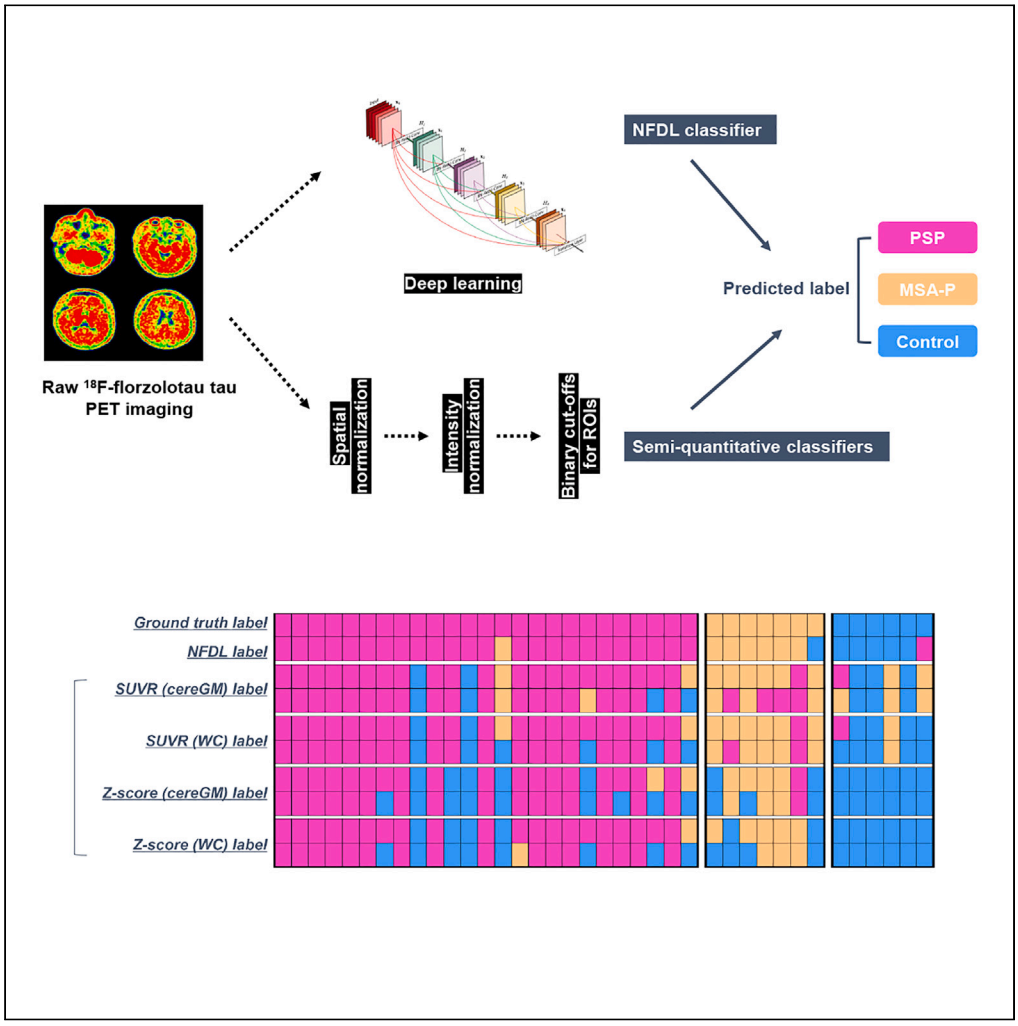
iScience

CellPress
OPEN ACCESS

Article

Improved interpretation of ^{18}F-florzolotau PET in progressive supranuclear palsy using a normalization-free deep-learning classifier



Jiaying Lu, Christoph Clement, Jimin Hong, ..., Qianhua Zhao, Kuangyu Shi, for the Progressive Supranuclear Palsy Neuroimage Initiative (PSPNI)

liufengtao@fudan.edu.cn (F.L.)
zuochuantao@fudan.edu.cn (C.Z.)

Highlights

A normalization-free deep-learning (NFDL) model was designed for ^{18}F-florzolotau PET

NFDL is promising for early and accurate differentiation of atypical parkinsonism

Regions driving the model's decision were aligned with disease-specific topographies

The high generalizability makes NFDL a potential tool for tau PET interpretation

Lu et al., iScience 26, 107426
August 18, 2023
https://doi.org/10.1016/j.isci.2023.107426

iScience

CellPress
OPEN ACCESS

Article

Improved interpretation of ^{18}F-florzolotau PET in progressive supranuclear palsy using a normalization-free deep-learning classifier

Jiaying Lu,[1,2,10] Christoph Clement,[2,10] Jimin Hong,[2] Min Wang,[3,4] Xinyi Li,[5] Lara Cavinato,[2,6] Tzu-Chen Yen,[7] Fangyang Jiao,[1] Ping Wu,[1] Jianjun Wu,[5] Jingjie Ge,[1] Yimin Sun,[5] Matthias Brendel,[8] Leonor Lopes,[2] Axel Rominger,[2] Jian Wang,[5] Fengtao Liu,[5,*] Chuantao Zuo,[1,9,11,*] Yihui Guan,[1] Qianhua Zhao,[5] Kuangyu Shi,[2,4] and for the Progressive Supranuclear Palsy Neuroimage Initiative (PSPNI)

SUMMARY

While ^{18}F-florzolotau tau PET is an emerging biomarker for progressive supranuclear palsy (PSP), its interpretation has been hindered by a lack of consensus on visual reading and potential biases in conventional semi-quantitative analysis. As clinical manifestations and regions of elevated ^{18}F-florzolotau binding are highly overlapping in PSP and the Parkinsonian type of multiple system atrophy (MSA-P), developing a reliable discriminative classifier for ^{18}F-florzolotau PET is urgently needed. Herein, we developed a normalization-free deep-learning (NFDL) model for ^{18}F-florzolotau PET, which achieved significantly higher accuracy for both PSP and MSA-P compared to semi-quantitative classifiers. Regions driving the NFDL classifier's decision were consistent with disease-specific topographies. NFDL-guided radiomic features correlated with clinical severity of PSP. This suggests that the NFDL model has the potential for early and accurate differentiation of atypical parkinsonism and that it can be applied in various scenarios due to not requiring subjective interpretation, MR-dependent, and reference-based preprocessing.

INTRODUCTION

Interpreting tau positron emission tomography (PET) imaging remains challenging in clinical and research settings. Visual assessment and semi-quantification analysis are the two main methods applied in clinical and research scenarios, respectively. The former was mainly restricted by the absence of widely accepted guidelines, and currently, methods for visual assessment are only available for ^{18}F-flortaucipir, ^{18}F-MK-6240, and ^{18}F-florzolotau (also known as ^{18}F-PM-PBB3 or ^{18}F-APN-1607) tau PET imaging in Alzheimer's disease (AD),[1–3] a secondary tauopathy. The disease heterogeneity reflected by the pathological topographies[4–8] also increases the difficulty of visual reading. Additionally, the significant interobserver variability, a common problem in subjective assessment, requires caution. On the other hand, semi-quantification analysis is mainly used in research settings as an alternative to visual reading. This involves calculating tracer uptake in selected regions of interest (ROIs) and dividing it by the signal in the reference region to obtain semi-quantitative parameters (i.e., standardized uptake value ratios, SUVRs),[9] which serve as relatively objective bases for tau PET-imaging interpretation. However, this methodology is limited by the need for spatial normalization,[10] potential inconsistencies in ROI selection, the involvement of the same ROIs in different diseases, the lack of widely accepted cut-offs for PET parameters,[11] and uncertainties in establishing the most suitable reference areas.[12] It is worth noting that selecting an optimal reference tissue is especially challenging in the most common primary tauopathy, progressive supranuclear palsy (PSP).[13] This is due to the diffused tau depositions with disease progression,[4] which have been reported in commonly used reference regions in PSP, such as the cerebellum and subcortical white matter.[4] Additionally, since initial tau accumulation is often seen in small anatomical areas (i.e., pallido-nigro-luysian axis in PSP[4] or locus coeruleus and entorhinal cortex in AD[14]), PET-based semi-quantitative assessment of regional tau burden is prone to noise and artifacts during spatial alignment of brain structures, which reduces the early diagnostic value of tau PET imaging. Although individual structural magnetic resonance (MR) image-based spatial normalization is considered the best standard for high-precision spatial

[1]Department of Nuclear Medicine & PET Center & National Center for Neurological Disorders & National Clinical Research Center for Aging and Medicine, Huashan Hospital, Fudan University, Shanghai 200235, China

[2]Department of Nuclear Medicine, Inselspital, Bern University Hospital, University of Bern, 3010 Bern, Switzerland

[3]Institute of Biomedical Engineering, School of Life Science, Shanghai University, Shanghai 200444, China

[4]Department of Informatics, Technical University of Munich, 80333 Munich, Germany

[5]Department of Neurology & National Center for Neurological Disorders & National Clinical Research Center for Aging and Medicine, Huashan Hospital, Fudan University, Shanghai 200400, China

[6]MOX - Modeling and Scientific Computing, Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy

[7]APRINOIA Therapeutics Co., Ltd, Suzhou 215122, China

[8]Department of Nuclear Medicine, University of Munich, 80539 Munich, Germany

[9]Human Phenome Institute, Fudan University, Shanghai 200433, China

[10]These authors contributed equally

[11]Lead contact

Continued

warping,[10] it is important to note that high-resolution MR images can be costly and time-consuming to obtain and cannot be performed in all patients (i.e., due to implanted electronic devices).

Recent advances in artificial intelligence, particularly in the field of deep learning (DL), have made it possible to detect complex patterns in large-imaging datasets. While a growing number of DL-based clinical applications have been proposed for PET imaging,[15] their use in tau PET imaging has been limited to ^{18}F-flortaucipir and ^{18}F-MK-6240 for AD.[16–18] To the best of our knowledge, no studies have investigated the performance of DL algorithms using raw PET images without spatial and intensity normalization as input. In the context of PSP, where elevated tracer uptake mainly occurs in small, deep brain structures such as the subcortical nucleus, Endo et al. recently proposed optimized semi-quantitative parameters for ^{18}F-florzolotau tau PET. Their approach utilized machine-learning techniques based on traditional imaging preprocessing. However, the discriminative performance of this novel method has only been examined in comparisons between PSP and healthy controls/AD.[19] In contrast to AD, where regions with increased tracer uptake are usually diffused (i.e., inferior lateral temporal, posterior cingulate, and lateral parietal regions),[20] disease hallmark regions with abnormal tracer uptake in PSP are pretty small (i.e., globus pallidus and subthalamic nucleus).[21,22] It is worth noting that such tiny and deep regions are more susceptible to inaccurate spatial normalization, making it more challenging for assessors, conventional semi-quantitative indices, and DL models to understand the disease-specific pattern in PSP. Consequently, the potential for DL-based algorithms to effectively differentiate PSP from subjects exhibiting similar patterns of increased tau tracer uptake remains to be explored.

Pathologically, PSP is the most common primary tauopathy (4-repeat form of tau),[13] and is clinically classified as an atypical parkinsonism syndrome that shares plenty of symptoms with Parkinson's disease (PD) as well as other atypical parkinsonism syndromes like the parkinsonian subtype of multiple system atrophy (MSA-P), especially at early stages.[23,24] The significant phenotypic overlap and clinical heterogeneity make an early and accurate diagnosis of these conditions challenging. New-generation tau ligands with high affinity to both 3-repeat and 4-repeat forms of tau show promise in ameliorating this situation.[21,22] First, pilot observation with different new-generation tau radiotracers, including ^{18}F-PI-2620[22] and ^{18}F-florzolotau[21,25,26] have consistently found abnormal signals in the pallido-nigro-luysian axis in patients with PSP. Second, the presence of abnormal tau tracer retention is not limited to tauopathies but may also occur in MSA-P.[27–30] Notably, in MSA-P, increased tracer uptake is predominantly located in the basal ganglia, close to or partially overlapping the core tau deposition areas in PSP (Figure 1). Therefore, differential diagnosis of PSP and MSA-P using tau PET imaging is still difficult for both visual reading and semi-quantitative analysis.

