Behavioural Brain Research 344 (2018) 103–109

Contents lists available at ScienceDirect

Behavioural Brain Research

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

Research report

Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging

Hongyoon Choi[a,*,1], Kyong Hwan Jin[b,c,**,1], for the Alzheimer's Disease Neuroimaging Initiative[2]

[a] Cheonan Public Health Center, Chungnam, Republic of Korea
[b] Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[c] Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

ARTICLE INFO

Keywords:
Brain PET
Amyloid
Deep learning
Convolutional neural network
Alzheimer's disease

ABSTRACT

For effective treatment of Alzheimer's disease (AD), it is important to identify subjects who are most likely to exhibit rapid cognitive decline. We aimed to develop an automatic image interpretation system based on a deep convolutional neural network (CNN) which can accurately predict future cognitive decline in mild cognitive impairment (MCI) patients using flurodeoxyglucose and florbetapir positron emission tomography (PET). PET images of 139 patients with AD, 171 patients with MCI and 182 normal subjects obtained from Alzheimer's Disease Neuroimaging Initiative database were used. Deep CNN was trained using 3-dimensional PET volumes of AD and normal controls as inputs. Manually defined image feature extraction such as quantification using predefined region-of-interests was unnecessary for our approach. Furthermore, it used minimally processed images without spatial normalization which has been commonly used in conventional quantitative analyses. Cognitive outcome of MCI subjects was predicted using this network. The prediction accuracy of the conversion of mild cognitive impairment to AD was compared with the conventional feature-based quantification approach. Accuracy of prediction (84.2%) for conversion to AD in MCI patients outperformed conventional feature-based quantification approaches. ROC analyses revealed that performance of CNN-based approach was significantly higher than that of the conventional quantification methods ($p < 0.05$). Output scores of the network were strongly correlated with the longitudinal change in cognitive measurements ($p < 0.05$). These results show the feasibility of deep learning as a practical tool for developing predictive neuroimaging biomarker.

1. Introduction

1.1. Backgrounds

Recent treatment strategies for Alzheimer's disease (AD) are aimed at slowing cognitive decline and are focused on the pre-dementia stage which includes mild cognitive impairment (MCI) [1]. However, as the etiology of MCI is heterogeneous, MCI patients show different rates of cognitive decline, and even some never convert to AD [2]. Thus, it is a matter of utmost importance to identify the patients with MCI who would benefit from treatment. So far, several studies have investigated a number of imaging biomarkers that can predict whether a patient with MCI will convert to AD. They include brain metabolism and amyloid load measured by ^{18}F-fluorodeoxyglucose (FDG) and ^{18}F-florbetapir (AV-45) positron emission tomography (PET), respectively [3–6]. Previous studies have used quantitative parameters or visual assessment of the brain images for predicting MCI patients who would covert to AD. However, visual analysis cannot provide quantitative and objective data and quantitative analyses commonly require complicated processing [7–10].

In this study, we showed a deep convolutional neural network (CNN) based method, a type of deep learning, could accurately predict

Abbreviations: AD, Alzheimer's disease; MCI, mildcognitive impairment; FDG, ^{18}F-fluorodeoxyglucose; AV-45, ^{18}F-florbetapir; CNN, convolutional neural network; ADNI, Alzheimer's Disease Neuroimaging Initiative; NC, normal control; CDR-SB, Clinical Dementia Rating Sum of Boxes; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive Subscale 13-item; FAQ, Functional Activities Questionnaire; MMSE, Mini-Metal Status Examination; ROC, receiver operating characteristic

* Corresponding author at: Cheonan Public Health Center, 234-1 Buldang-Dong, Seobuk-Gu, Cheonan, Republic of Korea.
** Corresponding author at: Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.
E-mail addresses: chy1000@snu.ac.kr (H. Choi), kyong.jin@epfl.ch (K.H. Jin).

[1] These authors contributed equally to this work.

[2] 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.

https://doi.org/10.1016/j.bbr.2018.02.017
Received 30 January 2018; Received in revised form 2 February 2018; Accepted 13 February 2018
Available online 14 February 2018
0166-4328/ © 2018 Elsevier B.V. All rights reserved.

cognitive decline. Recent advances in CNN have dramatically improved image recognition field [11]. We applied CNN to FDG and AV-45 PET images to predict cognitive decline in MCI patients. Our method was designed to discriminate patients' groups classified according to the cognitive outcome with minimized image processing. In addition, it could provide a quantitative biomarker combining both the FDG and AV-45 PET information. We showed that the CNN-based biomarker strongly correlated with future cognitive decline. Our major contribution was 1) to predict future MCI patients' outcome by applying the deep CNN model trained for discriminating AD from controls, and 2) to achieve prediction with minimally processed multimodal neuroimage data due to the benefits of the deep CNN.

1.2. Related works

Recently, deep learning has been applied to brain images to classify patients' disorders. One of the initial paper that used deep learning for identifying AD using brain images employed Deep Boltzmann Machine as a feature extractor [12]. Discriminative features were extracted from 3D patches of FDG PET and MRI for the further classification task. Another study for extracting deep learning-based discriminative features from multimodal images used randomized denoising autoencoder and showed correlation between the output score of the classification model and clinical outcomes [13]. Another study used conventional image quantification and deep learning. Multimodal image features extracted by predefined region-of-interests were combined as inputs of a neural network for classifying AD [14]. These studies focused on discriminative feature extraction from multimodal image data, however, required routinely performed image processing. Furthermore, the input of the neural network was high-dimensional vector or manually extracted features instead of raw 3-dimensional voxels as CNN models. Instead of feature extracting using autoencoder or Boltzmann machine, 3D CNN models exploiting all voxels as inputs have been applied to classifying disorders. A 3D CNN model was applied to structural MRI to classify AD, MCI and controls [15]. Another 3D CNN model for structural MRI was built upon pretraining process performed by convolutional autoencoder [16]. fMRI-based classification model was reported using 2D CNN by stacking processed fMRI images across axial and time axes [17]. So far, no studies focused on exploiting multimodal 3D PET images without normalization as inputs for CNN despite ease of clinical applicability of minimally processed images. Furthermore, independent future outcome prediction of MCI patients using CNN model trained by AD and controls has not been reported.

