A 3D convolutional neural network to classify subjects as Alzheimer's disease, frontotemporal dementia or healthy controls using brain 18F-FDG PET.

Antoine Rogeau, Florent Hives, Cecile Bordier, Hélène Lahousse, Vincent Roca, Thibaud Lebouvier, Florence Pasquier, Damien Huglo, Franck Semah, Renaud Lopes

NeuroImage 288 (2024) 120530

Contents lists available at ScienceDirect

NeuroImage

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

A 3D convolutional neural network to classify subjects as Alzheimer's disease, frontotemporal dementia or healthy controls using brain 18F-FDG PET

Antoine Rogeau [a,b,*], Florent Hives [a], Cécile Bordier [c,d], Hélène Lahousse [a], Vincent Roca [c,d], Thibaud Lebouvier [c,e], Florence Pasquier [c,e], Damien Huglo [a,f], Franck Semah [a,c], Renaud Lopes [c,d]

[a] Department of Nuclear Medicine, Lille University Hospitals, Lille, France
[b] Institute of Nuclear Medicine, University College London Hospitals NHS Foundation Trust, London, UK
[c] University of Lille, Inserm, CHU Lille, U1172 - LilNCog - Lille Neuroscience & Cognition, Lille, France
[d] Institut Pasteur de Lille, University of Lille, CNRS, Inserm, CHU Lille, US 41 - UAR 2014 - PLBS, Lille F-59000, France
[e] Memory Clinic, Lille University Hospitals, Lille, France
[f] Inserm, CHU Lille, University of Lille, U1189 OncoTHAI, Lille, France

ARTICLE INFO

Keywords:
FDG PET
Deep learning
Alzheimer's disease
Frontotemporal dementia
Convolutional neural network

ABSTRACT

With the arrival of disease-modifying drugs, neurodegenerative diseases will require an accurate diagnosis for optimal treatment. Convolutional neural networks are powerful deep learning techniques that can provide great help to physicians in image analysis. The purpose of this study is to introduce and validate a 3D neural network for classification of Alzheimer's disease (AD), frontotemporal dementia (FTD) or cognitively normal (CN) subjects based on brain glucose metabolism. Retrospective [18F]-FDG-PET scans of 199 CE, 192 FTD and 200 CN subjects were collected from our local database, Alzheimer's disease and frontotemporal lobar degeneration neuroimaging initiatives. Training and test sets were created using randomization on a 90 %-10 % basis, and training of a 3D VGG16-like neural network was performed using data augmentation and cross-validation. Performance was compared to clinical interpretation by three specialists in the independent test set. Regions determining classification were identified in an occlusion experiment and Gradient-weighted Class Activation Mapping. Test set subjects were age- and sex-matched across categories. The model achieved an overall 89.8 % accuracy in predicting the class of test scans. Areas under the ROC curves were 93.3 % for AD, 95.3 % for FTD, and 99.9 % for CN. The physicians' consensus showed a 69.5 % accuracy, and there was substantial agreement between them (kappa = 0.61, 95 % CI: 0.49–0.73). To our knowledge, this is the first study to introduce a deep learning model able to discriminate AD and FTD based on [18F]-FDG PET scans, and to isolate CN subjects with excellent accuracy. These initial results are promising and hint at the potential for generalization to data from other centers.

1. Introduction

Alzheimer's disease (AD) is a debilitating condition that affects millions of people worldwide. It is the fourth leading cause of disability-adjusted life-year (DALY) in older adults and its prevalence is set to increase dramatically due to population aging (Nichols et al., 2022). This reiterates the urgent need for effective disease-modifying drugs to treat neurodegenerative disorders.

An accurate diagnosis is paramount to the development of effective treatments because of the variety of pathological mechanisms at play in different neurodegenerative diseases. This is even more challenging in diseases where clinical pictures overlap such as AD and the frontotemporal dementia (FTD) spectrum, given the fact that gold standard diagnosis relies on pathological findings, only possible after the patient's death, and the importance of an early diagnosis to change the course of the disease.

Current diagnostic criteria rely on a compatible clinical presentation, cerebrospinal fluid (CSF) biomarkers, magnetic resonance imaging

* Corresponding author at: Department of Nuclear Medicine, Lille University Hospitals, Lille, France.
E-mail address: antoinerogeau2109@gmail.com (A. Rogeau).

https://doi.org/10.1016/j.neuroimage.2024.120530
Received 29 August 2023; Received in revised form 1 February 2024; Accepted 2 February 2024
Available online 3 February 2024

(MRI) and positron emission tomography (PET), giving to brain glucose metabolism a foremost position (Dubois et al., 2014; Gorno-Tempini et al., 2011; Rascovsky et al., 2011). However, in our clinical experience, interpretation of metabolic brain patterns can be challenging even for experts, as illustrated by the moderate inter-observer agreement (Brucher et al., 2015). Using standalone FDG-PET, approximately 20 % of subjects with dementia are incorrectly labeled AD (Bloudek et al., 2011). This is particularly problematic in cases with overlapping metabolic patterns as seen in the frontal variant of AD and behavioral variant FTD, or in logopenic primary progressive aphasia (PPA), related to AD pathology, and other PPA (Minoshima et al., 2021). With the anticipated approval of anti-amyloid therapy, which will certainly require a highly certain diagnosis for AD, there is a need to develop more accurate and objective interpretation tools to diagnose patients with precision.

Artificial intelligence (AI) techniques have been successfully applied in radiology and are increasingly popular in neuroimaging, especially for structural imaging (Hu et al., 2020; Jo et al., 2019; Nemoto et al., 2021). FDG-PET exhibits a fair level of evidence in detecting AD or FTD and is therefore an essential tool for diagnosis, a pivotal milestone in a patient's clinical pathway (Arbizu et al., 2018). Additionally, it has the ability to reclassify misdiagnosed subjects, which makes a strong case for applying AI techniques to metabolic imaging (Jack et al., 2010; Perini et al., 2021). A remarkable example of which is a recent work that classified with high accuracy FDG-PET scans, using a 3D convolutional neural network (CNN), into dementia with Lewy bodies (DLB), Alzheimer's spectrum or cognitively normal (Etminani et al., 2022). Regarding AD vs FTD classification using FDG-PET, supervised machine learning approaches, such as decision trees, support vector machine (SVM), or principal component analysis, have been employed, achieving accuracies ranging from 80 to 95 % (Perovnik et al., 2022; Sadeghi et al., 2008; Xia et al., 2014). Some studies have also incorporated additional information, such as clinical data or structural MRI (Dukart et al., 2013; Garcia-Gutierrez et al., 2022). Crucially, deep learning methods, which have the ability to autonomously learn complex features from data, have never been applied to AD vs FTD classification using brain glucose metabolism.

Therefore, the primary objective of the current study is to serve as a proof-of-concept, showcasing the potential of a 3D CNN approach in aiding the diagnosis of AD, FTD or normal aging through [18F]-FDG-PET. The research premise is that, following training, this model will display strong classification performance on test data acquired in the same conditions as training data, with results comparable to those achieved by specialists relying solely on [18F]-FDG-PET scans for diagnosis. Secondary objectives were to explore the model's underlying mechanisms through testing on a sample of DLB subjects, which is the third most prevalent dementing disorder and also exhibits cortical hypometabolism on FDG-PET (Minoshima et al., 2022). Additionally,

