

# Machine Learning Model to Predict Diagnosis of Mild Cognitive Impairment by Using Radiomic and Amyloid Brain PET

Andrea Ciarmiello, MD,\* Elisabetta Giovannini, MD,\* Sara Pastorino, PhD,\* Ornella Ferrando, PhD,†  
Franca Foppiano,† Antonio Mannironi, MD,‡ Antonio Tartaglione, MD,§  
Giampiero Giovacchini, MD, PhD,\* and The Alzheimer's Disease Neuroimaging Initiative

**Purpose:** The study aimed to develop a deep learning model for predicting amnesic mild cognitive impairment (aMCI) diagnosis using radiomic features and amyloid brain PET.

**Patients and Methods:** Subjects (n = 328) from the Alzheimer's Disease Neuroimaging Initiative database and the EudraCT 2015-001184-39 trial (159 males, 169 females), with a mean age of  $72 \pm 7.4$  years, underwent PET/CT with  $^{18}\text{F}$ -florbetaben. The study cohort consisted of normal controls (n = 149) and subjects with aMCI (n = 179). Thirteen gray-level run-length matrix radiomic features and amyloid loads were extracted from 27 cortical brain areas. The least absolute shrinkage and selection operator regression was used to select features with the highest predictive value. A feed-forward neural multilayer network was trained, validated, and tested on 70%, 15%, and 15% of the sample, respectively. Accuracy, precision, F1-score, and area under the curve were used to assess model performance. SUV performance in predicting the diagnosis of aMCI was also assessed and compared with that obtained from the machine learning model.

**Results:** The machine learning model achieved an area under the receiver operating characteristic curve of 90% (95% confidence interval, 89.4–90.4) on the test set, with 80% and 78% for accuracy and F1-score, respectively. The deep learning model outperformed SUV performance (area under the curve, 71%; 95% confidence interval, 69.7–71.4; 57% accuracy, 48% F1-score).

**Conclusions:** Using radiomic and amyloid PET load, the machine learning model identified MCI subjects with 84% specificity at 81% sensitivity. These findings show that a deep learning algorithm based on radiomic data and amyloid load obtained from brain PET images improves the prediction of MCI diagnosis compared with SUV alone.

**Key Words:** aMCI, amyloid PET, neural network

(*Clin Nucl Med* 2023;48: 1–7)

Alzheimer disease (AD) is the most frequent cause of dementia and is characterized by progressive deterioration of cognitive functions, especially episodic memory. Because of its influence on the normal lives of patients and caregivers, AD has now become

a relevant medical and economic burden on society.<sup>1</sup> Mild cognitive impairment (MCI) is a transitional stage between age-related cognitive decline and AD, and the earliest clinically detectable stage of progression toward AD.<sup>2</sup> Identifying risk factors for progression from MCI to AD is critical for early therapy. Several biomarkers have recently been identified and are divided into 3 categories based on the nature of the underlying pathophysiology.<sup>3</sup> Biomarkers of fibrillary  $\beta$ -amyloid ( $\text{A}\beta$ ) deposition are high ligand retention on amyloid PET or low cerebrospinal fluid (CSF)  $\text{A}\beta_{42}$ . Biomarkers of tau pathology (neurofibrillary tangles) are elevated CSF phosphorylated tau and tau PET ligands. Biomarkers of AD-like neurodegeneration or neuronal injury are CSF total tau,  $^{18}\text{F}$ -FDG PET hypometabolism, and atrophy on structural MRI in regions characteristic of AD.<sup>3</sup>

Interest in using radiomics to diagnose and predict outcomes for several diseases including neurodegenerative disorders has been steadily increasing. Radiomics is a method that extracts many features from medical images using data characterization algorithms. Starting from acquired images, radiomics uses image segmentation, data extraction, data reduction, and data modeling. The extracted features—termed radiomic features—can uncover disease characteristics that cannot be detected by the naked eye. The hypothesis of radiomics is that the distinctive imaging features of disease forms may be useful for predicting prognosis and therapeutic response for various conditions, thus providing valuable information for personalized therapy.<sup>4</sup>

Most radiomics studies of AD have been successfully applied with MRI. For example, radiomic features of the hippocampal area have allowed researchers to accurately distinguish AD patients from normal controls (NCs).<sup>5</sup> The machine learning framework has demonstrated its capability in distinguishing stable MCI patients from progressive MCI patients.<sup>6</sup> Finally, a texture analysis study found texture differences in the corpus callosum and thalamus between brain MRI scans of AD and MCI patients.<sup>7</sup>

PET studies, in contrast, are so far more limited. Zhou et al<sup>8</sup> showed that fused MRI/FDG images increased the accuracy of the classification of AD and MCI patients compared with the clinical model. Li et al<sup>9</sup> used FDG PET to study a large sample size of patients from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database and patients from the Huashan Hospital, China. The authors extracted 168 stable radiomic features for cortical regions of interest.<sup>9</sup> The classification experiment led to maximal accuracies of 91.5%, 83.1%, and 85.9% for classifying AD versus NCs, MCI versus NC, and AD versus MCI, respectively.

