Alzheimer's Disease Prediction Using Long Short-Term Memory with Early-Phase ^{18}F-Florbetaben Imaging Data

Hyeon Kang
Institute of Convergence BioHealth, Dong-A University, Busan, Republic of Korea
Kyung Won Park
Department of Neurology, Dong-A University College of Medicine, Busan, Republic of Korea
Do-Young Kang (dykang@dau.ac.kr)
Department of Nuclear Medicine, Dong-A University College of Medicine, Busan, Republic of Korea

Research Article

Keywords: Alzheimer's disease, amyloid-β, blood perfusion, functional neuroimaging, machine learning, neural network

Posted Date: February 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-242396/v1

Abstract

Single amyloid-beta (Aβ) imaging test is not enough to rise to the challenge of making AD diagnosis because of Aβ-negative AD or positive cognitively normal (CN). We aimed to distinguish AD from CN with dual-phase ^{18}F-Florbetaben (FBB) via machine learning algorithms and evaluate the AD positivity scores compared to delay-phase FBB (dFBB) which is currently adopted for AD diagnosis.

A total of 264 patients (74 CN and 190 AD), who underwent FBB imaging test and neuropsychological tests were retrospectively analyzed. We compared three kinds of machine learning-based models and evaluated their performance with 4-fold cross validation.

AD positivity scores estimated from dual-phase FBB showed better accuracy (ACC) and area under the receiver operating characteristic curve (AUROC) for AD detection (ACC: 84.091 %, AUROC: 0.900) than those from dFBB imaging (ACC: 81.364 %, AUROC: 0.890). The association between predicted AD positivity and the AD occurrence were compared, the use of dual-phase FBB was highest (OR: 56.333), followed by dFBB (OR: 35.182).

These results show that the combined model which interpret dual-phase FBB with long short-term memory can be used to provide a more accurate AD positivity score, which shows a closer association with AD, than the prediction with only single-phase FBB.

Introduction

Approximately 50 million people worldwide suffer from dementia, and nearly 10 million new cases occur every year. The total population with such dementia is expected to be 82 million by 2030 and 152 million by 2050[1]. Alzheimer's disease (AD), the most common cause of dementia, is complex and multi-factorial in elucidating the continuum of conditions leading to asymptomatic, mild cognitive impairment, and dementia. Amyloid-β (Aβ), which can be measured through positron emission tomography (PET) scan or cerebrospinal fluid analysis, is one of those defining the pathology of AD, and is known as the earliest sign among AD biomarkers. Therefore, Aβ-related biomarkers have studied for a clinical diagnostic index as well as for early diagnosis or prediction[2-4]. However, as AD is known to be affected by neurofibrillary tangles aggregated by hyperphosphorylated tau protein, genetics, and environmental influences as well[5], both Aβ-negative AD and Aβ-positive CN inevitably exist[6]. In addition, it is difficult to monitor the patient's condition because Aβ plaques are already saturated by the time cognitive function clinically declines[7]. These facts remind us how an additional AD biomarkers is required to understand and respond to AD. 18F-Fluorodeoxyglucose (FDG), which is radiopharmaceutical that enables imaging of changes in glucose metabolism in brain tissue, is another one of the representative AD biomarkers. The hypometabolism, which is measured using FDG-PET, is known to be associated with neurodegeneration and cognitive decline[8]. However, such a series of PET imaging tests have drawbacks that make patients who need a diagnosis or longitudinal studies for AD undergo relatively frequent radiation exposure and high financial expenditure.

Aβ uptake in early phase Aβ-PET is known to be a potential perfusion imaging modality that reflects cerebral blood flow[9-11]. Reference[4] reviewed the coupled relationship between hypoperfusion that causes deleterious changes in neurons and cerebral hypometabolism that underlies neuronal/synaptic dysfunction with the respective associations with cognitive impairment. Given an adequate evaluation of neuronal function and Aβ load from dual-phase Aβ-PET imaging, we may be able to provide patients with a more accurate AD diagnosis and prognostic evaluation without compromising patient convenience. Compared to delay phase Aβ-PET, however, there is no consensus or a well-established guide regarding how to interpret and evaluate the potential perfusion imaging for AD.

In the field of imaging biomarkers, various efforts have been made to provide an improved quality of medical services continuously. In particular, the latest technologies incorporating artificial intelligence have been reported to show a consistent inference and classification performance comparable to human doctors. Such technologies are excellent at not only reducing a portion of labor for human doctors but also addressing inter-observation problems[12, 13]. In addition, machine learning-based studies on imaging for AD biomarkers are also actively reported[14]. Existing machine learning-based studies for AD have commonly suggested a predictive model that learns single or more than two kinds of imaging data such as magnetic resonance imaging, FDG, or Aβ-PET. Those attempts using a variety of information for AD detection could be appropriate designs that address the complex and heterogenous characteristics of AD.

In this study, we attempted to develop and evaluate an improved imaging biomarker in the machine learning algorithm by engaging with dynamic early phase Aβ-PET as well as single delay phase Aβ-PET conventionally used for AD diagnosis. The method included (1) extracting the mean of the standard uptake value ratio (SUVr) with a consistent area and reference region from individual dual-phase Aβ-PET imaging, (2) developing a machine learning-based predictive model, such as logistic regression (LR), support vector machine (SVM), and neural network (NN), which estimates the AD positivity score, and (3) comparing the classification performance among models and evaluating the association between predicted AD positivity scores and cognitive function or occurrence of AD.

Materials And Methods

Participants

We adopted FBB PET as Aβ-PET for this experiment and retrospectively recruited subjects who visited the Department of Neurology and Nuclear Medicine of the Dong-A University Hospital (DAUH) and underwent dual-phase FBB from November 2015 to June 2020. The total number of subjects was 264, consisting of 74 cognitive normal (CN) and 190 AD. Detailed demographic data of the participants are presented in Figure 1. All CN cases had normal age-, gender-, and education-adjusted performance on standardized cognitive tests. The AD participants met the following inclusion criteria: 1) criteria for dementia according to the Diagnostic and Statistical Manual of Mental Disorders 4th Edition (DSM-IV-TR)[15], and 2) the criteria for probable AD according to the NIA-AA core clinical criteria[16]. The individual FBB PET imaging for Aβ load was visually evaluated by brain Aβ plaque load (BAPL) scoring system, which defines a BAPL score

of 1 (no Aβ load), 2 (minor Aβ load), and 3 (significant Aβ load)[17]. Dong-A University Hospital Institutional Review Board (DAUHIRB) reviewed this study with the member who participated in Institutional Review Board Membership List and finally approved this study protocol (DAUHIRB-17-108). All procedures for data acquisition were in accordance with the ethical standards of DAUHIRB with 1964 Helsinki declaration and its later amendments or comparable ethical standards. We guarantee that informed consent was obtained from all participants for this study.

