FDG PET,” *J. Nucl. Med.* **46**(10), 1625–1632 (2005).
- 14 D. Renard, R. Vandenberghe, L. Collombier, *et al.*, “Glucose metabolism in nine patients with probable sporadic Creutzfeldt-Jakob disease: FDG-PET study using SPM and individual patient analysis,” *J. Neurol.* **260**(12), 3055–3064 (2013).
- 15 K. J. Friston, A. P. Holmes, K. J. Worsley, *et al.*, “Statistical parametric maps in functional imaging: A general linear approach,” *Hum. Brain Mapp.* **2**(4), 189–210 (1994).
- 16 J. C. Patterson, D. L. Lilien, A. Takalkar, *et al.*, “Early Detection of Brain Pathology Suggestive of Early AD Using Objective Evaluation of FDG-PET Scans,” *Int. J. Alzheimer’s Disease* (2011).

- 17 C. Cerami, A. Dodich, L. Greco, *et al.*, “The Role of Single-Subject Brain Metabolic Patterns in the Early Differential Diagnosis of Primary Progressive Aphasias and in Prediction of Progression to Dementia,” *J. Alzheimers Dis.* **55**(1), 183–197 (2016).
- 18 N. Burgos, M. J. Cardoso, A. Mendelson, *et al.*, “Subject-specific models for the analysis of pathological FDG PET data,” in *Medical Image Computing and Computer-Assisted Intervention*, 651–658 (2015).
- 19 N. Burgos, J. Samper-González, A. Bertrand, *et al.*, “Individual analysis of molecular brain imaging data through automatic identification of abnormality patterns,” in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, 13–22 (2017).
- 20 M. Modat, D. M. Cash, P. Daga, *et al.*, “A symmetric block-matching framework for global registration,” *Proc. SPIE Medical Imaging* **9034** (2014).
- 21 D. G. M. Murphy, C. DeCarli, A. R. McIntosh, *et al.*, “Sex Differences in Human Brain Morphometry and Metabolism: An In Vivo Quantitative Magnetic Resonance Imaging and Positron Emission Tomography Study on the Effect of Aging,” *Arch. Gen. Psychiatry* **53**(7), 585–594 (1996).
- 22 K. P. Cosgrove, C. M. Mazure, and J. K. Staley, “Evolving Knowledge of Sex Differences in Brain Structure, Function, and Chemistry,” *Biol. Psychiatry* **62**(8), 847–855 (2007).
- 23 G. Kalpouzos, G. Chételat, J.-C. Baron, *et al.*, “Voxel-based mapping of brain gray matter volume and glucose metabolism profiles in normal aging,” *Neurobiol. Aging* **30**(1), 112–124 (2009).
- 24 P. K. Curiati, J. H. Tamashiro-Duran, F. L. S. Duran, *et al.*, “Age-Related Metabolic Profiles in Cognitively Healthy Elders: Results from a Voxel-Based [18F]Fluorodeoxyglucose-Positron-Emission Tomography Study with Partial Volume Effects Correction,” *Am. J. Neuroradiology* **32**(3), 560–565 (2011).
- 25 D. S. Knopman, C. R. Jack, H. J. Wiste, *et al.*, “18f-fluorodeoxyglucose positron emission tomography, aging, and apolipoprotein E genotype in cognitively normal persons,” *Neurobiol. Aging* **35**(9), 2096–2106 (2014).
- 26 M. Modat, G. R. Ridgway, Z. A. Taylor, *et al.*, “Fast free-form deformation using graphics processing units,” *Comput. Methods Programs Biomed.* **98**(3), 278–84 (2010).
- 27 Z. Wang, A. C. Bovik, H. R. Sheikh, *et al.*, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Proc.* **13**(4), 600–612 (2004).
- 28 P. Cachier, E. Babinet, D. Dormont, *et al.*, “Iconic feature based nonrigid registration: the PASHA algorithm,” *Comput. Vision Image Understanding* **89**(2-3), 272–298 (2003).
- 29 N. Burgos, F. Guerreiro, J. McClelland, *et al.*, “Iterative framework for the joint segmentation and CT synthesis of MR images: application to MRI-only radiotherapy treatment planning,” *Phys. Med. Biol.* **62**(11), 4237 (2017).
- 30 N. Burgos, *Image Synthesis for the Attenuation Correction and Analysis of PET/MR Data*. Doctoral, UCL (University College London) (2016).
- 31 N. Burgos, M. J. Cardoso, K. Thielemans, *et al.*, “Attenuation correction synthesis for hybrid PET-MR scanners: Application to brain studies,” *IEEE Trans. Med. Imaging* **33**(12), 2332–2341 (2014).
- 32 R. C. Petersen, P. S. Aisen, L. A. Beckett, *et al.*, “Alzheimer’s disease neuroimaging initiative (ADNI): clinical characterization,” *Neurology* **74**(3), 201–209 (2010).