High brain uptake of the amyloid ligand  $^{18}\text{F}$ -florbetaben (FBB) has been associated with memory decline in MCI and AD.<sup>10–12</sup> Textural feature analysis is a promising approach to quantitatively assess cortical uptake and was reported to be useful in assessing amyloid status.<sup>9</sup> Many parameters can be extracted from the images, and whether and which parameters are useful in improving disease prediction and in supporting clinical decisions must still be assessed.

PET studies with amyloid tracers assessing radiomics in AD are limited. Ben Bouallegue et al,<sup>13</sup> using PET and  $^{18}\text{F}$ -florbetapir in AD patients and in subjects with MCI or significant memory concern,

Received for publication March 4, 2022; revision accepted September 3, 2022. From the \*Nuclear Medicine Unit, †Medical Physics Unit, and ‡Neurology Unit, S. Andrea Hospital; and §Memory Laboratory CNS-ONLUS, La Spezia, Italy. Conflicts of interest and sources of funding: none declared.

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

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal's Web site ([www.nuclearmed.com](http://www.nuclearmed.com)).

Correspondence to: Andrea Ciarmiello, MD, Nuclear Medicine Unit, S. Andrea Hospital, Via Vittorio Veneto, 197, 19124 La Spezia, Italy. E-mail: [andrea.ciarmiello@asl5.liguria.it](mailto:andrea.ciarmiello@asl5.liguria.it)

Copyright © 2022 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0363-9762/23/4801-0001

DOI: 10.1097/RLU.0000000000004433

showed that appropriate amyloid textural and shape features predict conversion to AD with at least as good accuracy as SUV ratio (SUVr). Nemmi et al<sup>14</sup> assessed changes in the white matter concentration of <sup>18</sup>F-florbetapir in patients with AD and healthy controls. White matter histogram analysis revealed significant differences between AD and healthy subjects that were not revealed by SUVr analysis.

The study aims were 2-fold: (1) to develop a deep learning model for predicting the diagnosis of amnesic MCI (aMCI) using radiomic features and amyloid brain load; and (2) to compare the diagnostic performance of the machine learning model with SUV predictive ability.

## PATIENTS AND METHODS

### Population

The study was based on pooled data collected from the ADNI database and the previous work of Ciarmiello et al.<sup>10,11</sup> The database's investigators contributed to the design and implementation of ADNI and provided data, but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

The ADNI was launched in 2003 as a public-private partnership led by principal investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see [www.adni-info.org](http://www.adni-info.org). The study by Ciarmiello et al<sup>10,11</sup> aimed to assess the extent to which amyloid load could be predictive of clinical progression in aMCI. Subject selection procedure and the study protocol can be found in previously published articles.<sup>10,11</sup> The final population consisted of 328 subjects studied with FBB-PET imaging. Among these patients, 179 were aMCI patients and 149 were NCs. Population demographic characteristics, enrollment criteria, and imaging procedures are described in detail elsewhere.<sup>10,11,15</sup>

### PET Imaging and Preprocessing Procedures

The ADNI database and EudraCT 2015-001184-39 PET imaging procedures are described elsewhere.<sup>10,11,15</sup> Image postprocessing was performed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/software/SPM12>) implemented under MATLAB 9.9 (MATLAB R2020b, Mathworks Inc, Natick, MA).

PET images were aligned to the MNI T1 template available in SPM12.<sup>11,16</sup> All images were visually inspected by expert neuroradiologists in the native and template spaces to avoid image artifacts or misalignments.

PET data were converted to SUV by scaling each image according to the relationship between the body weight of each subject and the injected dose. The SUVr was generated by dividing all regional SUVs by the cerebellar gray matter SUV. To directly compare SUVs across imaging centers and devices, images were smoothed with a full-width at half-maximum 8 mm isotropic Gaussian kernel reported as a suitable value for reliable data harmonization.<sup>17</sup>

Consistent with previous reports, we used the brain regions considered to play a key role in the development and progression of AD.<sup>18-21</sup> The volumes of interest (VOIs) used in this analysis consisted of 27 areas encompassing frontal, parietal, temporal, occipital, cingulate gyrus, caudate nucleus, and putamen regions.

The VOIs were extracted from the automated anatomical labeling atlas<sup>22</sup> using the PickAtlas SPM toolbox.<sup>23</sup> The left and right sides of each structure were merged into a single VOI and transferred onto subject PET images to extract the mean SUVr and radiomic feature values.

### Image Analysis and Texture Feature Extraction

In recent years, machine learning and deep learning techniques have been increasingly used for whole-body and brain image analysis and in neuroimaging to improve diagnostic accuracy in several psychiatric and neurodegenerative disorders.<sup>24-27</sup> Within the aim of the present study, a deep learning model was used to predict the diagnosis of MCI using radiomic features and SUVs derived from amyloid PET images. Spatial and intensity heterogeneity of amyloid PET images were evaluated with first- and second-level textural features that were extracted as described by Vallières et al<sup>28</sup> after gray level rebinning into 64 levels via intensity histogram equalization.<sup>28-30</sup> Texture features were derived from all PET images under MATLAB R2020b using available radiomic analysis software (<https://github.com/mvallieres/radiomics/>).

