

Original Article

Heuristic scoring method utilizing FDG-PET statistical parametric mapping in the evaluation of suspected Alzheimer disease and frontotemporal lobar degeneration

Jeremy N Ford¹, Elizabeth M Sweeney², Myrto Skafida³, Shannon Glynn³, Michael Amoashiy⁴, Dale J Lange⁵, Eaton Lin³, Gloria C Chiang³, Joseph R Osborne³, Silky Pahlajani⁴, Mony J de Leon³, Jana Ivanidze³

¹Department of Radiology, Massachusetts General Hospital, Boston, MA, United States; ²Department of Population Health Sciences, Division of Biostatistics and Epidemiology, Weill Cornell Medical College, New York, NY, United States; ³Departments of ³Radiology, ⁴Neurology, Weill Cornell Medical College, New York, NY, United States; ⁵Department of Neurology, Hospital for Special Surgery, New York, NY, United States

Received April 21, 2021; Accepted July 21, 2021; Epub August 15, 2021; Published August 30, 2021

Abstract: Distinguishing frontotemporal lobar degeneration (FTLD) and Alzheimer Disease (AD) on FDG-PET based on qualitative review alone can pose a diagnostic challenge. SPM has been shown to improve diagnostic performance in research settings, but translation to clinical practice has been lacking. Our purpose was to create a heuristic scoring method based on statistical parametric mapping z-scores. We aimed to compare the performance of the scoring method to the initial qualitative read and a machine learning (ML)-based method as benchmarks. FDG-PET/CT or PET/MRI of 65 patients with suspected dementia were processed using SPM software, yielding z-scores from either whole brain (W) or cerebellar (C) normalization relative to a healthy cohort. A non-ML, heuristic scoring system was applied using region counts below a preset z-score cutoff. W z-scores, C z-scores, or WC z-scores (z-scores from both W and C normalization) served as features to build random forest models. The neurological diagnosis was used as the gold standard. The sensitivity of the non-ML scoring system and the random forest models to detect AD was higher than the initial qualitative read of the standard FDG-PET [0.89-1.00 vs. 0.22 (95% CI, 0-0.33)]. A categorical random forest model to distinguish AD, FTLD, and normal cases had similar accuracy than the non-ML scoring model (0.63 vs. 0.61). Our non-ML-based scoring system of SPM z-scores approximated the diagnostic performance of a ML-based method and demonstrated higher sensitivity in the detection of AD compared to qualitative reads. This approach may improve the diagnostic performance.

Keywords: Dementia, Alzheimer, FTLD, SPM, FDG-PET

Introduction

Alzheimer Disease (AD) is the most common cause of cognitive impairment and dementia, accounting for 50-60% of cases and currently affecting over 5 million individuals in the US. At age 45, the estimated lifetime risk is 10-20% [1]. In current clinical practice, assessment of region-specific patterns of cortical glucose metabolism with 2-[¹⁸F]fluoro-2-deoxy-d-glucose (FDG) PET represents a mainstay of neuroimaging assessment in clinically suspected dementia. Specifically, FDG-PET can aid in the differential diagnosis of AD versus other types

of dementia such as frontotemporal lobar degeneration (FTLD), which includes frontotemporal dementia (FTD) and primary progressive aphasia (PPA) [2]. The differentiation of AD from FTLD remains challenging, with a false positive rate of up to 60% [3]. With the recent FDA approval of aducanumab [4], the distinction between AD and FTLD in the clinical setting will be increasingly important, given that FTLD would hypothetically not respond to an amyloid-targeting agent.

For this indication, FDG-PET outperforms CSF-based biomarkers as well as MRI and CT [5, 6].

Qualitatively, patients with FTLD demonstrate reduced FDG avidity in the anterior cingulate cortex, frontal lobes, basal ganglia, and anterior temporal lobes, whereas patients with AD demonstrate decreased FDG avidity in the posterior temporoparietal lobes and posterior cingulate cortex with preservation of FDG avidity in the anterior cingulate cortex [7-9]. However, advanced AD can involve both the frontal and anterior temporal lobes, which can present a differential diagnostic challenge.

Diagnostic accuracy of FDG-PET is further improved utilizing statistical parametric mapping (SPM), a technique in which standardized uptake values (SUVs) within select regions are compared to cognitively normal subjects [10, 11]. An important consideration for SPM is whether normalization of FDG avidity throughout the subject's brain is performed relative to an intensity composite of the whole brain (W) or the cerebellum (C). It is not well established which normalization method should be routinely used. While cerebellar normalization seems to be more sensitive for early diagnosis, cerebral global normalization might be superior for differential diagnostic purposes in dementia syndromes [12].

While SPM has been well-validated by multiple research groups in dementia imaging [11, 13-15], adoption among radiologists in clinical practice is lagging, who often still rely upon diagnostically inferior, pre-processed FDG-PET to render diagnoses. In fact, the American College of Radiology (ACR) Appropriateness Criteria does not specifically endorse the use of SPM in clinical practice [16]. One potential cause of low adoption could be due to the lack of standardized methods when faced with a large dataset of region-specific z-scores for each patient. In fact, a lack of standardization in interpreting FDG-PET has led some experts to dissuade the routine use of FDG-PET in suspected early dementia [17].

Machine learning (ML) is an emerging approach that could assist radiologists in interpreting either standard FDG-PET images or SPM data. Different techniques have been attempted, ranging from convolutional neural networks to random forests, incorporating different components of FDG-PET [18, 19]. However, it is

unclear which particular features from FDG-PET yield the most robust ML models given that redundancy and irrelevancy can lead to overfitting [20]. Moreover, ML approaches for image interpretation currently face multiple barriers to widespread adoption, including divergence between research and real-world datasets, costs incurred for regulatory approval, heterogeneity in the clinical environments, and concerns over data ownership [21]. There is also the practical consideration that no clinically-oriented, commercially-available ML toolkit is available that classifies dementia subtypes based on FDG-PET patterns. SPM software, on the other hand, is readily available, and it is possible that a non-ML heuristic approach could approximate the performance of ML.

The purpose of this study is to benchmark the performance of a non-ML scoring system based on regions below a preset z-score cutoff against random forest models trained on raw SPM z-scores. Findings will inform whether a heuristic approach to quantitative SPM analysis can approximate the performance of a ML method. This practical scoring system can leverage the important validation of SPM in research setting and encourage SPM's adoption into clinical practice.

Methods

A summary of patient retrieval, image processing, and model validation is outlined in **Figure 1**.