- 33 S. M. Landau, M. A. Mintun, A. D. Joshi, *et al.*, “Amyloid deposition, hypometabolism, and longitudinal cognitive decline,” *Ann. Neurol.* **72**(4), 578–586 (2012).
- 34 C. R. Jack, M. A. Bernstein, B. J. Borowski, *et al.*, “Update on the magnetic resonance imaging core of the Alzheimer’s disease neuroimaging initiative,” *Alzheimer’s & Dementia* **6**(3), 212–220 (2010).
- 35 W. J. Jagust, S. M. Landau, R. A. Koeppe, *et al.*, “The Alzheimer’s disease neuroimaging initiative 2 PET core: 2015,” *Alzheimer’s & Dementia* **11**(7), 757–771 (2015).
- 36 N. J. Tustison, B. B. Avants, P. A. Cook, *et al.*, “N4ITK: improved N3 bias correction,” *IEEE Trans. Med. Imaging* **29**(6), 1310–1320 (2010).
- 37 M. Cardoso, M. Modat, R. Wolz, *et al.*, “Geodesic information flows: Spatially-variant graphs and their application to segmentation and fusion,” *IEEE Trans. Med. Imaging* **34**(9), 1976–1988 (2015).
- 38 S. Minoshima, K. A. Frey, N. L. Foster, *et al.*, “Preserved pontine glucose metabolism in alzheimer disease: a reference region for functional brain image (pet) analysis,” *J. Comp. Assist. Tomo.* **19**(4), 541–547 (1995).
- 39 C. Hutton, J. Declerck, M. A. Mintun, *et al.*, “Quantification of <sup>18</sup>F-florbetapir PET: comparison of two analysis methods,” *Eur. J. Nucl. Med. Mol. Imaging* **42**(5), 725–732 (2015).
- 40 J. Samper-González, N. Burgos, S. Bottani, *et al.*, “Reproducible evaluation of classification methods in Alzheimer’s disease: Framework and application to MRI and PET data,” *NeuroImage* **183**, 504–521 (2018).
- 41 F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine learning in python,” *J. Machine Learning Research* **12**(Oct), 2825–2830 (2011).
- 42 A. Forsberg, H. Engler, O. Almkvist, *et al.*, “PET imaging of amyloid deposition in patients with mild cognitive impairment,” *Neurobiol. Aging* **29**(10), 1456–1465 (2008).
- 43 W. Jagust, “Positron emission tomography and magnetic resonance imaging in the diagnosis and prediction of dementia,” *Alzheimer’s & Dementia* **2**(1), 36–42 (2006).
- 44 D. S. Knopman, J. H. Kramer, B. F. Boeve, *et al.*, “Development of methodology for conducting clinical trials in frontotemporal lobar degeneration,” *Brain* **131**(11), 2957–2968 (2008).
- 45 A. M. Staffaroni, P. A. Ljubenkov, J. Kornak, *et al.*, “Longitudinal multimodal imaging and clinical endpoints for frontotemporal dementia clinical trials,” *Brain* **142**(2), 443–459 (2019).
- 46 J. L. Whitwell, S. A. Przybelski, S. D. Weigand, *et al.*, “Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: a cluster analysis study,” *Brain* **132**(11), 2932–2946 (2009).
- 47 J. D. Rohrer and H. J. Rosen, “Neuroimaging in frontotemporal dementia,” *Int. Rev. Psychiatry* **25**(2), 221–229 (2013).
- 48 G. D. Rabinovici, W. J. Jagust, A. J. Furst, *et al.*, “A $\beta$  amyloid and glucose metabolism in three variants of primary progressive aphasia,” *Ann. Neurol.* **64**(4), 388–401 (2008).
- 49 T. Grimmer, J. Diehl, A. Drzezga, *et al.*, “Region-Specific Decline of Cerebral Glucose Metabolism in Patients with Frontotemporal Dementia: A Prospective 18f-FDG-PET Study,” *Dement. Geriatr. Cogn. Disord.* **18**(1), 32–36 (2004).
- 50 Y. Jeong, S. S. Cho, J. M. Park, *et al.*, “18F-FDG PET Findings in Frontotemporal Dementia: An SPM Analysis of 29 Patients,” *J. Nucl. Med.* **46**(2), 233–239 (2005).