PET acquisition

All FBB PET imaging for this experiment was performed using a Biograph 40mCT Flow PET/CT scanner (Siemens Healthcare, Knoxville, TN, USA) and reconstructed through UltraHD-PET (TrueX-TOF). An FBB dose of 300 MBq was administered to the participants as an intravenous bolus injection from 0 to 20 min and from 90 to 120 min post-injection after helical CT with a 0.5 sec rotation time at 100 kVp and 228 mAs. The image acquisition time for dual-phase FBB PET was determined by related studies to sufficiently include the peak of Aβ uptake for early phase FBB PET (eFBB) and the manufacturer's recommendations and for delay-phase FBB PET (dFBB)[10, 17, 18]. The acquired dynamic eFBB and static dFBB was 27 frames of 128 × 128 × 110 (3.19 mm × 3.19 mm × 1.5 mm) resliced from a field of view of 408 mm × 408 mm × 165 mm, and one frame of 400 × 400 × 110 (1.02 mm × 1.02 mm × 1.5 mm) resliced from a field of view of 408 mm × 408 mm × 165 mm, respectively. Static eFBB to evaluate potential perfusion was made by averaging the frames corresponding to 2–7mins from dynamic eFBB. The optimal time period required to obtain static eFBB was internally determined using the approach in Reference[10].

Pre-processing

The pre-processing procedures applied to extract regional mean SUVr for dynamic eFBB or static eFBB/dFBB, respectively, were as follows. For the spatial normalization of all PET images, we used an in-house eFBB PET template[19], which averaged 8 CN and 8 AD randomly selected from the collective spatially normalized FBB data pool in Montreal Neurological Institute (MNI) space[20]. Each static eFBB was spatially non-linearly registered to the template space. For dynamic eFBB, the deformation field corresponding to the template space from the mean of the total number of frames was used. The deformation fields for a dFBB were identical to those estimated in the spatial normalization process of the matched static eFBB[21]. As a result, the spatially registered imaging was in a voxel space of 95 × 79 × 68 (height × width × depth). We merged Hammers atlas[22] into nine representative regions (frontal lobe, temporal lobe, parietal lobe, occipital lobe, posterior cingulate, caudate, putamen, and thalamus) for the reference region for count normalization and volume of interest for estimating the mean SUVr. After spatial normalization, the intensities of each image were normalized with respect to the mean uptake of the whole cerebellar region as a reference region. Finally, for dynamic eFBB and static eFBB/dFBB, regional mean SUVr of 8 × 1 and regional time-activity curve (TAC) data of 8 × 27 (number of target regions × temporal length) were obtained, respectively.

Calculation of AD positivity score based on brain blood flow and amyloid-β plaque

To calculate the AD positivity score, we aimed to build a proper machine learning-based classification model to predict the probability of whether the given regional TAC or mean SUVr data belong to the CN or AD distribution. Fig. 1 shows the structure of the adopted models for this study. Fig. 1a is a combined model that predicts AD using both dynamic eFBB and dFBB extracted from dual-phase FBB, and Fig. 1b shows the structure of the model using only one of either eFBB or dFBB (single-phase FBB). The combined model ($NN_{aggregated}$) in Fig. 1a consists of long short-term memory ($LSTM_{eFBB}$) model, which is used to extract temporal features such as dynamic eFBB, feedforward neural network (NN_{dFBB}) for dFBB, and last NN to make a final diagnosis decision (NN_{Dx}) using the concatenated features (eFBB and dFBB) delivered from the preceding layers. Each $LSTM_{eFBB}$ and NN_{dFBB} and NN_{Dx} have an output layer with two nodes leading to the softmax function to interpret the model output as the probability for diagnostic label, and their model parameters were trained to minimize the cross-entropy loss between the predicted probability and one-hot encoded actual label. To train the weights of each internal model ($LSTM_{eFBB}$, NN_{dFBB}, and NN_{Dx}) constituting the combined model, we alternately trained the internal models in a 1:1:2 ratio of epochs. The models depicted in Fig. 1b are a logistic regression (LR) and support vector machine (SVM), or an NN-based model, and were used as a baseline to evaluate the feasibility of the combined model that interprets dual-phase FBB.

Detailed parameters for model selection and model evaluation

For the purpose of this experiment, we focused on showing that the model with dual-phase FBB is more useful for estimating AD positivity than a model with only dFBB. Therefore, we tried to simplify and unify the model structure and detailed parameters of each model as much as possible. NN-based models, including LSTM, have two hidden layers, eight nodes of each hidden layer, and L2 regularization was applied with a weight of 0.01. The learning curves of all models were set to be trained up to 10000 epochs but were stopped if the validation loss was not updated more than 200 times. The learning rate was 0.0005, and the Adam optimizer[23] was used for each setting. If the validation loss was not updated more than 100 times at a point, 0.001 of decay rate was applied to the learning rate of the point.

The SVM used in the experiment used a linear kernel as a kernel function. A radial basis function or polynomial kernel was also tested in an internal experiment but no meaningful difference was observed, and a simpler model was finally adopted to prevent overfitting.

The software used in this experiment was the SPM12 library and MATLAB R2020a for the data pre-processing, including spatial normalization, count normalization, and for calculating regional mean SUVr based on the Hammers atlas[22]. Keras 2.2.4 library and Python 3.6.9 were used to select and evaluate a model for estimating AD positivity. The experimental tool was implemented and tested on Linux Ubuntu 16.04 LTS with an Intel Core i7-6800K CPU and two GPUs (NVIDIA GeForce GTX 1080).