Patient selection

Following institutional review board approval, we retrospectively identified patients in whom AD or FTLD was clinically suspected by searching clinical statements from radiology reports using our institutional radiology report database. Consent was not obtained due to retrospective design and data was analyzed anonymously. The following search terms were used: "Alzheimer", "dementia", "frontotemporal dementia", "semantic dementia", and "primary progressive aphasia". Patients with a history of cerebrovascular accident were excluded from analysis. This yielded 99 cases, of whom 3 were excluded for a history of CVA and 31 were excluded due to a lack of available documentation from a treating neurologist. CVA patients

FDG-PET SPM scoring method for AD and FTLD

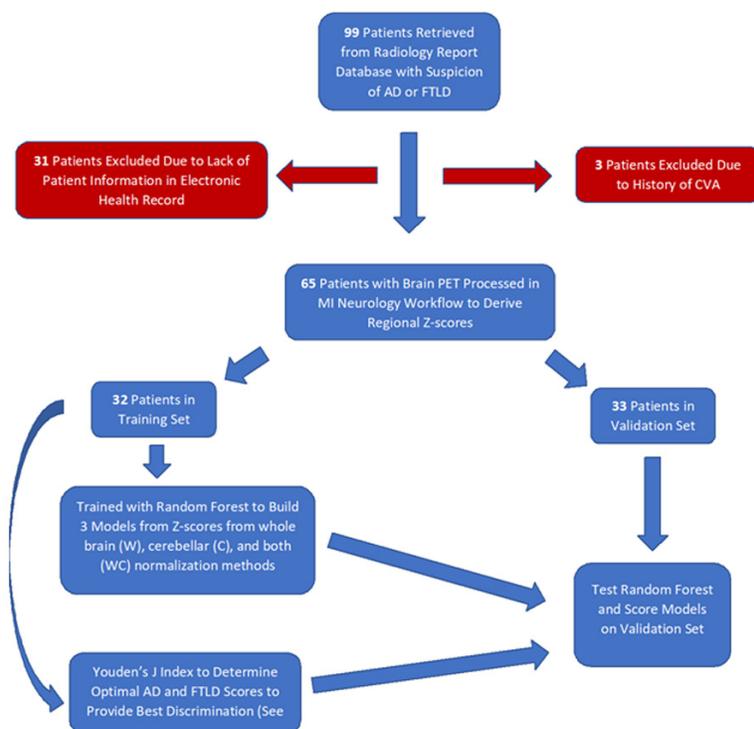


Figure 1. Flowchart of patient retrieval, image processing, and model validation.

were excluded given that their cognitive or behavioral findings were most likely due to their prior infarction, which could affect regional FDG avidity and calculated z-scores. This yielded a total of 65 patients scanned from 10/9/2013 to 11/28/2018. Cerebrospinal fluid (CSF) testing for AD biomarkers and results from cognitive testing were reviewed, when available.

Image acquisition and processing

PET/CT images were acquired on Siemens Biograph mCT PET/CT 64 slice and PET/MR was acquired on 3T Siemens Biograph mMR hybrid scanner according to standard departmental protocol for dementia imaging. FDG-PET data from PET/MR (N = 20) and PET/CT (N = 45) cases were post-processed using the syngo.via (Siemens Healthineers, Erlangen, Germany) MI Neurology Workflow. This involves alignment of PET to the MRI or CT using a rigid registration algorithm. Adequate co-registration PET and anatomical data was confirmed visually by JF, a senior radiology resident and JI, a board-certified radiologist with subspecialty in training in neuroradiology and nuclear medicine.

Attenuation correction for PET/MR studies was acquired by segmenting the Dixon images into 5 compartments: air, lung, fat, bone, and soft tissue, which is standard on Siemens Biograph mMR scanners [22, 23]. The PET/CT and PET/MR scanners were not harmonized given the retrospective study design and real-world clinical cohort. For the purposes of this study, SPM-derived z-scores from either PET/CT or PET/MR were considered directly comparable given that in clinical settings, PET/CT and PET/MR with optimal attenuation correction yield similar SUV measurements and diagnostic performance [24-26].

Multiple automated approaches for alignment to the database were employed, including linear affine registration

to account for global position and scaling differences as well as a deformable registration algorithm to allow for localized adjustments. Smoothing was accomplished with an isotropic Gaussian filter of size 12 mm fullwidth at half-maximum. This was followed by intensity normalization using an automated whole brain or cerebellar protocol [27].

Z-scores were generated by comparison to a database within the syngo.via MI Neurology Workflow software derived from 33 healthy controls (age 46-79, 22 female, 11 male), converting the volume to the Montreal Neurological Institute (MNI) standard space and parsing into Automated Anatomical Labeling (AAL) regions, which have been described previously [28]. A z-score (Patient SUV-Healthy Atlas SUV)/Healthy Atlas Standard Deviation) was generated for each AAL region.

Z-score-based random forest models

Random forest was the machine learning method selected for comparison given the relatively limited number of features and small sample size. Random forest is a method of combining decision trees and is especially useful when

there is limited number of correlated predictors, such as our dataset of region-specific z-scores. A comprehensive explanation of random forest methodology is described elsewhere [29]. Due to the limited sample size for this analysis, we restricted ourselves to using this classical machine learning algorithm instead of a more complex algorithm, such as those from deep learning. As we were using a small number of derived features from the FDG-PET, the gains from more flexible and complex machine learning algorithms would be expected to be small. For the random forest, we used the default parameters from the R implementation, as it has been shown that there are limited gains for tuning these parameters and that the biggest performance gains are made with the first 100 trees [30]. We were also limited in the amount of data in the training set to do this parameter tuning on.

All machine learning was performed with the R statistical and computing software, Version 3.6 using the random Forest R package. The default parameters for the random forest model were used: 500 trees and the number of variables available for splitting at each tree node is equal to the square root of the number of predictors in each model. More details about the training and validation set and assessing the variability of model performance can be found in the 'Diagnostic Performance and Statistical Analysis' section below.

After randomly assigning patients to training ($n = 32$) and validation ($n = 33$) sets, random forest models were created to evaluate for AD using SPM-generated z-scores in brain regions selected *a priori* that would have minimal overlap between FTLD and AD [7, 31]: pre-cuneus, posterior cingulate gyri, and posterior parietal cortices (a composite of the following AAL regions: angular gyrus, inferior parietal, and supramarginal gyrus) for AD; anterior cingulate gyri, anterior temporal cortices, and basal ganglia for FTLD. Separate models from z-scores from whole brain (W) and cerebellar (C) intensity normalization were created. An additional combined model (WC) was built using data z-scores from both normalization methods. The same approach was used to generate models to diagnose FTLD. The same training and validation partition was applied in all models.

Categorical random forest models were created by training on FTLD, AD, and non-AD/FTLD cases, with separate models trained on z-scores from whole brain, cerebellar, and combined whole brain/cerebellar normalization.

Region count scoring system (non-machine learning)

Using the same training and validation sets, an additional model (Score) was generated. Each subject received an AD score (0-6) or FTLD score (0-6), counting the number of bilateral AD and FTLD regions below the $z = -2.0$ cutoff. A z-score cutoff of -2.0 was selected given its prior validation in the workup of dementia subtypes [32], utility in stereotactic surface projection visualization [7], and its proximity to $z = -1.65$ and -1.96 , which have previously been used as thresholds to distinguish normal and pathologic values [33, 34]. **Figure 2** demonstrates a use case of how these scores are tabulated for each patient. Models to diagnose AD and FTLD were trained and validated separately. Youden's J index was used to determine the optimal AD or FTLD score to provide the best discrimination.

A third Score model was applied to simultaneously distinguish AD, FTLD, and non-AD/FTLD cases. Subjects only achieving the AD Score threshold were classified as AD, FTLD Score threshold were classified as FTLD, and those that met both AD and FTLD Score thresholds were distinguished based on the relative z-scores of the cingulate cortex (anterior cingulate < posterior cingulate = FTLD; posterior cingulate < anterior cingulate = AD). This particular tiebreaker was used given the anterior cingulate typically has lower FDG-avidity in FTLD [35], and the posterior cingulate has lower FDG-avidity in AD [36].

