ORIGINAL ARTICLE

Detection of Alzheimer's disease using ECD SPECT images by transfer learning from FDG PET

Yu-Ching Ni[1,2] · Fan-Pin Tseng[1] · Ming-Chyi Pai[3,4] · Ing-Tsung Hsiao[5,6] · Kun-Ju Lin[5,6] · Zhi-Kun Lin[1] · Wen-Bin Lin[1] · Pai-Yi Chiu[7] · Guang-Uei Hung[8] · Chiung-Chih Chang[9] · Ya-Ting Chang[10] · Keh-Shih Chuang[2] · For the Alzheimer's Disease Neuroimaging Initiative

Abstract

Objective To develop a practical method to rapidly utilize a deep learning model to automatically extract image features based on a small number of SPECT brain perfusion images in general clinics to objectively evaluate Alzheimer's disease (AD).

Methods For the properties of low cost and convenient access in general clinics, Tc-99-ECD SPECT imaging data in brain perfusion detection was used in this study for AD detection. Two-stage transfer learning based on the Inception v3 network model was performed using the ImageNet dataset and ADNI database. To improve training accuracy, the three-dimensional image was reorganized into three sets of two-dimensional images for data augmentation and ensemble learning. The effect of pre-training parameters for Tc-99m-ECD SPECT image to distinguish AD from normal cognition (NC) was investigated, as well as the effect of the sample size of F-18-FDG PET images used in pre-training. The same model was also fine-tuned for the prediction of the MMSE score from the Tc-99m-ECD SPECT image.

Results The AUC values of w/wo pre-training parameters for Tc-99m-ECD SPECT image to distinguish AD from NC were 0.86 and 0.90. The sensitivity, specificity, precision, accuracy, and F1 score were 100%, 75%, 76%, 86%, and 86%, respectively for the training model with 1000 cases of F-18-FDG PET image for pre-training. The AUC values for various sample sizes of the training dataset (100, 200, 400, 800, 1000 cases) for pre-training were 0.86, 0.91, 0.95, 0.97, and 0.97. Regardless of the pre-training condition ECD dataset used, the AUC value was greater than 0.85. Finally, predicting cognitive scores and MMSE scores correlated ($R^2 = 0.7072$).

Conclusions With the ADNI pre-trained model, the sensitivity and accuracy of the proposed deep learning model using SPECT ECD perfusion images to differentiate AD from NC were increased by approximately 30% and 10%, respectively. Our study indicated that the model trained on PET FDG metabolic imaging for the same disease could be transferred to a small sample of SPECT cerebral perfusion images. This model will contribute to the practicality of SPECT cerebral perfusion images using deep learning technology to objectively recognize AD.

Keywords ECD SPECT images · Transfer learning · Alzheimer's disease

✉ Yu-Ching Ni
janet@iner.gov.tw

✉ Ming-Chyi Pai
pair@mail.ncku.edu.tw

Extended author information available on the last page of the article

Introduction

Alzheimer's disease (AD) is a neurodegenerative disease of the brain and the most common type of dementia, accounting for more than 60% of all dementia cases [1]. SPECT and PET have been used to study brain regional cerebral blood flow (rCBF) and regional cerebral glucose metabolism (rCGM) in AD patients and other brain neurodegenerative diseases, showing that the characteristic patterns of perfusion and metabolic abnormalities can distinguish AD from other types of dementia. Hence, as a marker of AD, SPECT and PET imaging can be used to

detect basic changes of brain blood perfusion and metabolism, as well as for the differential diagnosis, monitoring of disease progression, and response to treatment. The studies of SPECT brain perfusion and PET metabolism are usually consistent in abnormal areas [2]. F-18-FDG (Fluorodeoxyglucose, FDG) PET, which is typically used in the West for brain glucose metabolism examination, is currently not covered by National Health Insurance in Taiwan [3]. Therefore, most nuclear medicine departments in Taiwan use SPECT cerebral perfusion imaging and the Tc-99m-ECD (ethyl cysteinate dimer, ECD) tracer. Although SPECT has a longer imaging time and poorer image resolution than PET, it is low cost and the tracer is easily accessible, hence, is widely used in domestic clinical practice.

In recent years, artificial intelligence (AI) and radiomics technology, such as image mining, have been applied to medical imaging to identify non-invasive features of diseases. Image mining is considered to have great potential clinical significance, as it can be used for non-invasive diagnosis, feature extraction, and outcome prediction under almost all medical conditions [4]. The multidisciplinary aspect of clinical neuroscience has begun to be influenced by deep learning and is moving toward the development of new diagnostic and prognostic tools. Deep learning technology is particularly promising in neuroscience because clinical diagnosis relies on subtle symptoms and complex neuroimaging methods [5].

Although deep learning technology can automatically extract features from the original data, a large amount of data needs to be prepared for deep learning model training, which is a major barrier for nuclear medicine imaging applications. At present, the world's largest open dataset of dementia images is ADNI (Alzheimer's Disease Neuroimaging Initiative), which has been used to support numerous deep learning training studies. Researchers have used brain glucose metabolism images to develop many algorithms for the early detection of AD, differentiation of dementia types, as well as the prediction of disease progression [6, 7], however, the research is restricted due to the lack of open datasets of cerebral blood flow images. Recent studies have found that the image features extracted after training with a large amount of data can be transferred to a small amount of data in other fields [8, 9], thus, providing an opportunity to overcome the barrier to deep learning using a small dataset of cerebral blood flow images.

This study aimed to develop a practical deep learning approach to differentiate images between AD and normal cognition (NC) and predict MMSE (Mini-mental State Examination) scores using cerebral blood flow images. The lack of a large dataset of Tc-99 m-ECD SPECT images was overcome via two-stage transfer learning technology, extracting the features from a larger dataset

(F-18-FDG PET images from ADNI) for transfer to other domains with a smaller dataset (such as Tc-99m-ECD SPECT images).