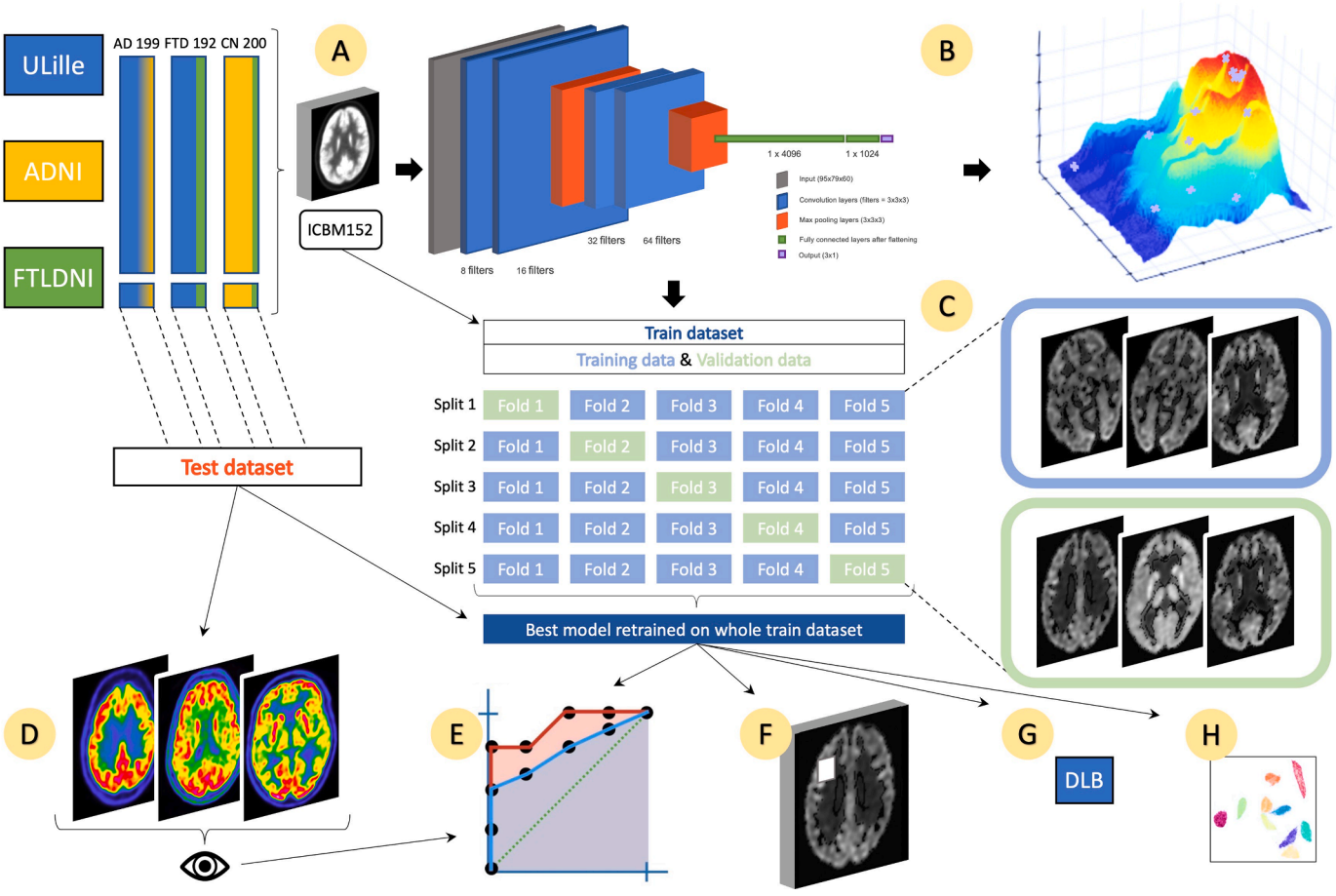

Fig. 1. Overview of methods. (A) Dataset of 199 CE, 192 FTD and 200 CN scans from ULille, ADNI and FTLDNI databases split between training & test sets on a 90 %−10 % basis. Scans were spatially normalized to ICBM152 template. (B) Network architecture and Bayes search to select hyperparameters. (C) Training on augmented data through 5-fold cross-validation, and best model kept for retraining on whole training set. (D) Clinical interpretation by 3 physicians of the native scans. (E) Performance comparison in the independent test set between model and physicians' consensual agreement. (F) Occlusion experiment and Grad-CAM. (G) Predictions of the model of DLB. (H) UMAP visualization. AD, Alzheimer's disease; ADNI, Alzheimer's disease neuroimaging initiative; CN, cognitively normal; DLB, dementia with Lewy bodies; FTD, frontotemporal dementia; FTLDNI, frontotemporal lobar degeneration neuroimaging initiative; ICBM, International Consortium for Brain Mapping; ULille, University of Lille.

visualization of brain regions driving classification and dimensionality reduction of extracted features were conducted.

2. Methods

2.1. Subjects

For an overall view of methods, see Fig. 1. We retrospectively collected [18F]-FDG-PET brain scans of AD, FTD and healthy controls from the University of Lille (ULille) Memory Clinic database acquired between 2015 and 2021, the Alzheimer's Disease Neuroimaging Initiative (ADNI, www.adni-info.org) 2 and 3 database, and the Frontotemporal Lobar Degeneration Neuroimaging Initiative (FTLDNI, http://memory.ucsf.edu/research/studies/nifd) database.

For each category, the ground truth label was based on a probable final clinical diagnosis as per consensual criteria (Gorno-Tempini et al., 2011; McKhann et al., 2011; Rascovsky et al., 2011). Subjects who showed conversion to another disorder were excluded. The AD category included 100 CE scans randomly selected from the ULille database and 104 scans from ADNI 2 & 3. ULille subjects also had typical CSF biomarkers (total tau $>$ 525 pg/mL, phospho-tau $>$ 73 pg/mL, Aβ1–42 $<$ 615 pg/mL and Aβ1–42/Aβ1–40 $<$ 5.6 %) and 16 subjects later demonstrated a confirmed diagnosis from autopsy or genetics, while ADNI subjects had a high-confidence diagnosis as reported by clinicians. All AD subjects were at the stage of dementia.

The FTD category comprised all 59 scans of the FTLDNI database who had FDG-PET, 4 of which were excluded due to absence of diagnosis, as well as all 137 FTD scans acquired at ULille. Among these 171 exhibited a behavioral presentation, 11 were classified as a semantic variant of PPA and 10 as an agrammatical variant of PPA. Sixteen ULille subjects later on demonstrated a certain diagnosis from autopsy or genetics. All FTD subjects were at the stage of dementia.

The CN category included 164 randomly selected baseline scans from ADNI 2 & 3 and all 36 control scans from FTLDNI (supplementary Fig. 1).

2.2. Data acquisition

PET-scans from ULille database were acquired on a hybrid 4-ring Biograph mCT-Flow PET/CT with 20-slice CT and 4 × 4 mm^2 lutetium oxyorthosilicate crystals (Siemens Medical Solutions USA, Inc., Molecular Imaging, Knoxville, TN, USA). Mean tracer dose was 177 MBq (SD = 19 MBq). Thirty minutes post injection, a low-dose CT scan of the brain was acquired for attenuation correction of the PET data, and 10-minute emission images were subsequently acquired. The PET data were reconstructed iteratively using an ordered subset expectation maximization algorithm with 8 iterations and 21 subsets. The reconstruction process included decay, random and scatter corrections and 2-mm full width at half-maximum Gaussian kernel smoothing. For each PET examination, the reconstructed images comprised a series of 109 axial slices with the following parameters: field of view = 408 × 408 × 221.3 mm3, matrix = 400 × 400 × 109, and voxel size = 1.02 × 1.02 × 2.03 mm3.

For ADNI acquisition the protocol can be found at https://adni.loni.usc.edu/wp-content/uploads/2010/05/ADNI2_PET_Tech_Manual_0142011.pdf. FDG tracer dose was 185 MBq (+/- 10 %), and between 30 and 60 min after injection 6 dynamic 3D scans of 5-min frames were acquired. A low-dose CT scan was acquired for attenuation correction, or for PET-only scanners an attenuation correction scan was acquired using rod sources.

All FTLDNI scans were acquired at the Mayo Clinic center on a GE Discovery RX PET/CT scanner. Participants were injected with 185 MBq (+/- 10 %) of FDG and acquisition started 30 min later, consisting of six 5-minute dynamic frames. A CT-scan, obtained prior to injection of FDG, was used for attenuation correction, and reconstruction used 3D filtered-back projection technique (Bejanin et al., 2020).

All images were reviewed by an expert (AR) for visual quality control.

2.3. Image preprocessing

Preprocessing was done using MATLAB R2020a (MathWorks, Natick, MA, USA) & the Statistical Parametric Mapping 12 (SPM12; Wellcome Trust Centre for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm). PET scans from ADNI and FTLDNI databases were downloaded in NIFTI format. Each of them consisted of 6 NIFTI files, which were realigned and averaged. PET scans from the ULille database were downloaded in DICOM format and subsequently converted to NIFTI with dcm2niix (Li et al., 2016). We spatially normalized each scan using default parameters of SPM12 normalization function and the International Consortium of Brain Mapping (ICBM) template. Subject-specific gray matter, white matter, CSF, bone, soft tissue and air probability maps were estimated from PET images using default parameters and tissue probability maps of SPM12 segmentation function. Masks were generated for voxels with a probability $>$ 0.7 of being gray matter, white matter or CSF. These individual masks were then merged into a unified mask, which was applied to spatially normalized volumes. As a result, voxels presumed to represent bone, soft tissue, or air were excluded. Feature-wise normalization was done through dividing each voxel by the maximum of the 3D scan it belonged to, so that all voxel values were between [0, 1]. No extra intensity normalization or smoothing was performed. Every scan was inspected in MRIcron (https://www.nitrc.org/projects/mricron) to ensure correct preprocessing. After visual inspection, 5 ADNI AD scans were excluded because of negative value voxels, which may have conflicted with the use of the ReLU function. Nineteen FTLDNI FTD, 6 ULille FTD, 1 ULille AD and 1 ADNI CN scans (N = 27, 4,6 %) were noticed to have segmentation issues, and were reprocessed changing the default clean up parameter from light cleanup to no cleanup. Preprocessing resulted in 3D volumes of 79×95×79 voxels (2 mm isotropic) along the standard x, y, z axes as used in SPM. Top and bottom slices without brain region information were removed, and final volume dimensions were 79×95×60.

2.4. Dataset and data augmentation

The whole dataset contained 591 preprocessed [18F]-FDG PET scans. Training and test sets were randomly built on a 90 % - 10 % basis for each of the three AD, FTD and CN groups in keeping with previous related works, resulting in 532 scans in the training set and 59 scans in the test set (Etminani et al., 2022; Nguyen et al., 2023). Dataset splitting was repeated randomly 30 times to improve stability of results. Due to human resource constraints and to uphold methodological validity, a single split was randomly selected for both comparison with specialist-based classification and secondary investigation of the model. Results are presented for this single split unless mentioned otherwise.

To prevent overfitting of the model, we augmented training data using a customized pipeline that generated batches and augmented data in real-time. Flipping along the sagittal plan, and +/- 10° random rotations across the 3 plans were performed in 50 % of cases to limit computational cost. Ten-percent translations across the 3 axes were performed for each scan, +/- 8, +/- 10 and +/- 6 voxels along the x, y and z axis, respectively.

2.5. Neural network building

Training was done at the Lille *In vivo* Imaging and Functional Exploration (LIIFE) research lab at Lille University Hospital, on a computer with Linux Ubuntu 20.04 operating system, 12 CPU Intel® Xeon® W-2133 3.60 GHz for a total of 102GB of memory, and an NVIDIA Quadro RTX 6000 with 24GB of memory. The network was inspired by the VGG16 architecture (Simonyan and Zisserman, 2014), and consisted of 2 blocks of two 3D convolution layers and a max pooling layer,

followed by a flattening layer and two dense layers (Fig. 1-B). The loss function used was cross-entropy. The input was preprocessed brain FDG-PET volumes, while the output was a probability value for AD, FTD and CN. To facilitate hyperparameter selection, we used Bayesian optimization for 200 iterations over the training set (supplementary Table 1). Following this, we retained Adagrad optimizer, learning rate = 0.0005, dropout rate = 0.5, batch size = 6, and performed end-to-end training on the augmented training set for 150 epochs using 5-fold cross-validation, early stopping with patience = 20 based on validation loss and saving the best model based on validation accuracy. Finally, the model with the highest validation accuracy within cross-validation was finetuned on the whole non-augmented training set with stochastic gradient descent with 0.0001 learning rate and 0.9 momentum for 50 epochs. This was repeated for each of the 30 random dataset splits.

In a complementary analysis, we repeated the procedure using the same hyperparameters while retaining only AD and FTD scans. This allowed us to assess the model's performance when it was exclusively trained on data of subjects with a neurodegenerative condition.