Gold standard and qualitative read

The gold standard was considered the clinical neurological diagnosis as ascertained from the neurologist's latest electronic medical record (EMR) outpatient clinical note or in consultation with two referring neurologists, one with 35 years (DL) and the other with 20 years of post-residency experience (MA), both blinded to outputs from random forest models and Score outputs. Fifteen cases in which the chart diagnosis was mild cognitive impairment (MCI) were

FDG-PET SPM scoring method for AD and FTLD

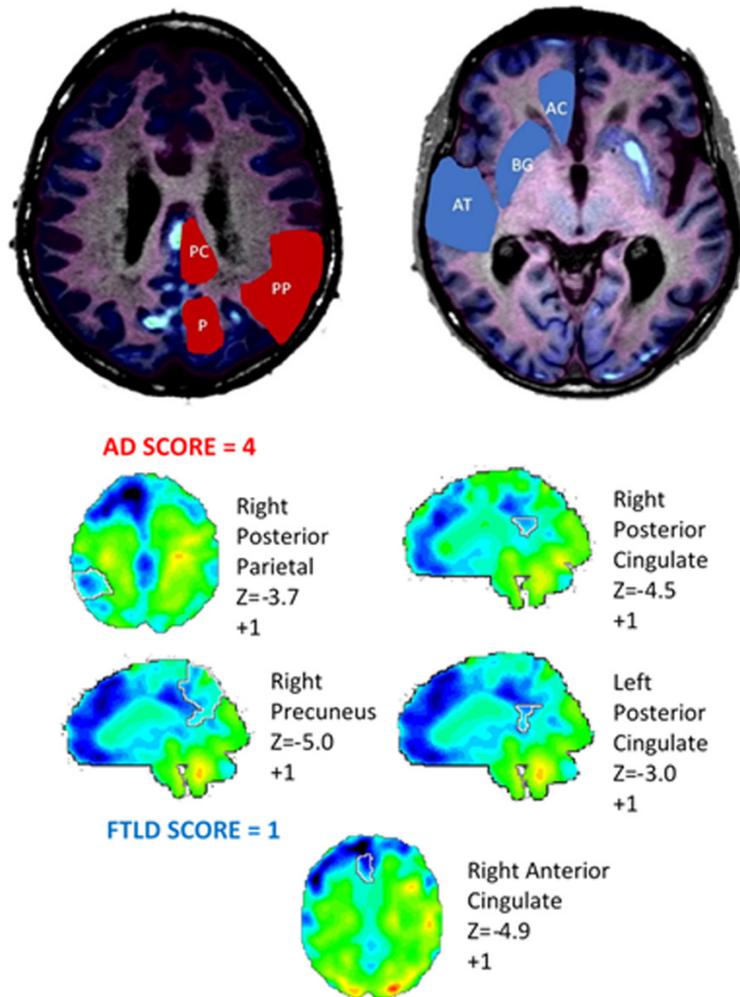


Figure 2. SPM-based scoring method. *Top images* demonstrates axial reformats of a 3D T1-weighted MPRAGE sequence from a representative normal subject at the level of the posterior parietal cortices and basal ganglia (B), with superimposed concurrently acquired FDG-PET (windowed at SUVMAX 0-15), as well as overlaying schematic representations of the regions of the Montreal Neurologic Institute standard space. Regions selected for the AD score include the bilateral posterior cingulate (PC), precuneus (P), and posterior parietal (PP). Regions selected for the FTLD score include the bilateral anterior cingulate (AC), basal ganglia (BG), and anterior temporal cortex (AT). *Bottom images* are SPM z-score maps in a patient with a neurologic diagnosis of AD and qualitative imaging diagnosis of FTLD. Applying a z-score cutoff of -2.0 for each region, four AD regions had z-scores below the cutoff (AD score = 4) and one FTLD region at z-scores below the cutoff.

reviewed by a neurologist with fellowship training in behavioral neurology and expertise in neurodegenerative disorders (SP). Blinded to random forest and Score outputs, SP reviewed the patients' chart, which included neuropsychological testing and occasionally CSF markers, classifying these patients as either AD (amnesic MCI on AD spectrum) or non-AD/FTLD (non-AD MCI). Defining categories, non-AD/FTLD includes both cognitively normal

patient and non-AD MCI, as ascertained by the final neurology note (gold standard). To assess the strength of the gold standard, SP also independently evaluated the clinical data of 10 randomly selected patients and yielded a separate diagnosis. A kappa value was subsequently calculated. For the purposes of this study, patients with diagnoses of PPA and FTD were considered FTLD for analysis purposes. All cases of PPA were semantic variant (svPPA), falls under the FTLD diagnostic category. While initially clinically divergent from a behavioral neurology perspective, PPA and FTD often have overlapping imaging appearances and are often grouped in neuroimaging and pathology studies [37]. Also, svPPA often converts to FTD as the disease progresses [38].

The given diagnosis from the impression from the initial radiology report was used as the qualitative read. These qualitative reads rendered interpreting radiologist within 48 hours of the FDG-PET acquisition. Qualitative reads were furnished by the radiologist without the use of any SPM software, interpreting only the raw SUV fused with either a CT or MRI of the brain. If multiple diagnoses were offered, the diagnosis that was most favored was used. If

multiple diagnoses were given without favoring one, the qualitative read was labeled "nonspecific". If no diagnosis was given, the qualitative read was considered "nonspecific".

Diagnostic performance and statistical analysis

All statistical analyses were performed with R statistical and computing software, Version

FDG-PET SPM scoring method for AD and FTLD

Table 1. Demographic and clinical characteristics of the study population

Neurologic Diagnosis	FTLD	AD	Non-AD/FTLD	Total
# of Cases	21 (15 FTD, 6 PPA)	23	21 (11 cognitively normal + 10 non-AD MCI)	65
Age (SD) Range	71.3 (8.0) 57-88	74.1 (8.5) 58-88	70.7 (9.6) 47-84	72.1 (8.7) 47-88
Number Female (%)	10 (48%)	10 (43%)	10 (48%)	30 (46%)
PET/CT	14	15	16	45
PET/MR	7	8	5	20

The majority of cases were PET/CT. Nearly half the patients were female. Mean age did not differ significantly between subgroups. FTLD = Frontotemporal Lobar Degeneration; FTD = Frontotemporal Dementia; PPA = Primary Progressive Aphasia; AD = Alzheimer Disease; MCI = Mild Cognitive Impairment.

Table 2. Diagnostic performance of qualitative reads, random forest models, and heuristic scoring model (Score) to distinguish AD and Non-AD/FTLD patients

	Qualitative Read	W	C	WC	Score
Accuracy (95% CI)	0.76 (0.61-0.79)	0.79 (0.61-0.88)	0.70 (0.61-0.88)	0.79 (0.64-0.91)	0.79 (0.70-0.88)
Sensitivity	0.22	1.00	0.89	0.89	0.89
Specificity	0.96	0.71	0.62	0.75	0.75
AUC (95% CI)	---	0.87 (0.66-0.90)	0.82 (0.65-0.90)	0.89 (0.66-0.91)	0.86 (0.75-0.94)

All random forest models (W, C, and WC) had higher sensitivity to detect AD than the qualitative read. While Score and WC models yielded the highest accuracy and WC had the highest AUC. However, confidence intervals were overlapping.

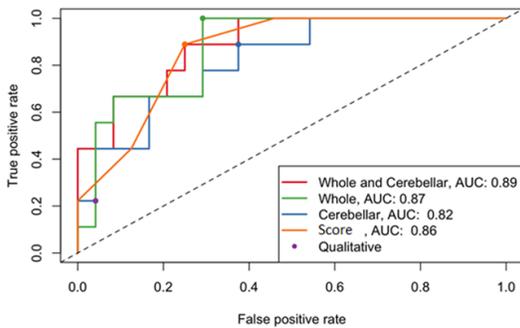


Figure 3. Receiver Operating Characteristics Curve for Models for AD. AUC for all random forest models were similar with overlapping confidence intervals.