Among the different available types of radiomic features, we used a gray-level run-length matrix for the texture analysis as it is reported to have low variability with different acquisition devices, reconstruction protocols, and image artifacts.<sup>31</sup>

The whole set of textural features consisted of the following 13 variables: short-run emphasis, long-run emphasis, gray-level nonuniformity, run-length nonuniformity, run percentage, low gray-level run emphasis, high gray-level run emphasis, short-run low gray-level emphasis, short-run high gray-level emphasis, long-run low gray-level emphasis, long-run high gray-level emphasis, gray-level variance, and run-length variance.

A detailed description of each of these features with acronyms and references has been reported by Vallières et al.<sup>28</sup>

### Feature Selection and Classification

We used least absolute shrinkage and selection operator (LASSO) regression with 10-fold cross-validation to reduce the features number by identifying only those with stronger classification performance.<sup>32</sup> This method is widely used in machine and deep learning to select the minimum number of features able to enhance prediction accuracy and to prevent overfitting in the model construction.<sup>33</sup> Shrink regression coefficients for LASSO were estimated using the penalty term called the L1-norm, which is calculated as the sum of the absolute coefficients. The penalty works by forcing the coefficient estimates with a minor contribution to the model to zero. The penalty term is tuned via a constant value called lambda ( $\lambda$ ). Because the performance of the model strongly depends on the value of  $\lambda$ , tuning parameters were selected using cross-validation. Cross-validation allows finding the  $\lambda$  value able to minimize the out-of-sample mean squared error of the predictions. After the LASSO regression, only features with nonzero coefficients were used as input into the neural network to classify normal and MCI subjects.

The feed-forward neural multilayer network (FNN) was used as a classifier. In recent years, several areas of medical imaging—such as oncology<sup>34</sup> and neuroscience<sup>35,36</sup>—have increasingly applied neural networks to classify clinical, genetic, or imaging patterns into disease classes and to support diagnosis prediction.

The network architecture is based on several connected levels (input, hidden, and output layers) consisting of a different number of nodes. The node number of the input layer is uniquely determined by the number of features of the input data set. The output layer neurons were selected according to the output vector dimension, which is equal to the number of patient groups.

The network architecture used in this study consisted of 2 hidden layers according to the universal approximation theorem stating that 2 layers are able to approximate any continuous function of  $n$ -dimensional input variables.<sup>37</sup> The maximum number of neurons per layer considering  $m$  output neurons to train  $N$  samples was given by  $N_{h1} = \sqrt{(m+2)N} + 2\sqrt{N/(m+2)}$  and  $N_{h2} = m\sqrt{(m+2)N}$  for the first and second hidden layers, respectively.<sup>38</sup>

**TABLE 1.** Clinical and Demographic Characteristics of the Sample

Demographic Data	All	NC	MCI	P
n	328	149	179	
Male	159	59	100	
Female	169	90	79	<0.004
Age, y	72.3 ± 7.4	70.8 ± 6.2	73.6 ± 8	<0.001
Neuropsychological battery				
MMSE	27.4 ± 2.9	27.2 ± 7.4	24.6 ± 6.5	<0.00001
CDR-SOB	1.2 ± 1.8	0 ± 0.1	2.1 ± 2	<0.00001

Performance was evaluated at each iteration with a 10-fold cross-validation scheme<sup>39</sup> in terms of mean squared error, and the model providing the lowest error was chosen for data analysis.

The network setting included the Levenberg-Marquardt backpropagation learning algorithm, with a sigmoid tangent transfer function for input and hidden layers and a sigmoid logistic function in the output layer. Network training was performed with a learning rate parameter of  $1.0e^{-02}$  for 30 epochs.

The LASSO logistic regression and neural network analysis were performed using the Deep Learning Toolbox running under MATLAB 2020b.

## Statistics

Continuous data were tested using independent *t* tests, with degrees of freedom adjusted for inequality of variance where appropriate. The  $\chi^2$  analysis was used for categorical variables.

Receiver operating characteristic (ROC) curves with the corresponding area under the curve (AUC) were used to assess the performance of model classification. In addition, the diagnostic performance of the SUV was evaluated with the ROC curve and the AUC, determining the optimal cutoff value based on metabolic parameters.

Significant differences between AUCs computed using different prediction models were observed using the Hanley and McNeil method.<sup>40</sup>

Accuracy, precision, F1-score, and AUC were used to assess the diagnostic performance and were reported for both deep learning model and SUV classification.

The differences between the mean scores of neuropsychological tests across groups classified with deep learning—selected radiomic features were determined with 1-way analysis of variance (ANOVA). Odds ratios and their 95% confidence intervals, positive predictive value, and negative predictive value were calculated to estimate how strongly the predicted diagnosis was associated with clinical status.