Materials and methods

Subjects

The Tc-99m-ECD SPECT images (total 247 subject; 113 AD; 134 NC) were part of the Taiwanese Nuclear Medicine Brain Image Database collected and built by the Institute of Nuclear Energy Research. All participants were evaluated by neurologists and clinical psychologists, and their education level was elementary school or above. People with normal cognitive function were assessed to rule out physical conditions that cannot be corrected and may cause dementia or delirium, such as poor vision, abnormal hearing, hypothyroidism, anemia, pneumonia, fever, dehydration, signs of abnormal liver function, abnormal renal function, signs of heart failure (NY class < 3), etc. Those with obvious head trauma, neurological diseases related to dysfunction of the extrapyramidal system or autonomic nervous systems, such as hydrocephalus, Parkinson's disease, cortical basal ganglia degeneration, and progressive supranuclear palsy, Vitamin B12 or folic acid deficiency caused by subacute combined degeneration, multiple system degeneration, and cerebrovascular diseases that may cause various local neurological symptoms were excluded. The systolic pressure of those with hypertension needed to be controlled below 160 mmHg, and the HbA1c of those with diabetes mellitus below 9.0. Those on medications that may cause cognitive dysfunction, such as anticholinergic drugs, hypnotics, or antipsychotics, were excluded. The Critical Mental Illness Scale (CHQ-12) score should be < 3, and all participants completed the clinical dementia rating (CDR) scale to determine the severity. Participants with clinically suspected AD received complete medical history inquiry (including important system and brain disease history and CDR), cognitive function (such as MMSE scores), and related examinations. Those who met the criteria further underwent Tc-99m-ECD SPECT imaging, and the images were interpreted by nuclear medicine experts. The demographic characteristics and clinical characteristics of the data are shown in Table 1. The Institutional Review Board (IRB) of National Cheng Kung University Hospital approved this study (serial number: NCKUH IRB B-BR-107–030).

The F-18-FDG PET images of AD and NC (total 1,267 subject; 638 AD; 629 NC) used for pre-training in this study were obtained from ADNI, a public database, and the demographic and clinical characteristics of the data are shown in Table 2.

Table 1 Demographic and clinical characteristics of the Tc-99m-ECD SPECT data

Characteristic	NC	AD
Number of subjects	134	113
Age at the time of SPECT (years)	67.0 ± 8.5	74.4 ± 7.0
Sex (F:M)	88:46	58:55
MMSE	27.5 ± 2.4	19.2 ± 5.3
CDR	0.22 ± 0.25	0.79 ± 0.39

Table 2 Demographic and clinical characteristics of F-18-FDG PET data

Characteristic	NC	AD	MCI
Number of subjects for image classification (MMSE $\geqq$15 for MMSE score prediction)	629 (613)	638 (599)	(654)
Age at the time of PET (years)	76.3 ± 5.8	76.2 ± 7.5	75.9 ± 7.8
Sex (F:M)	263:366	253:385	236:418
MMSE	28.5 ± 4.0	21.7 ± 5.4	27.1 ± 2.1
CDR	0.03 ± 0.16	0.86 ± 0.43	0.49 ± 0.10

Image acquisition and processing

F-18-FDG PET images were downloaded from the ADNI database (http://adni.loni.usc.edu 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 magnetic resonance imaging (MRI), PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and AD.). The images provided various levels of pre-processing and were averaged by a dynamic PET 3D scan (6 frames taken 30 to 60 mins after injection 185 MBq FDG), aligned and resampled so that the images had the same pixel size, with the intensity of the image values normalized to global brain FDG uptake. Each image was normalized to the MNI (Montreal Neurological Institute) space by SPM (Statistical Parametric Mapping). The image dimension was $91 \times 109 \times 91$ changing to $95 \times 95 \times 48$ after cropping, padding, and removal of the image slices above the skull and below the cerebellum to retain most of the brain parenchymal area.

Tc-99m-ECD SPECT images were acquired from four medical institutions and obtained by E-CAM, Symbia T16, and Symbia T2 SPECT equipment (Siemens Medical Solutions, Malvern, PA, USA) with LEHR (low energy high resolution) and fan beam collimators. Fifteen minutes after intravenous injection of 925 MBq Tc-99m-ECD,

SPECT images were acquired for 30 to 40 mins, and the image matrix size was 128×128. The images were reconstructed by filtered back projection (FBP) with Metz filter and ordered subsets expectation and maximization (OSEM) method using Chang's attenuation correction (attenuation coefficient is 0.1 cm^{-1}). The original image was processed by SPM (SPM8, University College of London, London, UK) software for spatial normalization and the Z-score method used for intensity standardization. The image values were scaled to a distribution with an average value of zero and a standard deviation of one. The image was resampled to $95 \times 95 \times 68$ with the voxel size $2 \times 2 \times 2$ mm^3, and the image slices above the cranium and below the cerebellum were removed to retain most of the parenchymal area, giving a final image dimension of $95 \times 95 \times 48$.

Deep learning model training

The design concept of this study first used the convolutional neural network (CNN) model, the architecture of which adopted the design of Inception v3, pre-trained by ImageNet data, a daily life image dataset containing approximately 14 million images labeled into 1000 categories to establish automatic extraction of low-level image features. Then, the model was transferred and re-trained using the F-18-FDG PET image from the ADNI database to learn the characteristics of nuclear medicine images from the F-18-FDG image. Finally, a suitable deep learning model for the Tc-99m-ECD SPECT images was transferred and the training weights of the aforementioned learning were applied to compensate for the limitation of the small amount of data.

Training a two-dimensional (2D) CNN model requires lower training computing resources and using 2D images occupies less memory compared to a 3D CNN model and images. To retain the information of the whole brain slices, the brain parenchymal area with image dimension $95 \times 95 \times 48$ was equally divided into 16 sections, with one image selected for each section, then 4×4 slices were reassembled to a 2D image. Sixteen slices were sorted in order from the caudal of the brain to the cranial as shown in Fig. 1a. The 48 slices were divided into three sets of 2D images with similar structures and positions in the order of [#1, 4, 7…], [#2, 5, 8…], [#3, 6, 9…]. Finally, a 3D F-18-FDG PET image was divided into three 2D images with a 380×380 matrix size. The Tc-99m-ECD SPECT images were also reassembled in the same way, as shown in Fig. 1b. The computing resources only required four cores of Intel Xeon 6230 2.1 GHz processor, 48 GB of DDR4 memory, and an NVIDIA 2080Ti computing card. In addition to using Tc-99m-ECD SPECT images to predict AD, the cognitive function score, MMSE, was also predicted.