3.6. Using Youden's J index, predicted probabilities from the random forest and scores model were thresholded to yield accuracies, sensitivities, and specificities in the validation set. In the random forest models, the Gini importance factors for each brain region were generated in the training set to identify which regions had the strongest discriminatory power in diagnosing AD and FTLD respectively. The Gini importance factors provide a relative ranking of the features and are a by-product in the training of the random forest classifier. Bootstrapped 95% confidence intervals for accuracy, sensitivity, specificity, and area under the curve (AUC) were calculated for both quantitative data and the qualitative reads.

Results

Clinical and demographic characteristics of the study population

Table 1 outlines the demographic and clinical features of our study population, including age, gender ratio, and scanning modality. Mean age was 72.1 years (range 47-88) and 46.1% were women. There were 67.7% with a significant neurological diagnosis, including AD (35.3%) and FTLD (32.3%). 32.3% of patients were deemed non-AD/FTLD (11 cognitively normal, 10 non-AD MCI). There were no significant differences in age among each clinical group.

Characteristics of radiologists providing qualitative reads

Radiologists who provided the qualitative read had a mean of 7.9 years of post-fellowship experience (ranging 0 to 32 years). Thirty-eight scans were read by five neuroradiologists, 19 by three nuclear medicine radiologists, and 8 by two body radiologists.

Diagnostic performance

SP's independent evaluation of the clinical data and comparison with the gold standard (the latest neurological diagnosis in the chart) yielded a kappa value of 0.66. **Table 2** outlines the performance random forest models in diagnosing

Table 3. Diagnostic performance of qualitative reads, random forest models, and heuristic scoring model (Score) to distinguish FTLD and Non-AD/FTLD patients

	Qualitative read	W	C	WC	Score
Accuracy (95% CI)	0.61 (0.64-0.85)	0.70 (0.52-0.85)	0.70 (0.58-0.88)	0.76 (0.55-0.88)	0.67 (0.55-0.76)
Sensitivity	0.08	0.67	0.83	0.50	1.00
Specificity	0.90	0.71	0.62	0.90	0.48
AUC (95% CI)	---	0.67 (0.52-0.82)	0.73 (0.56-0.88)	0.72 (0.53-0.85)	0.72 (0.59-0.83)

All random forest models (W, C, and WC) had higher sensitivity to detect FTLD than the qualitative read. While WC model yielded the highest accuracy, C model had the highest AUC. However, confidence intervals were overlapping.

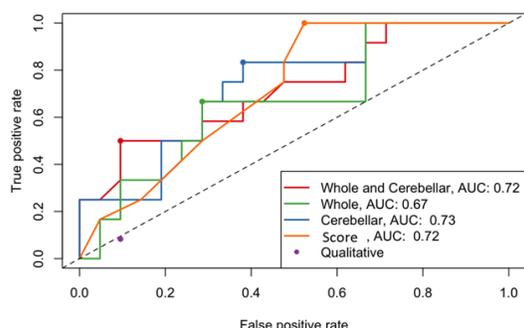


Figure 4. Receiver Operating Characteristics Curve for Models for FTLD. AUC for all random forest models were similar with overlapping confidence intervals.

AD. All models had a sensitivity of 0.89 or higher, surpassing the 0.22 (95% CI, 0-0.33) sensitivity of qualitative reads. There was a clear tradeoff with specificity, as ML models and Score model were less specific than qualitative reads (0.62-75 vs. 0.96). Nonetheless, all random forest models except for model C were more accurate than the qualitative read, although 95% confidence intervals were overlapping. Among the models, WC and Score had the highest accuracy, and the WC model had the highest AUC value, the latter of which is illustrated in **Figure 3**.

Table 3 and **Figure 4** illustrate random forest model performance in diagnosing FTLD. All models performed similarly in accuracy, sensitivity, and specificity with the qualitative read with overlapping 95% confidence intervals. Similar to the AD dataset, sensitivity among qualitative reads was low at 0.08, with near 1.00 sensitivity using the Score method. Again, there was a trade-off between sensitivity and specificity, which decreased with the Score model to 0.48 and increased among the qualitative reads to 0.90. In contrast the WC model had higher specificity than sensitivity (0.90 vs. 0.50). Among the models, WC had the highest

Table 4. Diagnostic performance of random forest models and heuristic scoring model (score) to distinguish AD, FTLD, and Non-AD/FTLD patients

	W	C	WC	Score
Accuracy	0.48	0.45	0.61	0.63

The random forest model that incorporated z-scores from both whole brain and cerebellar normalization (WC) had the highest accuracy, similar to the Score model.

accuracy (0.76) and C at the highest AUC (0.73), although confidence intervals were overlapping. Within the validation set, there were four cases in which all four models were discordant for AD and two cases in which all four models were discordant for FTLD.

Among categorical random forest models to distinguish AD, FTLD, and Non-AD/FTLD patients, the WC model had the highest accuracy of 0.61, followed by the C model (0.52), and W model (0.48). The Score model to distinguish all three categories performed similarly to the WC random forest, with an accuracy of 0.63 (**Table 4**).

CSF AD biomarkers and cognitive testing

AD biomarkers (t-tau, p-tau, and Aβ-42) from CSF samples were available in 5 patients and mental status exam scores were available in 6 patients, which were reported as either Montreal Cognitive Assessment (MoCA) or Folstein. Only one patient had positive CSF AD biomarkers. These data are summarized in **Table 5**. No statistical analysis was performed due to the limited number cases with CSF biomarker data.

Discussion

Given the growing incidence of AD and the emergence of disease-modifying therapies, our

FDG-PET SPM scoring method for AD and FTLD

Table 5. Findings for patients with CSF AD biomarkers and cognitive assessments

	CSF AD Biomarker	Neurological Diagnosis	Qualitative read	Score Diagnosis	Cognitive Exam Score
Patient 1	Negative	FTD	AD	Non-AD/FTLD	MoCA 24
Patient 2	Positive	AD	Non-AD/FTLD	AD	Folstein 22
Patient 3	Negative	AD	Nonspecific	AD	MoCA 16
Patient 4	-----	PPA	FTLD	FTLD	Folstein 25
Patient 5	Negative	AD	Non-AD/FTLD	Non-AD/FTLD	MoCA 14
Patient 6	-----	Non-AD/FTLD	Non-AD/FTLD	Non-AD/FTLD	Folstein 27
Patient 7	Negative	FTD	Normal	FTLD	-----

Of the 5 patients in our dataset, only 1 had positive biomarkers for AD, as determined from p-tau, t-tau, and A β -42. In this patient, the initial qualitative read interpreted the PET/CT as normal, whereas our Score method classified this patient as AD, concordant with the final neurological diagnosis. Scores on the MoCA range from zero to 30, with a score of 26 and higher generally considered normal. Scores on the Folstein are as follows: 25-30 points: normal cognition 21-24 points: mild dementia 10-20 points: moderate dementia 9 points or lower: severe dementia.

goal was to develop an easy-to-implement scoring system based on SPM z-scores that could approximate the performance of a machine learning approach. Our scoring system performed similarly to automated or machine learning based approaches in the literature when diagnosing AD in isolation, but performance degraded when distinguishing AD and FTLD.