To this end, to improve the clinical utility of ^{18}F-florzolotau tau PET imaging in the classification of PSP and MSA-P, we trained a normalization-free DL (NFDL) classifier on a set of tau PET images from patients with PSP, MSA-P, and controls without incorporating potential confounders such as spatial normalization, reference region, ROIs, or binary cut-offs. We included the traditional semi-quantitative parameter SUVR and its Z-transferred score using different reference tissues (cerebellar grey matter (cereGM), whole cerebellum (WC), subject-specific white matter (subWM), and subject-specific grey matter (subGM))[21,22,26,31–33] as a comparison to evaluate the performance of the NFDL classifier. We investigated the decision mechanism of the NFDL classifier using an occlusion sensitivity experiment and used NFDL-guided radiomic features (first-order radiomic features) from the weighted tau PET imaging to assess the model's ability to reflect disease severity.

RESULTS

Demographics

As shown in Table 1, 228 participants (mean age, 63.9years ±7.5 [SD]; 107 women) and 35 subjects (mean age, 65.3years ±8.2 [SD]; 22 women) in Cohort I and II were enrolled, respectively. No significant differences were observed in demographic and clinical variables across folds in Cohort I (Table S1).

The optimal NFDL classifier

The highest mean validation F1 score of 0.948 was achieved by applying intensity scaling and a combination of three random data augmentations (random Gaussian smoothing, random coarse dropout regularization,[34] and random elastic transformations) to the input images, the 3D DenseNet121 network,[35] cross-entropy loss, and Adam optimizer.[36] This classifier was selected as the most appropriate for further analysis (Figure 2A).

*Correspondence:
liufengtao@fudan.edu.cn
(F.L.),
zuochuantao@fudan.edu.cn
(C.Z.)

https://doi.org/10.1016/j.isci.2023.107426

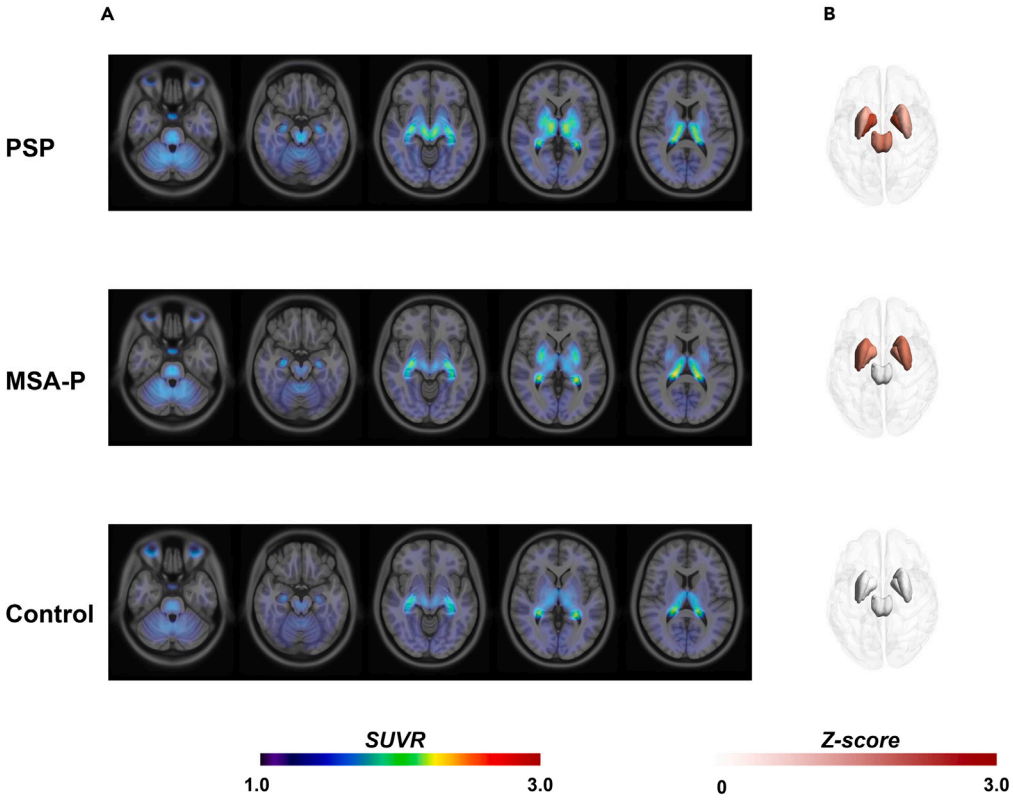


Figure 1. Characteristic ^{18}F-florzolotau-binding patterns in different clinical entities

(A and B) (A) Mean SUVR maps (reference region: cerebellar grey matter), and (B) mean heatmaps (*Z* score) based on our previously reported data.[21,27] Abbreviations: PSP, progressive supranuclear palsy; MSA-P, multiple system atrophy, parkinsonian type; SUVR, standardized uptake value ratio.

The semi-quantitative classifiers

For SUVR-based classifiers, the optimal cut-off values that defined positivity based on regional SUVR values were determined by maximizing Youden's statistic in the training/validation sets, as shown in Tables 2 and S2. For Z-score-based classifiers, the regional positivity cut-off was 2, as described in the STAR Methods section (Figure 2B).

Classification performance in test sets

For the Test$_1$ set (Figure 3A), the accuracies for all labels using the NFDL, SUVR, and *Z* score classifiers were 0.92, 0.79–0.87, and 0.50–0.95, respectively. That is, given the availability of MR images for traditional semi-quantitative metrics, the NFDL classifier outperformed all semi-quantitative classifiers except the *Z* score using subWM as the reference region for intensity normalization (accuracy = 0.95). Similar findings were found when assessing the classification performances for each label separately (Table 3).

For the additional Test$_2$ set (MR images were absent; Figure 3C), the NFDL model consistently outperformed all semi-quantitative classifiers. The accuracies for all labels using the NFDL, SUVR, and *Z* score classifiers were 0.86, 0.60–0.66, and 0.43–0.49, respectively. The details for the classification performances of each label are presented in Table S3.

Consistency with the application of different spatial normalization approaches for the conventional semi-quantitative analysis

At the group level in the training/validation sets, the consistency between the SUVRs varied among ROIs as well as reference regions (intraclass correlation coefficients (ICC) = 0.75–0.93), as the Bland-Altman plot (Figure 4A) shows. The one-sample t-test for the mean difference of SUVRs found significant differences in all ROIs ($p < 0.03$) except the subthalamic nucleus with WC as the reference region ($p = 0.055$).

Table 1. General characteristics of the study participants

Clinical diagnosis	Number of cases	Age, years	Proportion of women	MMSE score	PSPrs score	MDS-UPDRS-III score	H&Y scale	Disease duration, years	Proportion of PSP-RS[a]	Level of diagnostic certainty (probable/possible/suggestive)[a]
Cohort I split (training/validation sets)										
PSP	123	65.5 (6.8)	0.46	23.3 (5.5)	30.5 (13.5)	41.2 (15.3)	3.3 (0.8)	3.9 (2.7)	0.54	101/17/5
MSA-P	37	61.3 (6.7)	0.54	24.6 (4.5)	–	54.3 (11.2)	3.5 (0.8)	2.4 (1.3)	–	–
Control	30	59.1 (9.2)	0.63	28.1 (1.5)	–	–	–	–	–	–
Entire set	190	63.7 (7.6)	0.51	24.3 (5.2)	–	–	–	–	–	–
Cohort I split ($Test_0$ & $Test_1$ sets)										
PSP	25	66.6 (6.6)	0.32	23.4 (5.7)	33.1 (15.7)	41.4 (13.3)	3.4 (0.8)	3.4 (2.2)	0.68	23/2/0
MSA-P	7	63.4 (5.7)	0.29	26.0 (2.3)	–	45.4 (10.6)	2.8 (0.8)	2.8 (1.2)	–	–
Control	6	61.7 (6.1)	0.17	27.6 (0.5)	–	–	–	–	–	–
Entire set	38	65.3 (6.5)	0.29	24.4 (5.1)	–	–	–	–	–	–
Cohort II ($Test_2$ set)										
PSP	18	66.8 (8.2)	0.50	21.4 (7.9)	35.0 (20.9)	42.4 (22.5)	3.1 (1.0)	2.6 (1.9)	0.67	15/2/1
MSA-P	8	64.5 (5.7)	0.75	–	–	–	–	2.3 (1.8)	–	–
Control	9	63.1 (10.2)	0.78	28.3 (1.5)	–	–	–	–	–	–
Entire set	35	65.3 (8.2)	0.63	–	–	–	–	–	–	–

Data are expressed as mean (standardized deviation) unless otherwise stated.

Abbreviations: PSP, progressive supranuclear palsy; PSP-RS, PSP with Richardson's syndrome; MSA-P, multiple system atrophy, parkinsonian type; MMSE, Mini-Mental State Examination; PSPrs, progressive supranuclear palsy rating scale; MDS-UPDRS-III, Movement Disorders Society Unified Parkinson's Disease Rating Scale, part III; H&Y, Hoehn and Yahr.

[a]Data applicable to PSP only.