- 51 J. Diehl-Schmid, T. Grimmer, A. Drzezga, *et al.*, “Decline of cerebral glucose metabolism in frontotemporal dementia: a longitudinal 18f-FDG-PET-study,” *Neurobiol. Aging* **28**(1), 42–50 (2007).
- 52 D. Tosun, N. Schuff, G. D. Rabinovici, *et al.*, “Diagnostic utility of ASL-MRI and FDG-PET in the behavioral variant of FTD and AD,” *Ann. Clin. Transl. Neurol.* **3**(10), 740–751 (2016).
- 53 T. Kanda, K. Ishii, T. Uemura, *et al.*, “Comparison of grey matter and metabolic reductions in frontotemporal dementia using FDG-PET and voxel-based morphometric MR studies,” *Eur. J. Nucl. Med. Mol. Imaging* **35**(12), 2227–2234 (2008).
- 54 T. T. Rogers, J. Hocking, U. Noppeney, *et al.*, “Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging,” *Cogn. Affect. Behav. Neurosci.* **6**(3), 201–213 (2006).
- 55 R. Li, W. Zhang, H.-I. Suk, *et al.*, “Deep learning based imaging data completion for improved brain disease diagnosis,” in *Medical Image Computing and Computer-Assisted Intervention*, 305–312 (2014).
- 56 L. Bi, J. Kim, A. Kumar, *et al.*, “Synthesis of Positron Emission Tomography (PET) Images via Multi-channel Generative Adversarial Networks (GANs),” in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, 43–51 (2017).
- 57 A. Sikka, S. V. Peri, and D. R. Bathula, “MRI to FDG-PET: Cross-Modal Synthesis Using 3d U-Net for Multi-modal Alzheimer’s Classification,” in *Simulation and Synthesis in Medical Imaging*, 80–89 (2018).
- 58 W. Wei, E. Poirion, B. Bodini, *et al.*, “Learning Myelin Content in Multiple Sclerosis from Multimodal MRI Through Adversarial Training,” in *Medical Image Computing and Computer Assisted Intervention*, 514–522 (2018).
- 59 W. Wei, E. Poirion, B. Bodini, *et al.*, “Predicting PET-derived Demyelination from Multimodal MRI using Sketcher-Refiner Adversarial Training for Multiple Sclerosis,” *Med. Image Anal.* , 101546 (2019).
- 60 T. Xia, A. Chatsias, and S. A. Tsaftaris, “Pseudo-healthy synthesis with pathology disentanglement and adversarial learning,” *Medical Image Analysis* **64**, 101719 (2020).
- 61 D. Nie, X. Cao, Y. Gao, *et al.*, “Estimating CT Image from MRI Data Using 3d Fully Convolutional Networks,” in *Deep Learning and Data Labeling for Medical Applications*, 170–178 (2016).
- 62 K. Kläser, P. Markiewicz, M. Ranzini, *et al.*, “Deep Boosted Regression for MR to CT Synthesis,” in *Simulation and Synthesis in Medical Imaging*, 61–70 (2018).
- 63 L. Xiang, Q. Wang, D. Nie, *et al.*, “Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image,” *Med. Image Anal.* **47**, 31–44 (2018).
- 64 D. H. Ye, D. Zikic, B. Glocker, *et al.*, “Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization,” in *Medical Image Computing and Computer-Assisted Intervention*, 606–613 (2013).
- 65 Y. Tsunoda, M. Moribe, H. Orii, *et al.*, “Pseudo-normal Image Synthesis from Chest Radiograph Database for Lung Nodule Detection,” in *Advanced Intelligent Systems*, 147–155 (2014).
- 66 C. Bowles, C. Qin, R. Guerrero, *et al.*, “Brain lesion segmentation through image synthesis and outlier detection,” *NeuroImage: Clinical* **16**, 643–658 (2017).