2.6. Specialist-based classification

Each scan of the test set was reviewed by 2 French board-accredited nuclear medicine physicians (HL & FH), and a resident in nuclear medicine (AR) with respectively 12, 9 years and 3 years of experience in the field of nuclear medicine. Non-normalized native brain volumes were visualized using ITK-SNAP (www.itksnap.org), without any clinical information (Yushkevich et al., 2006). Due to physicians' preference, a French rainbow lookup table (LUT) was used for visualization (supplementary Fig. 2). Each scan was classified by specialists into the following categories: AD, FTD, or CN. Scans for which there was disagreement were reviewed to reach consensual agreement. Fleiss' and Cohen's kappa coefficients were calculated to evaluate interobserver agreement and consensus/model agreement.

2.7. Model visualization with saliency maps

Building on the assumption that prediction probability for the real class of a scan will substantially decreased when voxels relevant for classification are occluded, we performed an occlusion experiment over the non-augmented training set of the specified random split (Zeiler and Fergus, 2014). A 5-voxel occluding cube was applied on each normalized scan with a stride of 2 for all 3 directions. The variations in prediction probability for the class of each scan are plotted as a function of the position of the occlusion cube, generating $38\times46\times29$ voxel volumes. To allow overlay, these volumes were normalized through dividing by their max, resized to $75\times91\times56$ voxel volumes through cubic spline interpolation, padded to match the $79\times95\times60$ dimensions, and averaged over each category. Thresholding was performed using the minimum voxel outside the brain. Similarly, Gradient-weighted Class Activation Mapping (Grad-CAM) heatmaps were calculated for the last convolution layer, resized to $75\times95\times60$ voxel volumes, and averaged over each class (Selvaraju et al., 2017). Finally, heatmaps are displayed over each category's averaged brain.

2.8. Voxel-based hypometabolism evaluation

For further exploration of metabolism between categories and for comparative purposes with the saliency maps, we conducted a voxelwise analysis over the training set of the specified random split to align with the occlusion experiment using Statistical Parametric Mapping 12 (SPM12, Wellcome Trust Centre for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm) in MATLAB R2020a. Two-sample *t*-tests were performed for AD vs CN scans, FTD vs CN scans and AD vs FTD scans. The modeling function utilized default parameters, and grand mean scaling was applied. Multiple analysis error was corrected employing peak-level family-wise error (pFWE) < 0.05.

2.9. Network extrapolability

With the purpose of testing the network on significantly different subjects, 20 DLB scans were included from the ULille database. Nineteen DLB scans had a probable diagnosis and 1 had a certain diagnosis as per the consensus diagnostic criteria (McKeith et al., 2017). Image preprocessing was done identically to scans from training/test sets.

2.10. Dimensionality reduction examination

We aimed to investigate how data separated based on class and acquisition center over the entire dataset. To achieve this, we utilized the Uniform Manifold Approximation and Projection (UMAP) technique to project the final dense layer of our model into a 2D space. This method was chosen for its ability to better preserve the global structure of data and faster processing compared to t-distributed stochastic neighbor embedding (McInnes et al., 2018).

3. Results

3.1. Demographics

Subject demographics are summarized in Table 1. ANOVA tests were computed on age and MMSE, while chi-square tests were performed for sex between categories within each dataset. There was no significant difference for age and sex across categories in the test set ($p > .05$). Conversely, MMSE scores were significantly different across categories in the test set, and post-hoc testing revealed CN subjects had a significantly higher MMSE than AD or FTD subjects ($p = .0007$ and 0.0001 respectively), but there was no significant difference between AD and FTD subjects ($p = .83$).

3.2. Model training

Training of 200 models using Bayesian optimization was completed in 49 h and 31 min. The results of the 20 best performing models can be found in supplementary Table 1. Cross-validation training of the final model with data augmentation on training set took 1 h and 58 min to complete for one random split. Within this split, model 1 was chosen for finetuning on the whole training set, as it demonstrated the highest validation accuracy of 91.5 % (Fig. 2).

3.3. Model evaluation

Model accuracy on the test set of the evaluated split was 89.8 %. Model classification is described in detail in Table 2. The AUC for predictions were 93.3 % for AD, 95.3 % for FTD and 99.9 % for CN (Fig. 3). The average AUC across all data splits were 94.9 %, 97.9 % and 99.3 %, respectively. The model performed best in detecting CN with 100 % sensitivity (all 20 CN cases were detected), 97 % specificity (38 out of 39 non-CN cases were correctly ruled out by the model), 95 % precision (20 out of 21 cases labeled as CN were correctly classified), and F1 score 98 %. Detailed metrics for AD and FTD can be found in Table 3. A typical example of error by the model can be found in Fig. 4.

In the complementary analysis with AD and FTD data only, model accuracy on the same test set was 87.2 %, and AUC was 91.3 %. All FTD patients were labeled as FTD and 75 % of AD patients were classified as AD. Average accuracy across all splits was 88.2 % and the average AUC was 95.7 % (Fig. 5).

3.4. Specialist-based classification

Across the test set, there was substantial agreement between physicians, as evidenced by a Fleiss' kappa coefficient of 0.61 (95 % CI: [0.49,

Table 1
Data demographics. Data are presented for the same split as used for comparison with specialist-based classification and secondary investigation. * Some subjects had several scans acquired at different timepoints; these are treated as different individuals for calculation of demographical variables. ** Two AD subjects and 12 FTD subjects did not have a MMSE score within a year of scanning. ANOVA used for age and MMSE, and chi-square for sex. AD, Alzheimer's disease; ADNI, Alzheimer's disease neuroimaging initiative; CN, cognitively normal; FTD, frontotemporal dementia; FTLDNI, frontotemporal lobar degeneration neuroimaging initiative; MCI, mild cognitive impairment; MMSE, mini-mental state examination; SD, standard deviation; ULille, University of Lille.

Train set				
	AD	FTD	CN	Total
No. subjects*	170	124	155	449
No. scans	179	173	180	532
Source				
ULille	50.3 %	71.1 %	0	40 %
ADNI	49.7 %	0	81.7 %	44.4 %
FTLDNI	0	28.9 %	18.3 %	15.6 %
Clinical data				
Age [mean (SD)]	68.2 (10.7)	68.4 (8.7)	71.5 (7.9)	$p = .0008$
Female Sex	53.6 %	37 %	56.7 %	$p = .0004$
MMSE [mean (SD)]	19.3 (6.2)	21.9 (6.5)	29.1 (1.2)	$p < .0001$
Test set				
	AD	FTD	CN	Total
No. subjects*	20	18	20	58
No. scans	20	19	20	59
Source				
ULille	50 %	73.7 %	0	40.7 %
ADNI	50 %	0	85 %	45.8 %
FTLDNI	0	26.3 %	15 %	13.5 %
Clinical data				
Age [mean (SD)]	69.9 (7.5)	68.6 (6.8)	72.1 (7.2)	$p = .31$
Female Sex	35 %	36.8 %	50 %	$p = .57$
MMSE [mean (SD)]	21 (6.3)	19.8 (9.1)	28.9 (1.0)	$p < .0001$
Overall dataset				
	AD	FTD	CN	Total
No. subjects*	190	134	172	496
No. scans	199	192	200	591
Source				
ULille	50.3 %	71.4 %	0	40.1 %
ADNI	49.7 %	0	82 %	44.5 %
FTLDNI	0	28.6 %	18 %	15.4 %
Clinical data				
Age [mean (SD)]	68.4 (10.4)	68.4 (8.5)	71.6 (7.8)	$p = .0002$
Female Sex	51.8 %	37 %	56 %	$p = .0004$
MMSE [mean (SD)]**	19.5 (6.2)	21.7 (6.8)	29.1 (1.2)	$p < .0001$

0.73]). Consensual physician labeling (Table 2) had a 69.5 % accuracy ($N = 41$). The highest specificity was observed for FTD at 92.5 %, but this contrasted with a low sensitivity of 47 %. Comparatively, CN showed the highest sensitivity of 90 % (Table 3). Further information on labeling and detailed metrics per physician can be found in supplementary Tables 2 and 3. Cohen's kappa between the consensus and the model for the whole test set was 0.59 (95 % CI: [0.48, 0.71]), and for incorrectly labeled cases by the consensus or the model was −0.08 (95 % CI: [−0.15, −0.02]).

3.5. Saliency maps

The most prominent regions in driving AD classification were found to be the precuneus and posterior cingulate cortex (PCC). The insular cortex and cerebellum were also found to be important in Grad-CAM. The FTD heatmaps showed that anterior regions such as the dorsolateral and mesial frontal regions were driving classification. The FTD Grad-CAM map also highlighted posterior associative regions, but this was less evident in the occlusion heatmap, whereas the occlusion map highlights the cerebellum. The CN occlusion heatmap shows a diffuse pattern while the Grad-CAM map highlights a few regions such as the cerebellum and precentral regions overlying a diffuse pattern (Figs. 6, 7).

3.6. Voxel-based hypometabolism evaluation

When performing the AD vs CN *t*-test, most hypometabolic regions were the precuneus/PCC region, the bilateral posterior temporoparietal cortex and the mesial temporal lobe (all pFWE < 0.001). When performing the FTD vs CN analysis, most significant voxels were found in the anterior insula, the whole prefrontal cortex, the anterior temporal

Table 2
Confusion matrix. AD, Alzheimer's disease; CN, cognitively normal; FTD, frontotemporal dementia.