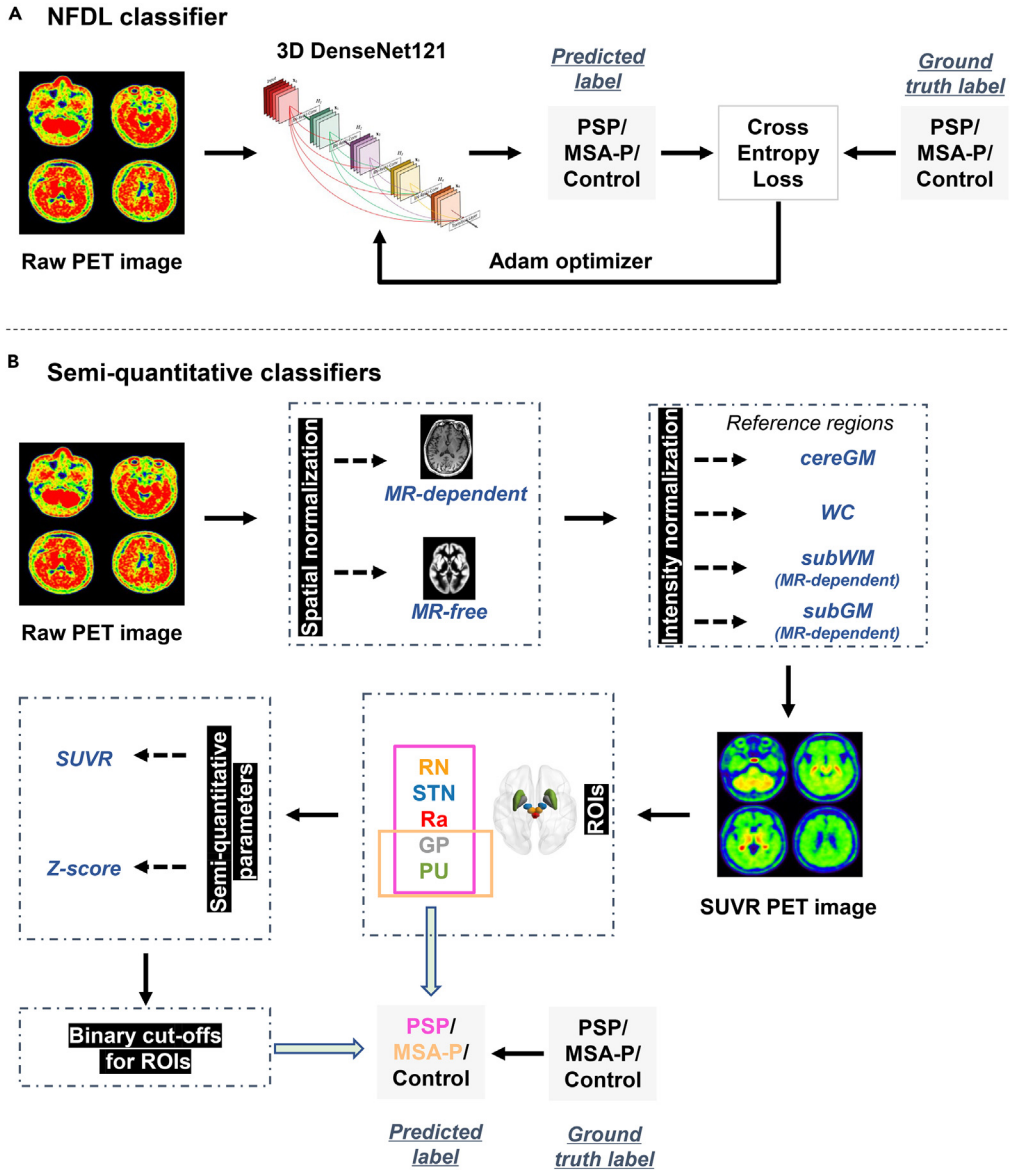


Figure 2. Design of different classifiers

(A) The design of the normalization-free deep-learning classifier.

(B) The design of the traditional semi-quantitative classifiers. Abbreviations: NFDL, normalization deep learning; PSP, progressive supranuclear palsy; MSA-P, multiple system atrophy, parkinsonian type; cereGM, cerebellar gray matter; WC, whole cerebellum; subWM, subject-specific white matter; subGM, subject-specific gray matter; SUVR, standardized uptake value ratio; ROIs, regions of interest; RN, red nucleus; STN, subthalamic nucleus; Ra, raphe nuclei; GP, globus pallidus; PU, putamen.

At the individual level in test sets (head-to-head comparisons of PET images that had undergone MR-dependent ($Test_1$) versus MR-free ($Test_{1add}$) spatial normalization; Figure 4B), a moderate agreement for diagnostic labels was found (Cohen's kappa coefficient = 0.67–0.78).[37] Compared to those in $Test_1$, the classification performances of semi-quantitative parameters consistently deteriorated in the $Test_{1add}$ set (Figure 3B and Table S4). Specifically, the accuracies for all labels using the SUVR and *Z* score classifiers in $Test_{1add}$ were 0.58–0.76 and 0.66, respectively. The only exceptions were SUVR (WC) for MSA-P (accuracy = 0.90 for $Test_1$, and 0.92 for $Test_{1add}$), *Z* score (cereGM) for MSA-P (accuracy = 0.87 for $Test_1$, and 0.90 for $Test_{1add}$), and *Z* score (WC) for MSA-P (accuracy = 0.84 for $Test_1$, and 0.87 for $Test_{1add}$). In general, the MR-dependent semi-quantitative parameters in $Test_1$ outperformed MR-free ones.

Table 2. Cut-off SUVRs for ^{18}F-florzolotau positivity at regional level using the MR-dependent spatial normalization method

ROIs	Optimal cut-off point	Area under curve	Sensitivity (%)	Specificity (%)
SUVR (cerebellar gray matter)				
Distinguishing patients with PSP from others				
Red nucleus	1.52	0.85	69.9	88.1
Subthalamic nucleus	1.58	0.89	73.2	94.0
Raphe nuclei	1.51	0.85	67.5	94.0
Distinguishing patients with PSP and MSA-P from controls				
Globus pallidus	1.40	0.92	93.3	79.4
Putamen	1.16	0.77	70.0	78.7
SUVR (whole cerebellum)				
Distinguishing patients with PSP from others				
Red nucleus	1.43	0.86	69.9	95.5
Subthalamic nucleus	1.44	0.91	79.7	94.0
Raphe nuclei	1.33	0.85	66.7	97.0
Distinguishing patients with PSP and MSA-P from controls				
Globus pallidus	1.29	0.93	90.0	85.0
Putamen	1.09	0.78	70.0	79.4
SUVR (subject-specific white matter)				
Distinguishing patients with PSP from others				
Red nucleus	1.42	0.87	76.4	85.1
Subthalamic nucleus	1.45	0.92	82.1	94.0
Raphe nuclei	1.40	0.88	75.6	89.6
Distinguishing patients with PSP and MSA-P from controls				
Globus pallidus	1.28	0.96	100	90.0
Putamen	1.10	0.91	83.3	91.2
SUVR (subject-specific gray matter)				
Distinguishing patients with PSP from others				
Red nucleus	1.43	0.81	71.5	85.1
Subthalamic nucleus	1.46	0.87	82.1	91.0
Raphe nuclei	1.40	0.82	73.2	88.1
Distinguishing patients with PSP and MSA-P from controls				
Globus pallidus	1.33	0.85	90.0	83.1
Putamen	1.14	0.73	63.3	78.7

The optimal cut-off values for defining ^{18}F-florzolotau positivity based on regional SUVR values were those that maximized Youden's statistic in the training/validation set. Abbreviations: PSP, progressive supranuclear palsy; MSA-P, multiple system atrophy, parkinsonian type; ROIs, regions of interest.

Interpretation of the NFDL model

The occlusion-sensitivity experiments revealed that the midbrain, thalamus, and globus pallidus were the most discriminating regions for PSP, while the putamen and globus pallidus were the most discriminating regions for MSA-P (Figure 5). For controls, subcortical regions, including midbrain, thalamus, globus pallidus, and putamen, provided the optimal discrimination ability for excluding positivity on tau PET images.

The correlation analysis found significant associations between NFDL-guided radiomic features extracted from the individually weighted tau PET imaging and scores of clinical disease severity in PSP (i.e., quantile 25%: PSP rating scale (PSPrs), $p = 0.010$; MDS Unified Parkinson's Disease Rating Scale-part III (MDS-UPDRS-III), $p = 0.041$; Mini-Mental State Examination (MMSE), $p = 0.028$; Table 4). Additionally, significant associations were identified between NFDL-guided radiomic features and SUVRs (i.e., median: SUVRs (cereGM) across all investigated ROIs, $p = 0.001$–0.016; Tables S5 and S6).

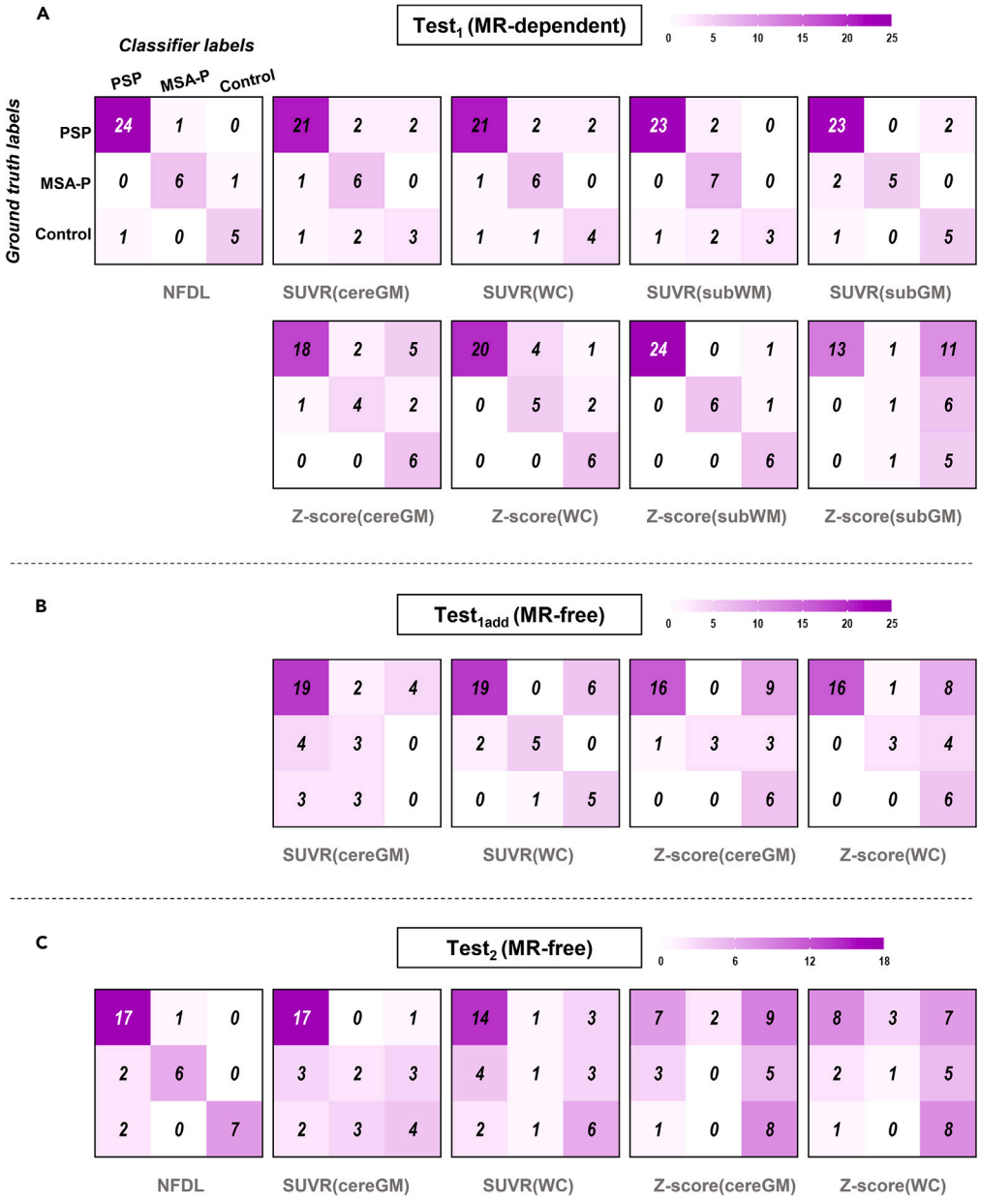


Figure 3. Classification performance in Test sets

Each classifier's name is displayed in gray text, with the numbers inside the confusion matrices representing the count of subjects. The highest number of subjects within a diagnostic group for each dataset is utilized to establish the color scale's maximum value.

(A) Confusion matrices in $test_1$ set: spatial normalization of all PET images was carried out based on concurrent MR images, following the MR-dependent approach.

(B) Confusion matrices in $test_{1add}$ set: this includes the same subjects as the $test_1$ set, but the PET images were spatially normalized using the unified algorithm from Statistical Parametric Mapping 12, following the MR-free approach.

(C) Confusion matrices in $test_2$ set: this set contained different subjects compared to $test_1$ set, and no concurrent MR images were available. Consequently, PET images were spatially normalized using the MR-free approach. Abbreviations: PSP, progressive supranuclear palsy; MSA-P, multiple system atrophy, parkinsonian type; NFDL, normalization-free deep learning; SUVR, standardized uptake value ratio; cereGM, cerebellar gray matter; WC, whole cerebellum; subWM, subject-specific white matter; subGM, subject-specific gray matter.