TABLE I
DEMOGRAPHICS OF EXPERIMENTAL DATA WITH DUAL-PHASE F18-FLORBETABEN IMAGING

Variables	CN	AD	Total	p-value
#	74	190	264	N/A
Sex (F/M)	49/25	102/88	151/113	0.065
Age	70.24 ± 7.41	71.95 ± 8.84	71.47 ± 8.48	0.142
FBB reading (Aβ (-) / Aβ (+))	59/15	32/158	91/173	<0.001
Education (y)	9.30 ± 4.09	10.10 ± 4.48	9.88 ± 4.39	0.188
K-MMSE	27.50 ± 1.71	19.52 ± 4.25	21.62 ± 5.14	<0.001

CN: Cognitively Normal; AD: Alzheimer's Disease; K-MMSE: Korean version of Mini Mental State Examination.

For model selection and evaluation, we used a nested cross-validation framework[24, 25]. In situations where there is a limited dataset, nested cross-validation could be an appropriate alternative when it is difficult to divide the dataset into training, validation, and testing sets. Its inner loop achieves model selection through a predefined hyper-parameter search space and wrapper function, and the outer loop is used to measure the generalization performance of the selected model. Although we used 4-fold cross-validation, instead of using the wrapper function, with the dataset corresponding to the inner loop, we manually tried to find a unified hyper-parameter that the NN-based model does not underfit or overfit. We assign the index to the total dataset for each fold to simulate nested cross-validation and to maintain the reproducibility of our experiment.

Statistical Analysis

The experimental data included only cases that underwent dual-phase FBB imaging obtained at the Department of Neurology or Nuclear Medicine of the DAUH, excluding cases with other types of dementia or inconsistent imaging protocols. As a result, a total of 264 cases, including 74 CN and 190 AD, were included in this analysis. We used independent-sample t-tests for numerical variables such as age, and education, and Pearson's Chi-square test for categorical variables such as sex, FBB reading, K-MMSE, CDR, and GDS to determine whether the characteristics of subjects in our experimental dataset are biased according to the diagnostic label. For the demographic analysis, we used IBM SPSS statistics version 23. In order to evaluate the classification performance of trained models, we calculated the accuracy (ACC) and area under the receiver operating characteristic curve (AUROC) for AD detection using DeLong's method[26] and Spearman correlation between predicted AD positivity scores and neuropsychological tests/actual diagnostic label. for these processes, we used MedCalc version 18.9.1 (MedCalc Software). In all tests, the statistical significance level was set at $p<0.001$ with a two-sided test.

Results

Data demographics

Table. I shows the statistics and test results for the characteristics of participants in the experimental data we used. There was no statistically significant difference between the CN and AD groups collected retrospectively by DAUH in age, sex, and education variables (Page < 0.142, Psex= 0.065, Peducation =

0.188). The results of K-MMSE, CDR, and GDS (which are the dominant variables in the diagnosis of AD, and reflect cognitive function) and dFBB readings (which reflect a state of Aβ plaque load) showed statistically significant differences between groups (PMMSE < 0.0001, PFBB_reading = 0.0001). Our experimental data included 20.83 % of Aβ-positive CN and 16.84 % of Aβ-negative AD.

Pre-processed Image 2

For the result of spatial registration, Fig. 2a shows static eFBB and dFBB registered in MNI space, which is randomly selected for each diagnostic label, compared with raw images of those in native space. As a result of pre-processing, it was confirmed that the spatial characteristics of individual imaging disappeared after they were transformed into MNI space but functional characteristics remained according to the diagnostic label.

In Fig. 2b, to check whether the functional information of eFBB on our pre-processing method and selected time-period is feasible, eFBB (2-7 min) was observed by t-contrast according to the diagnostic label. The functional information of dFBB was omitted because the results have already been verified through previous studies[21]. As a result of t-contrast, Aβ uptake of AD relatively lower CN in eFBB was not dominant, except for the cerebellar area. On the other hand, the relatively higher Aβ uptake of CN than AD is clustered throughout the brain tissue area.

AD classification performance

Table. Ⅱ shows the AD classification performance of ML-based predictive models measured by outer 4-fold cross-validation. LSTM (ACC : 78.182, AUROC: 0.846) was the best model for dynamic eFBB. SVM (ACC: 76.364, AUROC: 0.828), LR (ACC: 74.546, AUROC: 0.843) followed. For the classifier of static dFBB, which is used for conventional FBB reading, SVM (ACC: 81.364, AUROC: 0.890) was the best model for AD detection with dFBB (SVM_{dFBB}) and NN (ACC: 81.364, AUROC: 0.864), and LR (ACC: 77.887, AUROC: 0.849) followed. In comparison among all kinds of FBB, the $NN_{aggregated}$ was the best model (ACC: 84.091, AUROC: 0.900), which trained dual-phase FBB, followed by SVM that learned dFBB and LSTM that learned eFBB.

TABLE II
COMPARISON OF PREDICTIVE PERFORMANCE OF SCORES FOR ALZHEIMER'S DISEASE

Ac. Phase	Model	Accuracy (%) (SD)	Specificity (%) (SD)	Sensitivity (%) (SD)	AUROC (%) (SD)
Early	LR	74.546 (4.454)	61.278 (12.502)	82.066 (4.648)	0.843 (0.010)
	SVM	76.364 (5.123)	59.108 (8.285)	85.647 (9.871)	0.828 (0.030)
	LSTM	78.182 (2.571)	67.872 (13.949)	84.191 (5.107)	0.846 (0.025)
Delay	LR	77.887 (2.100)	64.849 (12.037)	84.867 (5.110)	0.849 (0.014)
	SVM	81.364 (3.928)	76.803 (18.000)	84.132 (3.908)	0.890 (0.018)
	NN	81.364 (4.545)	72.250 (2.790)	86.138 (6.432)	0.846 (0.014)
Dual	LSTM+NN	84.091 (4.545)	71.201 (18.726)	91.052 (3.347)	0.900 (0.032)

Ac. Phase: Acquisition Phase, AUROC : Area Under Receiver Operating Characteristic

AD positivities of total participants measured by three models ($NN_{aggregated}$, $LSTM_{eFBB}$, and SVM_{dFBB}) representing each kind of FBB (dual-phase FBB, dynamic eFBB, and static dFBB) in the total data are

presented in Table. ⍰.