2. Methods

2.1. Subjects

The data used in this study included subjects recruited in Alzheimer's Disease Neuroimaging Initiative-II (ADNI-2) with available baseline data on FDG and AV-45 PET (http://adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. Subjects have been recruited from over 50 sites across the US and Canada. The primary purpose 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 http://www.adni-info.org. Written informed consent to cognitive testing and neuroimaging prior to participation was obtained, approved by the institutional review boards of all participating institutions.

For diagnostic classification and learning the CNN, we firstly selected patients with AD and healthy subjects who had baseline FDG and AV-45 PET scans. We also selected MCI patients who had baseline FDG and AV-45 PET scans and 3-year follow-up clinical evaluation. This resulted in 182 normal controls (NCs), 139 AD patients and 171 MCI subjects. Based on whether the MCI patients would convert to AD within 3 years, MCI patients were grouped as MCI converters and nonconverters.

Cognitive function of the subjects was evaluated using Clinical Dementia Rating Sum of Boxes (CDR-SB), Alzheimer's Disease Assessment Scale-Cognitive Subscale 13-item (ADAS-Cog), Functional Activities Questionnaire (FAQ), and Mini-Metal Status Examination (MMSE). As we interested in whether CNN-based baseline PET biomarkers would be associated with longitudinal cognitive decline, longitudinal changes of cognitive measurements were also assessed at 1 year and 3 years after the baseline study.

2.2. FDG PET and AV-45 PET

All the PET images were downloaded from ADNI database at the most advanced preprocessing stage. FDG PET images were acquired 30 to 60 min and AV-45 PET images were acquired 50 to 70 min after the injection. The FDG and AV-45 PET images were co-registered to each other, averaged across the time frames, standardized to have same voxel size ($1.5 \times 1.5 \times 1.5$ mm). PET images were acquired in the 57 sites participating in ADNI, scanner-specific smoothing was additionally applied [18]. As the processing did not include nonlinear spatial warping, size and shape of the brains were not changed after the preprocessing. These preprocessed images could be downloaded from ADNI database and we used them for deep CNN training and testing as they are.

2.3. Study design

The main purpose of this paper was to develop a deep CNN-based method for prediction of cognitive decline and selection of subjects who would eventually convert to AD. Before the testing of MCI conversion, the deep CNN was trained using PET images of AD and NC subjects. For discrimination between MCI converter and nonconverter, the network trained by AD/NC data was directly transferred as the imaging features of MCI converter would be similar with those of AD. The nodes of the output layer were only reassigned to MCI converter and nonconverter. All the PET images of MCI subjects were tested whether they would convert to AD or not. Therefore, our deep CNN was a classifier independent from the training data to discriminate between MCI converter and nonconverter (Fig. 1). In addition, we also obtained a quantitative score for MCI converter. The quantitative value of the output of the network before the final activation layer (i.e. softmax) was defined as ConvScore, a score that indicates how close inputted baseline images are to AD. The score can be expected to be utilized for a predictive biomarker.

The CNN was designed using MatConvNet (Version 1.0-beta 16) [19]. Additional information about the network architecture is available in the Supplementary Methods and Supplementary Table 1.

Sensitivity, specificity and accuracy of the classification between AD and NC were calculated using cross-validation. Additional information about training and validation of the classification between AD and NC is available in the Supplementary Methods.

2.4. Prediction of cognitive decline in MCI subjects

Using the network trained by PET images of AD and NC, PET images of MCI subjects were tested whether they would convert to AD or not. The labels of the output nodes of the CNN, AD and NC, were changed to MCI converter and nonconverter, as aforementioned above. As the classification of AD/NC, we classified a patient as a predictive MCI converter if the probability of the network output layer was higher than 0.5. Sensitivity, specificity and accuracy were measured and ROC analysis was also performed using ConvScore.