Table 3. Classification performance of different classifiers for each label in $Test_1$ set

Classifier	Ground truth label	Area under curve	Accuracy (%)	Sensitivity (%)	Specificity (%)
NFDL	PSP	0.95	94.7	**96.0**	92.3
	MSA-P	0.97	94.7	85.7	96.8
	Control	0.88	**94.7**	83.3	**96.9**
SUVR (cereGM)	PSP	0.83	84.2	84.0	84.6
	MSA-P	0.85	86.8	85.7	87.1
	Control	0.81	86.8	50.0	93.8
SUVR (WC)	PSP	0.83	84.2	84.0	84.6
	MSA-P	0.87	89.5	85.7	90.3
	Control	0.84	89.5	66.7	93.8
SUVR (subWM)	PSP	0.93	92.1	92.0	92.3
	MSA-P	0.94	89.5	**100**	87.1
	Control	0.85	92.1	50.0	**100**
SUVR (subGM)	PSP	0.83	86.8	92.0	76.9
	MSA-P	0.66	94.7	71.4	**100**
	Control	0.87	92.1	83.3	93.8
Z score (cereGM)	PSP	0.82	78.9	72.0	92.3
	MSA-P	0.81	86.8	57.1	93.5
	Control	0.89	81.6	**100**	78.1
Z score (WC)	PSP	0.88	86.8	80.0	**100**
	MSA-P	0.93	84.2	71.4	87.1
	Control	0.91	92.1	**100**	90.6
Z score (subWM)	PSP	**0.97**	**97.4**	**96.0**	**100**
	MSA-P	**0.98**	**97.4**	85.7	**100**
	Control	**0.97**	**94.7**	**100**	93.8
Z score (subGM)	PSP	0.74	68.4	52.0	**100**
	MSA-P	0.69	79.0	14.3	93.6
	Control	0.70	52.6	83.3	46.9

The highest diagnostic performance parameters for the same ground truth label are marked in bold.
Abbreviations: PSP, progressive supranuclear palsy; MSA-P, multiple system atrophy, parkinsonian type; NFDL, normalization-free deep learning; SUVR, standardized uptake value ratio; cereGM, cerebellar gray matter; WC, whole cerebellum; subWM, subject-specific white matter; subGM, subject-specific gray matter.

DISCUSSION

We developed an NFDL classifier for ^{18}F-florzolotau PET images that does not require spatial normalization or reference region-based intensity normalization and is capable of discriminating between PSP, MSA-P, and control. The regions with the highest discriminative power for diagnostic purposes were consistent with disease-specific tau topography[4] and/or previous pilot observations.[21,27] Since our model does not rely on spatial normalization, it has the potential to be used in extended clinical scenarios where structural MR images are unavailable. Albeit preliminary, the significant associations between NFDL-guided radiomic features and scores of clinical severities in PSP indicate its potential in disease severity evaluation.

Spatial normalization[10] and reference-region-based intensity normalization[12,38] are two essential preprocessing steps in the semi-quantitative analysis of brain PET images. While the interpretation of ^{18}F-fluorodeoxyglucose PET scans does not necessarily require the corresponding MR images during spatial normalization, this approach is not ideal for amyloid and tau PET imaging. To investigate this issue further, we compared standard MR-dependent versus MR-free spatial normalizations at both group and individual levels. The presence of non-negligible variations in several study participants, along with moderate classification agreement, suggested that MR-free spatial normalization should not be recommended for ^{18}F-florzolotau PET. With the development of tracer-specific templates for amyloid (^{18}F-florbetapir and

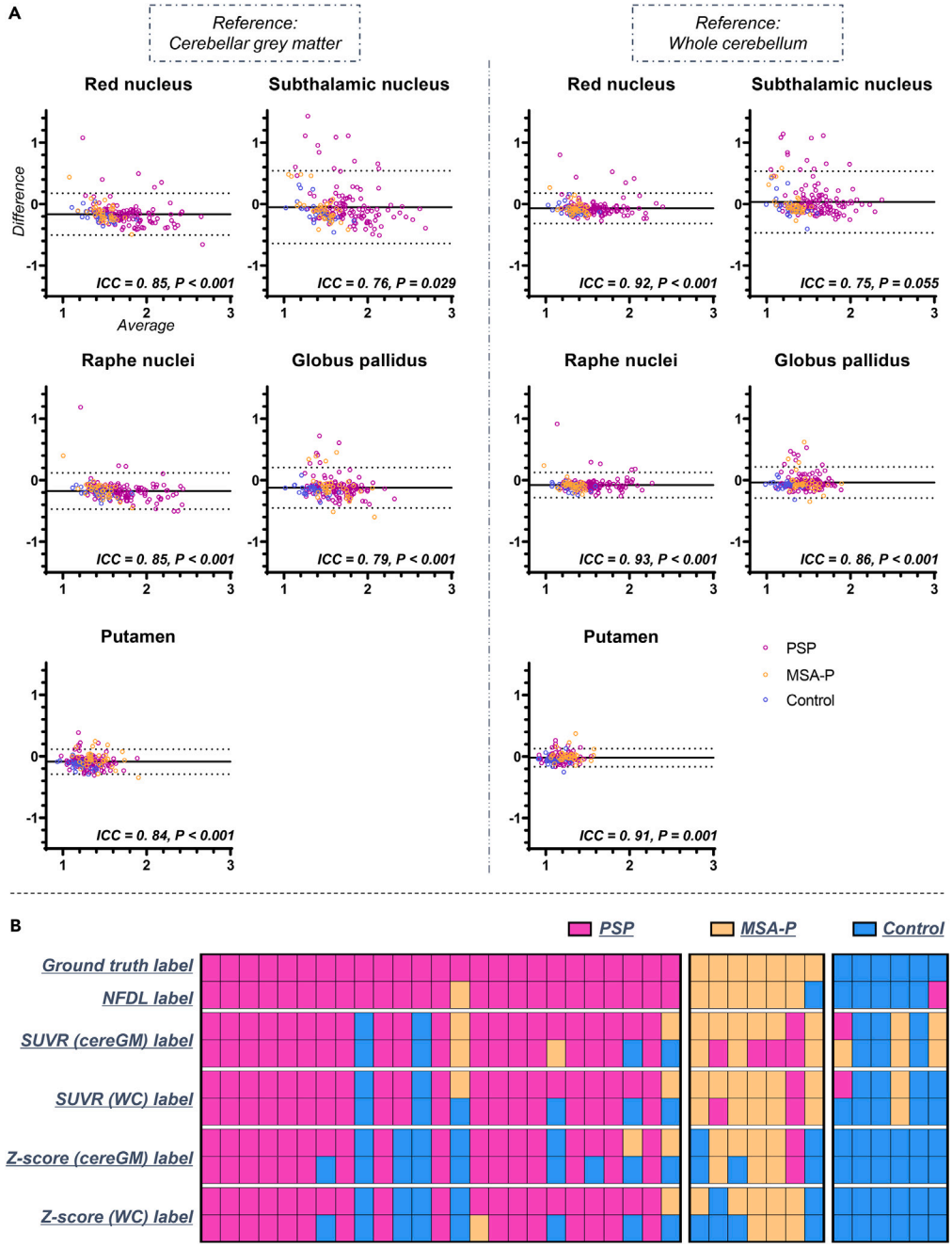


Figure 4. Consistency with the application of different spatial normalization approaches for the conventional semi-quantitative analysis

(A) Consistency at the group level (training/validation sets) evaluated through Bland-Altman analysis. The x axis of the plot displays the average SUVR of the MR-dependent and MR-free spatial normalization methods, and the y axis displays the difference in SUVRs between them (MR-free minus MR-dependent). Dots indicate individual SUVR. The solid line represents the average difference in SUVR between them, while the two dashed lines represent the 95% confidence interval limits for the average difference. The intraclass correlation coefficient (ICC from the Bland-Altman analysis is represented as "ICC", and "P" stands for the outcome of a one-sample t-test evaluating the mean difference in SUVRs. (B) Consistency at the individual level assessed through predicted labels. Each semi-quantitative classifier displays two rows of predicted labels: the first row represents the labels from the test1 set (MR-dependent), and the second row corresponds to the labels from the test1add set (MR-free).Abbreviations: NFDL, normalization-free deep learning; PSP, progressive supranuclear palsy; MSA-P, multiple system atrophy, parkinsonian type; SUVR, standardized uptake value ratio; cereGM, cerebellar gray matter; WC, whole cerebellum; subWM, subject-specific white matter.

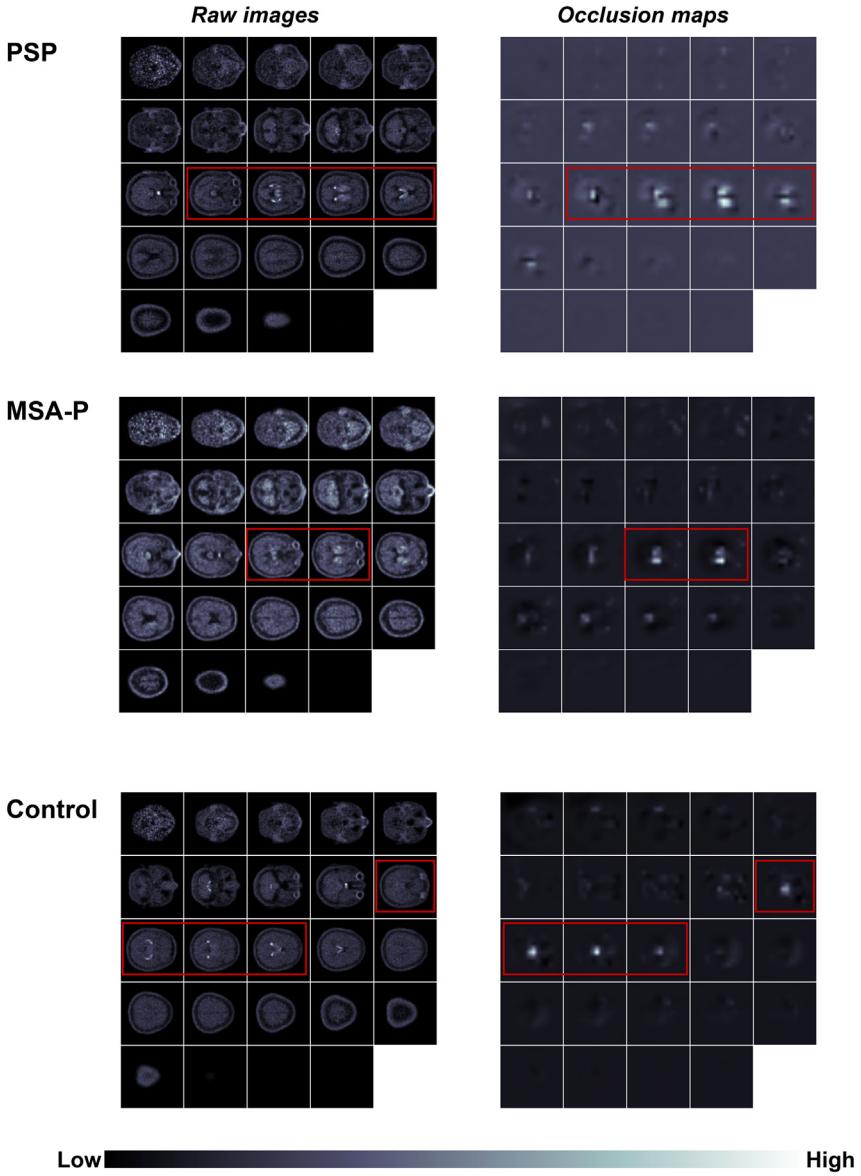


Figure 5. Representative occlusion sensitivity maps for all labels

The color scale indicates the impact of the region in terms of classification.Abbreviations: PSP, progressive supranuclear palsy. MSA-P, multiple system atrophy, parkinsonian type.