The development environments were all executed under Python 3 using Keras 2.2.5 to build neural networks and
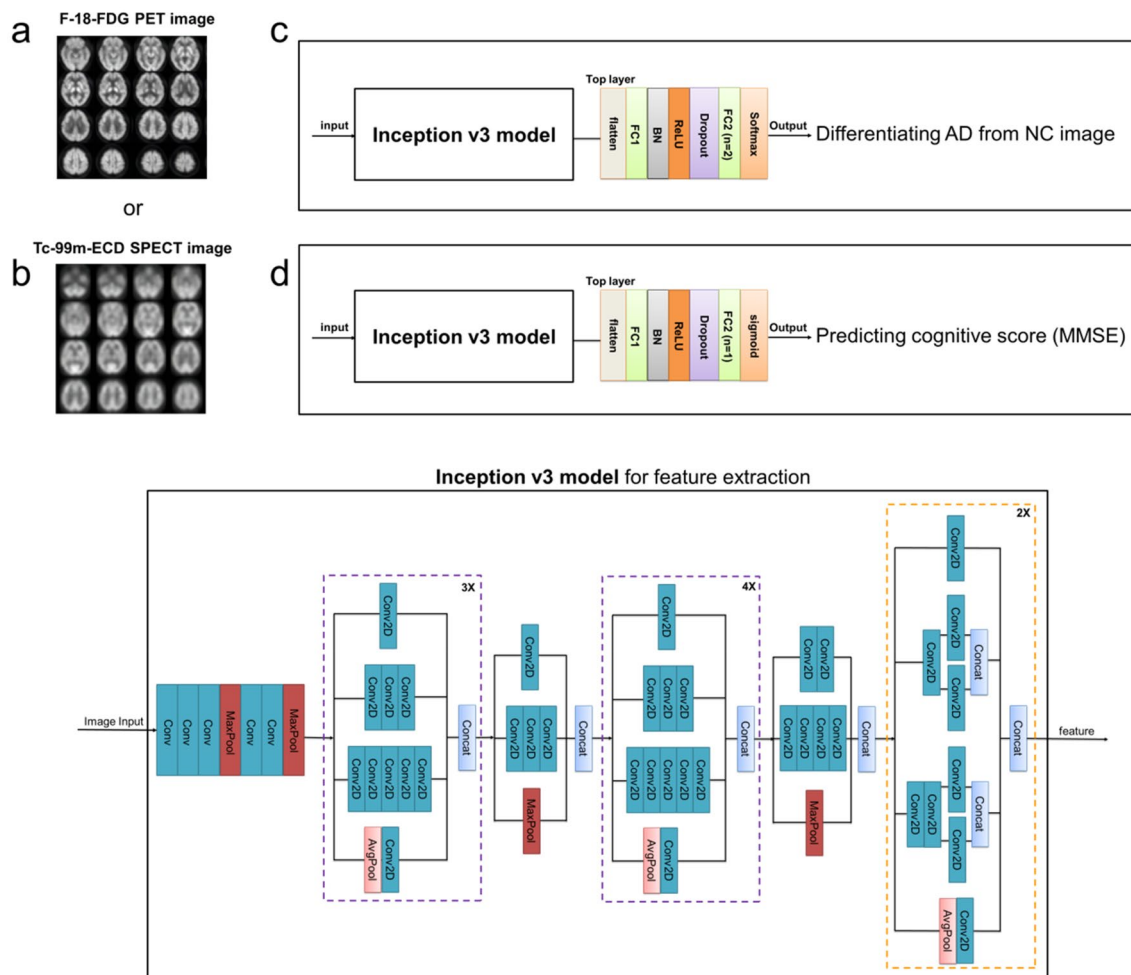
Fig. 1 a, b The two-dimensional input image reassembled by 4×4 slices of F-18-FDG PET and Tc-99m-ECD SPECT images respectively. **c** The CNN architecture, Inception v3 model was used in this study. There were 11 modules and each module was concatenated with different numbers of convolutional layers and average pooling or max pooling. A fully connected layer (FC) with a length of 128/256 was connected to the top layer of the Inception v3, and the batch normalization (BN) followed by activation function (ReLU) and dropout layer added after the FC. Finally, the Softmax function was used for classifying AD and NC. **d** The model for the predicting cognitive score was almost the same as **c** except the last layer used was sigmoid

import pre-trained models, and the backend runs as Tensor-Flow 1.15.2 (Google, Mountain View, Calif).

Pre-trained model using FDG PET images

To investigate the effect of pre-training with different numbers of images on the performance of subsequent transfer learning, 1000 F-18-FDG PET images (AD = 511; NC = 489) were divided into several training datasets, 100, 200, 400, 800, and 1000 cases respectively, with 80% of the data in each training dataset used for training and 20% for validation. The range of random width and height shift of data augmentation was 0–0.03% and the range of zooming was 1–1.03%. The CNN architecture is shown in Fig. 1c. A fully connected layer (FC) with a length of 256 was connected to the top layer of the Inception v3 model, and the batch normalization (BN) and dropout layer were added after the FC. The dropout layer was set to 0.5, as a form of regulation, to avoid neural network coadaptation by randomly removing nodes for a more robust model.

The loss function used categorical cross-entropy, and the optimization algorithm used Adaptive Moment Estimation (Adam) [10], the learning rate was set to 0.0000005, and the batch size was set to 8 for model training. The early stopping mechanism was used to judge the stop and choose a suitable epoch. The trained model was tested with an independent dataset ($n = 267$; AD = 118; NC = 149) and its performance was evaluated by accuracy.

Transferred AI model for ECD SPECT images

Fifty images (AD = 22; NC = 28) were randomly selected as an independent test set and 80% of the remaining 197 images (AD = 91; NC = 106) were used for training and 20% for validation. The training model was the same except the FC changed the length size to 128. The loss function and optimization algorithm selection were the same as above.

Transferred AI model for MMSE Prediction

For predicting the MMSE scores of Tc-99m-ECD SPECT images, F-18-FDG PET images of AD and NC were used for pre-train, also including the MCI data and only considering an MMSE score ≥ 15. The model was the same as the image classification task but used regression for predicting the output of the FC2 shown in Fig. 1d. The model training settings were similar except the loss function was Mean Squared Error, the learning rate was set to 0.0000005, and the learning performance evaluated by Mean Absolute Error.