ConvScore was correlated with the longitudinal changes of

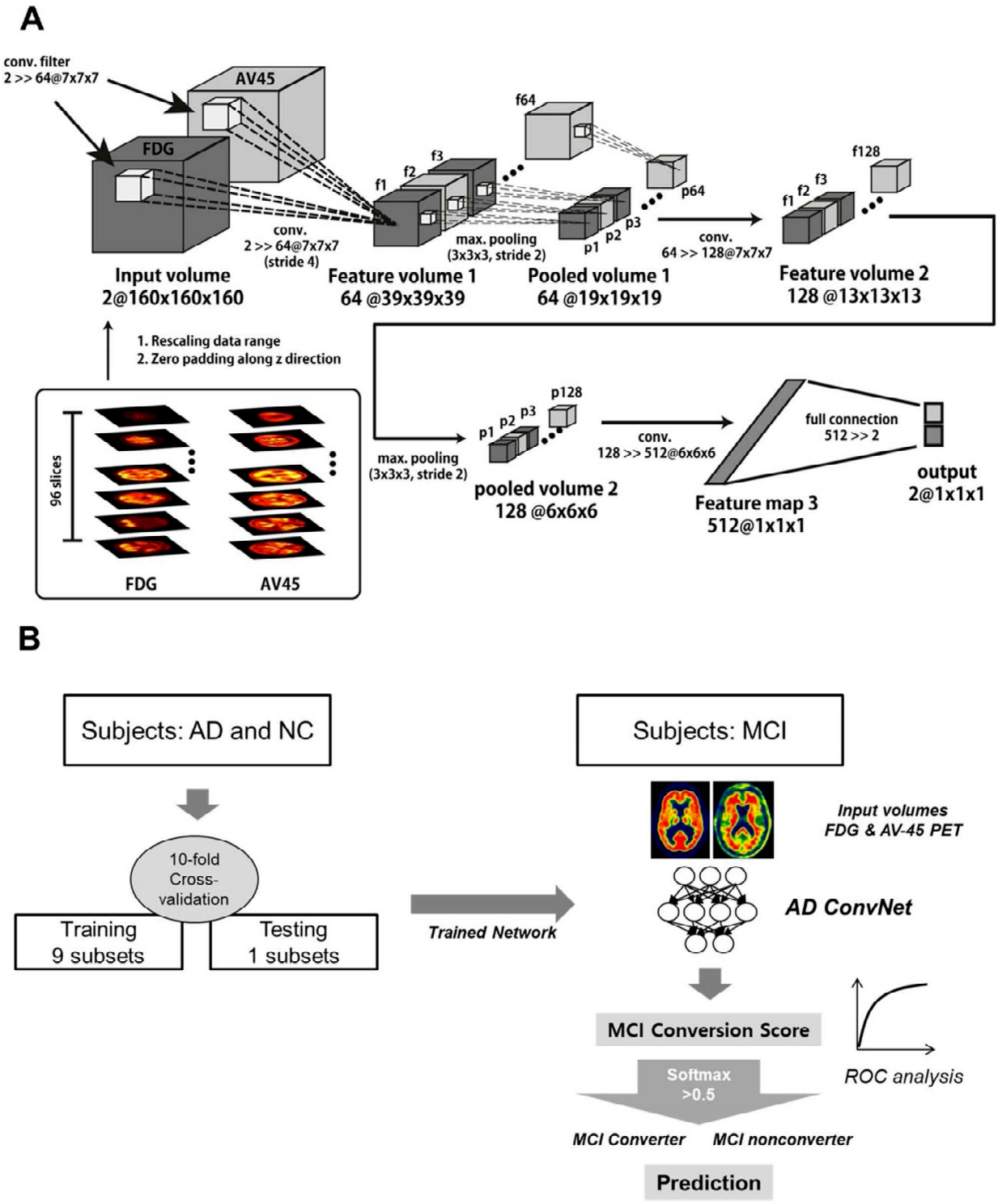

Fig. 1. Framework for predicting cognitive decline in mild cognitive impairment patients. (A) Deep convolutional neural network architecture is applied to the two PET images, FDG and florbetapir (AV-45). Each layer, features can be extracted by 3-dimensional convolution and activation (ReLU) function. Multilayer convolutions yield 1-dimensional output and the last layer have two nodes, which correspond to Alzheimer's disease (AD) and normal control (NC). (B) Deep CNN was trained from PET data of AD and NC. 10-fold cross validation was used. After the training, the trained network was directly used for the classification between mild cognitive impairment (MCI) converter and nonconverter. Thus, the network was independent from the PET data of MCI patients. Prediction accuracy for MCI conversion was assessed. Receiver operating characteristic (ROC) analysis was also performed.

cognitive measurements including CDR-SB, ADAS-Cog, FAQ and MMSE. Cognitive measurements of 1-year and 3-year follow-up visits were compared with those of baseline studies to calculate longitudinal changes. Pearson correlation was used for the correlation analysis.

2.5. Feature volume of interests based analysis

To compare CNN-based biomarker with conventional quantification methods, feature volume of interests (VOIs)-based analyses for FDG and

AV-45 PET were carried out. The images were processed and quantified by feature VOI as described previously [20–22]. In brief, FDG PET volumes were spatially normalized with nonlinear warping process. Average FDG uptake from the angular (right/left), temporal (right/left) and posterior cingulate cortices were calculated. The quantification data were expressed relative to the mean uptake of a reference regions, pons and cerebellar vermis [20]. For AV-45 PET, cortical uptake of AV-45 was calculated. Cortical regions were segmented with FreeSurfer, and used to extract mean AV-45 uptake in frontal, anterior and posterior cingulate, lateral parietal, and lateral temporal regions. The overall cortical mean uptake value expressed relative to uptake in the whole cerebellum was used [21,22].

2.6. Voxelwise feature extraction and support vector machine

Our proposed CNN-based approach was additionally compared with a conventional machine learning approach using voxelwise feature extraction. This approach used combined information of FDG and AV-45 PET images and did not use predefined VOI. Thus, it was a data-driven method as our CNN-based approach that was independent from previous knowledge of significant VOIs relevant to the diagnosis. This approach was performed by three steps: spatial normalization of PET images, feature extraction using principal component analysis (PCA) and support vector machine (SVM) classification. SVM classifier using feature extraction from PET images has been studied several times and showed good performance for the diagnosis of AD [23–25]. We trained PCA-SVM classier using FDG and AV-45 PET data of AD/NC and tested it for MCI patients as training and testing process of the CNN-based approach. Details are described in Supplementary Methods.