The literature shows that visual inspection of SPM images alone bolsters diagnostic performance of FDG-PET in dementia workup. Perani et al. achieved an AUC of 0.67 when the reader was evaluated SPM brain maps, compared to 0.50 when evaluating standard FDG images [11]. Deep learning approaches further enhance diagnostic performance, for example, Ding et al. trained a convolutional neural network on over a thousand patient FDG-PET studies from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and yielded AUC = 0.92 for AD and AUC = 0.63 for MCI due to AD [19]. Our AD random forest models and non-ML scoring system trained on a smaller dataset (n = 32) and on SPM z-scores rather than raw images featured similar levels of diagnostic performance for AD (AUC = 0.82-0.87), all with a statistically higher sensitivity for AD than the initial reader analyzing standard FDG-PET images.

While our heuristic scoring method with SPM improved sensitivity in distinguishing AD and FTLD patients from non-AD/FTLD patients, there appeared to be a clear tradeoff with specificity, given that specificity of the original qualitative read was higher (0.96 for AD, 0.90 for

FTLD). The relatively lower specificity of the scoring heuristic could be multifactorial. One is the inherent limitation of our gold standard, as discussed below. Its possible subset of our cases counted as false positives were in fact true positives and may have eventually converted to dementia. Due to the retrospective design of the study, our ability to follow-up on subjects beyond the EMR was restricted. Another possibility is that the maximum age in our study group was 88, whereas the maximum age in the SPM database was 79. Normal age-related declines in FDG-avidity are well-documented [39, 40]. Therefore, it is conceivable that the SPM scoring heuristic and random forest models misclassified normal aging-related changes in FDG-avidity as AD or FTLD, yielding false positives and lower specificity. Another factor may be the z-score cutoff selection of -2.0, although lower magnitude cutoffs have been used [33, 34].

The majority of machine learning studies using FDG-PET to diagnose dementia focuses on discriminating AD from normal patients [41-43]. However, according to the Centers for Medicare & Medicaid Services (CMS), the first condition to reimburse FDG-PET scan is that the "onset, clinical presentation, or course of cognitive impairment is atypical for AD, and FTD is suspected as an alternative neurodegenerative cause of the cognitive decline" [44]. Therefore, FDG-PET scans in the clinical setting are not intended to distinguish cognitively normal patients from patients with AD, but rather to distinguish AD and FTLD when there is already likely a clinical diagnosis of dementia.

Our non-ML scoring method meant to distinguish AD, FTLD, and non-AD/FTLD patients had lower accuracy (0.63) relative to the separate binary models described above (e.g., AD versus normal). The scoring method was also slightly better than most of the categorical random forest models to distinguish AD, FTLD, and normal, the latter with accuracies ranging from 0.48 to 0.61. This was likely due to lower dimensionality of data, relatively small sample size, and use of a real-world clinical cohort rather than curated, validated databases. Compared to previously applied machine learning techniques, Davatzikos et al. incorporating highly dimensional data from MRI scans yielded an accuracy of 0.84 on a similar size as our sample (AD $n = 37$, FTD $n = 12$), with the authors noting that the differentiation between AD and FTLD being more challenging than distinguishing between AD and normal subjects [45]. Applying stepwise forward logistic regression, Mosconi et al. were able to develop a SPM-based tool to distinguish five diagnostic groups, including Dementia with Lewy Bodies, attaining sensitivities from 71% to 99% and specificities from 68% to 98% [46]. An ongoing multisite effort is underway to merge data from the ADNI cohort with multiple emerging FTLD databases [47], which will probably result in even stronger discriminatory power.

A shared shortcoming of machine learning approaches is that they may have limited deployability in the community, especially in areas remote from academic medical centers without the software or expertise to implement them into clinical workflow. For example, in our study, random forest models were generated and interpreted with the statistical software package R by a statistician with expertise in R software development and machine learning.

These approaches also face the same barriers to implementation machine learning has across radiology, including concerns over data ownership, regulatory constraints, challenges with inter-institutional data transfer, and the degree to which the test set reflects real-world patients [21]. On the other hand, a ready to implement approach similar to the scoring method described above could enhance diagnostic performance today with commercially available software, particularly improving sensitivity in the detection of AD.

However, as machine learning image analysis tools transition from research to clinical settings, the level of user expertise required and interoperability issues will likely become diminishing factors. Multiple neuroimaging suites have already arrived to market, largely focused on large vessel occlusions in cerebral ischemia [48]. While these ready-to-use software packages are more practical than developing software *de novo*, additional training is often needed for physicians, technologists, and IT specialists to successfully deploy the AI system [49]. Moreover, options for commercial AI software for the evaluation of dementia remain limited.

Amyloid targeted PET offers an alternative to FDG-PET to help clinicians distinguish FTLD and AD, with the advantage imaging one of the pathologic proteins of AD rather than the emergent phenomenon of altered brain metabolism. However, amyloid targeted PET has similar discriminatory power between AD and FTLD relative to FDG-PET [50]. and in at least one study, FDG-PET was a better predictor of cognitive performance than ^{18}F florbetapir in the workup of AD and MCI [51]. The Imaging Dementia-Evidence for Amyloid Scanning (IDEAS) study offers seemingly contrary evidence, the primary diagnosis changed with amyloid PET in 35.6% of study participants. Moreover, compared to FDG-PET, IDEAS demonstrated higher sensitivity and negative predictive value with amyloid PET in the diagnosis of AD [52]. These mixed findings could be explained by how the clinical question is framed. Inclusion into the IDEAS study required that participants meet appropriateness criteria as outlined by the 2013 Amyloid Imaging Taskforce, outlining three “appropriate” clinical scenarios, none of which explicitly state that FTLD must be a diagnostic consideration [53]. In the 2011 study by Rabinovici et al. showing similar accuracy between amyloid PET and FDG-PET, participants had to have met clinical criteria for either AD or FTLD, a narrower clinical question than the IDEAS study and similar to the inclusion criteria of our study.

Despite the positive findings in the IDEAS study, amyloid-targeted PET is still not currently covered under Medicare outside of a clinical trial, as determined by the Centers for Medicare & Medicaid Services (CMS), leaving patients and families to cover the high costs [54]. FDG-PET, on the other hand, is often covered by Medicare

under most appropriate clinical scenarios, specifically when the primary diagnostic considerations are AD and FTLD [44]. While accumulating evidence from the New IDEAS study may sway policymakers, FDG-PET will likely remain the standard of care in the near term [55].

It is possible the lessons learned from the heuristic, quantitative SPM-based method described herein may have relevance for the broad use of amyloid PET in the future. Quantitative analysis of amyloid PET has already shown promise on a limited basis. In 55 patients scanned with ^{18}F florbetapir, quantitative analysis of PET images changed the initial visual read in 9.7% of interpretations [56]. A quantitative approach, such as our SPM method for FDG-PET, could similarly augment amyloid PET reads.

Tau imaging is also an emerging approach to diagnose AD and has the potential to better characterize the severity of disease, given that tau pathology has been found to more closely correlate with AD symptoms than amyloid [57]. However, tau imaging is currently primarily a research tool, not currently reimbursed or widely available.

This study also is among the first to utilize the commercially available syngo.via MI Neurology Workflow for dementia diagnosis in a research setting, one group using the tool to correlate findings between FDG-PET and EEG [58], and another exploring the utility or combined DTI/PET in FTLD workup [59]. Hitherto, MI Neurology Workflow has been predominately applied to epilepsy research [25, 60, 61]. Another easy-to-use software package optimized for FDG-PET analysis in the clinical setting is MIMneuro, which was recently successfully employed to compare volumetric and metabolic profiles in patients with dementia [62]. What these clinically focused FDG-PET software packages have in common is they are seamlessly integrated with hospital Picture Archiving and Communication Systems (PACS) and can be navigated with relatively little training. This contrasts with research-focused FDG-PET analysis software, such as PMOD, which is not optimized for PACS integration or easy use by the clinician. Nonetheless, tools such as PMOD are extensively validated, and in one study achieved an accuracy of 95.8% distinguishing AD from

normal patients with the support of deep learning [63].