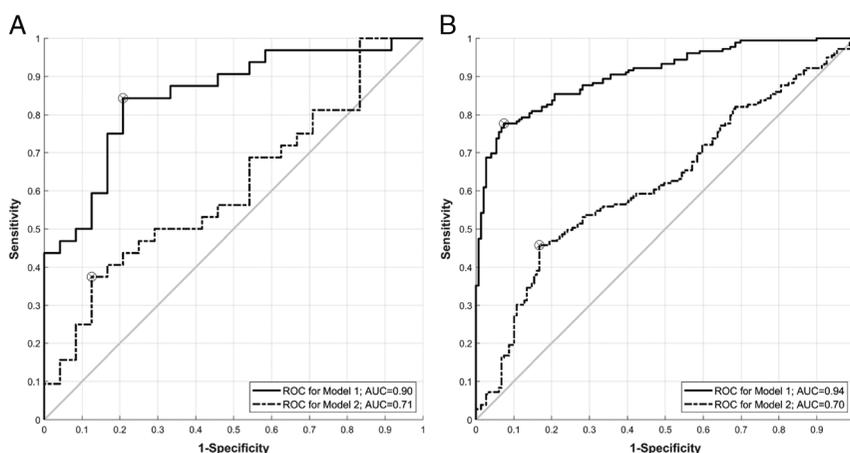
The assumption of normality was tested using the Kolmogorov-Smirnov statistic. Two-tailed *P* values of less than 0.05 were considered statistically significant.

## RESULTS

Demographic and clinical data are summarized in Table 1. The study sample consisted of 328 subjects with an average age of  $70.8 \pm 6.2$  years for healthy controls and  $73.6 \pm 8$  years for patients with MCI. The overall percentage of men was 48% (159 of 328). As expected, the study population included a slightly higher percentage of women (Table 1). The mean age of MCI subjects was significantly higher than for NCs ( $P = 0.001$ ). Predictably, MCI individuals had greater cognitive deficits than controls on Mini-Mental State Examination (MMSE) and Clinical Dementia Rating Scale Sum of Boxes (CDR-SOB) (Table 1).

Thirteen textural features and SUVrs were extracted from each of the 27 brain regions selected for this study. The complete set consisted of 351 textural and 27 metabolic variables for a total of 378 predictors. All variables were processed with the LASSO regression method to select only those with higher predictive values. Feature dimension reduction via the LASSO algorithm resulted in a predictors drop from 378 to 61, of which 34 were textural and 27 were metabolic variables (Supplemental Figure S1, Supplemental Digital Content, <http://links.lww.com/CNM/A390>; Supplemental Table S1, Supplemental Digital Content, <http://links.lww.com/CNM/A390>).

The network model used for this article consists of an input buffer with 63 input nodes, 2 hidden layers with, respectively, 24 and 8 nodes and 2 output nodes. The input nodes consist of 61 features identified by the LASSO regression plus age and gender. The output nodes represent the clinical diagnosis (Supplemental Figure S3,



**FIGURE 1.** Receiver operating characteristic curves of FNN model trained on 70% of the study set, validated on 15%, and tested on the remaining 15% of the study sample. **A**, ROC curves of trained deep learning model tested on the 15% of sample. ROC curve labeled model 1 represents the model performance for distinguishing MCI from NCs based on radiomic features and SUV used as FNN input. Model 2 represents performance based on SUV only. **B**, ROC curves showing model performance tested on all cases. Model 1 performs statistically significantly better at recognizing MCI subjects from healthy controls in test set as well as on overall sample.

**TABLE 2.** Performance Comparison of Machine Learning Algorithm and SUV

Parameter	Model 1*	Model 2†	Hanley P
Accuracy (%)	80 (70.0–90.8)	57 (44.2–70.1)	
Precision (%)	74 (62.6–85.6)	50 (36.9–63.1)	
F1-score (%)	78 (67.7–89.2)	48 (34.7–60.9)	
AUC (%)	90 (89.4–90.4)	71 (69.7–71.4)	0.02

Probability of belonging to normal or patient class, determined with SUV and radiomic features (\*) or SUV alone (†). Hanley test indicates significant difference of deep learning AUC compared with SUV AUC on independent test set.

Supplemental Digital Content, <http://links.lww.com/CNM/A390>). Model performance in terms of mean squared error for train, validation, and test data set is showed in Supplemental Figure S2, Supplemental Digital Content, <http://links.lww.com/CNM/A390>. Model diagnostic performance evaluated on AUCs was 94%, 91%, and 90% for train, validation, and test set, respectively (Hanley  $P > 0.05$ ).

The deep learning algorithm was trained on 70% of the study set, validated on 15%, and tested on the remaining 15% of the study sample (Supplemental Figure S2, Supplemental Digital Content, <http://links.lww.com/CNM/A390>). Figure 1 shows the ROC curves obtained on the test set and on the whole population using the FNN model.

The AUC for the prediction of MCI was 90% on the test set and 94% on the whole sample. These findings indicate that the neural network model can distinguish MCI subjects from NCs. As shown in Table 2, the accuracy, precision, and F1-score were 80%, 74%, and 78%, respectively.

The ROC curves obtained from SUVr alone are shown in Figure 1. The AUC for prediction of MCI from SUV was 71% on the test set and 70% on the whole sample. Using SUVr, the accuracy, precision, and F1-score were 57%, 50%, and 48%, respectively (Table 2).