2.7. Statistics

We compared the diagnostic and prediction accuracy of CNN with those of conventional feature-based approaches with McNemar's non-parametric test. ROC analysis with area under the curve (AUC) measurement was performed for ConvScore and feature VOI-based parameters. The AUCs were compared using a nonparametric test of DeLong for comparison of two correlated ROC curves [26]. The performance comparison results for predicting MCI converters were obtained by the model which showed the best accuracy for discriminating AD in the cross-validation. The performance of discriminating AD from NC was measured by the combined results of cross-validation. ConvScore was correlated with longitudinal changes of cognitive measurements using Pearson's correlation. All statistical analysis was performed by using the MATLAB software. P-value < 0.05 was considered significant.

3. Results

3.1. Results of AD classification and MCI conversion prediction

We included a total of 492 subjects in this study. Among them, 139 patients were AD, 171 patients were MCI and 182 were NC. Demographic data and cognitive measurements of each group were summarized in Table 1.

The classification accuracy of the CNN-based approach was compared with feature VOI-based analysis of FDG and AV-45 PET as conventional quantitative analyses. In addition, it was compared with a SVM classifier based on PCA features of FDG and AV-45 PET images. The PCA features were shown in Supplementary Fig. 1. Sensitivity, specificity and accuracy of CNN-based approach for classification between AD and NC were 93.5%, 97.8% and 96.0%, respectively. They were significantly higher than those of the SVM classifier as well as VOI-based analyses (p-values were summarized in Table 2). The performance of SVM classifiers with different kernels are summarized in Supplementary Table 2. Sensitivity, specificity and accuracy of the CNN-based approach for the prediction of MCI conversion was 81.0%, 87.0% and 84.2%, respectively. Accuracy of the deep CNN was significantly higher than VOI-based analysis of FDG PET and the SVM classifier. Specificity of deep CNN was significantly higher than VOI-based analysis of AV-45 PET and the SVM classifier. Sensitivity was also higher compared with conventional methods, though it did not reach statistical significance (*p*-values were summarized in Table 2).

In addition to the group classification, we calculated quantitative score, ConvScore, for predicting MCI converters. It was obtained from the value of the last layer of the network. Using ConvScore, ROC curves were drawn and AUC were calculated (Fig. 2, Table 2). AUC of ConvScore was significantly higher than feature-VOI based analysis for AD classification ($p < 0.001$ for deep CNN vs. FDG and vs. AV-45) and prediction of MCI conversion ($p < 0.01$ for deep CNN vs. FDG and $p < 0.05$ for deep CNN vs. AV-45).

3.2. Correlation of CNN-based biomarker with cognitive outcomes

ConvScore calculated from baseline PET images of MCI patients was significantly correlated with the longitudinal change of cognitive measurements at 1 year and 3 years (Fig. 3). ConvScore was significantly positively correlated with longitudinal change of CDR-SB ($r = 0.37$, $p < 0.0001$ at 1 year and $r = 0.63$, $p < 0.0001$ at 3 years), ADAS-Cog ($r = 0.29$, $p = 0.0001$ at 1 year and $r = 0.24$, $p = 0.004$ at 3 year) and FAQ ($r = 0.40$, $p < 0.0001$ at 1 year and $r = 0.67$, $p < 0.0001$ at 3 year). It was significantly negatively correlated with MMSE ($r = -0.30$, $p < 0.0001$ at 1 year and $r = -0.61$, $p < 0.0001$ at 3 year). Note that the relatively weak correlation between ConvScore and longitudinal changes in cognitive scores at 1 year was found. The changes were steeper for the measurements at 3 years compared with 1 year. ConvScore of MCI converters was significantly higher than that of MCI nonconverters (2.40 ± 1.49 and -0.13 ± 1.36, respectively. $p < 0.0001$) (Supplementary Fig. 2). ConvScore could be used as a quantitative biomarker for predicting longitudinal cognitive measurements decline in MCI patients as well as conversion to AD.

4. Discussion

In the present study, we developed a deep learning-based diagnostic method for predicting cognitive decline in MCI patients. According to our knowledge, we firstly applied recent deep CNN model to multimodal PET images to predict cognitive outcome. Deep CNN could accurately classify patients' diagnostic group with the minimal steps of image processing and provide a quantitative biomarker for predicting cognitive outcome. Accordingly, the prediction of cognitive decline in

Table 1
Demographic data.

Variables	Normal (n = 182)	AD (n = 139)	*p-value*	MCI converter (n = 79)	MCI nonconverter (n = 92)	*p-value*
Age (years)	73.4 ± 6.3	74.3 ± 8.2	n.s.	72.3 ± 7.2	70.3 ± 6.3	n.s.
Sex (M/F)	88/94	80/59	n.s.	43/36	51/41	n.s.
Education (years)	16.6 ± 2.5	15.8 ± 2.7	< 0.01	16.3 ± 2.6	16.4 ± 2.7	n.s.
MMSE	29.0 ± 1.2	23.0 ± 2.1	< 0.001	27.1 ± 1.9	28.1 ± 1.7	< 0.001
ADAS	9.1 ± 4.5	31.0 ± 8.6	< 0.001	22.0 ± 6.8	12.9 ± 5.5	< 0.001
APOE ε4 (−/+)	130/52 (28.6%)	44/95 (68.3%)	< 0.001	22/57 (72.2%)	58/34 (37.0%)	< 0.001

Table 2
Sensitivity, specificity and accuracy for the discrimination between AD and normal controls and the prediction for MCI converters.