		Actual labels			
		AD	FTD	CN	Total
Model labels	AD	15	1	0	16
	FTD	4	18	0	22
	CN	1	0	20	21
	Total	20	19	20	59
Physician labels	AD	14	3	2	19
	FTD	3	9	0	12
	CN	3	7	18	28
	Total	20	19	20	59

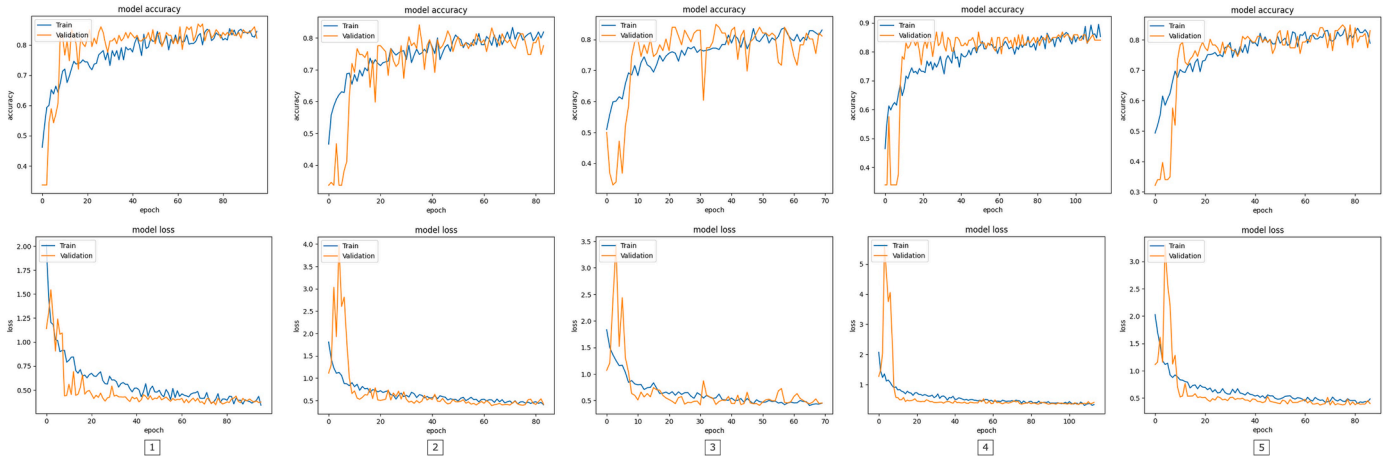

Fig. 2. Model training using 5-fold cross validation. Top graphs show training and validation accuracies. Bottom graphs show train and validation losses. Model 1 had the highest validation accuracy and was selected for further training.

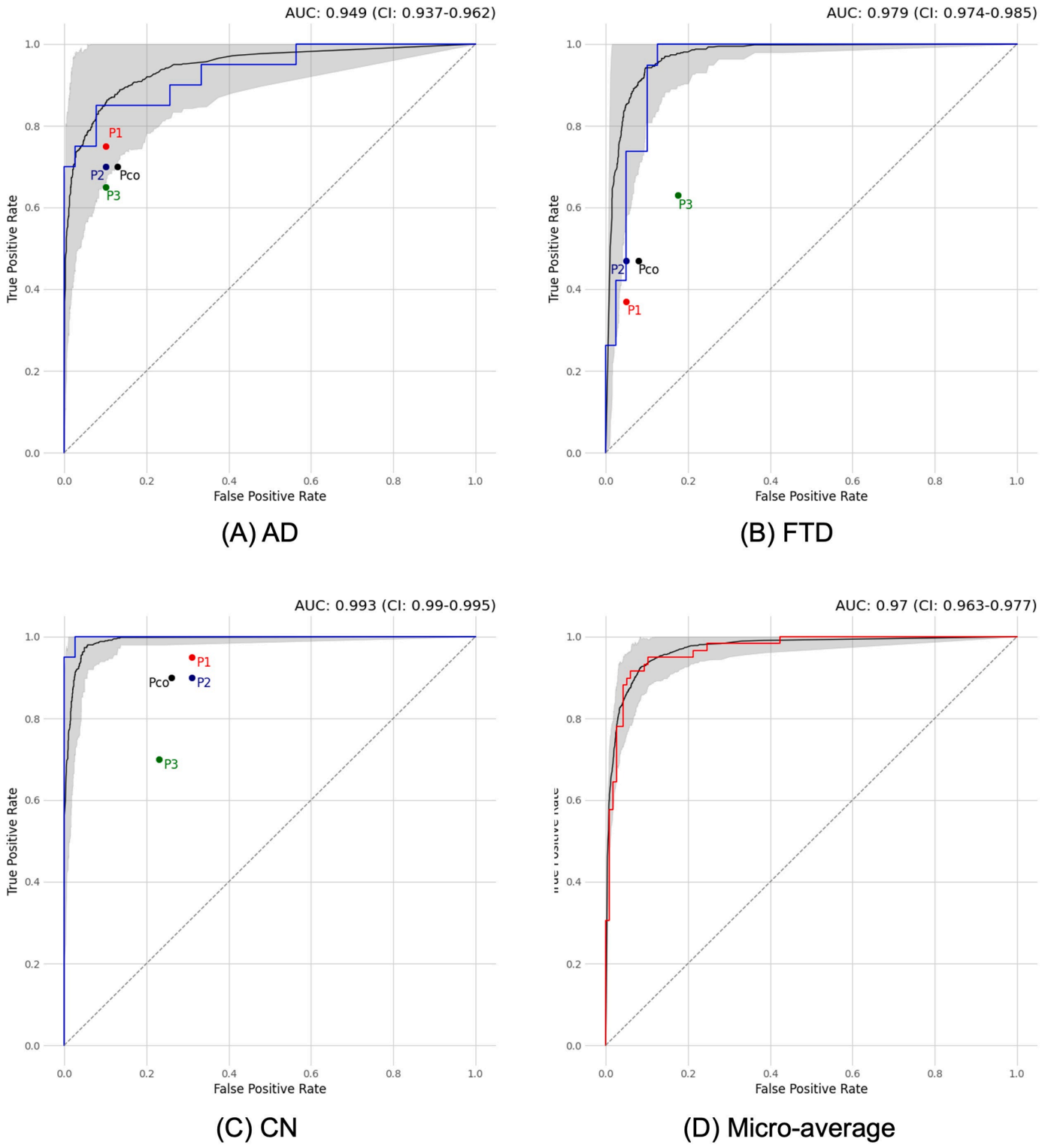

Fig. 3. ROC curves. Blue and red lines, selected iteration. Black lines, average of all iterations. AD, Alzheimer's disease; CN, cognitively normal; FTD, fronto-temporal; P1, physician 1; P2, physician 2; P3, physician 3; Pco, consensus between physicians; ROC, receiver operating characteristic. AUC and confidence intervals are presented for the averaging across the 30 dataset splits.

lobes, the basal ganglia and to a lesser extent the precuneus/PCC and posterior temporoparietal regions (all pFWE < 0.001). When performing the AD vs FTD analysis, a few voxels in the posterior temporoparietal and precuneus/PCC regions were found hypometabolic in AD subjects (pFWE < 0.01), while FTD subjects displayed hypometabolism of the ACC, prefrontal mesial cortex, anterior insula and OFPFC (pFWE < 0.01). Thresholded maps with pFWE < 0.05 are shown in Fig. 8.

3.7. Network extrapolability

Demographics of subjects with DLB can be found in supplementary Table 3. Thirteen DLB scans were classified as FTD, 7 as AD and 0 as CN.

Table 3
Metrics. AD, Alzheimer's disease; CN, cognitively normal; FTD, frontotemporal dementia; Se, sensitivity; Sp, specificity.

		Metrics				
		Se	Sp	Precision	F1-score	Support
Model	AD	75 % (15/20)	97 % (38/39)	94 % (15/16)	83 %	20
	FTD	95 % (18/19)	90 % (36/40)	82 % (18/22)	88 %	19
	CN	100 % (20/20)	97 % (38/39)	95 % (20/21)	98 %	20
		Se	Sp	Precision	F1-score	Support
Physician	AD	70 % (14/20)	87 % (34/39)	74 % (14/19)	72 %	20
	FTD	47 % (9/19)	92 % (37/40)	75 % (9/12)	58 %	19
	CN	90 % (18/20)	74 % (29/39)	64 % (18/28)	75 %	20

3.8. Dimensionality reduction examination

As shown in shown in Fig. 9, there appears to be a slight center separation in the CN class, where FTLDNI cases cluster on one side, although still displaying substantial overlap with ADNI cases. On the other hand, there is no evident separation by center for AD and FTD classes.

4. Discussion

The future of nuclear medicine in neurodegenerative diseases may require multiple scans over time for a single patient. In this study, we introduce a deep learning model able to classify [18F]-FDG-PET scans into AD, FTD or CN subjects with high accuracy, doing better than a consensus reached by physicians specialized in nuclear medicine in all metrics. Such a tool would be of great assistance to physicians to break free from intra- and inter-observer variability in clinical strategies involving repeated PET scans.

We extend the previous work of Etminani et al. by presenting a CNN model highly efficient in differentiating AD and FTD-spectrum diseases (Etminani et al., 2022). They reported in their study 96.4 %, 71.4 %, 96.2 % and 94.7 % AUC for AD, MCI, DLB and CN, respectively. Our model tends to show slightly better performances, but this should be put into perspective with the fact that we did not include MCI subjects, which, understandably, was the most misclassified category in their work. In addition, in the AD vs FTD analysis which better reflects the standard situation met in clinical settings, we found an average accuracy of 88.2 %, and the average AUC was 95.7 %. Regarding specialist-based diagnosis, while the consensus accuracy is similar at 69.5 % in the present study (compared to 57 % in Etminani et al. with an additional category), our Fleiss' kappa of 0.61 (compared to 0.19) indicates stronger agreement. This low variance in physician performance reflects a shared understanding of the test set and suggests that it is unlikely for another physician to outperform the model. On the other hand, the model and the physicians did not commit the same errors as shown by the absence of agreement (Cohen's kappa = - 0.08). Furthermore, we also extend classical machine learning approaches that aimed to

