Dement Neurocogn Disord. 2023 Apr;22(2):61-68
https://doi.org/10.12779/dnd.2023.22.2.61
pISSN 1738-1495·eISSN 2384-0757

DND Dementia and Neurocognitive Disorder

Original Article

Check for updates

Classification of Aβ State From Brain Amyloid PET Images Using Machine Learning Algorithm

Chanda Simfukwe,[1] **Reeree Lee**,[2] **Young Chul Youn**,[1] **Alzheimer's Disease and Related Dementias in Zambia (ADDIZ) Group**

[1]Department of Neurology, Chung-Ang University College of Medicine, Seoul, Korea
[2]Department of Nuclear Medicine, Chung-Ang University College of Medicine, Seoul, Korea

OPEN ACCESS

Received: Mar 29, 2023
Revised: Apr 19, 2023
Accepted: Apr 23, 2023
Published online: Apr 30, 2023

Correspondence to
Young Chul Youn
Department of Neurology, Chung-Ang University College of Medicine, 102 Heukseok-ro, Dongjak-gu, Seoul 06973, Korea.
Email: neudoc@cau.ac.kr

A preprint has previously been published in Research Square (Chanda Simfukwe, Reeree Lee, and Young Chul Youn. Prediction of Aβ State from Brain Amyloid PET Images Using Machine Learning Algorithm. 2022).

ORCID iDs
Chanda Simfukwe
https://orcid.org/0000-0003-2845-8016
Reeree Lee
https://orcid.org/0000-0001-8811-3699
Young Chul Youn
https://orcid.org/0000-0002-2742-1759

ABSTRACT

Background and Purpose: Analyzing brain amyloid positron emission tomography (PET) images to access the occurrence of β-amyloid (Aβ) deposition in Alzheimer's patients requires much time and effort from physicians, while the variation of each interpreter may differ. For these reasons, a machine learning model was developed using a convolutional neural network (CNN) as an objective decision to classify the Aβ positive and Aβ negative status from brain amyloid PET images.
Methods: A total of 7,344 PET images of 144 subjects were used in this study. The 18F-florbetaben PET was administered to all participants, and the criteria for differentiating Aβ positive and Aβ negative state was based on brain amyloid plaque load score (BAPL) that depended on the visual assessment of PET images by the physicians. We applied the CNN algorithm trained in batches of 51 PET images per subject directory from 2 classes: Aβ positive and Aβ negative states, based on the BAPL scores.
Results: The binary classification of the model average performance matrices was evaluated after 40 epochs of three trials based on test datasets. The model accuracy for classifying Aβ positivity and Aβ negativity was (95.00±0.02) in the test dataset. The sensitivity and specificity were (96.00±0.02) and (94.00±0.02), respectively, with an area under the curve of (87.00±0.03).
Conclusions: Based on this study, the designed CNN model has the potential to be used clinically to screen amyloid PET images.

Keywords: Amyloid; Supervised Machine Learning; Algorithms; PET Scan

INTRODUCTION

Alzheimer's disease (AD) is a progressive, irreversible neurodegenerative disease characterized by cognitive decline that can interfere with the individual's ability to function independently in daily activities.[1] Detecting the early stage of neurodegenerative disease is essential for developing future treatments and biomarkers.[1,2] The accumulation of β-amyloid (Aβ) has been considered an essential biomarker for diagnosing AD and predicting the prognosis.[3,4] Among biomarkers reflecting cerebral amyloidosis, 18F-florbetaben (FBB) amyloid positron emission tomography (PET) has been a widely used tool.[5,6]

Funding

The publication costs, design of the study, data management, and writing the manuscript for this article were supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (project number: NRF-2017S1A6A3A01078538).

Conflict of Interest

The authors have no financial conflicts of interest.

Author Contributions

Conceptualization: Simfukwe C, Youn YC; Data curation: Lee R; Formal analysis: Simfukwe C; Funding acquisition: Youn YC; Investigation: Youn YC; Methodology: Simfukwe C; Project administration: Youn YC; Resources: Youn YC; Software: Simfukwe C; Supervision: Youn YC; Validation: Simfukwe C; Writing - original draft: Simfukwe C; Writing - review & editing: Lee R, Youn YC.