	Feature VOI-based		SVM	Deep CNN based approach	*p-value*		
	FDG	AV45			*vs. FDG*	*vs. AV45*	*vs. SVM*
AD vs. Normal							
Sensitivity,%	84.2	81.8	84.2	93.5	0.004	< 0.001	0.001
Specificity,%	86.3	80.8	88.5	97.8	< 0.001	< 0.001	< 0.001
Accuracy,%	85.4	80.7	86.6	96.0	< 0.001	< 0.001	< 0.001
ROC AUC	0.91	0.84	–	0.98	< 0.001	< 0.001	
MCI Conversion Prediction							
Sensitivity,%	70.9	86.1	79.7	81.0	*0.08*	*0.29*	*0.99*
Specificity,%	79.3	75.0	72.8	87.0	*0.19*	*0.007*	< 0.001
Accuracy,%	75.4	80.1	76.0	84.2	*0.02*	*0.21*	*0.02*
ROC AUC	0.82	0.83	–	0.89	*0.006*	*0.03*	

ROC: receiver operating characteristic; AUC: area under curve.

our method could be automatically made by simply inputting subjects' images. Output value (ConvScore) was significantly correlated with the longitudinal change of cognitive measurements.

Previous PET imaging biomarkers relied on the uptake value in a set of regions of interest developed *a priori* or whole-cortical uptake value [6,9,10]. To obtain these values, several image processing steps such as spatial normalization and cortical segmentation using structural MRI were required. However, such processing was not standardized and nonlinear image transformation could introduce a potential source of errors particularly in morphological alterations [27,28]. Our proposed method used spatially unnormalized baseline image data of AD and NC. It suggests that our approach is simply able to be utilized to redesign the voxelwise brain imaging processing pipeline which has routinely implemented normalization to template space.

Accuracy of differentiation between MCI converter and non-converter (84.2%) outperformed the conventional machine learning approach using SVM with voxelwise feature selection as well as conventional feature-VOI based methods. ROC comparison results also revealed that the accuracy of prediction was significantly higher than other methods. Note that the feature-VOI based methods relied on prior information about important regions and partly employs MR information for cortical parcellation [20–22], while deep CNN-based and SVM-based approaches solely used PET information. Nonetheless, the prediction accuracy of our approach was higher than the VOI-based methods with a priori and structural information. That is because deep CNN has the benefit that could automatically discover the optimal features for image classification [11]. Furthermore, the accuracy of our approach also outperformed other state-of-the-art machine-learning algorithms based on feature selection other than deep learning for this differentiation problem [29–32], though those studies used different imaging modalities and clinical variables. For example, Cheng, et al. suggested a machine learning classifer using FDG PET, MRI and CSF biomarkers, which showed 80.1% accuracy for discriminating MCI converters [29]. Deep CNN models using MRI showed more than 95% accuracy for discriminating AD from normal which was similar with our model though they were not tested for identifying MCI converters [15,16]. The major benefit of deep CNN compared with other conventional machine learning methods is the end-to-end training [11]. Because 3d medical images have high dimensions, previous methods have commonly required dimension reduction by feature selection. However, deep CNN model uses all voxels of image data as inputs as convolutional filters make sparse connection compared with fully-connected neural networks [33]. It eventually reduces learning parameters of the model and enables end-to-end training.

Our approach could provide a quantitative variable, ConvScore, to be used as a fusion biomarker for multimodal images. As an imaging biomarker, low glucose metabolism and high amyloid deposit in the cortex at baseline can predict the longitudinal decline of cognitive

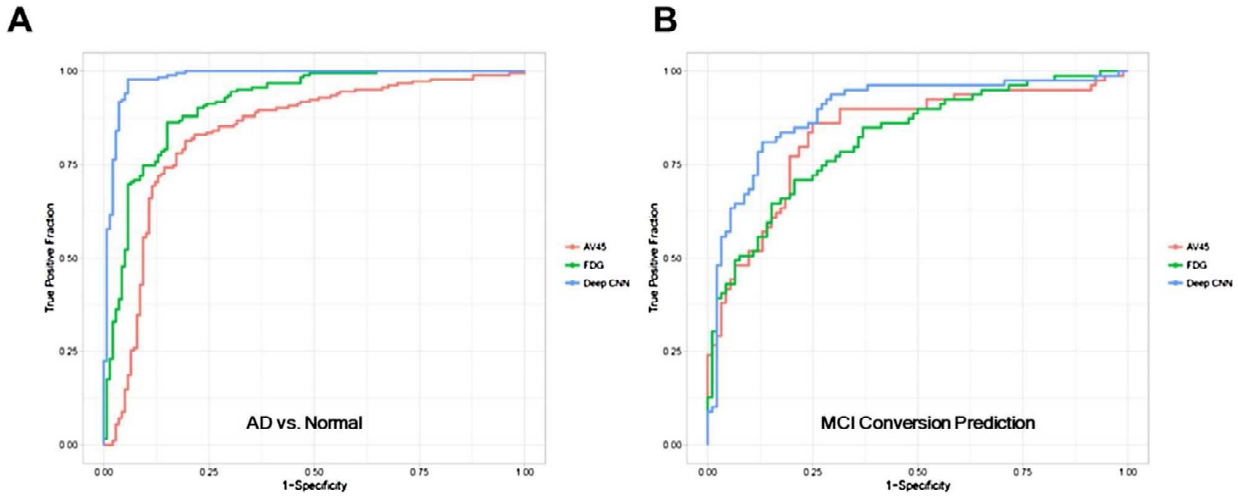

Fig. 2. ROC curves for deep CNN. ROC analyses were performed for the classification of AD (A) and the prediction of MCI conversion (B). ROC curves of feature VOI-based approaches using FDG and AV-45 PET were also drawn. Areas under curve (AUC) values were calculated. (A) AUC of ConvScore was significantly higher than feature-VOI based analysis for AD classification (AUC = 0.98, 0.91 and 0.84 for ConvScore, FDG and AV-45, respectively; $p < 0.001$ for deep CNN vs. FDG and vs. AV-45) (B) AUC of ConvScore was also significantly higher than feature VOI-based analysis for prediction of MCI conversion (AUC = 0.89, 0.82 and 0.83 for ConvScore, FDG and AV-45, respectively; $p < 0.01$ for deep CNN vs. FDG and $p < 0.05$ for deep CNN vs. AV-45).