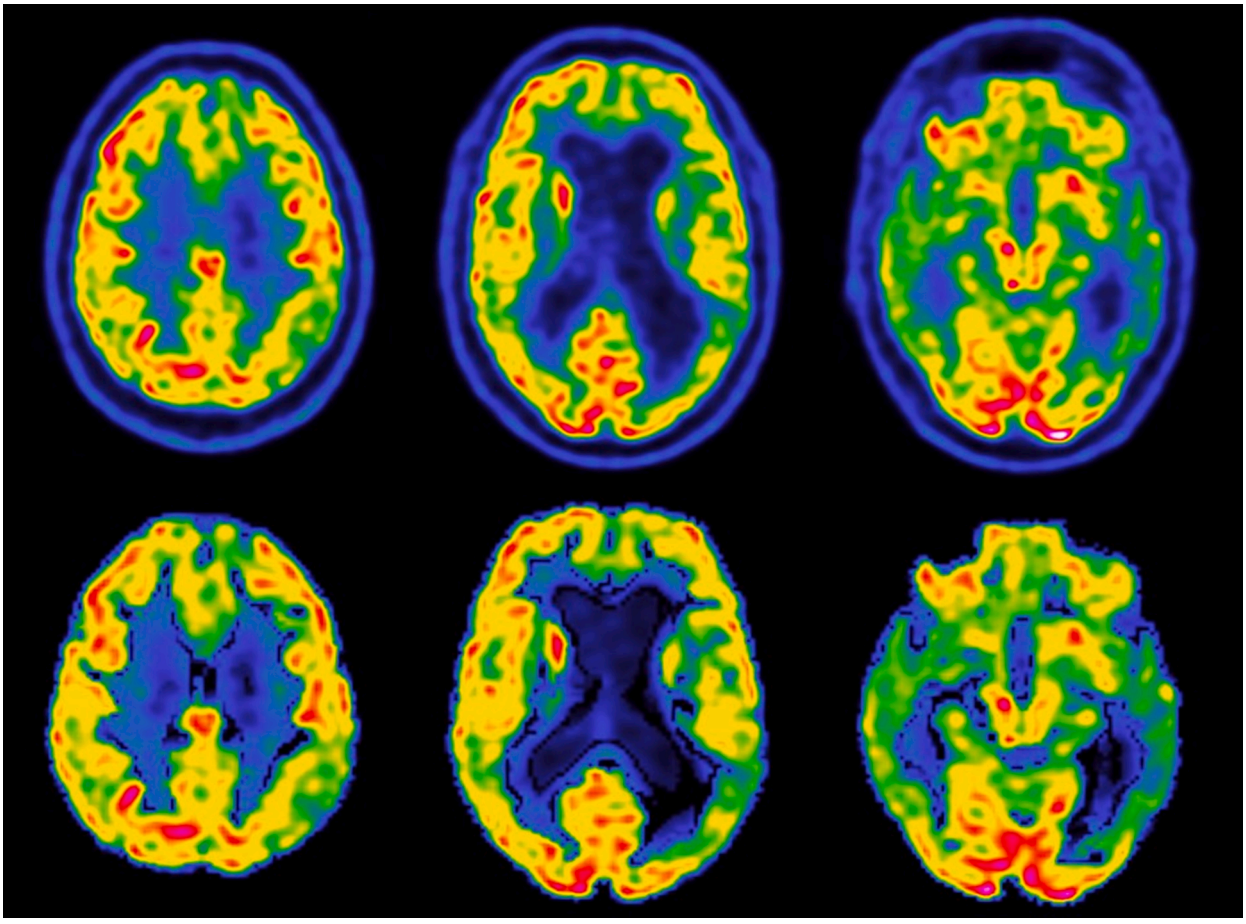

Fig. 4. AD scan labeled FTD. Top: native scan, bottom: preprocessed data. Hypometabolism in parietal associative posterior regions is observed and to a lesser degree in frontal mesial, whereas the posterior cingulate and precuneus appear relatively normal.

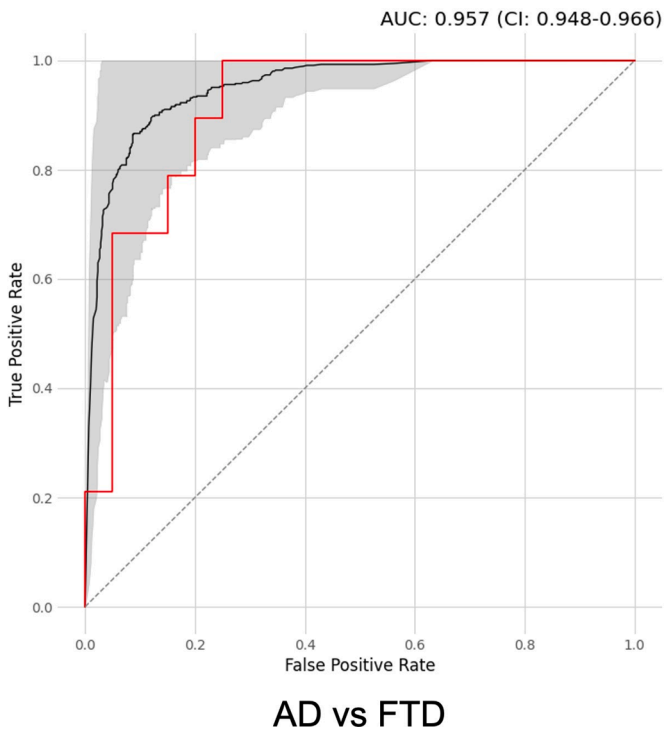

Fig. 5. ROC curve for AD vs. FTD analysis. Red line, selected iteration. Black lines, average of all iterations. ROC, receiver operating characteristic. AUC and confidence intervals are presented for the averaging across the 30 dataset splits.

distinguish AD and FTD using FDG-PET. An early work with spatial decision trees yielded around 90 % accuracy. However, the use of brain regions drawn empirically may have led to overfitting, and a high rating accuracy by experts, also approaching 90 %, indicates that cases might have been more straightforward to classify (Sadeghi et al., 2008). Another method using a multiple kernel algorithm demonstrated accuracy near 95 %, but once more empiric selection of the most efficient features may have resulted in some overfitting, and sample size was limited (Xia et al., 2014). More recently, an SVM approach yielded an overall accuracy of 78–80 % compared to expert accuracy of 71 % (Perovnik et al., 2022).

The main strength of our model lies in its exceptional accuracy to isolate CN subjects, and this could be envisioned as a first AI clinical application for screening brain scans and present them to physicians as high-probability to be normal/abnormal (Hao et al., 2022; McKinney et al., 2020). Our model accurately classified all CN subjects and erroneously labeled only 1 CE scan as CN. None of CN test subjects were reported to have converted to MCI or dementia since scanning. The clinician consensus had a high sensitivity too at 90 % (vs 100 %), but specificity was moderate at 74 % (vs 97 %). Interestingly, MMSE scores were significantly higher for dementia scans classified CN by physicians compared to true positives (supplementary Table 5). This brings support to the fact that the model may detect metabolic abnormalities still indiscernible to the human eye.

In the occlusion experiment and Grad-CAM analysis, we found salient regions consistent with the traditional hypometabolism pattern described in visual interpretation and illustrated in Fig. 8. The PCC was found to be the most important region by large to drive AD classification, but in clinical practice the emphasis tends to be put on other posterior associative areas such as temporoparietal junctions, and therefore it reiterates the need to scrupulously examine the PCC/precuneus when facing a possible diagnosis of AD. The PCC is an important area for cognition showing decreased [18F]-FDG uptake in early AD (Minoshima et al., 1997) and also, although inconsistently, decreased uptake in FTD (Scheltens et al., 2018). Contrarily to the aforementioned work, in which the PCC was a common driver for classification into AD, MCI or DLB (Etminani et al., 2022), herein the PCC was not found to guide classification towards FTD, and consequently should not be considered as a hallmark of neurodegenerative disease. This may also account for the misclassification of a few AD scans without hypometabolic PCC as FTD (Fig. 4). As expected, frontal regions were highlighted in both FTD occlusion and Grad-CAM maps. The CN heatmaps showed a diffuse pattern. Surprisingly, the cerebellum was a decisive region in several heatmaps driving classification in FTD occlusion and highlighted in both AD/CN Grad-CAM. Although it was not confirmed on the corrected maps of the voxelwise analysis, lower cerebellar metabolism was seen on the uncorrected maps (using $p < .001$) when comparing FTD to AD or CN. The cerebellum is debated to show altered cerebellar glucose metabolism in C9orf72 mutations ($N = 10$ in the training set) vs non-C9orf72 FTD (Castelnovo et al., 2019; Diehl-Schmid et al., 2019). However, it is considered intact in most neurodegenerative disorders and recommended as a reference region for intensity normalization by the European Association of Nuclear Medicine (Guedj et al., 2022). This calls for

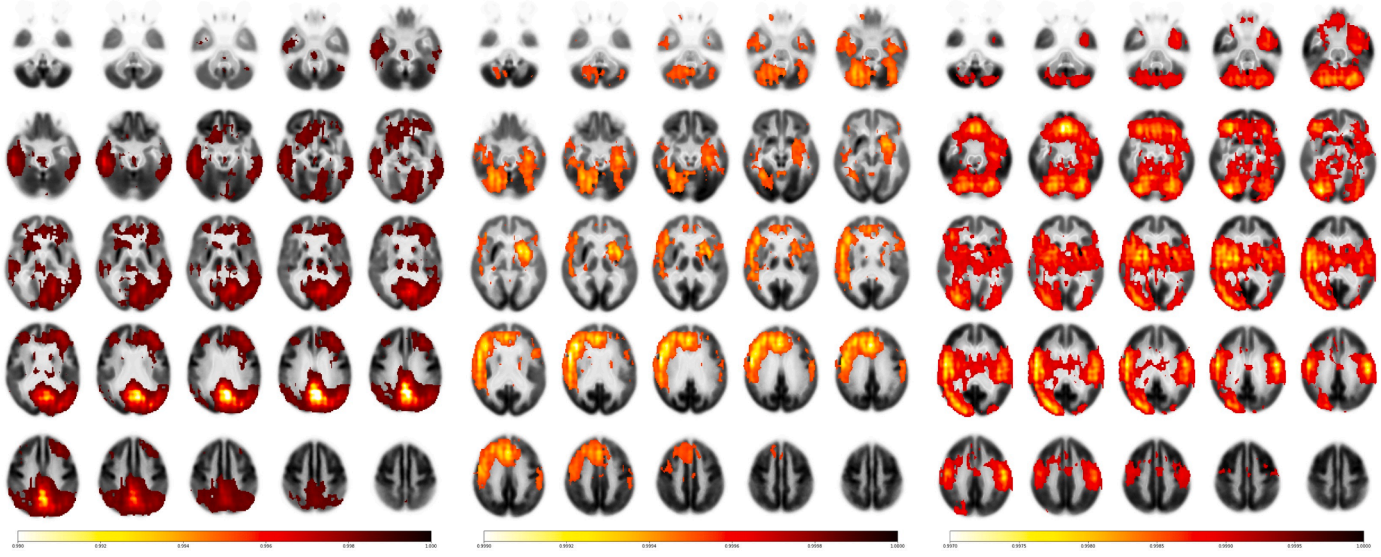

Fig. 6. Occlusion heatmaps. Left: Alzheimer's disease, middle: frontotemporal dementia, right: cognitively normal. Color-bars show the variation in prediction certitude.