TABLE III
COMPARISON OF AUROC OF TEST SCORES FOR ALZHEIMER'S DISEASE ACCORDING TO SPECIFIC DISTRIBUTION

Pop. / Test (Ac. Phase-Model)	AUROC (%) (SE)	Statistical significance[1]
Total		
K-MMSE	0.966 (0.024)	** (b), ** (c)
Early FBB (LSTM)	0.891 (0.024)	N/A
Delay FBB (SVM)	0.881 (0.026)	N/A
Dual FBB (LSTM+NN)	0.942 (0.016)	* (b), ** (c)
Aβ (−) CN vs Aβ (−) AD		
K-MMSE	0.979 (0.014)	** (b), *** (c)
Early FBB (LSTM)	0.843 (0.048)	N/A
Delay FBB (SVM)	0.790 (0.055)	N/A
Dual FBB (LSTM+NN)	0.927 (0.028)	* (b), ** (c)
Aβ (+) CN vs Aβ (+) AD		
K-MMSE	0.950 (0.021)	* (b), ** (c)
Early FBB (LSTM)	0.867 (0.037)	N/A
Delay FBB (SVM)	0.765 (0.075)	N/A
Dual FBB (LSTM+NN)	0.901 (0.036)	* (c)
Aβ (−) CN vs Aβ (+) AD		
K-MMSE	0.967 (0.012)	N/A
Early FBB (LSTM)	0.925 (0.023)	N/A
Delay FBB (SVM)	0.981 (0.009)	* (c)
Dual FBB (LSTM+NN)	0.981 (0.009)	** (b)

1 : (a) MMSE, (b) eFBB, (c) dFBB, and (d) Dual-phase FBB tested with Delong test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$), and an existence of the significant difference is presented on the larger side.
2 : Korean version of Mini Mental State Examination

$NN_{aggregated}$ (AUROC: 0.942) trained dual-phase FBB was able to detect AD better than $LSTM_{eFBB}$ (AUROC: 0.881) and SVM_{dFBB} (AUROC: 0.891). Specifically, unlike the model trained on single-phase FBB, the model trained on dual-phase FBB had no statistically significant difference from MMSE (AUROC: 0.966). A comparison of AUROC in Aβ-negative distribution (Aβ (−) NC vs. Aβ (−) AD), the $NN_{aggregated}$ (AUROC: 0.927) was the best, followed by $LSTM_{eFBB}$ (AUROC: 0.843) and SVM_{dFBB} (AUROC: 0.790). In Aβ-positive distribution (Aβ (+) CN vs. Aβ (+) AD), the $NN_{aggregated}$ (AUROC: 0.867) was the best as well, followed by $LSTM_{eFBB}$ (AUROC: 0.867) and SVM_{dFBB} (AUROC: 0.765). In the comparison between the distributions (Aβ (−) CN vs. Aβ (+) AD) where Aβ distribution best explains the occurrence of AD, the $NN_{aggregated}$ (AUROC: 0.981) and SVM_{dFBB} (AUROC: 0.981) were the best, followed by $LSTM_{eFBB}$ (AUROC: 0.925).

Input and Feature distribution according to the visual reading of dFBB and diagnostic label

In Fig. 3, the distributions of inputs and features in the last hidden layer of the NN-based model according to the phase of FBB are shown in a two-dimensional space using t-SNE[27]. Fig. 3c and 3e show the distribution of mean SUVr and features extracted from dFBB, and those do not seem to fully explain Aβ-positive CN and Aβ-negative AD. The distribution of mean SUVr and features extracted from eFBB shown in Fig. 3a, b, and d appears to be that Aβ-negative AD distribution is closer to Aβ-positive AD distribution compared to those extracted from dFBB. However, it is observed that the Aβ-positive CN distribution is still close to that of Aβ-positive AD. On the other hand, in Fig. 3f, the feature distribution extracted from

dual-phase FBB showed a continuously distributed pattern rather than clustered, and relatively well separating Aβ-negative and Aβ-positive AD.

Association between AD positivity score and neuropsychological test/occurrence of AD

Fig. 4 shows the correlation between AD positivity estimated from each phase of FBB imaging and the neuropsychological test results. In the correlation analysis, the scores estimated from eFBB (ρMMSE: −0.527, ρCDR: 0.477, ρGDS: 0.564) correlated with neuropsychological tests results better than dFBB (ρMMSE: −0.174, ρCDR: 0.158, ρGDS: 0.064). AD positivity scores estimated from dual-phase FBB showed the correlation coefficients (ρMMSE: −0.556, ρCDR: 0.386, ρGDS: 0.533) similar with those of eFBB rather than dFBB. In comparison of odds ratio, AD positivity estimated from dual-phase FBB (odds ratio: 56.333, $P < 0.0001$) was most associated with AD, followed by dFBB (odds ratio: 35.182, $P < 0.001$) and eFBB (odds ratio: 27.679, $P < 0.001$). AD positivity scores estimated from eFBB and dual-phase FBB showed statistically significance in Spearman's rank correlation coefficients for all neuropsychological tests, but dFBB showed the statistically significance for only K-MMSE.

Discussion

We designed an experimental model to successfully improve the conventional imaging biomarkers with only static dFBB by engaging in dynamic eFBB based on the following two assumptions: (1) The potential blood flow information included in eFBB is sufficiently distinguished from dFBB and they provide complementary information with respect to AD diagnosis. (2) The temporal information included in dynamic eFBB can be summarized as a feature vector representing blood flow information by the LSTM model. In the remaining paragraphs, we will elucidate the experimental results or related problems with respect to the hypotheses above.

Compared with the use of only dFBB in a conventional context, to improve the accuracy of AD classification by analyzing dual-phase FBB, eFBB and dFBB must contain sufficient complementary information regarding AD, that is, eFBB should be able to sufficiently explain AD in different aspects from dFBB. As shown in Fig. 2 and Table. , we tried to confirm whether the potential perfusion information of eFBB is suitable for this experiment. Even though the deformation field used for registration in eFBB was applied to dFBB, both eFBB and dFBB could be located in the MNI space, and the Aβ load pattern in a region of gray matter was still observed in each preprocessed dFBB. In voxel-based analysis, hypo-perfusion was observed in the AD group regardless of the Aβ distribution (Fig. 2). From the comparison of characteristics between the same Aβ distribution in Table. , it was observed that the AD positivity calculated from eFBB relatively explained AD distribution relatively well, which was difficult to discriminate the diagnostic label with dFBB in a same Aβ distribution. Therefore, the dynamic or static eFBB which acquired from our experimental protocol is meant to be complementary to the uptake of dFBB for AD detection, and the improved classification performance of the $NN_{aggregated}$ with dual-phase FBB could be based on the additional potential blood flow data.