^{18}F-florbetaben) and tau (^{18}F-flortaucipir and ^{18}F-florzolotau) PET imaging, two recent studies in AD demonstrated high concordance between semi-quantitative results obtained with the MR-free strategy and those of MR-dependent normalization.[39,40] However, it remains unclear whether this strategy is applicable to the analysis of ^{18}F-florzolotau PET imaging in PSP. The areas of tau accumulation in PSP are smaller than in AD, which makes it more susceptible to inaccurate spatial normalization when measuring semi-quantitative parameters of tau deposition in PSP. Unlike traditional semi-quantitative approaches and previous DL models in this field, our NFDL classifier achieved optimal classification performance without requiring spatial normalization. Therefore, our proposed method has the potential for wider application in practice, independent of structural MR images.

Reference-region-based intensity normalization is one of the pivotal confounding factors in the semi-quantitative analysis of tau PET images and is particularly challenging in PSP due to the diffused tau distribution during progression.[4] Inappropriate reference areas can lead to underestimation of tau deposition. Our initial findings

Table 4. Correlations between NFDL-guided radiomic features derived from individually weighted tau PET images and clinical severity scores in PSP patients

		Median	Mean	Standard deviation	Variance	Minimum	Maximum	Skew	Kurtosis	Range	Quantile_25	Quantile_75
MMSE	R	−0.111	−0.046	−0.027	−0.036	−0.008	−0.066	−0.068	−0.092	−0.066	−0.204*	−0.012
	P	0.236	0.622	0.774	0.701	0.935	0.481	0.471	0.326	0.480	0.028	0.897
MDS UPDRS-III	R	0.095	0.138	0.122	0.135	−0.054	0.137	0.026	0.027	0.139	0.187*	0.124
	P	0.305	0.135	0.187	0.144	0.561	0.137	0.778	0.775	0.133	0.041	0.179
H&Y scale	R	0.045	0.069	0.073	0.112	−0.09	0.101	0.107	0.102	0.103	0.103	0.043
	P	0.628	0.457	0.428	0.222	0.326	0.273	0.245	0.269	0.263	0.262	0.644
PSPrs	R	0.157	0.042	0.002	−0.003	−0.035	0.066	−0.037	0.021	0.067	0.236*	0.013
	P	0.089	0.649	0.986	0.971	0.71	0.479	0.691	0.818	0.473	0.010	0.892

The correlation between variables was evaluated using Pearson's correlation coefficient. Statistically significant p values and corresponding correlation coefficients (R) are denoted with an asterisk (*, $p < 0.05$).
Abbreviations: MMSE, Mini-Mental State Examination; MDS-UPDRS-III, Movement Disorders Society Unified Parkinson's Disease Rating Scale, part III; PSPrs, progressive supranuclear palsy rating scale; H&Y, Hoehn and Yahr.

indicate that the subject-specific reference region subWM surpassed atlas-based methods in terms of differentiation. This observation aligns with earlier research on AD, which showed that SUVRs derived from parametric estimation of reference signal intensity (PERSI) exhibit superior discriminative capabilities.[31,32,41] It is crucial to highlight that the fundamental rationale behind the efficacy of PERSI in PSP and MSA-P varies from its effectiveness in AD. In PSP and MSA-P, the primary advantage stems from the removal of potential specific binding events in the cerebellum. Conversely, in AD, the benefit arises from the enhanced stability of quantification values attributed to the more expansive PERSI reference voxels in comparison to those in the cerebellum. However, the subject-specific reference region subGM in the current study exhibited unsatisfactory performance comparing to the subject-specific reference region subWM. This discrepancy may be attributed to the progression of tau deposition in gray matter regions, such as the frontal lobe in PSP. Therefore, the utilization of subGM as a reference region based on PERSI in PSP for interpreting tau PET imaging warrants further investigation. Notably, the classification performances of all reference region-based semi-quantitative indices were inferior to the reference-free NFDL model. The only exception was *Z* score (subWM) which was slightly superior to the NFDL model. However, the generalizability of this approach in practice is relatively limited due to the reliance on the individual MR image.

Another problem is the use of binary cut-offs for defining positive regions. While dichotomization is relevant for patient-level diagnosis, values below the cut-off do not necessarily imply the absence of pathological tau accumulation. Previous studies in PSP have shown that the use of different cut-offs and inconsistent criteria for classifying regional tau positivity can impact accuracy results.[21,22] In this study, the NFDL model demonstrated a lower false-negative rate (i.e., 4.0% and 5.6% for PSP, 14.3% and 25.0% for MSA-P in $Test_1$ and $Test_2$) than conventional methods without MR images (i.e., 24.0–36.0% and 5.6–61.1% for PSP, 28.6–57.1% and 75.0–100% MSA-P in $Test_{1add}$ and $Test_2$), indicating its potential for identifying tau deposition below the detection thresholds of visual and/or semiquantitative analyses. Ultimately, this offers the possibility of identifying early stage disease.

Moreover, the differential diagnosis would be challenging for semi-quantitative parameters when an elevated signal is present in overlapping regions (i.e., putamen and globus pallidum in PSP and MSA-P). In the current study, a non-negligible percentage of patients with MSA-P were misdiagnosed as PSP by semi-quantitative indices, especially when using the MR-free approach. This situation was significantly improved when applying the proposed NFDL model. Notably, the patterns of tracer accumulation extracted from the model were consistent with both previous data obtained in MSA-P[27] and the reconstructed dynamic maps of tau propagation reported in *postmortem* studies in PSP.[4] Although the reasons underlying the retention of a tau PET tracer in an α-synucleinopathy remain unclear, this imaging modality shows promise for improving the differential diagnosis between MSA-P and PSP.

Convolutional neural networks (CNNs) are well-established DL algorithms and have been widely applied in medical-imaging classification. One advantage of CNNs is their ability to learn filters during the

preprocessing phase automatically. By directly inputting raw PET images into the model, significant sources of potential bias, such as spatial normalization, reference regions, selection of ROIs, and optimal cutoffs, are eliminated. Therefore, the NFDL classifier described in our study meets the growing need for an objective interpretation of tau PET images for clinical and research applications. Our model can also be successfully applied in settings where experienced readers and MR imaging facilities are lacking.

One of the prominent advantages of *in vivo* tau PET imaging is that it serves as both a qualitative biomarker for disease diagnosis and a quantitative biomarker for severity monitoring. Although the primary aim of the proposed NFDL model was classification, the preliminary findings indicate that NFDL-guided radiomic features significantly correlate with clinical scores, suggesting its potential additional role in disease severity assessment. Further studies with larger sample sizes are essential to validate and strengthen the reliability of these preliminary findings. Future exploration with larger data incorporating clinical and pathological phenotypes may further improve the results, and longitudinal data will be collected to validate its efficiency in the near future.

Midbrain atrophy, as observed on structural MRI, is regarded as a supportive imaging finding for diagnosing PSP[42] and holds potential for the distinguishment of PSP and MSA-P.[43] However, it is essential to recognize that midbrain atrophy in PSP is thought to be a late-stage outcome of earlier molecular processes and could occur in diseases other than PSP. Moreover, normal elderly individuals may also exhibit midbrain atrophy.[44] Recent research indicates that ^{18}F-PI-2620 tau PET has the potential to enhance the differentiation of PSP from age-matched healthy controls through both visual and quantitative evaluations.[45] Given the importance of cost-effectiveness in daily practice, it is crucial to compare the discriminative performance of these two imaging modalities in PSP and MSA-P cases.

In conclusion, the proposed NFDL method can serve as a practical and efficient assistant for the interpretation of ^{18}F-florzolotau PET imaging in PSP, as it allows simple and qualitative analysis without complex procedures such as PERSI and related programs.

Limitations of the study

There are several limitations to this study. The single-center design and the focus on one PET tracer may limit the external validity of the results. Another caveat is that diagnoses were made on clinical grounds, and no *postmortem* findings are available. Although the clinical misdiagnosis rates of MSA as PSP (i.e., 0%,[46] 2.9%[47]) and vice versa (i.e., 0%,[46] 3.9%,[48] 10.0%[47]) were relatively low, it is crucial to exercise caution when interpreting the findings of this study. Despite the random augmentation procedure, overfitting is still a concern due to the small sample size. In addition, we did not include visual readings due to the lack of consensus among assessors. Finally, in analyzing different forms of atypical parkinsonism, we did not include Lewy body dementia since tau deposition in Lewy body dementia can be attributed to AD-related pathology.[49] Furthermore, we did not include corticobasal degeneration because it is clinically challenging to distinguish corticobasal degeneration from corticobasal syndrome associated with other distinct pathologies.[50]

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Clinical assessment
  - Imaging acquisition
  - Imaging preprocessing
  - Normalization-free deep learning classifiers
  - Conventional semi-quantitative classifiers
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107426.

ACKNOWLEDGMENTS

We extend our sincere gratitude to all the participants and their families for their invaluable contribution to this study. We would also like to express our appreciation to APRINOIA Therapeutics Co., Ltd. (Suzhou, China) for providing the tosylate precursor of ^{18}F-florzolotau. This work was supported by the National Natural Science Foundation of China (grants 82272039, 82021002, 81971641, 82171252, 81701250, 82071200, and 81902282), the China Scholarship Council (grant 202006100181), the Research Project of Shanghai Health Commission (grant 2020YJZX0111), the Clinical Research Plan of SHDC (grants SHDC2020CR1038B and SHDC2020CR4007), the Science and Technology Innovation 2030 Major Project of China (grant 2022ZD0211600), National Key R&D Program of China (grants 2022YFC2009902, 2022YFC2009900), Medical Innovation Research Project of Shanghai Science and Technology Commission (grant 21Y11903300).

AUTHOR CONTRIBUTIONS

Conceptualization, C.Z., K.S., F.L., J.L., and C.C.; Methodology, C.C., J.L., M.W., J.H., L.C., and L.L.; Formal Analysis, C.C. and J.L.; Investigation, J.L., X.L., F.J., P.W., J.Wu, J.G., Y.S., J.Wang, C.Z., F.L., Q.Z., and Y.G.; Resources, C.Z. and J.Wang; Writing—Original Draft, J.L. and C.C.; Writing—Review & Editing, C.Z., K.S., T.-C.Y., F.L., Q.Z., M.B., and A.R.; Visualization, J.L. and C.C.; Supervision, C.Z., K.S., and F.L.; Project Administration, C.Z., K.S., F.L., Y.G., and A.R.; Funding Acquisition, C.Z., F.L., Q.Z., J.L., and J.G.

DECLARATION OF INTERESTS

T.-C.Y. is an employee of APRINOIA Therapeutics Co., Ltd. (Suzhou, China). K.S. and A.R. received research support from Novartis and Siemens Healthineers. M.B. received speaker honoraria from GE Healthcare, Roche, and Life Molecular Imaging and is an advisor of Life Molecular Imaging. All other authors declare no competing interests.

Received: March 30, 2023
Revised: May 28, 2023
Accepted: July 17, 2023
Published: July 20, 2023

REFERENCES

1. Sonni, I., Lesman Segev, O.H., Baker, S.L., Iaccarino, L., Korman, D., Rabinovici, G.D., Jagust, W.J., Landau, S.M., and La Joie, R.; Alzheimer's Disease Neuroimaging Initiative (2020). Evaluation of a visual interpretation method for tau-PET with 18F-flortaucipir. Alzheimers Dement. *12*, e12133. https://doi.org/10.1002/dad2.12133.