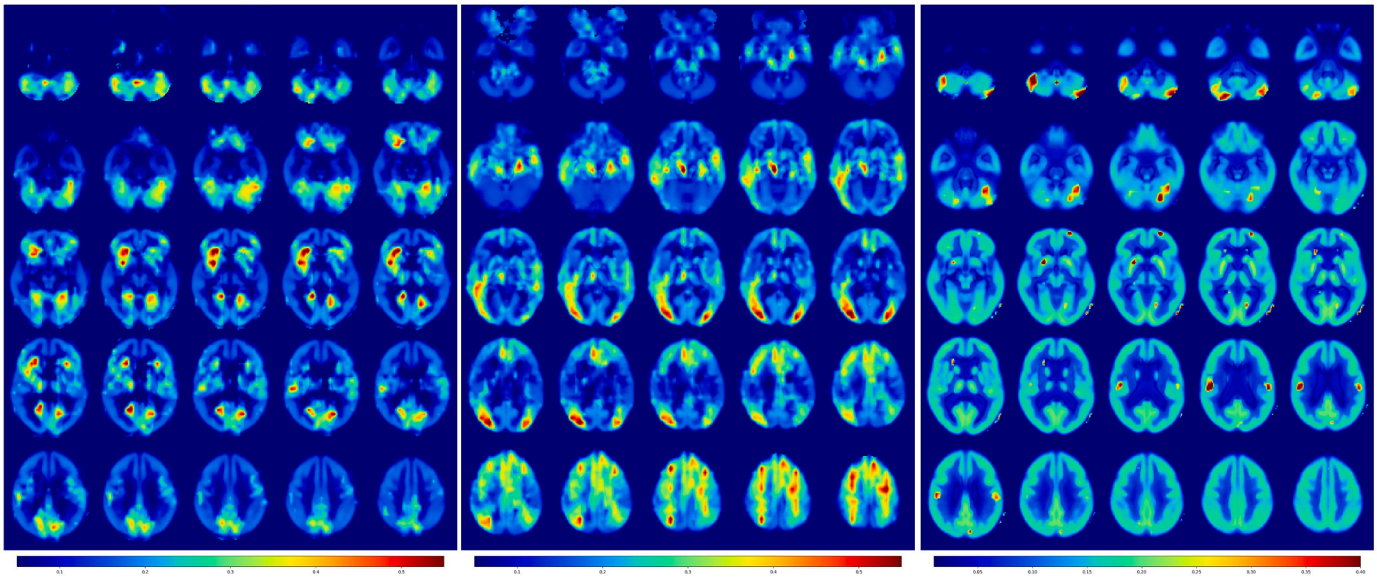

Fig. 7. Grad-CAM heatmaps. Left: Alzheimer's disease, middle: frontotemporal dementia, right: cognitively normal. Color-bars reflect weights from the final convolution layer for each class.

further investigation, and it may be interesting in the future to perform a voxelwise analysis between non-C9orf72 FTD and AD using a cerebellar mask.

In testing extrapolability on DLB, we found that all of them were classified as AD or FTD. This highlights the coherence of our model in successfully ruling out subjects with another neurodegenerative disease as healthy subjects, which would be the most prejudicious situation in a clinical setting. This is also in line with the findings of Etminani et al., in which the dimensionality reduction visualization clearly separated CN and DLB subjects.

Our UMAP dimensionality reduction revealed a slight center separation within the CN class, where FTLDNI cases cluster on one side, perhaps due to all FTLDNI scans being acquired on a single PET camera. However, there was no evident separation by center for AD and FTD.

4.1. Strengths

As shown in Fig. 2, we successfully prevented overfitting during the cross-validated training using several regularization methods such as data augmentation, batch normalization, dropout and early stopping. This, we believe, shows that the network could learn robust features during training that would allow its extrapolation to new brain scans with a different acquisition protocol. Similarly, training over a large multicentric dataset, well-balanced across the different categories would also facilitate utilization in different conditions. We also demonstrated in a supplementary analysis that although acquisition parameters varied between centers, there was no significant difference in accuracy between them (supplementary).

Another strength is that we could include subjects who received a certain diagnosis ($N = 30$). This is of interest since there is a large overlap in radiological presentations of neurodegenerative diseases (Olney et al., 2017), and protein-targeting drugs will require an unequivocal diagnosis. In addition, the rest of our subject had a probable diagnosis. Longitudinal analyses have shown individuals fulfilling FTD probable criteria will continue to do so over time or move to the certain category after postmortem analysis (Devenney et al., 2015). This is, however, to put into perspective with the fact that most often neurodegenerative pathologies overlap, as pure AD is thought to only represent 30 % of cases and in 30 % of cases associates with TDP-43 pathology (Villain and Dubois, 2019).

4.2. Limitations

Unfortunately, a same center could not provide subjects to all 3 categories, which would ensure the network did not learn any acquisition specificities. For example, the CN category did not include any ULille subjects. This is because our database, built from clinical data, does not include any healthy control subjects. This could also be viewed as a wider limit to the development of AI in clinical practice since healthy controls are not commonly found among hospital patients. On the same topic, data augmentation can theoretically increase bias between classes (Balestriero et al., 2022). However, it was required to expand data during training, and the absence of underfitting gives confidence in the absence of significant bias (Fig. 2).

It could be argued that subjects having several scans may have led to inflated performance of the model. However, except 1 subject who had 1-month interval scans, there was a minimum of 6 months between scans for other subjects which is enough to see progressive metabolic changes (Forster et al., 2011). Additionally, we repeated the classification analyses using only one scan for each patient and average accuracy was at 84.1 % still substantially higher than physician accuracy (supplementary Fig. 3).

Physicians reviewed studies in their native space, while the model labeled spatially and feature-wise normalized volumes. This decision aligned with physician preferences for examining scans and ensure that the scans they reviewed displayed the highest resolution. Therefore, if this introduced a bias, it is more likely that it contributed to increasing physician performance rather than the opposite.

Probably because of regions showing severe hypometabolism, SPM segmentation basic parameters first considered large cortical areas as not gray matter, and these scans ($N = 27$) required another segmentation treatment. This is a minor issue in the current study, but it reflects what may happen when developing an AI pipeline. Even if AI is successfully incorporated in imaging departments, this highlights the need for quality checks to ensure coherent results.

Finally, in the age of emerging blood biomarkers and new radiotracers, some may question the clinical relevance of using FDG-PET, which might seem an out-of-fashion approach to neurodegenerative classification. However, we argue that this is not the case. Lumbar puncture remains an invasive procedure, blood biomarkers are promising but have yet to be seen in clinical routine and amyloid PET can be twice as expensive (Contador et al., 2023; Teunissen et al., 2022).

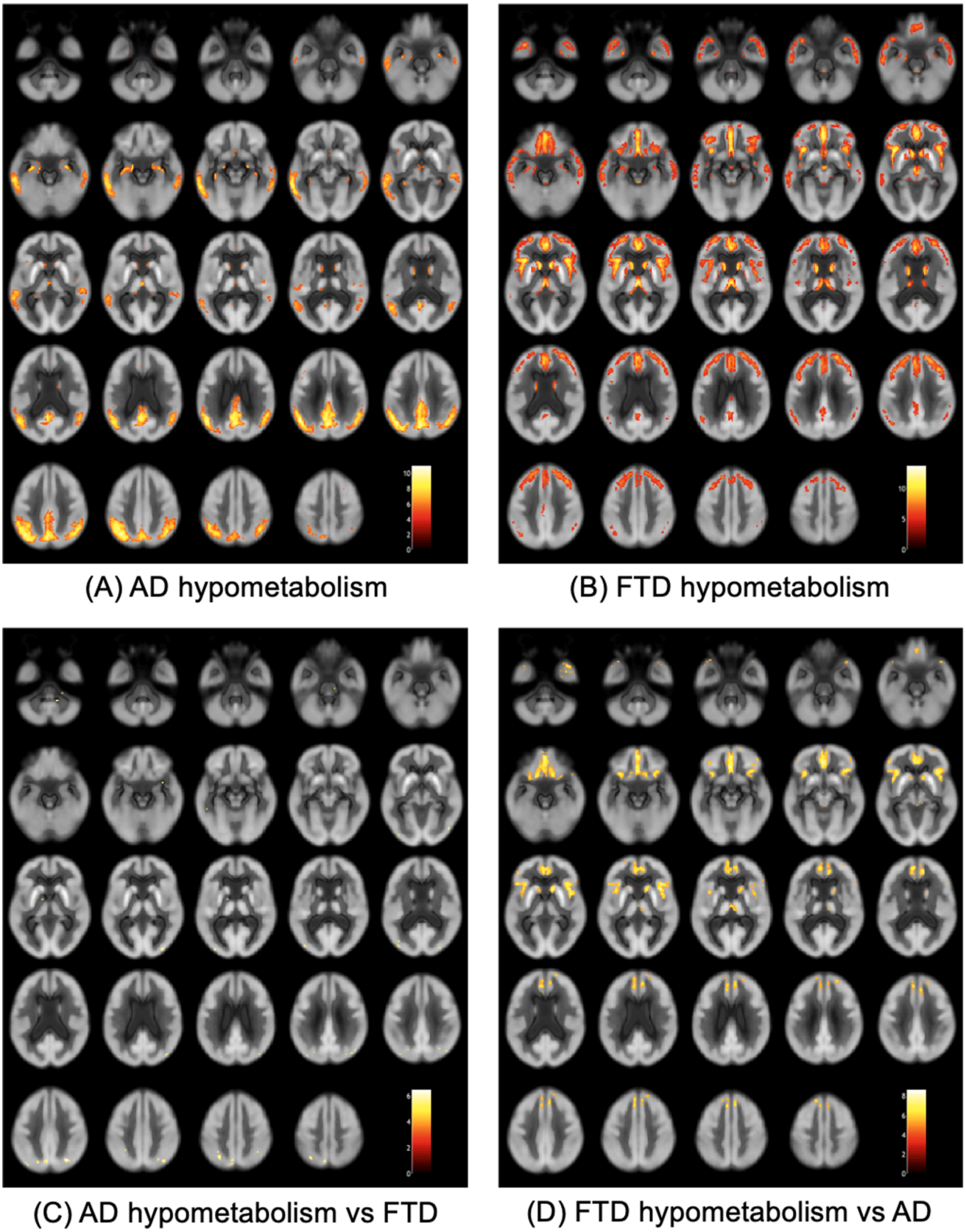

Fig. 8. Hypometabolism maps. A: AD hypometabolism compared to CN. B: FTD hypometabolism compared to CN. C: AD hypometabolism compared to FTD. D: FTD hypometabolism compared to AD. AD, Alzheimer's disease; CN, cognitively normal; FTD, frontemporal dementia. Color-bars show T-score values.