The Hanley test, used to measure the significance of the difference between the AUCs obtained with the FNN and the SUVr, resulted in a  $P = 0.015$ . These findings showed that the differences are statistically significant and that FNN performs better at differentiating MCI subjects from healthy controls in the test set and overall sample (Table 2).

Table 3 indicates the comparison between aMCI patients and NCs based on SUVr and the network classification. Feed-forward neural multilayer network classifier identified individuals with

cognitive impairment with a sensitivity of 81% and a specificity of 84%, as compared with NCs ( $\chi^2 = 138$ ;  $P < 0.000001$ ). The positive and negative predictive values were 81% and 84%, respectively (Table 3), whereas SUV-based classification identified aMCI subjects with a sensitivity of 73% and specificity of 50% ( $\chi^2 = 19$ ;  $P < 0.0003$ ). All statistical measures considered— $\chi^2$ , specificity, sensitivity, odds ratio, positive predictive values, and negative predictive values—indicated a higher statistical significance for the neural network classifier than the SUV-only classification.

Differences between mean MMSE and the CDR-SOB scores were assessed in the predicted classes by 1-way ANOVA (Figs. 2, 3).

The classes of normal and MCI-affected subjects were predicted using the score derived from the machine learning model. To compare the performance of the neural model with the classification derived from the SUVs, the optimal threshold value predefined with the ROC analysis of the SUV (Youden index = 1.37) was used.

Analysis of results showed a significant difference in MMSE ( $28.9 \pm 1.5$  vs  $26.2 \pm 3.2$ ,  $P = 2.5e^{-17}$ ) and CDR-SOB ( $0.35 \pm 1.0$  vs  $1.98 \pm 2.0$ ,  $P = 6.0e^{-16}$ ) scores between positive and negative groups, as predicted by the deep learning model. Classification based on the SUV also showed a significant difference between the means of neurocognitive test scores of the sample classes, although the  $F$  statistic showed less significance. Table 4 provides the results of the ANOVA analysis for subject classification predicted with both models.

## DISCUSSION

This study assessed the role of FNN and the use of radiomic features extracted by amyloid PET images in improving the prediction of cognitive deficits. Overall, results confirmed that radiomic features (objective measures of tissue heterogeneity) may be useful in reliably predicting cognitive function deficits in aMCI patients. In particular, the analyses identified 27 radiomic features and SUVr's as the optimal variable set providing the best predictive, combined (textural and nontextural) model to distinguish between subjects with and without cognitive deficits. Specificity and sensitivity had higher statistical significance for the FNN outcome than for SUVr prediction (Table 3).

The diagnostic performance of FNN evaluated as the area under the ROC curve compares favorably with the results obtained in other radiomic studies in patients with cognitive impairment.<sup>36–38</sup> The SUVr-based diagnostic performance assessed in our sample was similar to that reported by other authors.<sup>39,40</sup> Note, however, that direct comparisons of AUC accuracy may be biased by variables

**TABLE 3.** Bivariate Analysis of Predicted Status on Clinical Diagnosis

Prediction Model	Predicted Class	True Class		$\chi^2$	SP (95% CI)	SE (95% CI)	OR (95% CI)	PPV (95% CI)	NPV (95% CI)
		NC	MCI						
Model 1*	NC	121 (81%)	29 (16%)	138‡	84 (78–89)	81 (75–87)	22.35 (12.62–39.6)	81 (74–87)	84 (79–90)
	MCI	28 (19%)	150 (84%)						
Model 2†	NC	109 (55%)	89 (68%)	19§	50 (43–58)	73 (66–80)	2.76 (1.73–4.39)	55 (48–62)	69 (61–77)
	MCI	40 (20%)	90 (69%)						

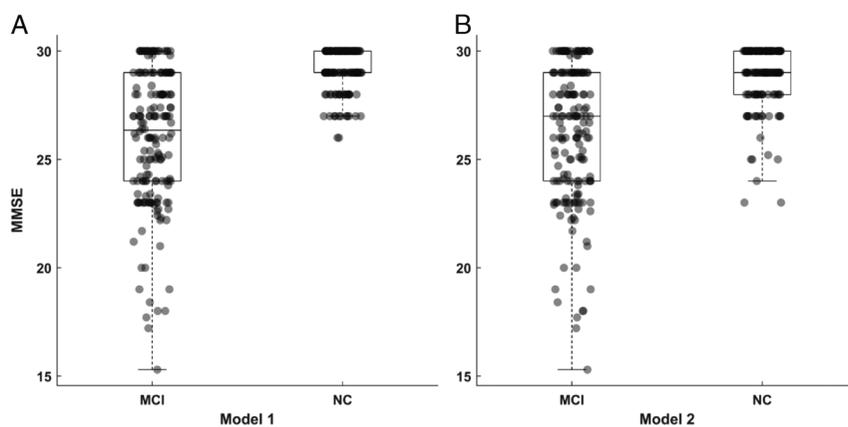
\*Probability of belonging to normal or patient class, determined with FNN using radiomic features plus SUV as input.