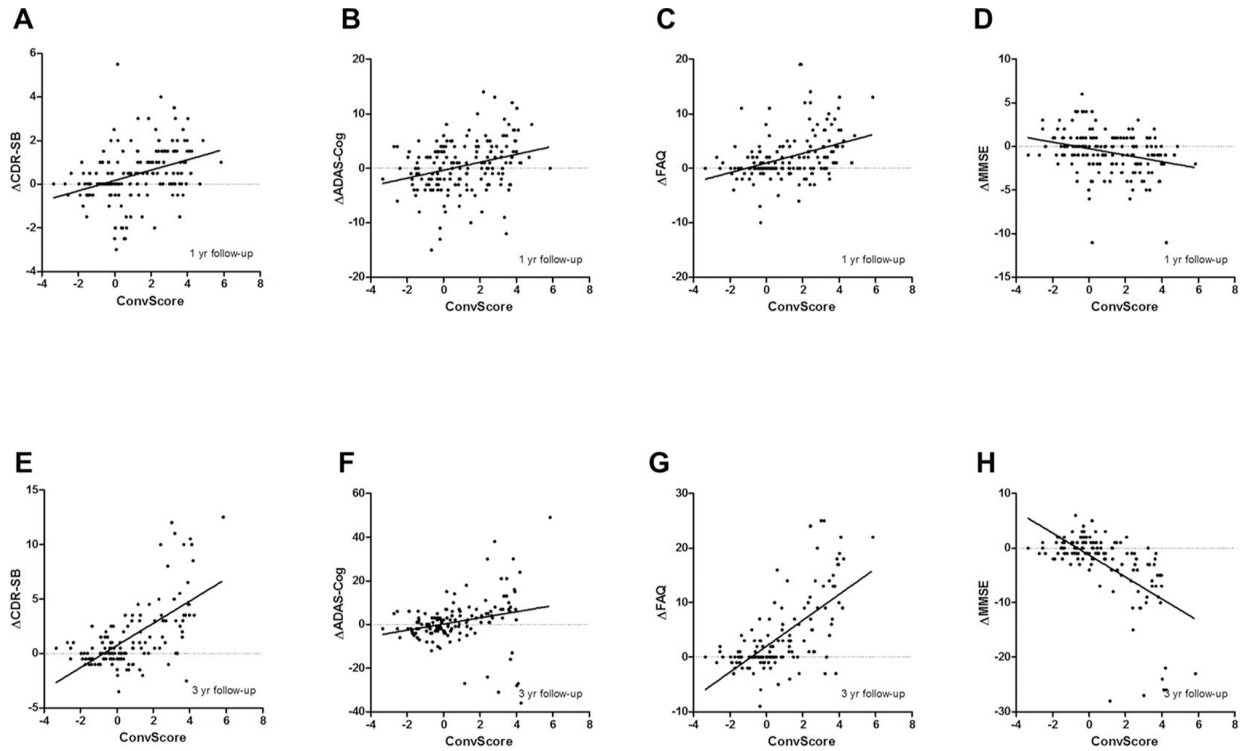

Fig. 3. Correlation between output of the network and longitudinal changes of cognitive measurements. The last layer provides an output scores for AD or MCI converter, which are defined as ConvScore. Cognitive measurements at baseline, 1 year, and 3 years were obtained for each MCI patient and ConvScore was correlated with the longitudinal change of them. (A–D) ConvScore was significantly correlated with the change of Cognitive Dementia Scaling Sum of Boxes (CDR-SB) ($r = 0.37$, $p < 0.0001$), Alzheimer Disease Assessment Scale-Cognitive Subtest (ADAS-Cog) ($r = 0.29$, $p = 0.0001$), Functional Activities Questionnaire (FAQ) ($r = 0.40$, $p < 0.0001$) and Mini-mental State Examination (MMSE) ($r = -0.30$, $p < 0.0001$) from baseline to 1 year follow-up. (E–H) The significant correlation between ConvScore and the change of the measurements from baseline to 3 years was also found ($r = 0.63$, $p < 0.0001$ for CDR-SB; $r = 0.24$, $p = 0.004$ for ADAS-Cog; $r = 0.67$, $p < 0.0001$ for FAQ; $r = -0.61$, $p < 0.0001$ for MMSE).

scores [20,21]. However, a combined parameter considering both metabolism and amyloid deposit has been needed. ConvScore is not only a fusion biomarker directly obtained by both PET images, but correlated with longitudinal cognitive measurements. It suggests cognitive functions of patients with high ConvScore at baseline could be rapidly deteriorated. Even though relatively weak correlation between ConvScore and cognitive measurements changes at 1 year (less than $r < 0.4$), we could be used ConvScore as a single parameter that reflects both PET image patterns for future cognitive decline particularly for long-term outcome. This correlation is an important observation that has impact on clinical trials for early treatment intervention in prodromal AD. ConvScore could help select the subjects who would benefit from treatment in the clinical trials.

In the clinical setting, most imaging studies are assessed by experts' visual analysis, because it is simpler and more practical than the quantitative assessment which needs time-consuming procedures. Deep CNN is motivated by human visual perception, which hierarchically processes recognized images in the cerebral cortex [34]. As hierarchical features of images are automatically trained by data, manual feature selection or image processing steps can be minimized in deep CNN [11,34]. Therefore, after the training, the network is automatically able to analyze patients' images by simply inputting subjects' images. Considering the ease of application, the CNN-based image interpretation system has a large potential to be used in development of biomarkers of several diseases including cancer, cardiovascular disorders as well as neurodegenerative diseases.