A limitation of our study compared to publications discussed above is that our scoring system and random forest models were trained to a small heterogeneous sample rather than a curated research cohort. The small sample size is also less than ideal for random forest training, and conceivably, a larger training set could have improved the performance of random forest model. Future applications of the random forest model on a larger dataset would confirm this.

Another limitation is that in cases of MCI in which the clinician did not denote AD versus non-AD, we relied on a neurologist with expertise in dementia to discern the neurological diagnosis retrospectively. MCI is a clinical heterogeneous entity, which includes prodromal AD, systemic diseases, neuroinflammatory conditions, medical side effects, metabolic dysfunction, and is therefore a less rigid diagnostic category than AD or FTLD [64]. Another consideration is that we considered the neurological diagnosis the ground truth for this study, which is an inherently limited gold standard given that interrater reliability among neurologists is weak to moderate globally, with Kappa values for multiple dementia disorders is consistently under 0.7 [65]. In our internal evaluation of the strength of the gold standard, we calculated a Kappa value of 0.66. This may partly explain the similar degrees of accuracy with other FDG-PET-based models using the clinical diagnosis as the gold standard [11, 19]. Of note, there were six patients in which all four models classified the patient discordantly, and it is therefore possible that some of these patients had mixed pathology or were misclassified by the neurologist. Additionally, 31 cases were eliminated from the sample due to lack of documentation in the EMR, which may have introduced a degree of selection bias.

This study serves as a proof-of-concept for a readily deployable scoring system based on SPM z-score maps generated from commercially available software. This non-ML approach mirrored the accuracy of a machine learning model along multiple metrics. Future efforts may include further validation on a larger, curated clinical cohort, such as the ADNI data-

set, as well as the forthcoming Frontotemporal Lobar Degeneration Neuroimaging Initiative dataset. In addition, a similar scoring system may be applied alongside prospective, longitudinal studies incorporating systematic neuropsychologic testing and CSF samples. If diagnostic performance is validated, radiologists in the community could have an easy-to-use tool at their disposal to assist neurologists in distinguishing AD from FTLD.

Acknowledgements

This work was primarily supported by the Ana-Maria and Stephen Kellen Foundation through Weill Cornell Medicine (PI:JI) and the JumpStart Research Career Development Program (PI:JI). Additionally, this work was supported in part by NIH NIA 1RF1AG057570 (PI: MJdL), R56 AG058913 (PI: MJdL), and NIH NIA R01AG068398 (PI:GCC).

Disclosures of conflicts of interest

None.

Address correspondence to: Jeremy N Ford, Department of Radiology, Massachusetts General Hospital, 55 Fruit Street, Boston, MA, United States. Tel: 919-922-5695; E-mail: jeremynford@gmail.com

References

- [1] Alzheimer's Association. 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2018; 14: 367-429.
- [2] Vieira RT, Caixeta L, Machado S, Silva AC, Nardi AE, Arias-Carrion O and Carta MG. Epidemiology of early-onset dementia: a review of the literature. *Clin Pract Epidemiol Ment Health* 2013; 9: 88-95.
- [3] Shinagawa S, Catindig JA, Block NR, Miller BL and Rankin KP. When a little knowledge can be dangerous: false-positive diagnosis of behavioral variant frontotemporal dementia among community clinicians. *Dement Geriatr Cogn Disord* 2016; 41: 99-108.
- [4] Cummings J, Aisen P, Lemere C, Atri A, Sabbagh M and Salloway S. Aducanumab produced a clinically meaningful benefit in association with amyloid lowering. *Alzheimers Res Ther* 2021; 13: 98.
- [5] Bloudek LM, Spackman DE, Blankenburg M and Sullivan SD. Review and meta-analysis of biomarkers and diagnostic imaging in Alzheimer's disease. *J Alzheimers Dis* 2011; 26: 627-645.
- [6] Dukart J, Mueller K, Horstmann A, Barthel H, Moller HE, Villringer A, Sabri O and Schroeter ML. Combined evaluation of FDG-PET and MRI improves detection and differentiation of dementia. *PLoS One* 2011; 6: e18111.
- [7] Brown RK, Bohnen NI, Wong KK, Minoshima S and Frey KA. Brain PET in suspected dementia: patterns of altered FDG metabolism. *RadioGraphics* 2014; 34: 684-701.
- [8] Ishii K, Sakamoto S, Sasaki M, Kitagaki H, Yamaji S, Hashimoto M, Imamura T, Shimomura T, Hirono N and Mori E. Cerebral glucose metabolism in patients with frontotemporal dementia. *J Nucl Med* 1998; 39: 1875-1878.
- [9] Jeong Y, Cho SS, Park JM, Kang SJ, Lee JS, Kang E, Na DL and Kim SE. 18F-FDG PET findings in frontotemporal dementia: an SPM analysis of 29 patients. *J Nucl Med* 2005; 46: 233-239.
- [10] Kim J, Cho SG, Song M, Kang SR, Kwon SY, Choi KH, Choi SM, Kim BC and Song HC. Usefulness of 3-dimensional stereotactic surface projection FDG PET images for the diagnosis of dementia. *Medicine (Baltimore)* 2016; 95: e5622.
- [11] Perani D, Della Rosa PA, Cerami C, Gallivanone F, Fallanca F, Vanoli EG, Panzacchi A, Nobili F, Pappata S, Marcone A, Garibotto V, Castiglioni I, Magnani G, Cappa SF and Gianolli L; EADC-PET Consortium. Validation of an optimized SPM procedure for FDG-PET in dementia diagnosis in a clinical setting. *Neuroimage Clin* 2014; 6: 445-454.
- [12] Dukart J, Mueller K, Horstmann A, Vogt B, Frisch S, Barthel H, Becker G, Moller HE, Villringer A, Sabri O and Schroeter ML. Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies. *Neuroimage* 2010; 49: 1490-1495.
- [13] Caminiti SP, Ballarini T, Sala A, Cerami C, Pre-sotto L, Santangelo R, Fallanca F, Vanoli EG, Gianolli L, Iannaccone S, Magnani G and Perani D; BIOMARKAPD Project. FDG-PET and CSF biomarker accuracy in prediction of conversion to different dementias in a large multicentre MCI cohort. *Neuroimage Clin* 2018; 18: 167-177.
- [14] Cerami C, Dodich A, Lettieri G, Iannaccone S, Magnani G, Marcone A, Gianolli L, Cappa SF and Perani D. Different FDG-PET metabolic patterns at single-subject level in the behavioral variant of fronto-temporal dementia. *Cortex* 2016; 83: 101-112.
- [15] Iaccarino L, Chiotis K, Alongi P, Almkvist O, Wall A, Cerami C, Bettinardi V, Gianolli L, Nordberg A and Perani D. A cross-validation of FDG- and amyloid-PET biomarkers in mild cognitive impairment for the risk prediction to dementia