†SUV prediction was based on 1.3 threshold value determined in previous articles. Both estimates included sex and age in the prediction model.

‡ $P < 0.000001$ .

§ $P < 0.0003$ .

SP, specificity; SE, sensitivity; OR, odds ratio; PPV, positive predictive value; NPV, negative predictive value.



**FIGURE 2.** Box plot of MMSE in MCI and NC predicted classes. Box plot showing MMSE values for predicted classes of MCI and NC. Box indicates median and interquartile range. Whiskers above and below the boxes indicate the 90th and 10th percentiles. Model 1 = radiomic features plus SUV, model 2 = SUV. Both prediction models included sex and age.

such as the severity of cognitive deficits, the radiotracer, and variables included in the model.

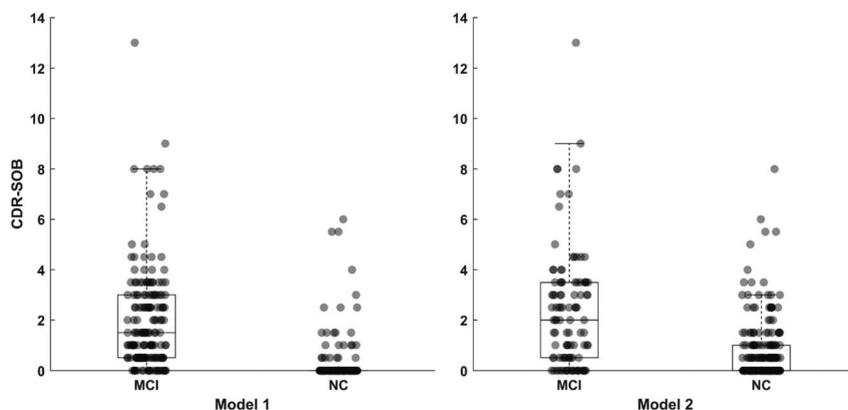
Moreover, findings support that the combined textural and nontextural model significantly predicted performance in cognitive tests, with a statistical significance that was similar or slightly higher than that of the model based only on SUVr. This finding has important clinical implications, supporting the usefulness of radiomic features in improving group stratification based on imaging techniques. Higher cortical SUVr in aMCI patients versus NCs is a consistent finding.<sup>10,11,41</sup>

Our results are consistent with and add new information to the results of 2 previous studies with PET and amyloid tracers. In 2015, Shokouhi et al<sup>42</sup> found a higher inverse correlation between the CSF amyloid- $\beta$  and a weighted 2-point correlation function derived from <sup>11</sup>C-PiB PET images—a measurement providing detailed information about image clustering—than between the CSF amyloid- $\beta$  and SUVr mean or median in NCs and MCI patients. More recently, Ben Bouallegue et al<sup>13</sup> showed that the variables with the highest prognostic value were composite SUVr, skewness, local minima, Geary's index, gradient norm maximal argument, and the support vector machine model, which somehow had higher accuracy. The novelty of our study lies in the combination of the radiopharmaceutical—the amyloid tracer FBB—with the data analysis

system or neural network. Neural networks are some of the most commonly used statistical models and have been successful in biomarker discovery studies in AD.<sup>43</sup> However, to the best of our knowledge, the FNN had never been used to differentiate aMCI patients from NCs, nor were they correlated with cognitive deficits.

In our study, the association between textural features and patients' clinical diagnosis was more significant than that observed for the SUVr metric (Table 3). This finding suggests that the spatial arrangement of similar signal intensity voxels can have a strong predictive value. Indeed, the association between low gray-level textural features with the clinical diagnosis suggests that contiguous regions with low signal intensity in FBB-PET images may predict cognitive impairment. Moreover, features measuring rapid gray-level changes in gray-level intensity within the target region are significantly related to clinical status and appear to reflect tracer uptake heterogeneity rather than the signal increase observed in cognitively impaired subjects.

In other studies, the radiomic model included a higher number of parameters showing that adding multiple variables in a study with an adequate sample size increases the accuracy of the model.<sup>44-46</sup> For example, Zhang et al<sup>45</sup> obtained an accuracy of 93.2% classifying AD from healthy controls when combining 3 modalities of biomarkers (MRI, FDG PET, and CSF biomarkers), compared with only 86.5%



**FIGURE 3.** Box plot of CDR-SOB in MCI and NC predicted classes. Box plot showing CDR-SOB values for predicted classes MCI and NC. Box indicates median and interquartile range. Whiskers above and below the boxes indicate the 90th and 10th percentiles. Model 1 = radiomic features plus SUV, model 2 = SUV. Both prediction models included sex and age.

**TABLE 4.** Analysis of Variance Across MMSE and CDR-SOB in Subjects Groups Classified According to SUV and Radiomic Features Plus SUV Prediction

Test	Classification	NC	MCI	F	P
CDR-SOB	Model 1*	0.35 ± 1.02	1.98 ± 2.04	73	6.0e <sup>-16</sup>
	Model 2†	0.72 ± 1.24	2.34 ± 2.35	62	6.7e <sup>-14</sup>
MMSE	Model 1*	28.87 ± 1.48	26.2 ± 3.23	81	2.5e <sup>-17</sup>
	Model 2†	28.21 ± 2.11	25.71 ± 3.55	60	1.4e <sup>-13</sup>

\*Probability of belonging to normal or patient class was determined with FNN using radiomic features plus SUV as input.