2. Seibyl, J.P., DuBois, J.M., Racine, A., Collins, J., Guo, Q., Wooten, D., Stage, E., Cheng, D., Gunn, R.N., Porat, L., et al. (2023). A visual interpretation algorithm for assessing brain tauopathy with 18-F MK-6240 positron emission tomography. J. Nucl. Med. *122*, 264371. https://doi.org/10.2967/jnumed.122.264371.

3. Lin, H.-C., Lin, K.J., Huang, K.-L., Chen, S.-H., Ho, T.-Y., Huang, C.-C., Hsu, J.-L., Chang, C.-C., and Hsiao, I.-T. (2023). Visual reading for [18F]Florzolotau ([18F]APN-1607) tau PET imaging in clinical assessment of Alzheimer's disease. Front. Neurosci. *17*, 1148054. https://doi.org/10.3389/fnins.2023.1148054.

4. Kovacs, G.G., Lukic, M.J., Irwin, D.J., Arzberger, T., Respondek, G., Lee, E.B., Coughlin, D., Giese, A., Grossman, M., Kurz, C., et al. (2020). Distribution patterns of tau pathology in progressive supranuclear palsy. Acta Neuropathol. *140*, 99–119. https://doi.org/10.1007/s00401-020-02158-2.

5. Vogel, J.W., Young, A.L., Oxtoby, N.P., Smith, R., Ossenkoppele, R., Strandberg, O.T., La Joie, R., Aksman, L.M., Grothe, M.J., Iturria-Medina, Y., et al. (2021). Four distinct trajectories of tau deposition identified in Alzheimer's disease. Nat. Med. *27*, 871–881. https://doi.org/10.1038/s41591-021-01309-6.

6. Young, C.B., Winer, J.R., Younes, K., Cody, K.A., Betthauser, T.J., Johnson, S.C., Schultz, A., Sperling, R.A., Greicius, M.D., Cobos, I., et al. (2022). Divergent cortical tau positron emission tomography patterns among patients with preclinical Alzheimer disease. JAMA Neurol. *79*, 592–603. https://doi.org/10.1001/jamaneurol.2022.0676.

7. Lu, J., Zhang, Z., Wu, P., Liang, X., Zhang, H., Hong, J., Clement, C., Yen, T.-C., Ding, S., Wang, M., et al. (2023). The heterogeneity of asymmetric tau distribution is associated with an early age at onset and poor prognosis in Alzheimer's disease. Neuroimage. Clin. *38*, 103416. https://doi.org/10.1016/j.nicl.2023.103416.

8. Liu, F.-T., Lu, J.-Y., Sun, Y.-M., Li, L., Yang, Y.-J., Zhao, J., Ge, J.-J., Wu, P., Jiang, J.-H., Wu, J.-J., et al. (2023). Dopaminergic dysfunction and glucose metabolism characteristics in parkin-induced early-onset Parkinson's disease compared to genetically undetermined early-onset Parkinson's disease. Phenomics *3*, 22–33. https://doi.org/10.1007/s43657-022-00077-8.

9. Lu, J., Wang, M., Wu, P., Yakushev, I., Zhang, H., Ziegler, S., Jiang, J., Förster, S., Wang, J., Schwaiger, M., et al. (2023). Adjustment for the age- and gender-related metabolic changes improves the differential diagnosis of Parkinsonism. Phenomics *3*, 50–63. https://doi.org/10.1007/s43657-022-00079-6.

10. Zhang, T., Wu, S., Zhang, X., Dai, Y., Wang, A., Zhang, H., and Tian, M. (2022). Spatial normalization and quantification approaches of PET imaging for neurological disorders. Eur. J. Nucl. Med. Mol. Imaging *49*, 3809–3829. https://doi.org/10.1007/s00259-022-05809-6.

11. Weigand, A.J., Maass, A., Eglit, G.L., and Bondi, M.W. (2022). What's the cut-point?: a systematic investigation of tau PET thresholding methods. Alzheimer's Res. Ther. *14*, 49. https://doi.org/10.1186/s13195-022-00986-w.

12. Young, C.B., Landau, S.M., Harrison, T.M., Poston, K.L., and Mormino, E.C.; ADNI (2021). Influence of common reference regions on regional tau patterns in cross-sectional and longitudinal [18F]-AV-1451 PET data. Neuroimage *243*, 118553. https://doi.org/10.1016/j.neuroimage.2021.118553.

13. Boxer, A.L., Yu, J.T., Golbe, L.I., Litvan, I., Lang, A.E., and Höglinger, G.U. (2017). Advances in progressive supranuclear palsy: new diagnostic criteria, biomarkers, and therapeutic approaches. Lancet Neurol. *16*, 552–563. https://doi.org/10.1016/S1474-4422(17)30157-6.

14. Goedert, M. (2015). Alzheimer's and Parkinson's diseases: the prion concept in relation to assembled Aβ, tau, and α-synuclein. Science *349*, 1255555. https://doi.org/10.1126/science.1255555.

15. Zaharchuk, G. (2019). Next generation research applications for hybrid PET/MR and PET/CT imaging using deep learning. Eur. J. Nucl. Med. Mol. Imaging *46*, 2700–2707. https://doi.org/10.1007/s00259-019-04374-9.

16. Alvén, J., Heurling, K., Smith, R., Strandberg, O., Schöll, M., Hansson, O., and Kahl, F. (2019). In A Deep Learning Approach to MR-less Spatial Normalization for Tau PET Images - Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, eds. (Springer International Publishing), pp. 355–363.

17. Jo, T., Nho, K., Risacher, S.L., and Saykin, A.J.; Alzheimer's Neuroimaging Initiative (2020). Deep learning detection of informative features in tau PET for Alzheimer's disease classification. BMC Bioinformatics *21*, 496. https://doi.org/10.1186/s12859-020-03848-0.

18. Zou, J., Park, D., Johnson, A., Feng, X., Pardo, M., France, J., Tomljanovic, Z., Brickman, A.M., Devanand, D.P., Luchsinger, J.A., et al. (2021). Deep learning improves utility of tau PET in the study of Alzheimer's disease. Alzheimers Dement. *13*, e12264. https://doi.org/10.1002/dad2.12264.

19. Endo, H., Tagai, K., Ono, M., Ikoma, Y., Oyama, A., Matsuoka, K., Kokubo, N., Hirata, K., Sano, Y., Oya, M., et al. (2022). A machine learning-based approach to discrimination of tauopathies using [18 F]PM-PBB3 PET images. Mov. Disord. *37*, 2236–2246. https://doi.org/10.1002/mds.29173.

20. Tian, M., Civelek, A.C., Carrio, I., Watanabe, Y., Kang, K.W., Murakami, K., Garibotto, V., Prior, J.O., Barthel, H., Zhou, R., et al. (2022). International consensus on the use of tau PET imaging agent 18 F-flortaucipir in Alzheimer's disease. Eur. J. Nucl. Med. Mol. Imaging *49*, 895–904. https://doi.org/10.1007/s00259-021-05673-w.

21. Li, L., Liu, F.T., Li, M., Lu, J.Y., Sun, Y.M., Liang, X., Bao, W., Chen, Q.S., Li, X.Y., Zhou, X.Y., et al. (2021). Clinical utility of 18F-APN-1607 Tau PET imaging in patients with progressive supranuclear palsy. Mov. Disord. *36*, 2314–2323. https://doi.org/10.1002/mds.28672.

22. Brendel, M., Barthel, H., Van Eimeren, T., Marek, K., Beyer, L., Song, M., Palleis, C., Gehmeyr, M., Fietzek, U., Respondek, G., et al. (2020). Assessment of 18F-PI-2620 as a biomarker in progressive supranuclear palsy. JAMA Neurol. *77*, 1408–1419. https://doi.org/10.1001/jamaneurol.2020.2526.

23. Deutschländer, A.B., Ross, O.A., Dickson, D.W., and Wszolek, Z.K. (2018). Atypical parkinsonian syndromes: a general neurologist's perspective. Eur. J. Neurol. *25*, 41–58. https://doi.org/10.1111/ene.13412.

24. Xu, Z., Shen, B., Tang, Y., Wu, J., and Wang, J. (2022). Deep clinical phenotyping of Parkinson's disease: towards a new era of research and clinical care. Phenomics *2*, 349–361. https://doi.org/10.1007/s43657-022-00051-4.

25. Tagai, K., Ono, M., Kubota, M., Kitamura, S., Takahata, K., Seki, C., Takado, Y., Shinotoh, H., Sano, Y., Yamamoto, Y., et al. (2021). High-contrast *in vivo* imaging of tau pathologies in Alzheimer's and Non-Alzheimer's disease tauopathies. Neuron *109*, 42–58.e8. https://doi.org/10.1016/j.neuron.2020.09.042.

26. Liu, F.-T., Lu, J.-Y., Li, X.-Y., Liang, X.-N., Jiao, F.-Y., Ge, J.-J., Wu, P., Li, G., Shen, B., Wu, B., et al. (2023). 18F-Florzolotau PET imaging captures the distribution patterns and regional vulnerability of tau pathology in progressive supranuclear palsy. Eur. J. Nucl. Med. Mol. Imaging *50*, 1395–1405. https://doi.org/10.1007/s00259-022-06104-0.

27. Liu, F.T., Li, X.Y., Lu, J.Y., Wu, P., Li, L., Liang, X.N., Ju, Z.Z., Jiao, F.Y., Chen, M.J., Ge, J.J., et al. (2022). 18F-florzolotau tau positron emission tomography imaging in patients with multiple system atrophy–Parkinsonian subtype. Mov. Disord. *37*, 1915–1923. https://doi.org/10.1002/mds.29159.

28. Cho, H., Choi, J.Y., Lee, S.H., Ryu, Y.H., Lee, M.S., and Lyoo, C.H. (2017). 18F-AV-1451 binds to putamen in multiple system atrophy. Mov. Disord. *32*, 171–173. https://doi.org/10.1002/mds.26857.

29. Schönecker, S., Brendel, M., Palleis, C., Beyer, L., Höglinger, G.U., Schuh, E., Rauchmann, B.S., Sauerbeck, J., Rohrer, G., Sonnenfeld, S., et al. (2019). PET imaging of astrogliosis and tau facilitates diagnosis of Parkinsonian syndromes. Front. Aging Neurosci. *11*, 1–9. https://doi.org/10.3389/fnagi.2019.00249.

30. Perez-Soriano, A., Arena, J.E., Dinelle, K., Miao, Q., McKenzie, J., Neilson, N., Puschmann, A., Schaffer, P., Shinotoh, H., Smith-Forrester, J., et al. (2017). PBB3 imaging in Parkinsonian disorders: evidence for binding to tau and other proteins. Mov. Disord. *32*, 1016–1024. https://doi.org/10.1002/mds.27029.

31. Southekal, S., Devous, M.D., Kennedy, I., Navitsky, M., Lu, M., Joshi, A.D., Pontecorvo, M.J., and Mintun, M.A. (2018). Flortaucipir F 18 quantitation using parametric estimation of reference signal intensity. J. Nucl. Med. *59*, 944–951. https://doi.org/10.2967/jnumed.117.200006.

32. Zhang, H., Wang, M., Lu, J., Bao, W., Li, L., Jiang, J., and Zuo, C.; Alzheimer's Disease Neuroimaging Initiative (2021). Parametric estimation of reference signal intensity for semi-quantification of tau deposition: a flortaucipir and [18F]-APN-1607 study. Front. Neurosci. *15*, 598234. https://doi.org/10.3389/fnins.2021.598234.