- due to Alzheimer's disease in a clinical setting. *J Alzheimers Dis* 2017; 59: 603-614.
- [16] Expert Panel on Neurological I, Moonis G, Subramaniam RM, Trofimova A, Burns J, Bykowski J, Chakraborty S, Holloway K, Ledbetter LN, Lee RK, Pannell JS, Pollock JM, Powers WJ, Roca RP, Rosenow JM, Shih RY, Utukuri PS and Corey AS. ACR Appropriateness Criteria(R) dementia. *J Am Coll Radiol* 2020; 17: S100-S112.
- [17] Smailagic N, Vacante M, Hyde C, Martin S, Ukoumunne O and Sachpekidis C. (1)(8)F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* 2015; 1: CD010632.
- [18] Cabral C and Silveira M; Alzheimer's Disease Neuroimaging Initiative. Classification of Alzheimer's disease from FDG-PET images using favourite class ensembles. *Conf Proc IEEE Eng Med Biol Soc* 2013; 2013: 2477-2480.
- [19] Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, Lituiev D, Copeland TP, Aboian MS, Mari Aparici C, Behr SC, Flavell RR, Huang SY, Zalocusky KA, Nardo L, Seo Y, Hawkins RA, Hernandez Pampaloni M, Hadley D and Franc BL. A deep learning model to predict a diagnosis of Alzheimer disease by using (18)F-FDG PET of the brain. *Radiology* 2019; 290: 456-464.
- [20] Tuv E, Borisov A, Runger G and Torkkola K. Feature selection with esembles, artificial variables, and redudancy elimination. *Journal of Machine Learning Research* 2009; 10: 1341-1366.
- [21] Taylor J and Fenner J. The challenge of clinical adoption-the insurmountable obstacle that will stop machine learning? *BJR Open* 2018; 1: 20180017.
- [22] Koesters T, Friedman KP, Fenchel M, Zhan Y, Hermosillo G, Babb J, Jelescu IO, Faul D, Boada FE and Shepherd TM. Dixon sequence with superimposed model-based bone compartment provides highly accurate PET/MR attenuation correction of the brain. *J Nucl Med* 2016; 57: 918-924.
- [23] Paulus DH, Quick HH, Geppert C, Fenchel M, Zhan Y, Hermosillo G, Faul D, Boada F, Friedman KP and Koesters T. Whole-body PET/MR imaging: quantitative evaluation of a novel model-based mr attenuation correction method including bone. *J Nucl Med* 2015; 56: 1061-1066.
- [24] Drzezga A, Souvatzoglou M, Eiber M, Beer AJ, Furst S, Martinez-Moller A, Nekolla SG, Ziegler S, Ganter C, Rummeny EJ and Schwaiger M. First clinical experience with integrated whole-body PET/MR: comparison to PET/CT in patients with oncologic diagnoses. *J Nucl Med* 2012; 53: 845-855.
- [25] Poirier SE, Kwan BYM, Jurkiewicz MT, Samar-gandy L, Iacobelli M, Steven DA, Lam Shin Cheung V, Moran G, Prato FS, Thompson RT, Burneo JG, Anazodo UC and Thiessen JD. An evaluation of the diagnostic equivalence of (18)F-FDG-PET between hybrid PET/MRI and PET/CT in drug-resistant epilepsy: a pilot study. *Epilepsy Res* 2021; 172: 106583.
- [26] Sekine T, Buck A, Delso G, Ter Voert EE, Huellner M, Veit-Haibach P and Warnock G. Evaluation of Atlas-based attenuation correction for integrated PET/MR in human brain: application of a head Atlas and comparison to true CT-based attenuation correction. *J Nucl Med* 2016; 57: 215-220.
- [27] Thomas W. Database comparison in MI neurology workflow (white paper). Siemens Medical Solutions 2014.
- [28] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B and Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 2002; 15: 273-289.
- [29] James G, Witten D, Hastie T and Tibshirani R. An introduction to statistical learning. Springer; 2013.
- [30] Probst P and Boulesteix AL. To tune or not to tune the number of trees in random forest. *J Mach Learn Res* 2017; 18.1: 6673-6690.
- [31] Ishii K. PET approaches for diagnosis of dementia. *AJNR Am J Neuroradiol* 2014; 35: 2030-2038.
- [32] Kono AK, Ishii K, Sofue K, Miyamoto N, Sakamoto S and Mori E. Fully automatic differential diagnosis system for dementia with Lewy bodies and Alzheimer's disease using FDG-PET and 3D-SSP. *Eur J Nucl Med Mol Imaging* 2007; 34: 1490-1497.
- [33] Wei L GK, Li Y, Guo Z, Gao C, Yuan M and Zhang M. Construction of a novel Chinese normal brain database using 18F-FDG PET images and MIMneuro software, the initial application in epilepsy. *Int J Neurosci* 2019; 129: 417-422.
- [34] Fällmar D, Lilja J, Danfors T, Kilander L, Iyer V, Lubberink M, Larsson EM and Sörensen J. Z-score maps from low-dose 18F-FDG PET of the brain in neurodegenerative dementia. *Am J Nucl Med Mol Imaging* 2018; 8: 239.
- [35] Womack KB, Diaz-Arrastia R, Aizenstein HJ, Arnold SE, Barbas NR, Boeve BF, Clark CM, DeCarli CS, Jagust WJ, Leverenz JB, Peskind ER, Turner RS, Zamrini EY, Heidebrink JL, Burke JR, DeKosky ST, Farlow MR, Gabel MJ, Higdon R, Kawas CH, Koeppe RA, Lipton AM and Foster NL. Temporoparietal hypometabolism in frontotemporal lobar degeneration and associated