†Model 2 reports SUV prediction based on 1.3 threshold value determined in previous articles.

when using the best individual modality of biomarkers. Similarly, the authors achieved a classification accuracy of 76.4% in distinguishing MCI and healthy controls using the 3 modalities of biomarkers, compared with only 72% using the best individual modality of biomarkers.<sup>45</sup> Young et al<sup>44</sup> also showed that the prediction of MCI conversion based on the combination of MRI, FDG PET, and CSF was substantially more accurate than any other modality alone.

Given the need for large sample sizes (generally consisting of several hundred subjects) to develop and assess predictive models, our analyses were performed on a total sample of 328 subjects undergoing FBB-PET, including 149 NC and 179 aMCI patients. Indeed, as a rough indication, at least an estimated 10 to 15 patients must be included in a radiomic study for each feature when statistical tests are used to test a specific hypothesis.<sup>47</sup> We used a total of 27 variables selected by the LASSO method for the permutation. Therefore, our sample appears to have been of adequate size for the desired statistical analysis.

Because of the transversal design of the study, we were able to compare radiomic features in 2 groups that differed in their cognitive profiles but were not able to assess whether radiomic features also predict cognitive decline. Choi and Jin<sup>48</sup> developed an automatic image interpretation system based on a deep convolutional neural network that accurately predicted future cognitive decline in MCI patients using FDG and <sup>18</sup>F-florbetapir PET. Receiver operating characteristic analysis revealed that the performance of the neural network-based approach was significantly superior to that of conventional quantification methods. Output scores of the network were significantly correlated with the neuropsychological measures of cognitive impairment.<sup>48</sup>

## CONCLUSIONS

In summary, a neural network is a promising approach for predicting diagnosis and cognitive performance in aMCI patients. The best predictive model was composed of SUVr and 34 textural features. Adding textural features to the prediction model provided by SUVr significantly increased the overall prediction of the AUC in all cerebral regions. The combined model significantly predicted performance in cognitive tests and had a higher statistical significance than the model based only on SUVr. Because a single biomarker is unlikely to improve the diagnosis of AD, machine learning by analyzing multiple parameters extracted from images and derived from a clinical setting will support the identification of a panel of markers capable of ensuring earlier and more accurate diagnosis.

## REFERENCES

- Reitz C, Mayeux R. Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochem Pharmacol.* 2014;88:640–651.

- Markesbery WR. Neuropathologic alterations in mild cognitive impairment: a review. *J Alzheimers Dis.* 2010;19:221–228.
- Jack CR Jr, Bennett DA, Blennow K, et al. A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology.* 2016;87:539–547.
- Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp.* 2018;2:36.
- Feng F, Wang P, Zhao K, et al. Radiomic features of hippocampal subregions in Alzheimer's disease and amnesic mild cognitive impairment. *Front Aging Neurosci.* 2018;10:290.
- Moradi E, Pepe A, Gaser C, et al. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage.* 2015;104:398–412.
- Zhang J, Yu C, Jiang G, et al. 3D texture analysis on MRI images of Alzheimer's disease. *Brain Imaging Behav.* 2012;6:61–69.
- Zhou H, Wang J, Lu J, et al. Dual-model radiomic biomarkers predict development of mild cognitive impairment progression to Alzheimer's disease. *Front Neurosci.* 2018;12:1045.
- Li Y, Jiang J, Lu J, et al. Radiomics: a novel feature extraction method for brain neuron degeneration disease using (18)F-FDG PET imaging and its implementation for Alzheimer's disease and mild cognitive impairment. *Ther Adv Neurol Disord.* 2019;12:1756286419838682.
- Ciarmiello A, Giovannini E, Riondato M, et al. Longitudinal cognitive decline in mild cognitive impairment subjects with early amyloid-beta neocortical deposition. *Eur J Nucl Med Mol Imaging.* 2019;46:2090–2098.
- Ciarmiello A, Tartaglione A, Giovannini E, et al. Amyloid burden identifies neuropsychological phenotypes at increased risk of progression to Alzheimer's disease in mild cognitive impairment patients. *Eur J Nucl Med Mol Imaging.* 2019;46:288–296.
- Florek L, Tiepolt S, Schroeter ML, et al. Dual time-point [<sup>18</sup>F]Florbetaben PET delivers dual biomarker information in mild cognitive impairment and Alzheimer's disease. *J Alzheimers Dis.* 2018;66:1105–1116.
- Ben Bouallegue F, Vauchot F, Mariano-Goulart D, et al. Diagnostic and prognostic value of amyloid PET textural and shape features: comparison with classical semi-quantitative rating in 760 patients from the ADNI-2 database. *Brain Imaging Behav.* 2019;13:111–125.
- Nemmi F, Saint-Aubert L, Adel D, et al. Insight on AV-45 binding in white and grey matter from histogram analysis: a study on early Alzheimer's disease patients and healthy subjects. *Eur J Nucl Med Mol Imaging.* 2014;41:1408–1418.
- ADNI. Defining Alzheimer's Disease: Procedure Manual. 2008. Available at: <http://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf>. Accessed March 17, 2020.
- Koole M, Nelissen N, Khela M, et al. Template comparison for spatial normalization of [<sup>18</sup>F]flutemetamol brain PET images. *J Nucl Med.* 2010;51:571–571.
- Tsutsui Y, Awamoto S, Himuro K, et al. Characteristics of smoothing filters to achieve the guideline recommended positron emission tomography image without harmonization. *Asia Ocean J Nucl Med Biol.* 2018;6:15–23.
- Barthel H, Gertz HJ, Dresel S, et al. Cerebral amyloid-β PET with florbetaben (<sup>18</sup>F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. *Lancet Neurol.* 2011;10:424–435.
- Ong K, Villemagne VL, Bahar-Fuchs A, et al. <sup>18</sup>F-florbetaben Aβ imaging in mild cognitive impairment. *Alzheimers Res Ther.* 2013;5:4.
- Tiepolt S, Hesse S, Patt M, et al. Early [(18)F]florbetaben and [(11)C]PiB PET images are a surrogate biomarker of neuronal injury in Alzheimer's disease. *Eur J Nucl Med Mol Imaging.* 2016;43:1700–1709.
- Pfeil J, Hoenig MC, Doering E, et al. Unique regional patterns of amyloid burden predict progression to prodromal and clinical stages of Alzheimer's disease. *Neurobiol Aging.* 2021;106:119–129.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage.* 2002;15:273–289.
- Maldjian JA, Laurienti PJ, Kraft RA, et al. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage.* 2003;19:1233–1239.
- Mateos-Perez JM, Dadar M, Lacalle-Aurioles M, et al. Structural neuroimaging as clinical predictor: a review of machine learning applications. *Neuroimage Clin.* 2018;20:506–522.
- Chen CL, Liang CK, Yin CH, et al. Effects of socioeconomic status on Alzheimer disease mortality in Taiwan. *Am J Geriatr Psychiatry.* 2020;28:205–216.