33. Tagai, K., Ikoma, Y., Endo, H., Debnath, O.B., Seki, C., Matsuoka, K., Matsumoto, H., Oya, M., Hirata, K., Shinotoh, H., et al. (2022). An optimized reference tissue method for quantification of tau protein depositions in diverse neurodegenerative disorders by PET with 18F-PM-PBB3 (18F-APN-1607). Neuroimage *264*, 119763. https://doi.org/10.1016/j.neuroimage.2022.119763.

34. DeVries, T., and Taylor, G.W. (2017). Improved regularization of convolutional neural networks with cutout. Preprint at arXiv. https://doi.org/10.48550/arxiv.1708.04552.

35. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.

36. Kingma, D.P., and Ba, J. (2014). Adam: a method for stochastic optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1412.6980.

37. McHugh, M.L. (2012). Interrater reliability: the kappa statistic. Biochem. Med. *22*, 276–282.

38. Zhang, H., Wu, P., Ziegler, S.I., Guan, Y., Wang, Y., Ge, J., Schwaiger, M., Huang, S.C., Zuo, C., Förster, S., et al. (2017). Data-driven identification of intensity normalization region based on longitudinal coherency of 18F-FDG metabolism in the healthy brain. Neuroimage *146*, 589–599. https://doi.org/10.1016/j.neuroimage.2016.09.031.

39. Landau, S.M., Ward, T.J., Murphy, A., Iaccarino, L., Harrison, T.M., La Joie, R., Baker, S., Koeppe, R.A., and Jagust, W.J.; Alzheimer's Disease Neuroimaging Initiative (2023). Quantification of amyloid beta and tau PET without a structural MRI. Alzheimer's Dement. *19*, 444–455. https://doi.org/10.1002/alz.12668.

40. Lu, J., Ju, Z., Wang, M., Sun, X., Jia, C., Li, L., Bao, W., Zhang, H., Jiao, F., Lin, H., et al. (2023). Feasibility of 18F-florzolotau quantification in patients with Alzheimer's disease based on an MRI-free tau PET template. Eur. Radiol. *33*, 4567–4579. https://doi.org/10.1007/s00330-023-09571-7.

41. Wang, M., Yan, Z., Zhang, H., Lu, J., Li, L., Yu, J., Wang, J., Matsuda, H., Zuo, C., and Jiang, J.; Alzheimer's Disease Neuroimaging Initiative (2021). Parametric estimation of reference signal intensity in the quantification of amyloid-beta deposition: An 18F-AV-45 study. Quant. Imaging Med. Surg. *11*, 249–263. https://doi.org/10.21037/QIMS-20-110.

42. Höglinger, G.U., Respondek, G., Stamelou, M., Kurz, C., Josephs, K.A., Lang, A.E.,

Mollenhauer, B., Müller, U., Nilsson, C., Whitwell, J.L., et al. (2017). Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria. Mov. Disord. *32*, 853–864. https://doi.org/10.1002/mds.26987.

43. Heim, B., Krismer, F., and Seppi, K. (2021). Differentiating PSP from MSA using MR planimetric measurements: a systematic review and meta-analysis. J. Neural. Transm. *128*, 1497–1505. https://doi.org/10.1007/s00702-021-02362-8.

44. Morelli, M., Arabia, G., Messina, D., Vescio, B., Salsone, M., Chiriaco, C., Perrotta, P., Rocca, F., Cascini, G.L., Barbagallo, G., et al. (2014). Effect of aging on magnetic resonance measures differentiating progressive supranuclear palsy from Parkinson's disease. Mov. Disord. *29*, 488–495. https://doi.org/10.1002/mds.25821.

45. Messerschmidt, K., Barthel, H., Brendel, M., Scherlach, C., Hoffmann, K.T., Rauchmann, B.S., Rullmann, M., Marek, K., Villemagne, V.L., Rumpf, J.J., et al. (2022). 18F-PI-2620 Tau PET improves the imaging diagnosis of progressive supranuclear palsy. J. Nucl. Med. *63*, 1754–1760. https://doi.org/10.2967/jnumed.121.262854.

46. Joutsa, J., Gardberg, M., Röyttä, M., and Kaasinen, V. (2014). Diagnostic accuracy of parkinsonism syndromes by general neurologists. Park. Relat. Disord. *20*, 840–844. https://doi.org/10.1016/j.parkreldis.2014.04.019.

47. Hughes, A.J., Daniel, S.E., Ben-Shlomo, Y., and Lees, A.J. (2002). The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain *125*, 861–870. https://doi.org/10.1093/brain/awf080.

48. Josephs, K.A., and Dickson, D.W. (2003). Diagnostic accuracy of progressive supranuclear palsy in the Society for Progressive Supranuclear Palsy brain bank. Mov. Disord. *18*, 1018–1026. https://doi.org/10.1002/mds.10488.

49. Lee, S.H., Cho, H., Choi, J.Y., Lee, J.H., Ryu, Y.H., Lee, M.S., and Lyoo, C.H. (2018). Distinct patterns of amyloid-dependent tau accumulation in Lewy body diseases. Mov. Disord. *33*, 262–272. https://doi.org/10.1002/mds.27252.

50. Koga, S., Josephs, K.A., Aiba, I., Yoshida, M., and Dickson, D.W. (2022). Neuropathology and emerging biomarkers in corticobasal syndrome. J. Neurol. Neurosurg. Psychiatry *93*, 919–929. https://doi.org/10.1136/jnnp-2021-328586.

51. Gilman, S., Wenning, G.K., Low, P.A., Brooks, D.J., Mathias, C.J., Trojanowski, J.Q., Wood, N.W., Colosimo, C., Dürr, A., Fowler, C.J., et al. (2008). Second consensus statement on the diagnosis of multiple system atrophy. Neurology *71*, 670–676. https://doi.org/10.1212/01.wnl.0000324625.00404.15.

52. Jessen, F., Amariglio, R.E., Van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., Dubois, B., Dufouil, C., Ellis, K.A., Van Der Flier, W.M., et al. (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimer's Dement. *10*, 844–852. https://doi.org/10.1016/j.jalz.2014.01.001.

53. Ossenkoppele, R., Rabinovici, G.D., Smith, R., Cho, H., Schöll, M., Strandberg, O., Palmqvist, S., Mattsson, N., Janelidze, S., Santillo, A., et al. (2018). Discriminative accuracy of [18F]flortaucipir positron emission tomography for Alzheimer disease vs other neurodegenerative disorders. JAMA *320*, 1151–1162. https://doi.org/10.1001/jama.2018.12917.

54. Tian, M., Zuo, C., Civelek, A.C., Carrio, I., Watanabe, Y., Kang, K.W., Murakami, K., Garibotto, V., Prior, J.O., Barthel, H., et al. (2022). International nuclear medicine consensus on the clinical use of amyloid positron emission tomography in Alzheimer's disease. Phenomics. https://doi.org/10.1007/s43657-022-00068-9.

55. Zhang, Y., Lu, J., Wang, M., Zuo, C., and Jiang, J. (2022). Influence of gender on tau precipitation in Alzheimer's Disease according to ATN research framework. Phenomics. https://doi.org/10.1007/s43657-022-00076-9.

56. Rolls, E.T., Huang, C.C., Lin, C.P., Feng, J., and Joliot, M. (2020). Automated anatomical labelling atlas 3. Neuroimage *206*, 116189. https://doi.org/10.1016/j.neuroimage.2019.116189.

57. Falcon, W. (2019). The PyTorch Lightning Team (PyTorch Lightning). https://doi.org/10.5281/zenodo.3828935.

58. MONAI (2022). MONAI: Medical Open Network for AI. https://doi.org/10.5281/zenodo.6114127.

59. Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999). Automated model-based tissue classification of MR images of the brain. IEEE Trans. Med. Imaging *18*, 897–908. https://doi.org/10.1109/42.811270.

60. Kang, G., Dong, X., Zheng, L., and Yang, Y. (2017). Patchshuffle regularization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1707.07103.

61. Tan, M., and Le, Q. (2019). Efficientnet: rethinking model scaling for convolutional neural networks. In International conference on machine learning (PMLR), pp. 6105–6114.

62. Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141.

63. Loshchilov, I., and Hutter, F. (2017). Decoupled weight decay regularization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1711.05101.

64. Ginsburg, B., Castonguay, P., Hrinchuk, O., Kuchaiev, O., Lavrukhin, V., Leary, R., Li, J., Nguyen, H., Zhang, Y., and Cohen, J.M. (2019). Stochastic gradient methods with layer-wise adaptive moments for training of deep networks. Preprint at arXiv. https://doi.org/10.48550/arXiv.1905.11286.

65. Suah, F.B.M., Zeiler, M.D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision (Springer), pp. 818–833. https://doi.org/10.1016/j.ancr.2017.02.001.

66. Xiao, Y., Fonov, V., Chakravarty, M.M., Beriault, S., Al Subaie, F., Sadikot, A., Pike, G.B., Bertrand, G., and Collins, D.L. (2017). A dataset of multi-contrast population-averaged brain MRI atlases of a Parkinson's disease cohort. Data Brief *12*, 370–379. https://doi.org/10.1016/j.dib.2017.04.013.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Imaging data and code	Current paper	Available from the corresponding author upon reasonable request
Software and algorithms		
MATLAB 9.5	MathWorks	https://www.mathworks.com/products/matlab.html
SPM 12	FIL	https://www.fil.ion.ucl.ac.uk/spm/software/spm12/
SPSS 23	IBM Corporation	https://www.ibm.com/products/spss-statistics
Python 3.9	Python Software Foundation	https://www.python.org/
PyTorch	The Linux Foundation	https://pytorch.org/get-started/locally/
MONAI	Project MONAI	https://monai.io/core.html

RESOURCE AVAILABILITY

Lead contact

Additional information, inquiries about resources, or requests for reagents should be directed to the corresponding author, Chuantao Zuo (zuochuantao@fudan.edu.cn).

Materials availability

During the course of this study, no new or unique reagents were generated. Analyses were performed using MATLAB and Python software. To facilitate data reanalysis from this article, the necessary code can be obtained by contacting the lead contact. Additionally, the lead contact is available for providing any supplemental information that may be essential for a comprehensive reevaluation of the reported data.

Data and code availability

The related dataset and code will be available to the scientific community upon completion of the non-disclosure agreement with the lead contact author according to international data protection regulations.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study protocol was approved by the institutional review board of Huashan Hospital, a leading tertiary care center in Shanghai, China. All participants or legal guardians provided written informed consent. The study subjects were consecutively registered in the Progressive Supranuclear Palsy Neuroimage Initiative.[26] Clinical diagnoses were made by experienced movement disorder specialists according to currently accepted guidelines (the 2017 MDS diagnostic criteria for PSP;[42] the second consensus statement on the diagnosis of MSA[51]) and in a blinded fashion with respect to tau PET imaging. The levels of diagnostic certainty (probable, possible, suggestive) and clinical predominance types (*i.e.*, PSP with Richardson's syndrome (PSP-RS), PSP with predominant gait freezing (PSP-PGF), PSP with predominant parkinsonism (PSP-P)) were determined in patients with PSP according to the 2017 MDS diagnostic criteria.[42] Controls were cognitively unimpaired and had no significant neurological or psychiatric illnesses. They were a mix of research volunteers recruited through advertisements without cognitive or movement complaints and persons visiting the memory clinic with cognitive complaints but normal performance at neuropsychological testing (*i.e.*, "subjective cognitive decline"[52]) with negative findings on amyloid PET imaging.[53,54] The details of controls are presented in Table S7.