The amyloid PET images were interpreted visually by objective quantitative evaluation with standardized uptake value ratio (SUVr), which is performed using the mean activity of six cortical regions of interest (ROIs) (frontal, temporal, parietal, precuneus, anterior cingulate, and posterior cingulate), normally using the cerebellum as a reference region.[7,8] An appropriate reference region is important, and has to be free of Aβ. If the reference region SUV is high due to accumulated Aβ, the SUVr of the target region will be low or false negative. While the SUVr plays a key role in clinical studies of AD, visual interpretation is the standard approach in clinical practice for amyloid PET.[9] Brain amyloid plaque load (BAPL) score is a three-grade scoring system that is visually determined by a clinician according to the cerebral Aβ loads using 18F-FBB.[10] The BAPL scoring system is divided into three stages: BAPL1, BAPL2, and BAPL3, in which BAPL1 indicates no Aβ load, BAPL2 indicates minor Aβ load, and BAPL3 indicates significant Aβ load.[10,11] This system allows Aβ loads to be interoperated without the magnetic resonance imaging (MRI) required for spatial normalization to PET, and while it is not affected by Aβ loads of the reference site, there may be limitations in objective evaluation. We expect that BAPL scoring-based machine learning algorithms will help overcome these limitations in predicting cerebral amyloidosis by allowing machines to predict amyloid loads.

In recent years, machine learning and deep learning algorithms have been used in the prediction of AD in MRI or PET brain images.[12] Sato et al.[13] predicted the BAPL score from brain FBB PET based on the convolutional neural network (CNN) algorithms with a joint discriminative loss function. Although Sato's method was able to classify BAPL scores, it was based on different algorithms that were trained with axial, sagittal, and coronal plane FBB PET images. In the prediction of the BAPL score, the coronal plane images showed higher accuracy, compared to axial plane images. However, in this study, we designed a CNN algorithm that classified Aβ status based on axial plane FBB PET images, with a batch of images per subject (**Fig. 1**). We evaluated the model's performance using the confusion matrix and performance metrics.[14]

METHODS

Amyloid PET dataset

The dataset was hospital-based data collected from all participants who underwent FBB PET regardless of diagnosis, and consisted of age, gender, and BAPL scores. The study involved a total of 144 subjects, who were categorized into 3 groups: BAPL1, BAPL2, and BAPL3. The BAPL scores of FBB PET images were obtained by a well-trained nuclear medicine specialist and a neurologist, respectively, by visualizing Aβ deposition in different parts of the brain

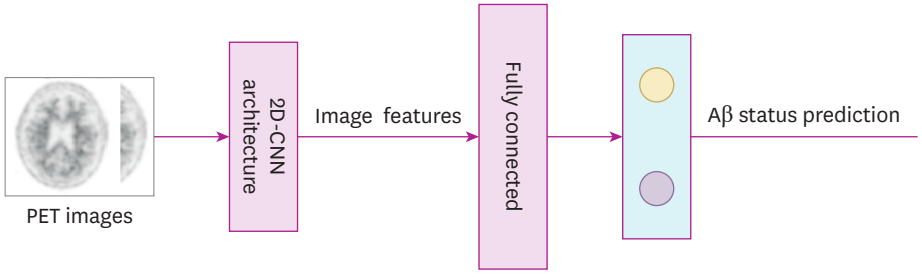

Fig. 1. Study framework.
PET: positron emission tomography, CNN: convolutional neural network, Aβ, β-amyloid.

https://doi.org/10.12779/dnd.2023.22.2.61

Table 1. Demographics of the dataset (n=144)

Characteristics	BAPL1 (n=63)	BAPL2 (n=8)	BAPL3 (n=73)
Age (yr)	73.02±9.22	77.25±4.59	72.13±8.45
Sex (male:female)	24:39	4:4	20:53

Ages are shown as mean ± standard deviation.
BAPL: brain amyloid plaque load; BAPL1: no amyloid-beta load; BAPL2: minor amyloid-beta load; BAPL3: significant amyloid-beta load.

regions: lateral temporal, posterior cingulate, precuneus, and parietal lobe. Then, in the case of inconsistency, the final score was decided by consultation and agreement. Participants with brain lesions, acute mental problems such as depression, neurological conditions like epilepsy, and psychiatric disorders like schizophrenia were excluded from the study. In the absence of Aβ deposition, it was determined as BAPL1; in the case of minor Aβ loads, BAPL2; and in the case of significant Aβ loads, BAPL3. Sixty-three subjects had a score of BAPL1, 8 subjects had a score of BAPL2, and 73 subjects had a score of BAPL3. **Table 1** presents the other demographics. The subjects with BAPL1 were categorized as Aβ negative, while those with BAPL2 and BAPL3 as Aβ positive.