Model interpretation and features visualization

The nonlinear dimensionality reduction algorithm t-distributed stochastic neighbor embedding (t-SNE) [11] is suitable for dimension reduction of high-dimensional data to two dimensions for visualization. In this study, image features extracted from each image (including AD and NC) through the CNN model were dimension reduced to two dimensions by t-SNE using package scikit-learn [12], allowing visual observation of the scattered location of each image to evaluate the similarity between the data.

To more intuitively observe how the CNN model deduces its classification results, we used Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize the regions of the image that were important for prediction as AD or NC from the model. Selecting the last concatenate layer of the Inception v3 model generated 2048 feature maps with a matrix size of 10×10. The class-feature heatmap of each image was a weighted combination of the weights, computing gradient backpropagation from the last layer to the above-mentioned concatenate layer, and 2048 feature maps. The result showed the distribution of the most relevant image pixels when the image was classified as AD or NC.

Model testing and result evaluation

The accuracy of the model was evaluated by receiver operating characteristic (ROC) curves and the area under the ROC curve (AUC). The ROC curve was plotted with 95% confidence intervals (CI) calculated using MATLAB (MATLAB R2020a, MathWorks, Natick, Massachusetts, USA) with 1,000 iterations of bootstrapping. Also, statistical analysis was performed on the results of the classification prediction, including the calculation of the sensitivity, specificity, precision, accuracy, and F1 score.

The MMSE scores predicted by the deep learning model and the actual MMSE scores of the image dataset were plotted as a scatter chart, with the linear relationship R^2 calculated to evaluate the performance of the prediction model. The training and testing datasets of F-18-FDG PET and Tc-99m-ECD SPECT images were processed respectively.

Results

Model interpretation and features visualization

The features extracted from each F-18-FDG PET training image via the Inception v3 model were displayed in t-SNE as shown in Fig. 2a. NC and AD data are observed as two clusters, indicating that the features of AD and NC can be distinguished after training. Figure 2b shows the feature of the images used for testing in the ADNI dataset, with AD data in the upper right and NC data in the lower left of the figure but there was a partial mixing of the clusters. The feature distributions of the Tc-99m-ECD SPECT image training and testing datasets are shown in Fig. 2c and 2d, with the distinction between the NC and AD data in the training dataset more obvious than in the testing dataset.

The Grad-CAM heatmaps of AD and NC cases with correct prediction results are shown in Fig. 3. ECD image features related to classifying AD or NC were localized on the individual ECD images. The regions of individuals highlighted by heatmaps were slightly different. Overall, the CNN model focuses on the regions of the parietal lobe and temporal lobe, which is consistent with the general clinical interpretation of AD in temporoparietal region hypoperfusion.

Recognition of AD

The ROC curves of the pre-trained model using F-18-FDG PET images with different numbers of image datasets are shown in Fig. 4a, with the AUC values and performance evaluation index listed in Table 3. The model trained by 1000 cases performed best. Furthermore, the ROC curves of models w/wo pre-training parameters for Tc-99m-ECD SPECT images are shown in Fig. 4b, with the AUC values and performance evaluation index listed in Table 4. Among them, the sensitivity, specificity, precision, accuracy, and F1 score were 77% (17/22), 79% (22/28), 74% (17/23), 78% (39/50), and 76 for the CNN model without ADNI pre-training; 100% (22/22), 75% (21/28), 76% (22/29), 86% (43/50) and 86 for CNN model with 1000 cases of F-18-FDG PET images for pre-training.
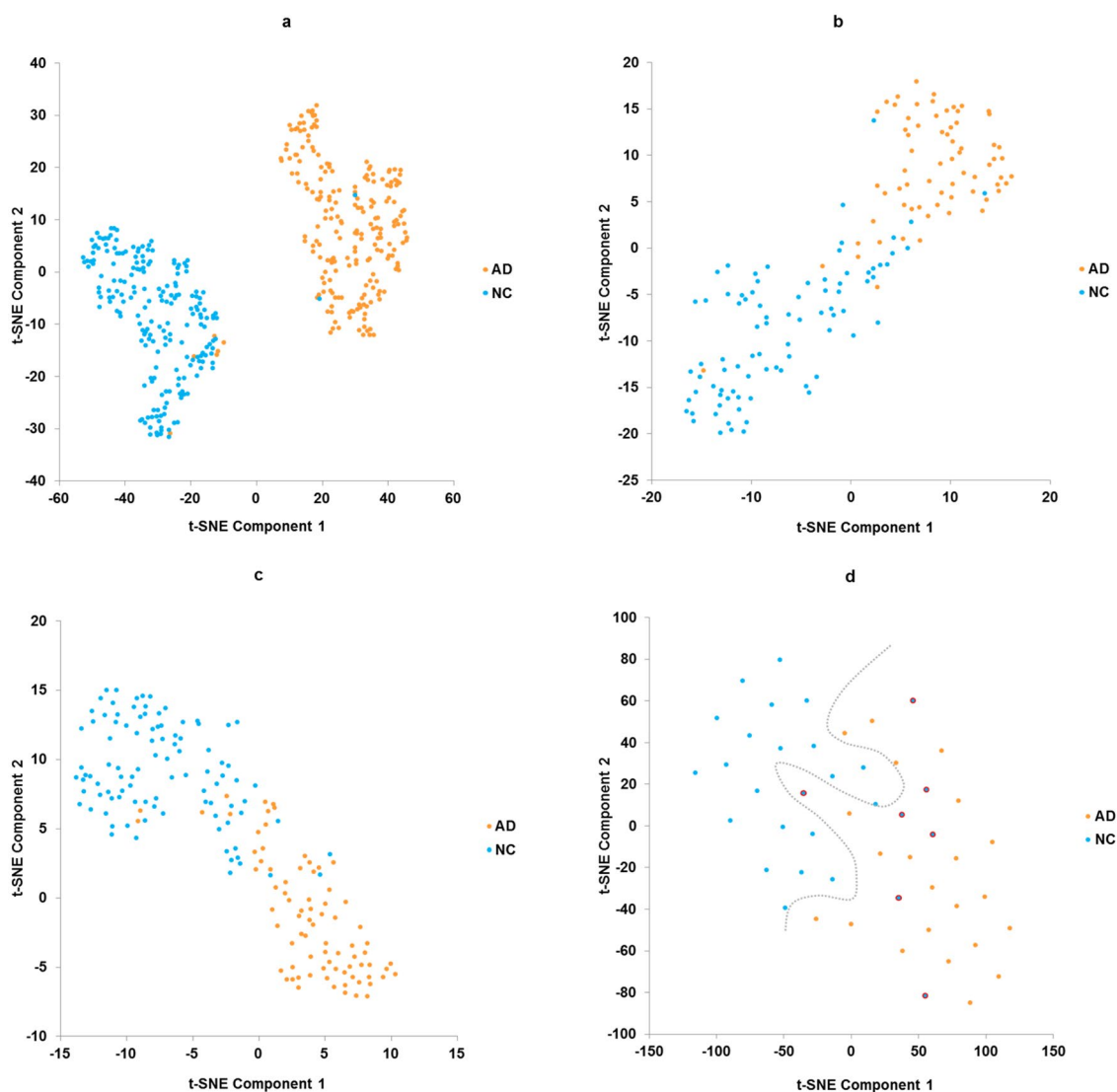
Fig. 2 **a**, **b** Image features (256 features of FC1 for each image) visualization of the ADNI training dataset (1,000 cases) and testing dataset (267 cases) after dimension reduction with t-SNE. The AD and NC images were distinctively clustered. **c**, **d** Image features (128 features of FC1 for each image) visualization of the ECD training dataset (197 cases) and testing dataset (50 cases) after dimension reduction with t-SNE. The AD cluster distinguished from the NC cluster slightly