LSTM is a representative NN for time series data that ultimately understands the long-term contextual information by managing the cell state necessary to determine the output from the input over time through input, output, and forget gates[28]. In terms of research on medical data, LSTM has been frequently used in EEG/ECG[29], imaging reports[30], electronic health records[31], and static or dynamic imaging data[32, 33], which include temporal information. A common delay-phase static PET image is acquired at the acquisition time determined by investigating the pseudo-equilibrium interval in which specific binding remains stable through TAC data and considering other parameters such as image quality and diagnostic accuracy. In the case of the FBB radiotracer, the manufacturer provides acquisition time for the delay phase, not for the early phase. In eFBB, the optimal acquisition time interval closest to perfusion cannot be found in the stable state owing to the curve that changes rapidly around the peak; therefore, the interval must be determined exploratory. Even the interval for ideal potential perfusion imaging is not deterministic and may vary from case to case. As a related work, it was mainly considered in studies that explored a specific acquisition time based on the similarity between eFBB and FDG images. They randomly selected an interval, including the peak uptake[11], or searched for a combination of the start time and time window to determine the acquisition time with the correlation most similar to FDG[10]. In Table. □, the LSTM model showed better performance than the NN model trained static eFBB at 2-7 min, which had a good correlation with FDG in our prior study[18]. These experimental results may indicate that the LSTM could understand the temporal features required for AD classification from potential perfusion information in dynamic eFBB and the calculation of optimal acquisition time could be omitted.

Fig. 5a and 5b show that eFBB and dFBB discriminate AD from CN using different features of each image. In Fig. 5b, most misclassifications occurred in the Aβ-positive CN and Aβ-negative AD populations, whereas, in Fig. 5a, the eFBB classifier consistently scores a proper AD positivity for CN or AD regardless of Aβ distribution. Therefore, it could be considered that the performance of dual-phase FBB classifier originates from the state of neuronal injury by comprehensively evaluating the degree of hypo-perfusion from eFBB and Aβ plaque deposition from dFBB, respectively (Fig. 5c). In Table □, AD positivity scores calculated by dual-phase FBB for the entire population showed the AUROC, which had no statistically significant difference compared to the MMSE, and better classification performance than those calculated using only dFBB regardless of Aβ distribution. These results may indicate that it is possible to improve the evaluation of the degree of neuronal damage in research or clinically when the AD positivity score of dual-phase FBB is provided. In addition, it could provide a quantitative index to nuclear medicine physicians to explain false negative/positive cases in FBB imaging tests. This quantitative method could be considered for application to other types of tracers or PET imaging where early phase PET reflects potential perfusion information.

This study proposes a quantitative method for the interpretation of dual-phase FBB at this point when the evaluation criteria for potential perfusion information of eFBB have not yet been established. Ultimately, it could help to reduce the radiation exposure and costs for patients with AD, and for a nuclear medicine physician, it could be a helpful tool in visual assessment for dual-phase FBB. On the other hand, as a

limitation of this study, the predictive model analyzing dual-phase FBB needs to be evaluated in terms of external validation or clinical validity in the future. As mentioned earlier, AD is associated with neurofibrillary tangles aggregated by phosphorylated tau, CSF biomarker, genetics, and environmental factors, in addition to Aβ plaque accumulation. Given additional clinical and laboratory data in the future, it would be possible to develop a predictive model that aggregates more various predictive factors for AD in addition to improving the performance of the quantitative model in this study.

Conclusion

In this paper, we report on how to interpret dual-phase FBB using ML-based models and their evaluation results. In comparison with the AD classification, the model trained on mean SUVr extracted from dual-phase FBB imaging (ACC: 84.091%, AUROC: 0.900) showed better AD classification than single-phase FBB, eFBB (ACC: 78.182%, AUROC: 0.846), or dFBB (ACC: 81.364%, AUROC: 0.890). In addition, the AD positivity score estimated by dual-phase FBB (ρMMSE: −0.556, OR: 59.333) showed a higher correlation with psychological test results and the association with AD occurrence compared to those with only dFBB (ρMMSE: −0.174, OR: 27.679). These experimental results show that the proposed method could be used to interpret eFBB in dual-phase FBB and that by reflecting eFBB into the current reading system, Aβ-PET reading, AD diagnosis, or the monitoring system could be improved.

Declarations

Acknowledgements

This research was supported by the National Research Foundation (NRF) of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2018 R1A2B2008178). We would like to thank Editage (www.editage.co.kr) for English language editing.

Competing interests

The authors declare no competing interests.

Author contributions

H.K. conceptualized and designed the study, investigated related works and wrote source code and manuscript.

K.W.P. provided neuropsychological test results as well as well-arranged participants demographic data and reviewed the manuscript.

D.Y.K. contributed to the conceptualization for the study, provided well-arranged ^{18}F-Florbetaben imaging data, reviewed the manuscript, and supervised the study as a project administrator.

Data availability

The data that used for the study are available on request from the corresponding author.