Recent remedies of CNN for achieving higher accuracy are increasing the depth of the network [35]. However, to learn a deeper neural network, more image data will be essential. As the network is trained by the larger data, the higher performance it shows. In our study, to overcome the limited number of imaging data, PET images were augmented by flipping image in left-right direction (Supplementary Methods). It was based on the previous knowledge that AD and MCI converters showed symmetrically decreased FDG uptake and increased AV-45 uptake in the cerebral cortices. The network trained without this augmentation process showed 89.4% and 81.3% accuracy for the differentiation between AD and NC and predicting MCI converters, respectively. Though the augmentation process increased the performance of the network, it might cause potential error in the classification because the two brain hemispheres have partly different functions. In the future, larger image data cohort and deeper network architecture could improve the CNN-based approach. Furthermore, we used only ADNI dataset which could be insufficient for the generalized validation as well as optimal training. Therefore, as a future work, the model needs to be validated in PET images acquired from independent centers and modified by larger datasets.

5. Conclusions

Our deep CNN-based approach could accurately predict cognitive decline in MCI patients by combining information of FDG and AV-45 PET images. For testing whether a MCI subject would convert to AD, baseline PET images without spatial transformation were needed as a feature extraction was automatically performed. As a future work, our approach may be additionally validated in clinical trials with independent large cohorts. As an accurate biomarker, we expect our approach will help select appropriate prodromal AD patients who benefit from early intervention.

Role of the funding source

The funders of the study had no role in the study design, data

collection, data analysis, data interpretation, or writing of the report. All authors could access all data and programming code used in this study for the analyses.

Conflict of Interest Disclosures

None declared.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.;Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

[1] R.A. Sperling, P.S. Aisen, L.A. Beckett, D.A. Bennett, S. Craft, A.M. Fagan, T. Iwatsubo, C.R. Jack Jr, J. Kaye, T.J. Montine, D.C. Park, E.M. Reiman, C.C. Rowe, E. Siemers, Y. Stern, K. Yaffe, M.C. Carrillo, B. Thies, M. Morrison-Bogorad, M.V. Wagster, C.H. Phelps, Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimer's Dement. 7 (3) (2011) 280–292.

[2] J.H. Grand, S. Caspar, S.W. Macdonald, Clinical features and multidisciplinary approaches to dementia care, J. Multidiscip. Healthc. 4 (2011) 125–147.

[3] R.J. Perrin, A.M. Fagan, D.M. Holtzman, Multimodal techniques for diagnosis and prognosis of Alzheimer's disease, Nature 461 (7266) (2009) 916–922.

[4] C.R. Jack Jr, D.S. Knopman, W.J. Jagust, R.C. Petersen, M.W. Weiner, P.S. Aisen, L.M. Shaw, P. Vemuri, H.J. Wiste, S.D. Weigand, T.G. Lesnick, V.S. Pankratz, M.C. Donohue, J.Q. Trojanowski, Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers, Lancet Neurol. 12 (2) (2013) 207–216.

[5] A. Forsberg, H. Engler, O. Almkvist, G. Blomquist, G. Hagman, A. Wall, A. Ringheim, B. Langstrom, A. Nordberg, PET imaging of amyloid deposition in patients with mild cognitive impairment, Neurobiol. Aging 29 (10) (2008) 1456–1465.

[6] C.M. Clark, J.A. Schneider, B.J. Bedell, T.G. Beach, W.B. Bilker, M.A. Mintun, M.J. Pontecorvo, F. Hefti, A.P. Carpenter, M.L. Flitter, M.J. Krautkramer, H.F. Kung, R.E. Coleman, P.M. Doraiswamy, A.S. Fleisher, M.N. Sabbagh, C.H. Sadowsky, E.P. Reiman, S.P. Zehntner, D.M. Skovronsky, A.A.S. Group, Use of florbetapir-PET for imaging beta-amyloid pathology, JAMA 305 (3) (2011) 275–283.

[7] S. Schreiber, S.M. Landau, A. Fero, F. Schreiber, W.J. Jagust, I. Alzheimer's Disease Neuroimaging, Comparison of visual and quantitative florbetapir F 18 positron emission tomography analysis in predicting mild cognitive impairment outcomes, JAMA Neurol. 72 (10) (2015) 1183–1190.

[8] S. Ng, V.L. Villemagne, S. Berlangieri, S.T. Lee, M. Cherk, S.J. Gong, U. Ackermann, T. Saunder, H. Tochon-Danguy, G. Jones, C. Smith, G. O'Keefe, C.L. Masters, C.C. Rowe, Visual assessment versus quantitative assessment of 11C-PIB PET and 18F-FDG PET for detection of Alzheimer's disease, J. Nucl. Med. 48 (4) (2007) 547–552.

[9] A. Drzezga, N. Lautenschlager, H. Siebner, M. Riemenschneider, F. Willoch, S. Minoshima, M. Schwaiger, A. Kurz, Cerebral metabolic changes accompanying conversion of mild cognitive impairment into Alzheimer's disease: a PET follow-up study, Eur. J. Nucl. Med. Mol. Imag. 30 (8) (2003) 1104–1113.

[10] G. Chetelat, B. Desgranges, V. de la Sayette, F. Viader, F. Eustache, J.C. Baron, Mild cognitive impairment: can FDG-PET predict who is to rapidly convert to Alzheimer's disease? Neurology 60 (8) (2003) 1374–1377.