Predicting MMSE score

F-18-FDG PET images and their respective MMSE scores were used to fine-tune the image classification model to predict the cognitive score. Figure 5a shows the prediction plotted with MMSE score for training dataset ($n = 1493$), and Fig. 5b shows the independent test dataset ($n = 373$), with R^2 values of 0.8967 ($p < 1 \times 10^{-4}$) for training data and 0.5129 for testing data ($p < 1 \times 10^{-4}$). Figure 5c and 5d show the scatter plot of predicting cognitive score and actual MMSE score using Tc-99m-ECD SPECT image, with R^2 values of 0.7072 ($p < 1 \times 10^{-4}$) for training data and 0.2225 for testing data ($p < 5.4 \times 10^{-4}$).

Discussion

The prerequisite for a deep learning model to automatically learn about disease features from data is to have a large amount of data to train the model. The lack of a large dataset of Tc-99m-ECD SPECT images was overcome by using two-stage transfer learning technology, extracting the features from a larger dataset (F-18-FDG PET images from ADNI) for transfer to other domains with a smaller dataset (such as Tc-99m-ECD SPECT images). This study using conventional hardware equipment and more than 200 cases of Tc-99m-ECD SPECT image data, reorganized 3D images into three sets of 2D images for data augmentation to

Fig. 3 The Grad-CAM heat-maps of AD and NC cases with correct prediction results. The number in parentheses was the probability of the image as judged by the CNN model to be within this classification. ECD image features related to classifying AD or NC were localized on individual ECD images
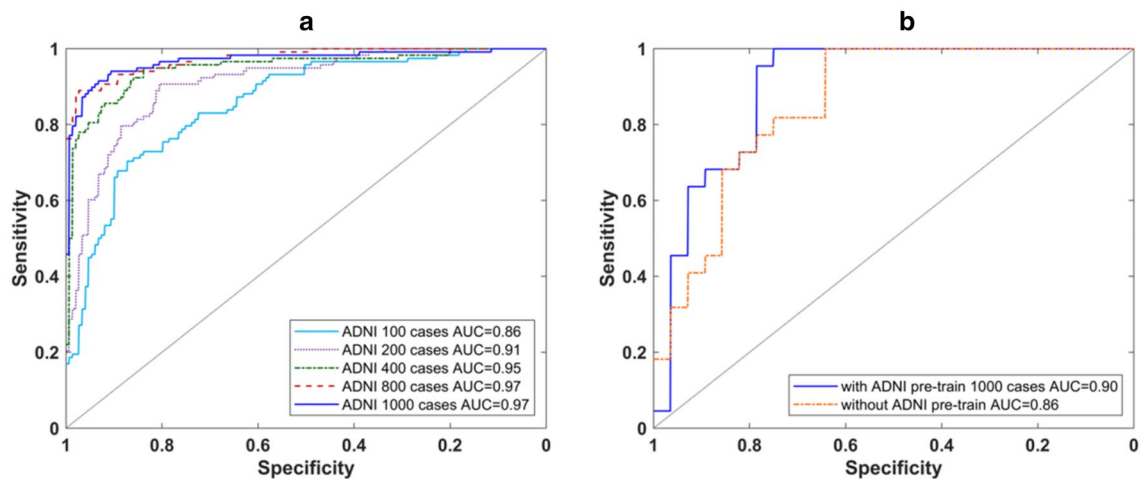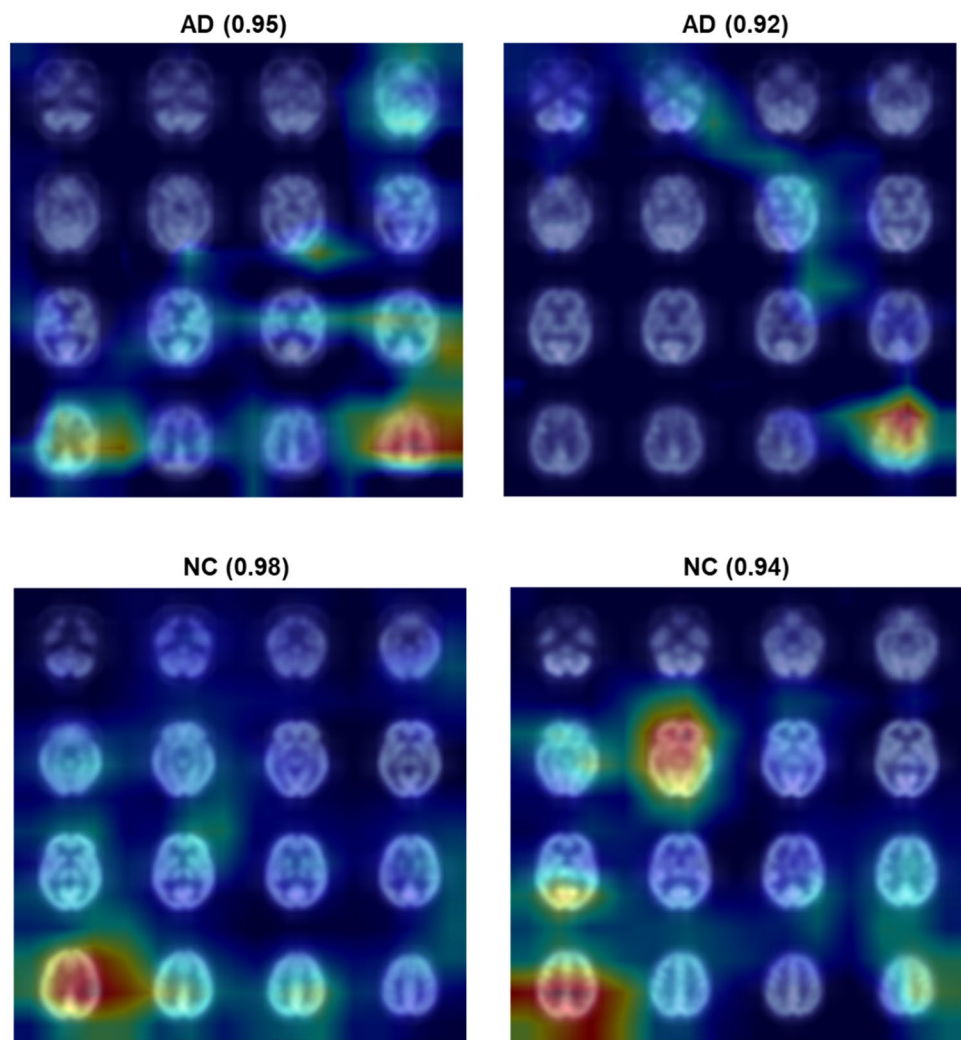