References

1. World Health Organization. Risk reduction of cognitive decline and dementia: WHO guidelines. (2019).
2. Villemagne, V. L., Rowe, C. C., Macfarlane, S., Novakovic, K. & Masters, C. L. Imaginem oblivionis: the prospects of neuroimaging for early detection of Alzheimer's disease. *Journal of Clinical Neuroscience* **12**, 221-230 (2005).
3. Villemagne, V. L. Amyloid imaging: past, present and future perspectives. *Ageing research reviews* **30**, 95-106 (2016).
4. Daulatzai, M. A. Cerebral hypoperfusion and glucose hypometabolism: key pathophysiological modulators promote neurodegeneration, cognitive impairment, and Alzheimer's disease. *J. Neurosci. Res.* **95**, 943-972 (2017).
5. Chételat, G. Aβ-independent processes—Rethinking preclinical AD. *Nature Reviews Neurology* **9**, 123-124 (2013).
6. Landau, S. M. *et al*. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Ann. Neurol.* **72**, 578-586 (2012).
7. Jack Jr, C. R. *et al*. An operational approach to National Institute on Aging–Alzheimer's Association criteria for preclinical Alzheimer disease. *Ann. Neurol.* **71**, 765-775 (2012).
8. Mosconi, L. *et al*. FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease. *European journal of nuclear medicine and molecular imaging* **36**, 811-822 (2009).
9. Rostomian, A. H., Madison, C., Rabinovici, G. D. & Jagust, W. J. Early 11C-PIB frames and 18F-FDG PET measures are comparable: a study validated in a cohort of AD and FTLD patients. *J. Nucl. Med.* **52**, 173-179 (2011).
10. Tiepolt, S. *et al*. Early [18 F] florbetaben and [11 C] PiB PET images are a surrogate biomarker of neuronal injury in Alzheimer's disease. *European journal of nuclear medicine and molecular imaging* **43**, 1700-1709 (2016).
11. Daerr, S. *et al*. Evaluation of early-phase [18F]-florbetaben PET acquisition in clinical routine cases. *NeuroImage: Clinical* **14**, 77-86 (2017).
12. Gulshan, V. *et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402-2410 (2016).
13. Lakhani, P. & Sundaram, B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* **284**, 574-582 (2017).
14. Jo, T., Nho, K. & Saykin, A. J. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Frontiers in aging neuroscience* **11**, 220 (2019).

15. American Psychiatric Association. American Psychiatric Association: Diagnostic and Statistical Manual ofMental Disorders. (1994).
16. McKhann, G. M. *et al*. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia* **7**, 263-269 (2011).
17. Barthel, H. *et al*. Cerebral amyloid-β PET with florbetaben (18F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. *The Lancet Neurology* **10**, 424-435 (2011).
18. Shin, H. J. *et al*. Optimal time frame for early-phase F-18-FBB brain PET compared to static F-18-FDG brain PET. *KSNM* **54**, 98-98 (2020).
19. Kang, H. & Kang, D. Y. Prediction of Alzheimer's Disease from early phase 18F-Florbetaben PET via LSTM. *KSNM* **54**, 105-105 (2020).
20. Hutton, C. *et al*. Quantification of 18 F-florbetapir PET: comparison of two analysis methods. *European journal of nuclear medicine and molecular imaging* **42**, 725-732 (2015).
21. Bae, S. *et al*. Spatial Normalization Using Early-Phase [18 F] FP-CIT PET for Quantification of Striatal Dopamine Transporter Binding. *Nuclear Medicine and Molecular Imaging* **54**, 305-314 (2020).
22. Hammers, A. *et al*. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum. Brain Mapp.* **19**, 224-247 (2003).
23. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
24. Varma, S. & Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* **7**, 1-8 (2006).
25. Taylor, J. C. & Fenner, J. W. Comparison of machine learning and semi-quantification algorithms for (I123) FP-CIT classification: the beginning of the end for semi-quantification? *EJNMMI physics* **4**, 1-20 (2017).
26. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837-845 (1988).
27. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. journal of Machine Learning Research 9. *Nov (2008)* (2008).
28. Schmidhuber, J. & Hochreiter, S. Long short-term memory. *Neural Comput.* **9**, 1735-1780 (1997).
29. Weng, W., Wagholikar, K. B., McCray, A. T., Szolovits, P. & Chueh, H. C. Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. *BMC medical informatics and decision making* **17**, 1-13 (2017).
30. Basiri, M. E., Chegeni, R. S., Karimvand, A. N. & Nemati, S. *Bidirectional LSTM Deep Model for Online Doctor Reviews Polarity Detection* (2020 6th International Conference on Web Research (ICWR), IEEE, 2020).

31. Lipton, Z. C., Kale, D. C., Elkan, C. & Wetzel, R. Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677* (2015).
32. Liang, D. et al. *Combining convolutional and recurrent neural networks for classification of focal liver lesions in multi-phase CT images* (International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018).
33. Yao, H., Zhang, X., Zhou, X. & Liu, S. Parallel structure deep neural network using CNN and RNN with an attention mechanism for breast cancer histology image classification. *Cancers* **11**, 1901 (2019).