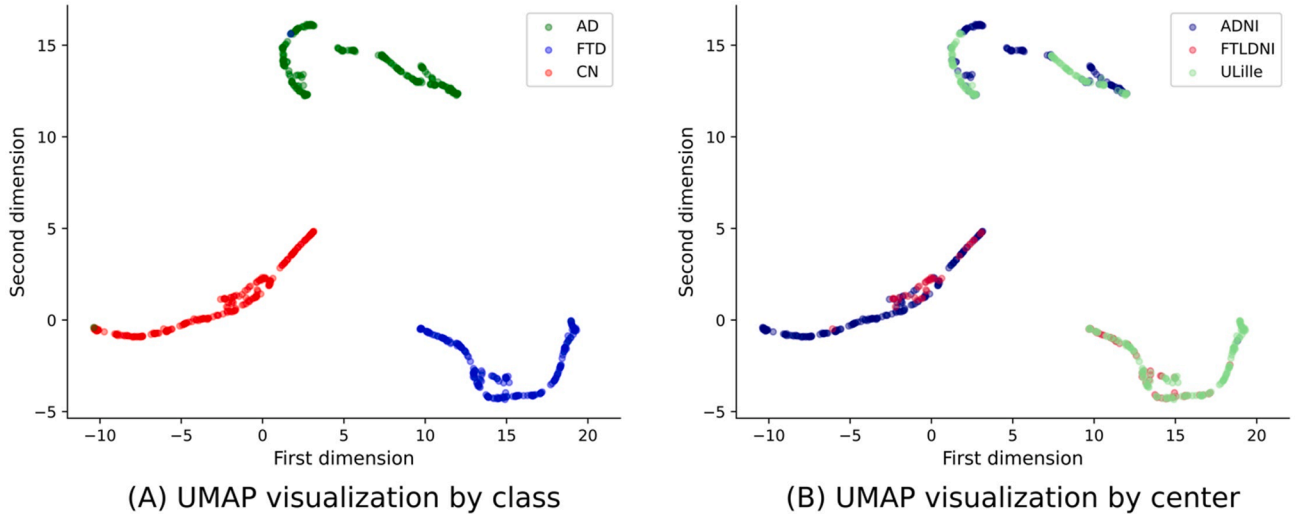

Fig. 9. Dimensionality reduction visualization. A: Visualization by class. B: Visualization by center. ADNI, Alzheimer's disease neuroimaging initiative; FTLDNI, frontotemporal lobar degeneration neuroimaging initiative; ULille, University of Lille; UMAP, uniform manifold approximation and projection.

Furthermore, a recent consortium has reaffirmed the importance of FDG-PET for dementia diagnosis, even prioritizing it over CSF markers in certain situations (Chetelat et al., 2020).

5. Conclusion

In this study, we demonstrated the ability of a tailor-made 3D CNN to accurately classify [18F]-FDG PET-scans between AD, FTD or CN subjects. Our results showed that this model outperforms clinical interpretation by experienced physicians and displays an excellent capability in identifying control subjects. These findings add to the growing field of AI in metabolic imaging and suggest our clinical practice may change in the nearby future integrating these tools. To our knowledge, this is the first work using deep learning techniques to classify AD vs FTD subjects based on brain glucose metabolism.

Data and code availability

ADNI and FTLDNI datasets are available on request at https://ida.loni.usc.edu/login.jsp. Scripts are publicly available on GitHub (https://github.com/rogeauA/NeuralNet_AD-FTD_FDG-PET).

Ethics statement

The present study adheres to the principles outlined in the reference methodology MR004 introduced by the Commission Nationale de l'Informatique et des Libertés (CNIL) for data protection and privacy. All procedures involving were conducted in accordance with the Declaration of Helsinki. Prior to the study, ULille participants provided written informed consent to the possible use of their medical data for research purposes. The study protocol was reviewed by the local ethics committee (*Service Protection des Données - Groupement Hospitalier de Territoire Lille Métropole Flandre Intérieure,* REF: DEC22–303). ADNI and FTLDNI subjects also provided written informed consent as per the enrolment protocols. To ensure the privacy and confidentiality of participants' data, all personal identifying information was anonymized and encrypted. Access to the data was strictly limited to the research team, and data storage and transfer followed secure protocols.

CRediT authorship contribution statement

Antoine Rogeau: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Florent Hives:** Writing – original draft, Supervision, Resources, Formal analysis, Conceptualization. **Cécile Bordier:** Writing – original draft, Software. **Hélène Lahousse:** Writing – original draft, Resources, Formal analysis. **Vincent Roca:** Writing – original draft. **Thibaud Lebouvier:** Writing – original draft, Resources, Conceptualization. **Florence Pasquier:** Writing – original draft, Resources. **Damien Huglo:** Writing – original draft, Conceptualization. **Franck Semah:** Writing – original draft, Resources, Conceptualization. **Renaud Lopes:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

ADNI is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California, and FTLDNI is coordinated by the University of California, San Francisco, Memory and Aging Center. ADNI and FTLDNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Public data collection and sharing was funded by ADNI (National Institutes of Health Grant U01 AG024904), DOD ADNI (Department of Defense award number W81XWH-12–2–0012), and FTLDNI (National Institutes of Health Grant R01 AG032306). 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. Canadian sites receive funds from the Canadian Institutes of Health Research. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2024.120530.

References

Arbizu, J., Festari, C., Altomare, D., Walker, Z., Bouwman, F., Rivolta, J., Orini, S., Barthel, H., Agosta, F., Drzezga, A., Nestor, P., Boccardi, M., Frisoni, G.B., Nobili, F., Disorders, E., 2018. Clinical utility of FDG-PET for the clinical diagnosis in MCI. Eur. J. Nucl. Med. Mol. ImAging 45, 1497–1508.

Balestriero, R., Bottou, L., LeCun, Y., 2022. The effects of regularization and data augmentation are class dependent. Adv. Neural Inf. Process. Syst. 35, 37878–37891.

Bejanin, A., Tammewar, G., Marx, G., Cobigo, Y., Iaccarino, L., Kornak, J., Staffaroni, A. M., Dickerson, B.C., Boeve, B.F., Knopman, D.S., Gorno-Tempini, M., Miller, B.L., Jagust, W.J., Boxer, A.L., Rosen, H.J., Rabinovici, G.D., 2020. Longitudinal structural and metabolic changes in frontotemporal dementia. Neurology 95, e140–e154.

Bloudek, L.M., Spackman, D.E., Blankenburg, M., Sullivan, S.D., 2011. Review and meta-analysis of biomarkers and diagnostic imaging in Alzheimer's disease. J. Alzheimers Dis. 26, 627–645.

Brucher, N., Mandegaran, R., Filleron, T., Wagner, T., 2015. Measurement of inter- and intra-observer variability in the routine clinical interpretation of brain 18-FDG PET-CT. Ann. Nucl. Med. 29, 233–239.

Castelnovo, V., Caminiti, S.P., Riva, N., Magnani, G., Silani, V., Perani, D., 2019. Heterogeneous brain FDG-PET metabolic patterns in patients with C9orf72 mutation. Neurol. Sci. 40, 515–521.

Chetelat, G., Arbizu, J., Barthel, H., Garibotto, V., Law, I., Morbelli, S., van de Giessen, E., Agosta, F., Barkhof, F., Brooks, D.J., Carrillo, M.C., Dubois, B., Fjell, A.M., Frisoni, G. B., Hansson, O., Herholz, K., Hutton, B.F., Jack Jr., C.R., Lammertsma, A.A., Landau, S.M., Minoshima, S., Nobili, F., Nordberg, A., Ossenkoppele, R., Oyen, W.J. G., Perani, D., Rabinovici, G.D., Scheltens, P., Villemagne, V.L., Zetterberg, H., Drzezga, A., 2020. Amyloid-PET and (18)F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias. Lancet Neurol. 19, 951–962.

Contador, J., Vargas-Martinez, A.M., Sanchez-Valle, R., Trapero-Bertran, M., Llado, A., 2023. Cost-effectiveness of Alzheimer's disease CSF biomarkers and amyloid-PET in early-onset cognitive impairment diagnosis. Eur. Arch. Psychiatry Clin. Neurosci. 273, 243–252.

Devenney, E., Bartley, L., Hoon, C., O'Callaghan, C., Kumfor, F., Hornberger, M., Kwok, J.B., Halliday, G.M., Kiernan, M.C., Piguet, O., Hodges, J.R., 2015. Progression in behavioral variant frontotemporal dementia: a longitudinal study. JAMa Neurol. 72, 1501–1509.

Diehl-Schmid, J., Licata, A., Goldhardt, O., Forstl, H., Yakushew, I., Otto, M., Anderl-Straub, S., Beer, A., Ludolph, A.C., Landwehrmeyer, G.B., Levin, J., Danek, A., Fliessbach, K., Spottke, A., Fassbender, K., Lyros, E., Prudlo, J., Krause, B.J., Volk, A., Edbauer, D., Schroeter, M.L., Drzezga, A., Kornhuber, J., Lauer, M., Group, F.T.S., Grimmer, T., 2019. FDG-PET underscores the key role of the thalamus in frontotemporal lobar degeneration caused by C9ORF72 mutations. Transl. Psychiatry 9, 54.

Dubois, B., Feldman, H.H., Jacova, C., Hampel, H., Molinuevo, J.L., Blennow, K., DeKosky, S.T., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S., Crutch, S., Engelborghs, S., Frisoni, G.B., Fox, N.C., Galasko, D., Habert, M.O., Jicha, G.A., Nordberg, A., Pasquier, F., Rabinovici, G., Robert, P., Rowe, C., Salloway, S., Sarazin, M., Epelbaum, S., de Souza, L.C., Vellas, B., Visser, P.J., Schneider, L., Stern, Y., Scheltens, P., Cummings, J.L., 2014. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol. 13, 614–629.