AD (0.95)   AD (0.92)

NC (0.98)   NC (0.94)

a

b

Fig. 4 **a** ROC curves of the deep learning model Inception v3 trained on the ADNI testing dataset to differentiate AD from NC, showing the effect of the size of the training dataset, with a larger dataset hav- ing a higher AUC value. **b** ROC curves of the deep learning model, Inception v3 combined w/wo pre-training parameters, trained on ECD testing datasets to differentiate AD from NC

Table 3 Comparison of the training performance of ADNI datasets

Cases number of ADNI dataset used for training	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)	F1 score (%)	AUC for AD/NC (95% CI)
100 cases	88 (104/118)	62 (93/149)	65 (104/160)	74 (197/267)	75	0.86 (0.81–0.90)
200 cases	91 (107/118)	80 (119/149)	78 (107/137)	85 (126/267)	84	0.91 (0.87–0.94)
400 cases	92 (108/118)	86 (128/149)	84 (108/129)	88 (136/267)	87	0.95 (0.91–0.97)
800 cases	92 (109/118)	89 (133/149)	87 (109/125)	91 (142/267)	90	0.97 (0.95–0.99)
1000 cases	93 (110/118)	91 (136/149)	89 (110/123)	92 (246/267)	91	0.97 (0.94–0.98)
636 cases (243 AD 393 NC) Choi et al. [9]	–	–	–	–	–	0.94 (0.89–0.98)
193 cases (93 AD 100 NC) Feng et al. [7]	93.33	91.26	–	92.23	91.8	0.9651
193 cases (93 AD 100 NC) Liu et al. [13]	91.4	91	–	91.2	–	0.953
1921 cases (AD, MCI, non-AD/MCI) Ding et al. [6]	81 (29/36)*	94 (143/152) *	76 (29/38) *	–	78*	0.92*

*In this paper, the main goal is to distinguish AD, MCI, and non-AD/MCI, here except only the results of recognizing AD

Table 4 Comparison of the training performance of ECD datasets in various pre-training models

Method	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)	F1 score (%)	AUC for AD/NC (95% CI)
Inception v3 model (without ADNI pre-train)	77 (17/22)	79 (22/28)	74 (17/23)	78 (39/50)	76	0.86 (0.72–0.93)
With ADNI pre-train 100 cases	91 (20/22)	68 (19/28)	69 (20/29)	78 (39/50)	78	0.85 (0.71–0.93)
With ADNI pre-train 200 cases	100 (22/22)	64 (18/28)	69 (22/32)	80 (40/50)	81	0.89 (0.75–0.95)
With ADNI pre-train 400 cases	100 (22/22)	68 (19/28)	71 (22/31)	82 (41/50)	83	0.86 (0.72–0.94)
With ADNI pre-train 800 cases	95 (21/22)	75 (21/28)	75 (21/28)	84 (42/50)	84	0.86 (0.72–0.94)
With ADNI pre-train 1000 cases	100 (22/22)	75 (21/28)	76 (22/29)	86 (43/50)	86	0.90 (0.77–0.97)
3 layers DNN*,+	95.12	75.0	–	83.51	–	–
Naïve Bayes+	68.29	91.07	–	81.44	–	–
Decision Trees+	78.05	85.71	–	82.47	–	–
SVM+	82.92	82.14	–	82.47	–	–

*Deep Neural Network +Segovia F et al. 2017 [14]

improve the accuracy of the training results. The respective prediction of three sets of 2D images from the same subject for ensemble learning improved the accuracy, which is helpful for deep learning training with a small amount of data. However, compared with 3D CNN training for complete 3D images, this study reassembled 16 slices of 3D images at equal intervals into 2D images for classification training. As a limitation, each subject's image should undergo spatial normalization processing to ensure that the location of the regional structure on each 2D image has consistent

features so that each slice image still has sufficient features to provide CNN models for learning to distinguish AD and NC. In addition, after the slices were rearranged, the order of the brain structure was destroyed, and the characteristics of inter-slice could not be learned from it, which may affect the performance of its deep learning training.