26. Kim JY, Oh D, Sung K, et al. Visual interpretation of [(18)F]Florbetaben PET supported by deep learning-based estimation of amyloid burden. *Eur J Nucl Med Mol Imaging*. 2020;48:1116–1123.
27. Choi H, Kim YK, Yoon EJ, et al. Cognitive signature of brain FDG PET based on deep learning: domain transfer from Alzheimer's disease to Parkinson's disease. *Eur J Nucl Med Mol Imaging*. 2020;47:403–412.
28. Vallières M, Freeman CR, Skamene SR, et al. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol*. 2015;60:5471–5496.
29. Hatt M, Majdoub M, Vallières M, et al. <sup>18</sup>F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med*. 2015;56:38–44.
30. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Phys Med Biol*. 2016;61:R150–R166.
31. Carles M, Fechter T, Marti-Bonmati L, et al. Experimental phantom evaluation to identify robust positron emission tomography (PET) radiomic features. *EJNMMI Phys*. 2021;8:46.
32. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc B Methodol*. 1996;58:267–288.
33. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1–22.
34. Landhuis E. Deep learning takes on tumours. *Nature*. 2020;580:551–553.
35. Savage N. How AI and neuroscience drive each other forwards. *Nature*. 2019;571:S15–S17.
36. Pedersen M, Verspoor K, Jenkinson M, et al. Artificial intelligence for clinical decision support in neurology. *Brain Commun*. 2020;2:fcaa096.
37. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput*. 2006;18:1527–1554.
38. Huang GB. Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Trans Neural Netw*. 2003;14:274–281.
39. Stone M. Cross-validated choice and assessment of statistical predictions (with discussion). *J R Stat Soc B Methodol*. 1976;38:102–102.
40. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148:839–843.
41. Kim JY, Lim JH, Jeong YJ, et al. The effect of clinical characteristics and subtypes on amyloid positivity in patients with amnesic mild cognitive impairment. *Dement Neurocogn Disord*. 2019;18:130–137.
42. Shokouhi S, Rogers BP, Kang H, et al. Modeling clustered activity increase in amyloid-beta positron emission tomographic images with statistical descriptors. *Clin Interv Aging*. 2015;10:759–770.
43. Zafeiris D, Rutella S, Ball GR. An artificial neural network integrated pipeline for biomarker discovery using Alzheimer's disease as a case study. *Comput Struct Biotechnol J*. 2018;16:77–87.
44. Young J, Modat M, Cardoso MJ, et al. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin*. 2013;2:735–745.
45. Zhang D, Wang Y, Zhou L, et al. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2011;55:856–867.
46. Madabhushi A, Agner S, Basavanthally A, et al. Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data. *Comput Med Imaging Graph*. 2011;35:506–514.
47. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278:563–577.
48. Choi H, Jin KH. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav Brain Res*. 2018;344:103–109.