- 67 X. Yang, X. Han, E. Park, *et al.*, “Registration of Pathological Images,” in *Simulation and Synthesis in Medical Imaging*, 97–107 (2016).
- 68 C. Baur, B. Wiestler, S. Albarqouni, *et al.*, “Deep autoencoding models for unsupervised anomaly segmentation in brain MR images,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 161–169 (2019).
- 69 H. Choi, S. Ha, H. Kang, *et al.*, “Deep learning only by normal brain PET identify unheralded brain anomalies,” *EBioMedicine* **43**, 447–453 (2019).
- 70 H. Uzunova, S. Schultz, H. Handels, *et al.*, “Unsupervised pathology detection in medical images using conditional variational autoencoders,” *Int. J. Comput. Assist. Radiol. Surg.* **14**(3), 451–461 (2019).
- 71 S. You, K. C. Tezcan, X. Chen, *et al.*, “Unsupervised Lesion Detection via Image Restoration with a Normative Prior,” in *International Conference on Medical Imaging with Deep Learning*, 540–556 (2019).
- 72 D. Zimmerer, F. Isensee, J. Petersen, *et al.*, “Unsupervised Anomaly Localization Using Variational Auto-Encoders,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, 289–297 (2019).
- 73 X. Chen, S. You, K. C. Tezcan, *et al.*, “Unsupervised lesion detection via image restoration with a normative prior,” *Medical Image Analysis* **64**, 101713 (2020).
- 74 C. Baur, S. Denner, B. Wiestler, *et al.*, “Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study,” *Medical Image Analysis* , 101952 (2021).
- 75 S. Andermatt, A. Horváth, S. Pezold, *et al.*, “Pathology Segmentation Using Distributional Differences to Images of Healthy Origin,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 228–238 (2019).
- 76 L. Sun, J. Wang, Y. Huang, *et al.*, “An Adversarial Learning Approach to Medical Image Synthesis for Lesion Detection,” *IEEE Journal of Biomedical and Health Informatics* **24**(8), 2303–2314 (2020).
- 77 T. Schlegl, P. Seeböck, S. M. Waldstein, *et al.*, “f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks,” *Med. Image Anal.* **54**, 30–44 (2019).
- 78 S. N. Yaakub, C. J. McGinnity, J. R. Clough, *et al.*, “Pseudo-normal PET Synthesis with Generative Adversarial Networks for Localising Hypometabolism in Epilepsies,” in *Simulation and Synthesis in Medical Imaging*, 42–51 (2019).
- 79 B. D. de Vos, F. F. Berendsen, M. A. Viergever, *et al.*, “A Deep Learning Framework for Unsupervised Affine and Deformable Image Registration,” *Med. Image Anal.* **52**, 128–143 (2019).
- 80 J. Fan, X. Cao, P.-T. Yap, *et al.*, “BIRNet: Brain image registration using dual-supervised fully convolutional networks,” *Med. Image Anal.* **54**, 193–206 (2019).
- 81 Z. Ding, G. Fleishman, X. Yang, *et al.*, “Fast Predictive Simple Geodesic Regression,” *Med. Image Anal.* **56**, 193–209 (2019).

## *Biographies of author*

**Ninon Burgos** is a CNRS researcher in the ARAMIS Lab, a joint laboratory between CNRS, Inria, Inserm and Sorbonne University within the Paris Brain Institute, France. She completed her PhD at University College London, UK, in 2016. Her research focuses on the development of computational imaging tools to improve the understanding and diagnosis of dementia.

**M. Jorge Cardoso** is Senior Lecturer in Artificial Medical Intelligence at King's College London, where he leads a research portfolio on big data analytics, quantitative radiology and value based healthcare. He is also the CTO of the new London Medical Imaging and AI Centre for Value-based Healthcare. Prior to King's, Dr Cardoso was a Lecturer at UCL, Technical Lead of the Quantitative Radiology Initiative at the National Hospital for Neurology and Neurosurgery (NHNN), and Engineering Lead of the Neuro-oncology Flagship Programme at UCL, Institute of Healthcare Engineering.

**Jorge Samper-González** is a data scientist currently working for Qynapse, a medical technology company that provides a cloud-based platform for the diagnosis, prognosis and drug efficacy measurement for central nervous system diseases. He did his PhD in the ARAMIS Lab, at the Paris Brain Institute, France, where he participated to the design and development of a software for neuroimaging studies, enabling reproducible classification and early prediction of Alzheimer's disease.

**Marie-Odile Habert** is a nuclear medicine physician at Pitié-Salpêtrière Hospital, senior lecturer in biophysics at Sorbonne Université, and affiliated to the LIB research lab (Biomedical Imaging Laboratory). She became a medical doctor in 1984, got a qualification in Nuclear Medicine in 1986 and master degrees in Biophysics and Medical Imaging between 1990 and 1996. Since 2015 she is also a scientific coordinator for PET/MR research at the Paris Brain Institute.

**Stanley Durrleman** is the co-head of joint Inria/ICM ARAMIS Lab at the Paris Brain Institute (ICM), France. He is also the coordinator of the ICM Center of Neuroinformatics, and scientific director of the ICM platform of biostatistics and bioinformatics. His work focuses on the development of new statistical and computational approaches for the analysis of image data and image-derived geometric data such a surface meshes.

**Sebastien Ourselin** is head of the School of Biomedical Engineering and Imaging Sciences and chair of Healthcare Engineering at King's College London. He was a professor at UCL, London, UK. Prior to that, he founded and led the CSIRO BioMediA Lab, Australia. He is an associate editor for Transactions on Medical Imaging, the Journal of Medical Imaging, Scientific Reports, and Medical Image Analysis. His research interests include image registration, segmentation, statistical shape modeling, surgical simulation, image-guided therapy, and minimally invasive surgery.

**Olivier Colliot** is a Research Director at CNRS and the co-head of the ARAMIS Lab, a joint laboratory between CNRS, Inria, Inserm and Sorbonne University within the Paris Brain Institute (ICM), France. He received his PhD in Signal Processing from Telecom ParisTech in 2003 and the Habilitation degree from University Paris-Sud in 2011. His research interests include machine learning, medical image analysis and their applications to neurological disorders.