Dukart, J., Mueller, K., Barthel, H., Villringer, A., Sabri, O., Schroeter, M.L., Alzheimer's Disease Neuroimaging, I., 2013. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. Psychiatry Res. 212, 230–236.

Etminani, K., Soliman, A., Davidsson, A., Chang, J.R., Martinez-Sanchis, B., Byttner, S., Camacho, V., Bauckneht, M., Stegeran, R., Ressner, M., Agudelo-Cifuentes, M., Chincarini, A., Brendel, M., Rominger, A., Bruffaerts, R., Vandenberghe, R., Kramberger, M.G., Trost, M., Nicastro, N., Frisoni, G.B., Lemstra, A.W., van Berckel, B.N.M., Pilotto, A., Padovani, A., Morbelli, S., Aarsland, D., Nobili, F., Garibotto, V., Ochoa-Figueroa, M., 2022. A 3D deep learning model to predict the diagnosis of dementia with Lewy bodies, Alzheimer's disease, and mild cognitive impairment using brain 18F-FDG PET. Eur. J. Nucl. Med. Mol. ImAging 49, 563–584.

Forster, S., Buschert, V.C., Buchholz, H.G., Teipel, S.J., Friese, U., Zach, C., la Fougere, C., Rominger, A., Drzezga, A., Hampel, H., Bartenstein, P., Buerger, K., 2011. Effects of a 6-month cognitive intervention program on brain metabolism in amnestic mild cognitive impairment and mild Alzheimer's disease. J. Alzheimers Dis. 25, 695–706.

Garcia-Gutierrez, F., Diaz-Alvarez, J., Matias-Guiu, J.A., Pytel, V., Matias-Guiu, J., Cabrera-Martin, M.N., Ayala, J.L., 2022. GA-MADRID: design and validation of a machine learning tool for the diagnosis of Alzheimer's disease and frontotemporal dementia using genetic algorithms. Med. Biol. Eng. Comput. 60, 2737–2756.

Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S.F., Ogar, J.M., Rohrer, J.D., Black, S., Boeve, B.F., Manes, F., Dronkers, N.F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B.L., Knopman, D.S., Hodges, J.R., Mesulam, M.M., Grossman, M., 2011. Classification of primary progressive aphasia and its variants. Neurology 76, 1006–1014.

Guedj, E., Varrone, A., Boellaard, R., Albert, N.L., Barthel, H., van Berckel, B., Brendel, M., Cecchin, D., Ekmekcioglu, O., Garibotto, V., Lammertsma, A.A., Law, I., Penuelas, I., Semah, F., Traub-Weidinger, T., van de Giessen, E., Van Weehaeghe, D., Morbelli, S., 2022. EANM procedure guidelines for brain PET imaging using [(18)F] FDG, version 3. Eur. J. Nucl. Med. Mol. ImAging 49, 632–651.

Hao, S., Liu, C., Li, N., Wu, Y., Li, D., Gao, Q., Yuan, Z., Li, G., Li, H., Yang, J., Fan, S., 2022. Clinical evaluation of AI-assisted screening for diabetic retinopathy in rural areas of midwest China. PLoS One 17, e0275983.

Hu, J., Qing, Z., Liu, R., Zhang, X., Lv, P., Wang, M., Wang, Y., He, K., Gao, Y., Zhang, B., 2020. Deep learning-based classification and voxel-based visualization of frontotemporal dementia and Alzheimer's disease. Front. Neurosci. 14, 626154.

Jack Jr., C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 9, 119–128.

Jo, T., Nho, K., Saykin, A.J., 2019. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front. Aging Neurosci. 11, 220.

Li, X., Morgan, P., Ashburner, J., Smith, J., Rorden, C., 2016. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J. Neurosci. Methods 264.

McInnes, L., Healy, J., Melville, J., 2018. Umap: uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

McKeith, I.G., Boeve, B.F., Dickson, D.W., Halliday, G., Taylor, J.P., Weintraub, D., Aarsland, D., Galvin, J., Attems, J., Ballard, C.G., Bayston, A., Beach, T.G., Blanc, F., Bohnen, N., Bonanni, L., Bras, J., Brundin, P., Burn, D., Chen-Plotkin, A., Duda, J.E., El-Agnaf, O., Feldman, H., Ferman, T.J., Ffytche, D., Fujishiro, H., Galasko, D., Goldman, J.G., Gomperts, S.N., Graff-Radford, N.R., Honig, L.S., Iranzo, A., Kantarci, K., Kaufer, D., Kukull, W., Lee, V.M.Y., Leverenz, J.B., Lewis, S., Lippa, C., Lunde, A., Masellis, M., Masliah, E., McLean, P., Mollenhauer, B., Montine, T.J., Moreno, E., Mori, E., Murray, M., O'Brien, J.T., Orimo, S., Postuma, R.B., Ramaswamy, S., Ross, O.A., Salmon, D.P., Singleton, A., Taylor, A., Thomas, A., Tiraboschi, P., Toledo, J.B., Trojanowski, J.Q., Tsuang, D., Walker, Z., Yamada, M., Kosaka, K., 2017. Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB consortium. Neurology 89, 88–100.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack Jr., C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carrillo, M.C., Thies, B., Weintraub, S., Phelps, C.H., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National institute on aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 263–269.

McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F.J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C.J., King, D., Ledsam, J.R., Melnick, D., Mostofi, H., Peng, L., Reicher, J.J., Romera-Paredes, B., Sidebottom, R., Suleyman, M., Tse, D., Young, K.C., De Fauw, J., Shetty, S., 2020. International evaluation of an AI system for breast cancer screening. Nature 577, 89–94.

Minoshima, S., Cross, D., Thientunyakit, T., Foster, N.L., Drzezga, A., 2022. (18)F-FDG PET imaging in neurodegenerative dementing disorders: insights into subtype classification, emerging disease categories, and mixed dementia with copathologies. J. Nucl. Med. 63, 2S–12S.

Minoshima, S., Giordani, B., Berent, S., Frey, K.A., Foster, N.L., Kuhl, D.E., 1997. Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. Ann. Neurol. 42, 85–94.

Minoshima, S., Mosci, K., Cross, D., Thientunyakit, T., 2021. Brain [F-18]FDG PET for clinical dementia workup: differential diagnosis of Alzheimer's disease and other types of dementing disorders. Semin. Nucl. Med. 51, 230–240.

Nemoto, K., Sakaguchi, H., Kasai, W., Hotta, M., Kamei, R., Noguchi, T., Minamimoto, R., Arai, T., Asada, T., 2021. Differentiating dementia with Lewy bodies and Alzheimer's disease by deep learning to structural MRI. J. Neuroimaging 31, 579–587.

Nguyen, H.D., Clement, M., Planche, V., Mansencal, B., Coupe, P., 2023. Deep grading for MRI-based differential diagnosis of Alzheimer's disease and Frontotemporal dementia. Artif. Intell. Med. 144, 102636.

Nichols, E., Steinmetz, J.D., Vollset, S.E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T.T., 2022. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 7, e105–e125.

Olney, N.T., Spina, S., Miller, B.L., 2017. Frontotemporal dementia. Neurol. Clin. 35, 339–374.

Perini, G., Rodriguez-Vieitez, E., Kadir, A., Sala, A., Savitcheva, I., Nordberg, A., 2021. Clinical impact of (18)F-FDG-PET among memory clinic patients with uncertain diagnosis. Eur. J. Nucl. Med. Mol. ImAging 48, 612–622.

Perovnik, M., Vo, A., Nguyen, N., Jamsek, J., Rus, T., Tang, C.C., Trost, M., Eidelberg, D., 2022. Automated differential diagnosis of dementia syndromes using FDG PET and machine learning. Front. Aging Neurosci. 14, 1005731.

Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., van Swieten, J.C., Seelaar, H., Dopper, E.G., Onyike, C.U., Hillis, A.E., Josephs, K.A.,

Boeve, B.F., Kertesz, A., Seeley, W.W., Rankin, K.P., Johnson, J.K., Gorno-Tempini, M.L., Rosen, H., Prioleau-Latham, C.E., Lee, A., Kipps, C.M., Lillo, P., Piguet, O., Rohrer, J.D., Rossor, M.N., Warren, J.D., Fox, N.C., Galasko, D., Salmon, D.P., Black, S.E., Mesulam, M., Weintraub, S., Dickerson, B.C., Diehl-Schmid, J., Pasquier, F., Deramecourt, V., Lebert, F., Pijnenburg, Y., Chow, T.W., Manes, F., Grafman, J., Cappa, S.F., Freedman, M., Grossman, M., Miller, B.L., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 134, 2456–2477.

Sadeghi, N., Foster, N.L., Wang, A.Y., Minoshima, S., Lieberman, A.P., Tasdizen, T., 2008. Automatic classification of Alzheimer's disease vs. Frontotemporal dementia: a spatial decision tree approach with FDG-PET. In: Proceedings of the 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, pp. 408–411.

Scheltens, N.M.E., van der Weijden, K., Adriaanse, S.M., van Assema, D., Oomen, P.P., Krudop, W.A., Lammertsma, A.A., Barkhof, F., Koene, T., Teunissen, C.E., Scheltens, P., van der Flier, W.M., Pijnenburg, Y.A.L., Yaqub, M., Ossenkoppele, R., van Berckel, B.N.M., 2018. Hypometabolism of the posterior cingulate cortex is not restricted to Alzheimer's disease. Neuroimage Clin. 19, 625–632.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Teunissen, C.E., Verberk, I.M.W., Thijssen, E.H., Vermunt, L., Hansson, O., Zetterberg, H., van der Flier, W.M., Mielke, M.M., Del Campo, M., 2022. Blood-based biomarkers for Alzheimer's disease: towards clinical implementation. Lancet Neurol. 21, 66–77.

Villain, N., Dubois, B., 2019. Alzheimer's disease including focal presentations. Semin. Neurol. 39, 213–226.

Xia, Y., Lu, S., Wen, L., Eberl, S., Fulham, M., Feng, D.D., 2014. Automated identification of dementia using FDG-PET imaging. Biomed Res. Int. 2014, 421743.

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31, 1116–1128.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 818–833.