- imaging diagnostic errors. *Arch Neurol* 2011; 68: 329-337.
- [36] Mosconi L, Mistur R, Switalski R, Tsui WH, Glodzik L, Li Y, Pirraglia E, De Santi S, Reisberg B, Wisniewski T and de Leon MJ. FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease. *Eur J Nucl Med Mol Imaging* 2009; 36: 811-822.
- [37] Rajagopalan V and Pioro EP. Longitudinal (18)F-FDG PET and MRI reveal evolving imaging pathology that corresponds to disease progression in a patient with ALS-FTD. *Front Neurol* 2019; 10: 234.
- [38] Matias-Guiu JA, Cabrera-Martin MN, Moreno-Ramos T, Garcia-Ramos R, Porta-Etessam J, Carreras JL and Matias-Guiu J. Clinical course of primary progressive aphasia: clinical and FDG-PET patterns. *J Neurol* 2015; 262: 570-577.
- [39] Loessner A, Alavi A, Lewandrowski K-U, Mozley D, Souder E and Gur R. Regional cerebral function determined by FDG-PET in healthy volunteers: normal patterns and changes with age. *J Nucl Med* 1995; 36: 1141-1149.
- [40] Zuendorf G, Kerrouche N, Herholz K and Baron JC. Efficient principal component analysis for multivariate 3D voxel-based mapping of brain functional imaging data sets as applied to FDG-PET and normal aging. *Hum Brain Mapp* 2003; 18: 13-21.
- [41] Lu D, Popuri K, Ding GW, Balachandar R and Beg MF; Alzheimer's Disease Neuroimaging Initiative. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Sci Rep* 2018; 8: 5697.
- [42] Sivapriya TR, Kamal AR and Thangaiah PR. Ensemble merit merge feature selection for enhanced multinomial classification in Alzheimer's dementia. *Comput Math Methods Med* 2015; 2015: 676129.
- [43] Toussaint PJ, Perlberg V, Bellec P, Desarnaud S, Lacomblez L, Doyon J, Habert MO and Benali H; Benali for the Alzheimer's Disease Neuroimaging Initiative. Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component analyses. *Neuroimage* 2012; 63: 936-946.
- [44] Phurrough S, Salive M, Richardson S and Cano C. Decision memo for positron emission tomography (FDG) and Other neuroimaging devices for suspected dementia. Centers for Medicare & Medicaid Services; 2004.
- [45] Davatzikos C, Resnick SM, Wu X, Parmpi P and Clark CM. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. *Neuroimage* 2008; 41: 1220-1227.
- [46] Mosconi L, Tsui WH, Herholz K, Pupi A, Drzezga A, Lucignani G, Reiman EM, Holthoff V, Kalbe E, Sorbi S, Diehl-Schmid J, Pernecky R, Clerici F, Caselli R, Beuthien-Baumann B, Kurz A, Minoshima S and de Leon MJ. Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *J Nucl Med* 2008; 49: 390-398.
- [47] Wang L, Heywood A, Stocks J, Bae J, Ma D, Popuri K, Toga AW, Kantarci K, Younes L, Mackenzie IR, Zhang F, Beg MF and Rosen H; Alzheimer's Disease Neuroimaging Initiative. Grant report on PREDICT-ADFTD: multimodal imaging prediction of ad/ftd and differential diagnosis. *J Psychiatr Brain Sci* 2019; 4: e190017.
- [48] Murray NM, Unberath M, Hager GD and Hui FK. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J Neurointerv Surg* 2020; 12: 156-164.
- [49] <https://www.rapidai.com/rapidu>.
- [50] Rabinovici GD, Rosen HJ, Alkalay A, Kornak J, Furst AJ, Agarwal N, Mormino EC, O'Neil JP, Janabi M, Karydas A, Growdon ME, Jang JY, Huang EJ, Dearmond SJ, Trojanowski JQ, Grinberg LT, Gorno-Tempini ML, Seeley WW, Miller BL and Jagust WJ. Amyloid vs. FDG-PET in the differential diagnosis of AD and FTLD. *Neurology* 2011; 77: 2034-2042.
- [51] Khosravi M, Peter J, Wintering NA, Serruya M, Shamchi SP, Werner TJ, Alavi A and Newberg AB. 18F-FDG is a superior indicator of cognitive performance compared to 18F-florbetapir in Alzheimer's disease and mild cognitive impairment evaluation: a global quantitative analysis. *J Alzheimers Dis* 2019; 70: 1197-1207.
- [52] Rabinovici GD, Gatsonis C, Apgar C, Chaudhary K, Gareen I, Hanna L, Hendrix J, Hillner BE, Olson C, Lesman-Segev OH, Romanoff J, Siegel BA, Whitmer RA and Carrillo MC. Association of amyloid positron emission tomography with subsequent change in clinical management among medicare beneficiaries with mild cognitive impairment or dementia. *JAMA* 2019; 321: 1286-1294.
- [53] Johnson KA, Minoshima S, Bohnen NI, Donohoe KJ, Foster NL, Herscovitch P, Karlawish JH, Rowe CC, Carrillo MC, Hartley DM, Hedrick S, Pappas V and Thies WH. Appropriate use criteria for amyloid PET: a report of the amyloid imaging task force, the society of nuclear medicine and molecular imaging, and the Alzheimer's Association. *J Nucl Med* 2013; 54: 476-490.
- [54] Cohen JP, Dong J, Lu CY and Chakravarthy R. Restricting access to florbetapir: Medicare coverage criteria for diagnostics and drugs are inconsistent. *BMJ* 2015; 351: h3333.

- [55] Rabinovici GD, Carrillo MC, Hillner BE, Siegel BA, Dilworth-Anderson P, Whitmer RA, Wilkins C, O'Bryant S, Rissman R and Gatsonis C. New IDEAS: imaging dementia-evidence for amyloid scanning study. A study to improve precision in amyloid PET coverage and patient care. Alzheimer's Association; 2020.
- [56] Harn NR, Hunt SL, Hill J, Vidoni E, Perry M and Burns JM. Augmenting amyloid PET interpretations with quantitative information improves consistency of early amyloid detection. *Clin Nucl Med* 2017; 42: 577-581.
- [57] Bateman RJ, Xiong C, Benzinger TL, Fagan AM, Goate A, Fox NC, Marcus DS, Cairns NJ, Xie X, Blazey TM, Holtzman DM, Santacruz A, Buckles V, Oliver A, Moulder K, Aisen PS, Ghetti B, Klunk WE, McDade E, Martins RN, Masters CL, Mayeux R, Ringman JM, Rossor MN, Schofield PR, Sperling RA, Salloway S and Morris JC; Dominantly Inherited Alzheimer Network. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med* 2012; 367: 795-804.
- [58] Smailovic U, Koenig T, Savitcheva I, Chiotis K, Nordberg A, Blennow K, Winblad B and Jelic V. Regional disconnection in Alzheimer dementia and amyloid-positive mild cognitive impairment: association between EEG functional connectivity and brain glucose metabolism. *Brain Connect* 2020; 10: 555-565.
- [59] Kramer J, Lueg G, Schiffler P, Vrachimis A, Weckesser M, Wenning C, Pawlowski M, Johnen A, Teuber A, Wersching H, Meuth SG and Duning T. Diagnostic value of diffusion tensor imaging and positron emission tomography in early stages of frontotemporal dementia. *J Alzheimers Dis* 2018; 63: 239-253.
- [60] Zhang M, Liu W, Huang P, Lin X, Huang X, Meng H, Wang J, Hu K, Li J, Lin M, Sun B, Zhan S and Li B. Utility of hybrid PET/MRI multiparametric imaging in navigating SEEG placement in refractory epilepsy. *Seizure* 2020; 81: 295-303.
- [61] Mendes Coelho VC, Morita ME, Amorim BJ, Ramos CD, Yasuda CL, Tedeschi H, Ghizoni E and Cendes F. Automated online quantification method for (18)F-FDG positron emission tomography/CT improves detection of the epileptogenic zone in patients with pharmacoresistant epilepsy. *Front Neurol* 2017; 8: 453.
- [62] Franceschi AM, Clifton MA, Naser-Tavakolian K, Ahmed O, Bangiyev L, Clouston S and Franceschi D. FDG PET/MRI for visual detection of crossed cerebellar diaschisis in patients with dementia. *AJR Am J Roentgenol* 2021; 216: 165-171.
- [63] De Carli F, Nobili F, Pagani M, Bauckneht M, Massa F, Grazzini M, Jonsson C, Peira E, Morbelli S and Arnaldi D; Alzheimer's Disease Neuroimaging Initiative. Accuracy and generalization capability of an automatic method for the detection of typical brain hypometabolism in prodromal Alzheimer disease. *Eur J Nucl Med Mol Imaging* 2019; 46: 334-347.
- [64] Karakaya T, Fusser F, Schroder J and Pantel J. Pharmacological treatment of mild cognitive impairment as a prodromal syndrome of Alzheimer's disease. *Curr Neuropharmacol* 2013; 11: 102-108.
- [65] Baldereschi M, Amato MP, Nencini P, Pracucci G, Lippi A, Gauthier LAS, Beatty L, Quiroga P, Klassen G, Galea A, Muscat P, Osuntokun B, Ogunniyi A, Portera-Sanchez A, Bermejo F, Hendrie H, Burdine V, Brashear A, Farlow M, Maggi S and Katzman R; WHO-PRA Age-Associated Dementia Working Group, WHO-Program for Research on Aging, Health of Elderly Program. Cross-national interrater agreement on the clinical diagnostic criteria for dementia. WHO-PRA Age-Associated Dementia Working Group, WHO-Program for research on aging, health of elderly program. *Neurology* 1994; 44: 239-242.