Data processing

Keras (https://keras.io/) and TensorFlow (https://www.tensorflow.org/) were used in a supervised learning CNN algorithm to train a model that distinguishes subjects with Aβ positive (BAPL2 and BAPL3) from Aβ negative status (BAPL1). We applied 51 axial PET images per subject of jpg format at (128×128) px to the binary classification algorithm of Aβ negative and Aβ positive. The model used batches of augmented images, and evaluated the accuracy, sensitivity, specificity, area under the curve (AUC), and confusion matrices of the models-based test dataset.

Firstly, the image data were allocated into 2 classes of Aβ negative and Aβ positive based on the BAPL score system criteria for binary classification, and the dataset was splinted into training, validation, and testing datasets using the function "flow_from_directory" from Keras library (https://keras.io/).

Secondly, the model was trained using CNN in an IDE PyCharm (JetBrains, Prague, Czechia; https://www.jetbrains.com/pycharm). The image data were imported in the jpg format. The training algorithm used a dataset comprising 70% of training data and 30% validation data. Only the training dataset was augmented 6 times. The performance of the algorithm was evaluated in the testing dataset consisting of 10 subjects each in Aβ negative and Aβ positive groups.

This artificial neural network architecture consisted of five 2D convolutional layers having input shapes of (128×128) and max-pooling layers with "relu" activation, while to prevent overfitting, 2 dropout rates of 0.2 and 0.4 were inserted into the convolutional layers (**Table 2**).

Since datasets were relatively small for machine learning, the ImageDataGenerator function from the "tensorflow.keras" module was used for image augmentation on the training dataset with parameters (https://keras.io/api/preprocessing/image/) (**Table 3**).

As an example, the code that predicted Aβ status in the PET images datasets is given in the GitHub repository link (GitHub, San Francisco, CA, USA; https://github.com/chanda1993/A-Status-Prediction-from-PET-Scans).

https://doi.org/10.12779/dnd.2023.22.2.61

Table 2. Details of parameters belonging to different layers of the developed CNN model

Layer Name	Input shape	Output shape	Activation	Regularization
Layer_1	128×128×3	126×126×16	Relu	
Max_Pooling_1	126×126×16	63×63×16		
Dropout_1	63×63×16	63×63×16		Dropout (0.4)
Layer_2	63×63×16	61×61×32	Relu	
Max_Pooling_2	61×61×32	30×30×32		
Dropout_2	30×30×32	30×30×32		Dropout (0.2)
Layer_3	30×30×32	28×28×64	Relu	
Max_Pooling_3	28×28×64	14×14×64		
Dropout_3	14×14×64	14×14×64		Dropout (0.2)
Layer_4	14×14×64	12×12×64		
Max_Pooling_4	12×12×64	6×6×64		
Layer_5	6×6×64	4×4×64	Relu	
Max_Pooling_5	4×4×64	2×2×64		
Dropout_5	2×2×64	2×2×64		Dropout (0.2)
Flatten	2×2×64	256		
Dense_1	256	128	Relu	
Dense_2	128	2	SoftMax	

Table 3. Image augmentation parameters on the training dataset

No.	Augmentation	Parameters
1	Width_shift_range	20
2	Height_shift_range	20
3	Fill_mode	"nearest"
4	ZCA_whitening	True
5	ZCA_epsilon	1e-06
6	Channel_shift_range	13

ZCA: zero-phase component analysis.

The loss was calculated with the "categorical_crossentropy" with the "RMSprop" optimizer having a learning rate of 0.001 during training. A batch size of 51 images and 40 epochs was applied to the algorithm that was performed three times with different shuffled datasets. All the parameters were tuned accordingly to the training set provided, giving the optimum training accuracy.

Finally, the algorithm was evaluated using accuracy, sensitivity, specificity, AUC, and confusion matrix in the test dataset using the "model.evaluate" function.

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (IRB) of Chung-Ang University Hospital (IRB number C2012049 [744]). Informed consent was waived by the IRB of Chung-Ang University, since this study was retrospective, and blinding of the personal information in the data was performed. All methods were carried out in accordance with the relevant guidelines and regulations.