[11] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[12] H.I. Suk, S.W. Lee, D. Shen, I. Alzheimer's Disease Neuroimaging, Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis, NeuroImage 101 (2014) 569–582.

[13] V.K. Ithapu, V. Singh, O.C. Okonkwo, R.J. Chappell, N.M. Dowling, S.C. Johnson, A.s.D.N. Initiative, Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment, Alzheimer's Dement. 11 (12) (2015) 1489–1499.

[14] F. Li, L. Tran, K.H. Thung, S. Ji, D. Shen, J. Li, A robust deep model for improved classification of AD/MCI patients, IEEE J. Biomed. Health Inf. 19 (5) (2015) 1610–1616.

[15] A. Payan, G. Montana, Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks, arXiv preprint arXiv:1502.02506, (2015).

[16] E. Hosseini-Asl, M. Ghazal, A. Mahmoud, A. Aslantas, A. Shalaby, M. Casanova, G. Barnes, G. Gimel'farb, R. Keynton, A. El-Baz, Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network, Front. Biosci. (Landmark Ed.) 23 (2018) 584.

[17] S. Sarraf, G. Tofighi, Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks, arXiv preprint arXiv:1603.08631, (2016).

[18] W.J. Jagust, S.M. Landau, R.A. Koeppe, E.M. Reiman, K. Chen, C.A. Mathis, J.C. Price, N.L. Foster, A.Y. Wang, The Alzheimer's disease neuroimaging initiative 2 PET core: 2015, Alzheimer's Dement. 11 (7) (2015) 757–771.

[19] A. Vedaldi, K. Lenc, MatConvNet: convolutional neural networks for matlab, Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, ACM, 2015, pp. 689–692.

[20] S.M. Landau, D. Harvey, C.M. Madison, R.A. Koeppe, E.M. Reiman, N.L. Foster, M.W. Weiner, W.J. Jagust, I. Alzheimer's Disease Neuroimaging, Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI, Neurobiol. Aging 32 (7) (2011) 1207–1218.

[21] S.M. Landau, M.A. Mintun, A.D. Joshi, R.A. Koeppe, R.C. Petersen, P.S. Aisen, M.W. Weiner, W.J. Jagust, I. Alzheimer's Disease Neuroimaging, Amyloid deposition, hypometabolism, and longitudinal cognitive decline, Ann. Neurol. 72 (4) (2012) 578–586.

[22] S.M. Landau, M. Lu, A.D. Joshi, M. Pontecorvo, M.A. Mintun, J.Q. Trojanowski, L.M. Shaw, W.J. Jagust, I. Alzheimer's Disease Neuroimaging, Comparing positron emission tomography imaging and cerebrospinal fluid measurements of beta-amyloid, Ann. Neurol. 74 (6) (2013) 826–836.

[23] P.-J. Toussaint, V. Perlbarg, P. Bellec, S. Desarnaud, L. Lacomblez, J. Doyon, M.-O. Habert, H. Benali, Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component analyses, NeuroImage 63 (2) (2012) 936–946.

[24] I. Illán, J. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, F. Segovia, R. Chaves, M. Gómez-Rio, C.G. Puntonet, A.s.D.N. Initiative, 18 F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis, Inf. Sci. 181 (4) (2011) 903–916.

[25] P. Padilla, M. López, J.M. Górriz, J. Ramirez, D. Salas-Gonzalez, I. Álvarez, NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease, IEEE Trans. Med. Imag. 31 (2) (2012) 207–216.

[26] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (3) (1988) 837–845.

[27] P. Ripolles, J. Marco-Pallares, R. de Diego-Balaguer, J. Miro, M. Falip, M. Juncadella, F. Rubio, A. Rodriguez-Fornells, Analysis of automated methods for spatial normalization of lesioned brains, NeuroImage 60 (2) (2012) 1296–1306.

[28] S. Reig, M. Penedo, J.D. Gispert, J. Pascau, J. Sanchez-Gonzalez, P. Garcia-Barreno, M. Desco, Impact of ventricular enlargement on the measurement of metabolic activity in spatially normalized PET, NeuroImage 35 (2) (2007) 748–758.

[29] B. Cheng, M. Liu, H.I. Suk, D. Shen, D. Zhang, I. Alzheimer's Disease Neuroimaging, Multimodal manifold-regularized transfer learning for MCI conversion prediction, Brain Imaging Behav. 9 (4) (2015) 913–926.

[30] C.Y. Wee, P.T. Yap, D. Shen, I. Alzheimer's Disease Neuroimaging, Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns, Hum. Brain. Mapp. 34 (12) (2013) 3411–3425.

[31] C. Zu, B. Jie, M. Liu, S. Chen, D. Shen, D. Zhang, I. Alzheimer's Disease Neuroimaging, Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment, Brain Imaging Behav. (2015).

[32] B. Jie, D. Zhang, B. Cheng, D. Shen, I. Alzheimer's Disease Neuroimaging, Manifold regularized multitask feature learning for multimodality disease classification, Hum. Brain. Mapp. 36 (2) (2015) 489–507.

[33] J.T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, Striving for simplicity: the all convolutional net, arXiv preprint arXiv:1412.6806, (2014).

[34] C.F. Cadieu, H. Hong, D.L. Yamins, N. Pinto, D. Ardila, E.A. Solomon, N.J. Majaj, J.J. DiCarlo, Deep neural networks rival the representation of primate IT cortex for core visual object recognition, PLoS Comput. Biol. 10 (12) (2014) e1003963.

[35] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, arXiv preprint arXiv:1409.4842, (2014).