Figures

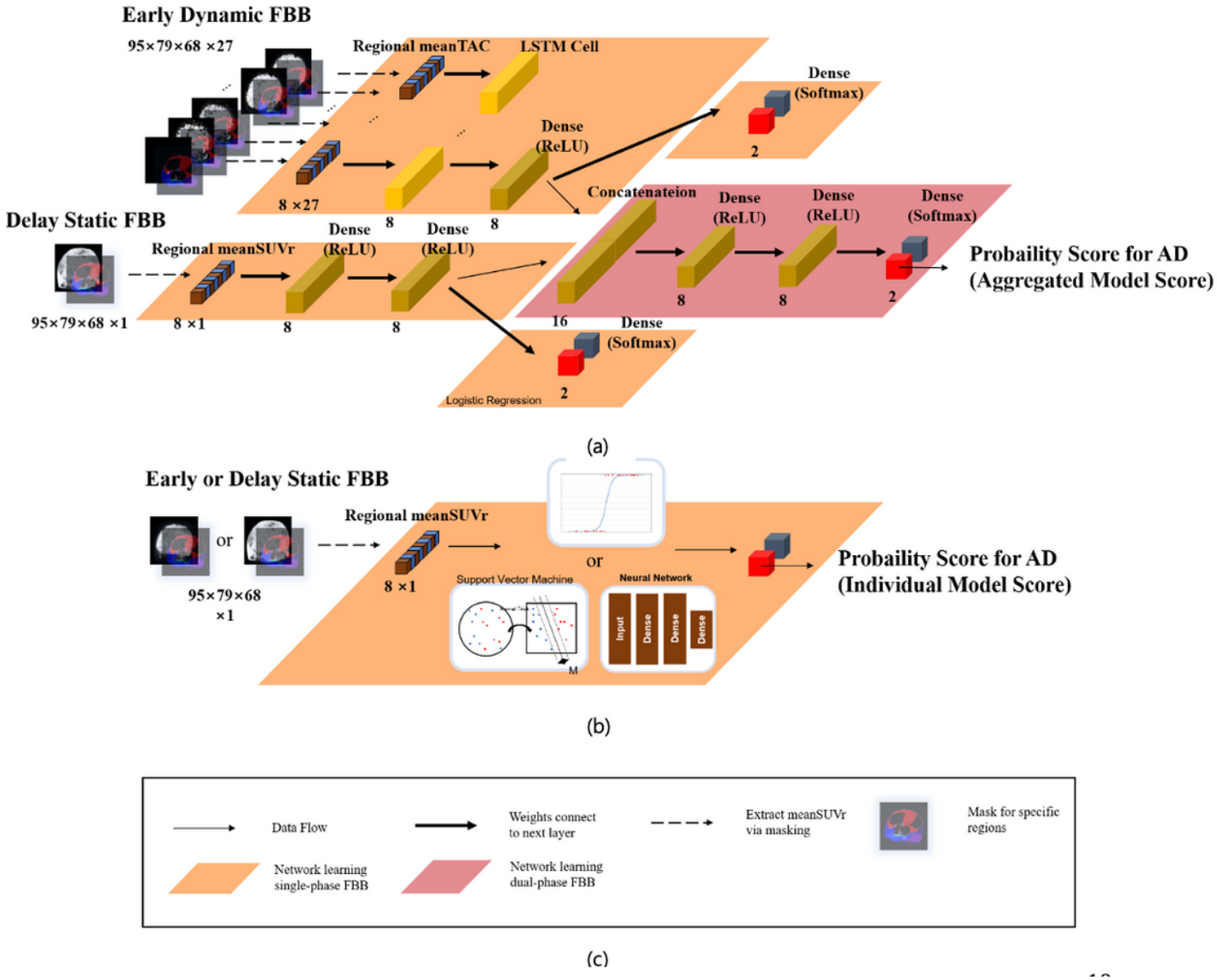

Figure 1

Structures of machine learning-based models to estimate AD positivity score with single or dual-phase 18F-Florbetaben (FBB) PET imaging. (a) shows a combined neural network that analyzes dual-phase FBB

(dynamic early phase and delay-phase FBB). (b) shows model structure that analyzes single-phase FBB. (c) shows a label for this figure. each component deBoth (a) and (b) use regional mean SUVr extracted from a preprocessed image as an input.

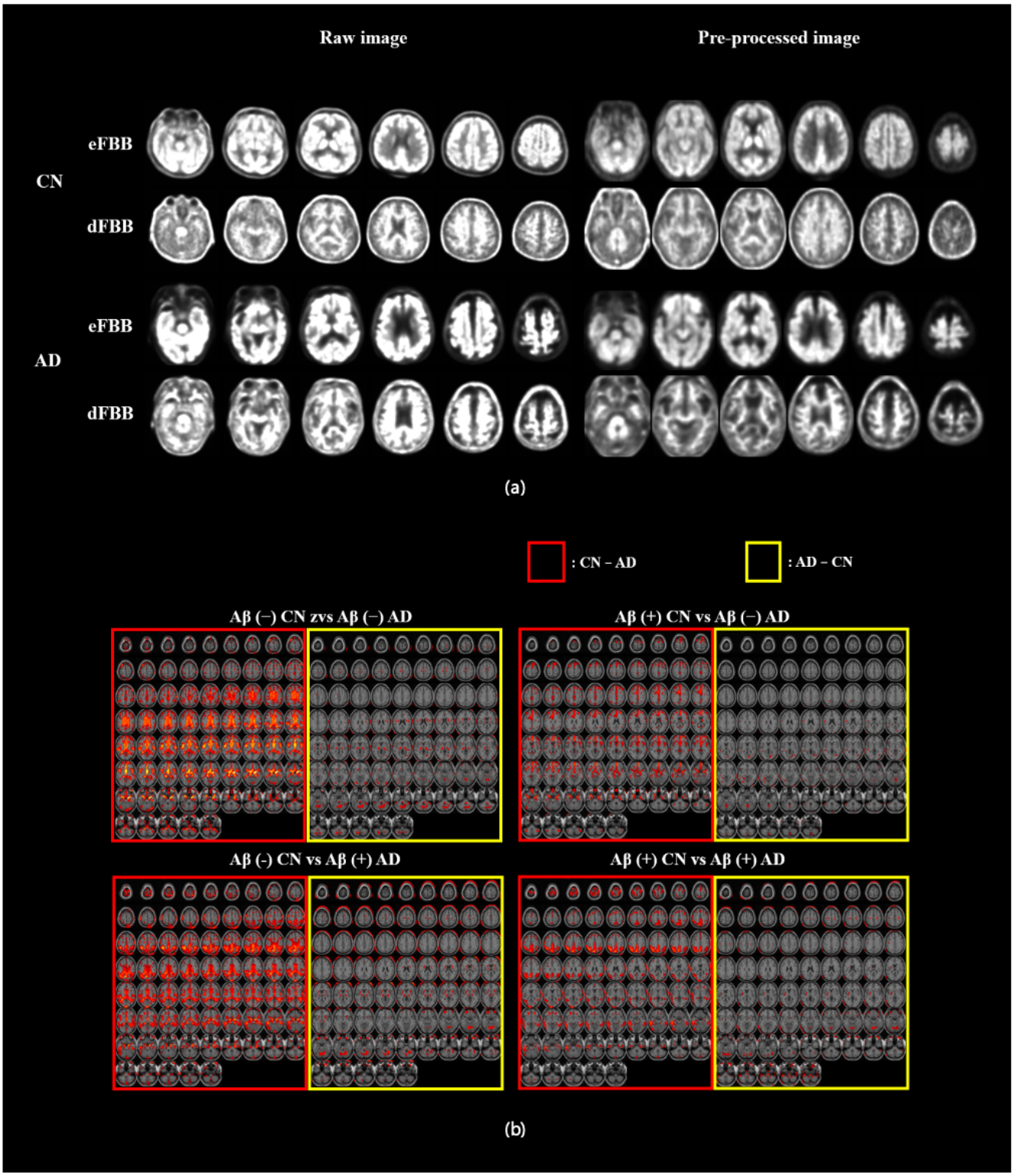

Figure 2

Pre-processed dual phase FBB PET and t-contrast of early phase F18-Florbetaben PET according to Aβ distribution.