The dataset was divided into two cohorts, depending on the availability of corresponding MR imaging (Cohort I with MR images, Cohort II without MR images). Cohort I, which consisted of 148 patients with PSP, 52 patients with MSA-P, and 45 controls, was subsequently divided into six splits (five for

training/validation and one for testing ($Test_1$)). The split was implemented with label stratification to minimize intergroup differences in terms of clinical diagnoses, demographic characteristics, and disease severity (see Table S1). Cohort II, which was used for additional testing ($Test_2$), comprised 18 patients with PSP, 8 patients with MSA-P, and 9 controls.

METHOD DETAILS

Clinical assessment

Patients with PSP and MSA-P who were taking antiparkinsonian medication were assessed after at least 12h from the last dose. Cognitive function was assessed using the MMSE. The severity of motor dysfunction was assessed using the MDS-UPDRS-III and the Hoehn and Yahr (H&Y) for patients with PSP and MSA-P. Disease severity in PSP was also characterized by PSPrs.

Imaging acquisition

^{18}F-florzolotau was prepared in Huashan Hospital using a nucleophilic substitution reaction followed by acid hydrolysis, utilizing with an ^{18}F-multifunction synthesizer (Beijing PET Technology Co., Ltd., Beijing, China). APRINOIA Therapeutics Co., Ltd (Suzhou, China) provided the tosylate precursor required for the radiosynthesis.

PET images were acquired on a Siemens mCT Flow PET/CT scanner (Siemens, Erlangen, Germany) in 3-dimensional (3D) mode. A low-dose CT transmission scan was performed for attenuation correction. ^{18}F-florzolotau was administered intravenously (370 MBq). ^{18}F-florzolotau PET imaging was performed over a 20-minute acquisition time (90–110 minutes). The images were reconstructed using a 3D ordered-subset expectation maximization algorithm (6 iterations; 21 subsets; Gaussian filter, 3.5 mm; zoom, 2). The reconstructed images had a matrix size of 256 × 256 × 148 and an effective voxel size of 1.59 × 1.59 × 1.50 mm.

High-resolution T1-weighted images were acquired in a 3.0-T horizontal magnet (Discovery MR750; GE Medical Systems, Milwaukee, WI) using the following parameters: TE = 3.2 ms, TR = 8.2 ms, TI = 450 ms, flip angle = 12°, acquisition matrix = 256 × 256 × 152, and voxel size = 1 × 1 × 1 mm.

Imaging preprocessing

For the development of the NFDL model, reconstructed PET images were directly used. For the semi-quantitative classifiers, spatial normalization and reference region-based intensity normalization were applied for preprocessing. For subjects in Cohort I, reconstructed ^{18}F-florzolotau PET scans were spatially normalized to the Montreal Neurosciences Institute (MNI) standard space. This was achieved through Statistical Statistical Parametric Mapping 12 (SPM12), implemented within MATLAB 9.5, using two distinct methods: MR-dependent and MR-free. While for subjects in Cohort II, only the MR-free approach was applied due to the absence of concurrent MR scans. More specifically, the MR-dependent preprocessing involved segmenting individual MR images into grey matter (GM), white matter (WM), and cerebrospinal fluid. Afterward, the reconstructed ^{18}F-florzolotau PET scans were resampled in the common space of the corresponding MR images and subsequently normalized to the MNI standard space by utilizing the transformation matrix derived from the MR segmentation.[55] In the MR-free spatial normalization approach, the unified algorithm in SPM12 (tissue probability map) was applied to the reconstructed PET images. All spatially normalized PET images were then smoothed using a Gaussian kernel with a full-width at half-maximum of 6 mm.[26]

In the next step, reference region-based intensity normalization was performed to obtain the SUVR maps. In detail, four different reference regions, cerebellar gray matter (cereGM), whole cerebellum (WC), subject-specific white matter (subWM), and subject-specific grey matter (subGM) were selected. The cereGM and WC regions were derived from the automated anatomical atlas three template (AAL3).[56] These two reference regions are widely used in AD[12] since the cerebellum is considered to contain minimal tau fibrils in AD. In PSP, the cereGM is a routinely selected reference region,[21,22,25] while WC is rarely used due to the potential deposition of tau in the dentate nucleus and cerebellar white matter as the disease advances.[4] The subWM region and subGM region were defined using PERSI,[31,32,41] a data-driven approach for count normalization that identifies subject-specific reference regions. These two approaches leverage the advantages of the lower variability of large reference regions while mitigating potential partial-volume effects.

The subWM region has been effectively utilized in AD research[31,32,41] while a recent study in PSP suggested that GM may offer more accurate semi-quantification in tau PET imaging compared to WM.[33] To generate a binary WM image, a threshold of 0.9 was applied to the WM probabilistic segmentation derived from individual T1 structural MR images. The spatially normalized PET images were then masked using each individual's binary WM image. Within this WM region, voxels were plotted as a histogram and fitted to a bimodal Gaussian distribution using a non-linear trust region reflective method. The PERSI method is employed to identify voxels with contamination in higher-intensity peaks based on counts from adjacent cortical tissues that exhibit confirmed ^{18}F-florzolotau uptake. In contrast, lower-intensity peaks represent a stable reference signal intensity. By excluding voxels found in higher-intensity peaks, the remaining voxels in lower-intensity peaks were utilized to establish a subject-specific reference region (subWM).[31,32,41] The process for acquiring a subject-specific reference region (subGM) is fundamentally similar, with only a minor variation. Instead of using WM probabilistic segmentation, GM probability probabilistic segmentation is employed, derived from individual T1-weighted MR images. As PERSI is dependent on MR, its application was limited exclusively to $Test_1$.

Normalization-free deep learning classifiers

The training pipeline was implemented using PyTorch Lightning,[57] whereas data transformations and network architectures were developed with MONAI.[58] Five-fold cross-validation was used to evaluate different design and hyperparameter options, and the model characterized by the highest validation F1 score was selected as the most appropriate classifier. The development process for the deep learning-based classifier unfolded as described below.

Training pipeline

The training pipeline was implemented using PyTorch Lightning.[57] With a child class of the *LightningDataModule*, the training, validation, and test datasets with their respective data transforms and augmentations were set up, and the corresponding *DataLoaders* were initialized. With a child class of the *LightningModule*, the network architecture, loss, metrics, and optimizer were set up, and the training, validation, and test steps were defined. We used MONAI,[58] a PyTorch based framework for medical imaging, for data transformations and implementing network architectures.

Optimal hyperparameter selection

To identify the optimal classifier with the highest F1 score (the harmonic mean of precision and recall), we evaluated different design and hyperparameter choices.

First, we tested two options for the intensity transform that was applied to the samples before feeding them into the network. The first option was intensity scaling where the input was scaled to values between 0.0 and 1.0. The second option was intensity normalization where the mean of the input is subtracted from the input and the result is divided by standard deviation of the input, leading to an output with a mean of 0.0 and a standard deviation of 1.0.

Second, to overcome overfitting on the training data, we compared the influence of several random data augmentations on the network's validation performance. The following random data augmentations were tested: addition of random Gaussian noise, random bias fields,[59] random contrast adjustment, random Gaussian smoothing, application of random affine or elastic transformations, and random coarse shuffle regularization.[60]

Third, we evaluated network architectures using different depths from the family of the DenseNets,[35] EfficientNets,[61] and Squeeze-and-Excitation Networks,[62] which were adapted for 3D inputs.

Finally, we tested three different optimizers, Adam,[36] AdamW,[63] and Novograd[64] implemented in MONAI.[58]

In addition to the previously described design options, we used the following fixed parameters during each run: a batch size of eight, a learning rate of 3e-4 and a weight decay of 1e-8. For each option, five-fold cross-validation was run to get more accurate results. The experiments were conducted on a server equipped with an AMD Ryzen Threadripper 3990X 64-core CPU, 256 GB RAM, and an NVIDIA GeForce RTX 3090 Graphics Card.

Occlusion sensitivity maps

Occlusion sensitivity maps[65] were computed to visualize which regions in the image are important for the model's classification decision. Additionally, NFDL-guided radiomic features (first-order statistical features), including median, mean, standard deviation, minimum, maximum, quantile, range, variance, skew, and kurtosis, were extracted from the individually weighted tau PET images calculated by multiplying the raw SUV image with the individual occlusion sensitivity map.

Conventional semi-quantitative classifiers

Regional level criteria for positive findings on ^{18}F-florzolotau PET imaging

The red nucleus, subthalamic nucleus, raphe nuclei, globus pallidus, and putamen were utilized to identify PSP.[4,21,26] The raphe nuclei was generated from AAL3,[56] and the others were defined according to MNI PD25 template.[66] The last two ROIs (globus pallidus, putamen) were also applied for MSA-P.[27] Receiver operating characteristic (ROC) curve analysis and Youden's *J* statistic (*J*=sensitivity+specificity-1) were used to determine the optimal cut-off points (*i.e.*, the SUVR value corresponding to the maximum *J* value) for ^{18}F-florzolotau positivity at the regional level in the training/validation datasets. Patients with PSP and MSA-P were merged to define ^{18}F-florzolotau positivity in the globus pallidus and putamen,[21,27] while MSA-P and control were merged to define ^{18}F-florzolotau negativity in all other ROIs. Regional SUVR Z-scores were calculated based on controls included in the training/validation sets of Cohort I using the following formula: [(individual SUVR - mean SUVR in controls)/standard deviation of the SUVR in controls], and a regional Z-score $\geq$ 2 was considered positive as suggested by prior research.[11,21,26]

Criteria for individual classification

Previously reported criteria for global ^{18}F-florzolotau positivity[21,22,26] were applied based on optimal cut-off SUVRs or a Z-score $\geq$ 2 for classification. Subjects were classified as PSP when regional positivity was found in at least one of the following three ROIs, that is red nucleus, subthalamic nucleus, and raphe nuclei. Subjects were classified as MSA-P when regional positivity was only observed in either the globus pallidus or the putamen or both. When all five ROIs were found to be negative, subjects were classified as controls.

QUANTIFICATION AND STATISTICAL ANALYSIS

Chi-square test, Kruskal-Wallis test and one-way analysis of variance (ANOVA) were used to examine the demographic and clinical variables differences among the training/validation sets and test set in Cohort I. Based on the individual classification labels derived from the methods mentioned earlier, we assessed the classification performance of each method by calculating the area under the ROC curves, accuracy, sensitivity, specificity, and confusion matrices. Specifically, ROC curve analyses were conducted to distinguish PSP from the combined group of MSA-P and controls, MSA-P from the combined group of PSP and controls, and controls from the combined group of PSP and MSA-P.

The discrepancy between semi-quantitative measures obtained with different spatial normalization methods was first examined in the training/validation sets at the group level using (1) ICC, and (2) the one-sample t-test for the mean difference of SUVRs. It was also examined in the test sets in a head-to-head manner at the individual level using Cohen's kappa coefficients for the classification ($Test_1$ and $Test_{1add}$). The $Test_{1add}$ set consisted of the same subjects in the $Test_1$ set but with MR-free spatial normalization.

The correlations between NFDL-guided radiomic features and clinical scores for severity, as well as the associations between NFDL-guided radiomic features and the SUVR values were assessed using Pearson's correlation analysis. The analyses were carried out in IBM SPSS Statistics for Windows, version 23 (IBM Corp., Armonk, NY, USA), and a P value smaller than 0.05 was considered significant.

ADDITIONAL RESOURCES

Not applicable.