Using t-SNE to display the feature distribution of the data after dimension reduction can intuitively help users understand the pros and cons of data grouping by category after deep learning model training, quickly identifying the
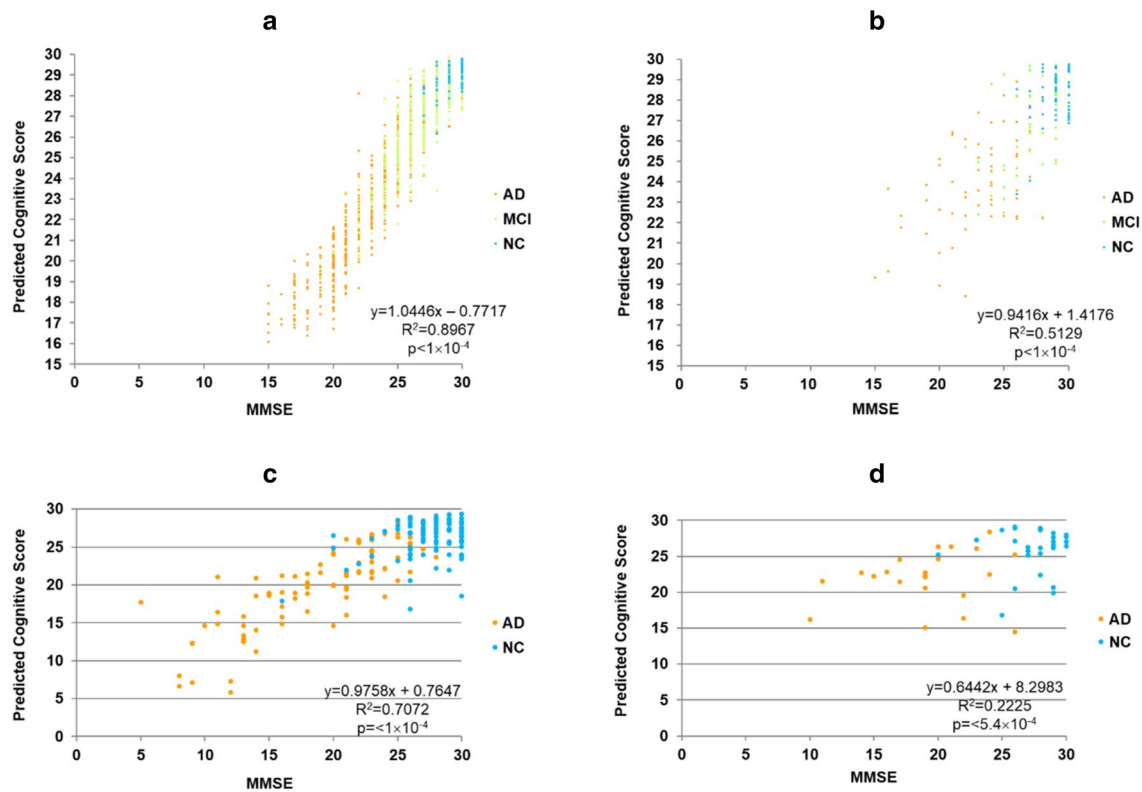
Fig. 5 Fine-tuning of the image classification model for predicting MMSE. **a**, **b** The last layer of the model was tuned to predict individual MMSE scores using the ADNI training and testing dataset. The output of the fine-tuned model was plotted with MMSE score, and was significantly correlated (training data: $R^2 = 0.8967$, $p < 1 \times 10^{-4}$;

testing data: $R^2 = 0.5129$, $p < 1 \times 10^{-4}$). **c**, **b** For ECD training and testing dataset, the output of the fine-tuned model was plotted with MMSE score and the output of the model correlated with the MMSE score (training data: $R^2 = 0.7072$, $p < 1 \times 10^{-4}$; testing data: $R^2 = 0.2225$, $p < 5.4 \times 10^{-4}$)

incorrectly predicted cases. For example, Fig. 2d shows the feature distribution of the ECD testing dataset trained by the Inception v3 model combined with the pre-training parameters, there were seven NC data points (blue dots with red borders) below the auxiliary line (gray dotted line) misjudged as AD. The age of these patients ranged from 61 to 84 years old, four of them had a CDR score of 0.5, and another person had obvious hypoperfusion in the left frontal lobe. These cases are difficult to evaluate by the model, consequently, changes in such cases require further follow-up.

The Grad-CAM heatmaps help us to explain the decision of the CNN model. Although the reassembled 2D lost the characteristics of inter-slice, whether AD-related or NC-related regions seemed to focus on the slices including the parietal lobe and temporal lobe. It should be noted that although these patterns are similar to the typical hypoperfusion area judged by ECD images as AD, they are still not directly regarded as abnormal image judgments from clinical use. The visualization of AD-related regions could indirectly prove the CNN model captured appropriate brain regions for decision and comparison with domain knowledge.

The overall comparison of the training performance of the ADNI and ECD datasets in various training models showed that for the ADNI database, the larger the amount of data used for training resulted in better training performance, with 1000 cases achieving the highest AUC value 0.97. In a similar study, the AUC value of AD/others was 0.92 [6] when 1921 cases of F-18-FDG PET images of AD, MCI, and non-AD/MCI were used. Also, 636 F-18-FDG PET images from the ADNI database used to train the 3D CNN model resulted in an AUC value of AD/NC of 0.94 [9]. In addition, literature [7] and [13] used a complex deep learning model of CNN and recurrent neural network (RNN) training 193 F-18-FDG PET images from the ADNI database, and their AUC values were higher than 0.95.

In summary, the developed model performed better than previously reported models (using about 2 times the data) [6], with high sensitivity (93 vs. 81%), specificity (91 vs 94%), precision (89 vs. 76%), accuracy (92%), and the F1 score (91 vs. 78) for F-18-FDG PET images. Other studies [7, 13] used a hybrid model with more complex methods to train 193 cases of F-18-FDG PET images, reporting a sensitivity of 93.33%, a specificity of 91.26%, accuracy of

92.23%, and an F1 score of 91.8 [7]. The performance of our model with a similar-sized training dataset (200 cases of F-18-FDG PET images) yielded a sensitivity of 91%, a specificity of 80%, accuracy of 85%, and an F1 score of 84. Although other studies [7, 13] used less data, the classification performance of the hybrid model of 3D CNN (without dimension reduction, more information was retained) and RNN (consider the information of inter-slice and multi-view) was better than our 2D CNN model (easier to approach and implement).