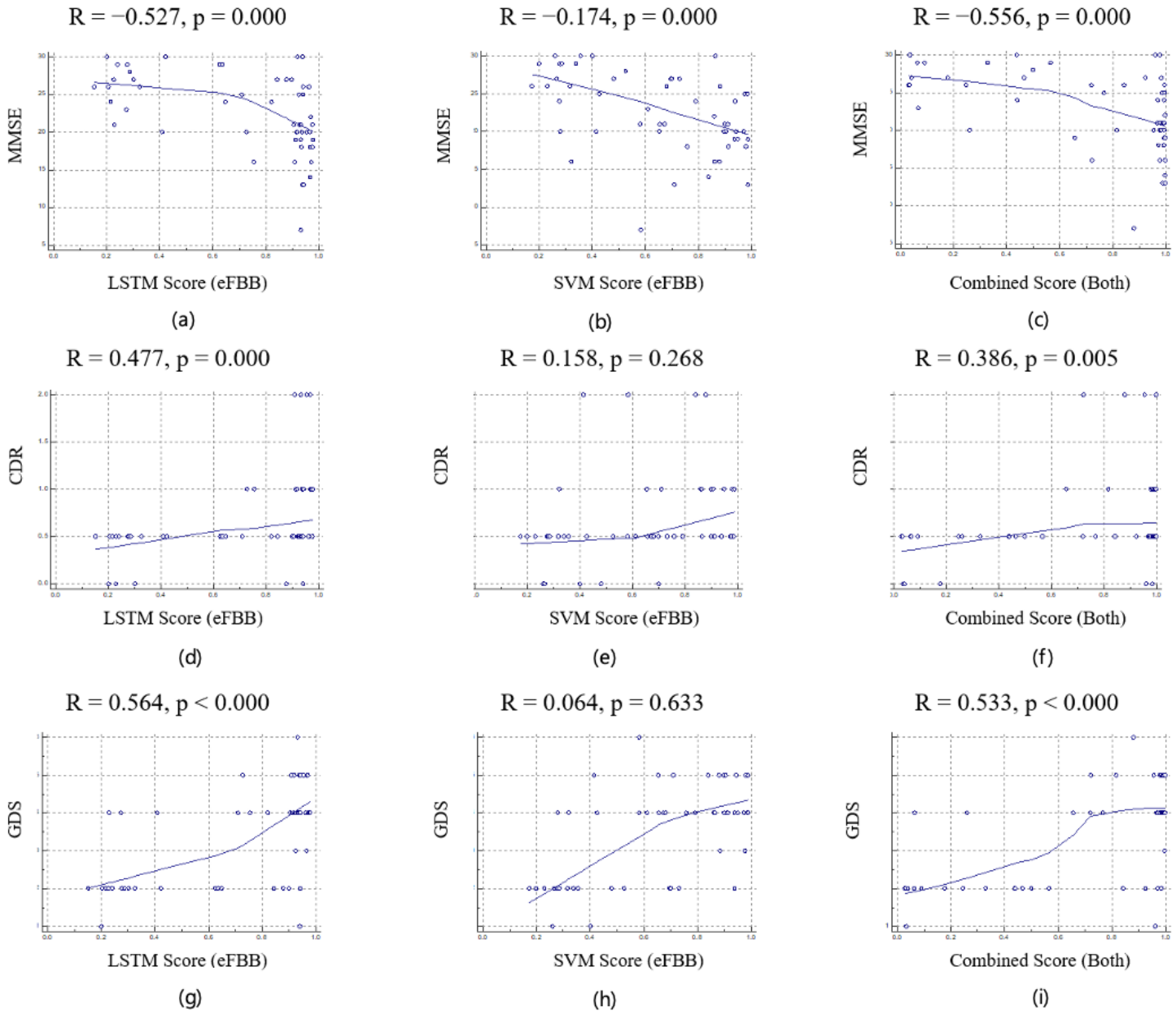

Figure 3

Distributions of model inputs and model features observed through t-SNE according to F18-Florbetaben reading label and diagnostic label. (a–c) show distributions of mean SUVr values used for a predictive model. (d–f) show distributions of feature vectors obtained from the last hidden layer of a neural network.

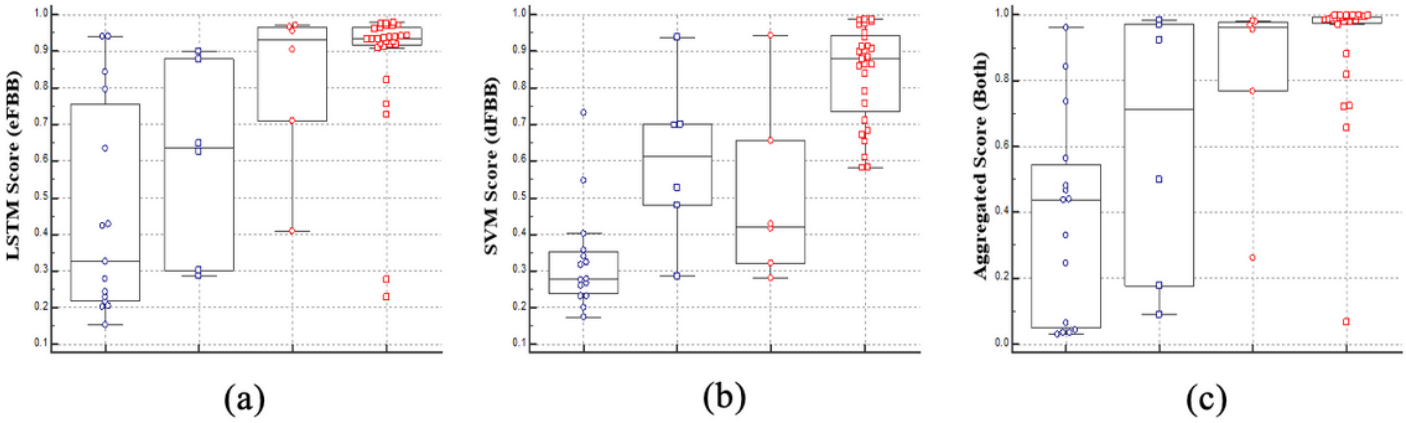

Figure 4

Correlation between predicted AD scores from machine learning-based models and neuropsychological tests.