Moreover, the training performance of the ECD dataset, the AUC value was higher than 0.85 regardless of training conditions. Although their confidence interval was wide, it matched the performance characteristics obtained with less data. In addition the sensitivity of 100%, a specificity of 75%, precision of 76%, accuracy of 86%, and F1 score of 86 for 1000 cases pre-training. Another study [14] used Tc-99m-ECD SPECT images for a deep learning method to diagnose AD, reporting a sensitivity of 95.12%, specificity of 75%, and accuracy of 83.51%. The results of our method were slightly better than those of the literature. Although it may be because the image quality of Tc-99m-ECD SPECT images is worse than F-18-FDG PET images, the error rate of judging NC is higher, resulting in specificity not being improved, and more test data is needed for further research and verification. Finally, in terms of the impact of transfer learning, the use of ADNI pre-training parameters resulted in better prediction results, increasing the sensitivity by 30%, and the accuracy by 10%.

In this study, apart from using a model trained by transferring a larger amount of ADNI F-18-FDG PET images to Tc-99m-ECD SPECT images to identify AD, it also used the same method to estimate the MMSE score of the Tc-99m-ECD SPECT images. This study demonstrated the clinical feasibility of transferring different biomarker characteristics based on deep learning by validating the model using Tc-99m-ECD SPECT images to identify AD. The problem with nuclear medicine images is that they are three-dimensional with different image characteristics from the clear-morphology image, so it was not appropriate to transfer the model trained by the ImageNet directly. It has been demonstrated that it is possible to transfer the model trained by relatively large F-18-FDG PET images consisting of AD and NC to Parkinson's disease patients [9]. Thus, the same metabolic characteristics of F-18-FDG PET images can be shared in different diseases. Our study indicated that the model trained using a large amount of PET metabolic imaging data for the same disease can be transferred to a small number of SPECT cerebral perfusion images. This model will contribute to the practicality of SPECT cerebral perfusion images and a relatively small dataset, using deep learning technology to objectively assess cognitive dysfunction. In the future, this research will be extended to investigate the use of deep learning technology in SPECT brain perfusion imaging to evaluate the cognitive function and severity of various neurodegenerative diseases, such as AD, vascular dementia, and dementia with Lewy bodies.

Conclusion

This study proposed two-stage transfer learning to develop a model to objectively assess AD. A large amount of PET metabolic imaging data was transferred to a small number of SPECT cerebral perfusion images, increasing the sensitivity and the accuracy of the model to recognize AD. This model will contribute to the application of SPECT cerebral perfusion images using deep learning technology.

Declarations

Ethical approval and informed consent This research plan has been approved by the Human Testing Committee and has been numbered as NCKUH IRB B-BR-107–030 before proceeding. All participants have been carefully explained by the researcher and signed informed consent.

Conflicts of interest There are no conflictsof interest to declare.

References

1. Taiwan Alzheimer Disease Association. Handbook of dementia diagnosis and treatment. Taipei City: Ministry of Health and Welfare; 2017.
2. De La Monte SM. The clinical spectrum of Alzheimer's disease-the charge toward comprehensive diagnostic and therapeutic strategies, chapter9. IntechOpen; 2011
3. Huang SH. Introduction of nuclear medicine brain scan. Chang Gung Med News. 2017;38(11):354–5.
4. Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: a systematic review on

artificial intelligence and radiomics. Eur J Nucl Med Mol Imaging. 2019;46(13):2656–72.

5. Valliani AA, Ranti D, Oermann EK. Deep learning and neurology: a systematic review. Neurol Ther. 2019;8(2):351–65.

6. Ding Y, Sohn JH, Kawczynski MG, et al. A deep learning model to predict a diagnosis of Alzheimer's disease by using 18 F-FDG PET of the brain. Radiology. 2019;290(2):456–64.

7. Feng C, Elazab A, Yang P, et al. Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM. IEEE Access. 2019;7:63605–18.

8. Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. Proc IEEE CVPR. 2014:1717–24.

9. Choi H, Kim YK, Yoon EJ, Lee JY, Lee DS. 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(2):403–12.

10. Kingma D, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv. 2014;1412.6980.

11. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9:2579–605.

12. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.

13. Liu M, Cheng D, Yan W. Alzheimer's disease neuroimaging initiative. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. Front Neuroinform. 2018;12(35):1–12.

14. Segovia F, García-Pérez M, Górriz JM, Ramírez J, Martínez-Murcia FJ. Assisting the diagnosis of neurodegenerative disorders using principal component analysis and tensorflow. In: Graña M, López-Guede JM, Etxaniz O, Herrero A, Quintián H, Corchado E, editors. International joint conference SOCO'16-CISIS'16-ICEUTE'16. San Sebastián: Springer; 2017. p. 43–52.

Authors and Affiliations

Yu-Ching Ni[1,2] · Fan-Pin Tseng[1] · Ming-Chyi Pai[3,4] · Ing-Tsung Hsiao[5,6] · Kun-Ju Lin[5,6] · Zhi-Kun Lin[1] · Wen-Bin Lin[1] · Pai-Yi Chiu[7] · Guang-Uei Hung[8] · Chiung-Chih Chang[9] · Ya-Ting Chang[10] · Keh-Shih Chuang[2] · For the Alzheimer's Disease Neuroimaging Initiative

1. Health Physics Division, Institute of Nuclear Energy Research, Atomic Energy Council, 1000 Wenhua Rd. Jiaan Village, Longtan District, Taoyuan City 325, Taiwan

2. Department of Biomedical Engineering and Environmental Sciences, National Tsing-Hua University, Hsinchu, Taiwan

3. Division of Behavioral Neurology, Department of Neurology, National Cheng Kung University Hospital, College of Medicine and Institute of Gerontology, National Cheng Kung University, 138, Sheng Li Road, North District, Tainan 704, Taiwan

4. Alzheimer's Disease Research Center, National Cheng Kung University Hospital, Tainan, Taiwan

5. Department of Medical Imaging and Radiological Sciences & Healthy Aging Center, Chang Gung University, Taoyuan, Taiwan

6. Department of Nuclear Medicine and Molecular Imaging Center, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan

7. Department of Neurology, Show Chwan Memorial Hospital, Changhua, Taiwan

8. Department of Nuclear Medicine, Chang Bing Show Chwan Memorial Hospital, Changhua, Taiwan

9. Department of Neurology, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan

10. Department of Neurology, Institute of Translational Research in Biomedicine, Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Kaohsiung, Taiwan