RESULTS

A total of 7,344 PET images of 144 subjects were allocated to the training, validation, and testing dataset. Every subject had 51 images per directory. The model was trained with 3,213 images of 63 subjects for Aβ negative and 4,131 images of 81 subjects for Aβ positive in the training and validation dataset. The training dataset was the only augmented dataset when

Table 4. Demographic of amyloid PET images included in this study and their classification in training (augmented), validation, and testing datasets according to Aβ negative and Aβ positive

Characteristic	Aβ negative	Aβ positive
No. of images (n=7,344)	3,213	4,131
Training dataset (augmented)	1,892 (11,352)	2,534 (15,204)
Validation dataset	811	1,087
Testing dataset	510	510

Note: The training dataset was augmented with the Image DataGenerator function from the Keras library with 6 arguments from **Table 3**.
PET: positron emission tomography, Aβ: β-amyloid.

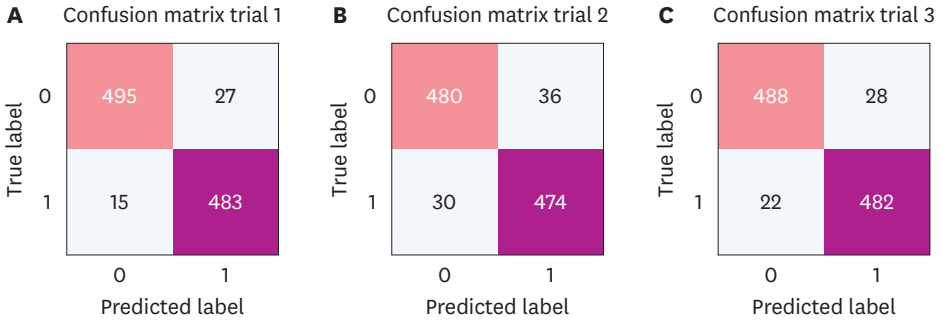

Fig. 2. The confusion matrixes of 3 trials predictors.
Confusion matrixes of 3 trials; β-amyloid positive (0) and β-amyloid negative (1).
Note: The confusion matrixes were obtained by using scikit.metrics function on each trial model classification performance with the test dataset.

Table 5. Mean accuracies of a designed convolutional neural network model classifying Aβ positive and Aβ negative on three trials

Characteristic	Trial 1	Trial 2	Trial 3	Mean ± SD
Accuracy	0.96	0.93	0.95	0.95±0.02
Loss	0.431	0.46	0.389	0.42±0.07
Sensitivity	0.97	0.94	0.96	0.96±0.02
Specificity	0.94	0.93	0.95	0.94±0.01
AUC	0.90	0.88	0.85	0.87±0.03

Aβ: β-amyloid, SD: standard deviation, AUC: area under the curve.

designing the model. The test dataset had 1,020 images of 20 subjects in Aβ negative and Aβ positive (**Table 4**).

Based on the BAPL scoring system, the model's accuracy for classifying Aβ positive and Aβ negative in the test dataset after three trials was (95.00±0.02). The model had a sensitivity of (96.00±0.02) for detecting Aβ positive, and a specificity of (94.00±0.01) for detecting Aβ negative, with an AUC of (87.00±0.03) (**Table 5**).

Confusion matrices were plotted to show the classification performance of the model based on the testing dataset on three trials; all trials showed good classification performance of detecting Aβ positive and Aβ negative (**Fig. 2**).

DISCUSSION

This study aimed to develop a machine-learning algorithm to differentiate patients with Aβ positive from Aβ negative on axial FBB PET images, thereby allowing physicians to focus on

https://doi.org/10.12779/dnd.2023.22.2.61

cases that require more medical attention. The CNN model was trained with three trials on 51 images per subject's directory from the test dataset in the present study, and an accuracy of 95% with loss, sensitivity, specificity, and AUC of 42%, 96%, 95%, and 85% were evaluated, respectively (**Table 5**). The high sensitivity of the tool is clinically useful for disease screening, because it can avoid missing actual pathological conditions.[15] Nuclear medicine experts can easily apply the developed model to pre-determine whether amyloid pathology is present.

Sato et al.[16] examined three state-of-the-art networks (i.e., VGG19, ResNet50, and DenseNet121) with various CNN layers and conventional loss functions in the prediction of Aβ state from PET images via machine learning trained with axial, sagittal, and coronal plane FBB PET images per patient. The accuracy for binary classification of Aβ negative and positive with balanced coronal plane images dataset was 96.97%, with an AUC of 0.97. Although the 5-layer CNN algorithm utilized only axial plane PET images of an imbalanced dataset and only augmented training dataset, this study showed predictive performance similar to that of the DenseNet121 algorithm in a previous study using a balanced dataset.

Various machine-learning algorithms (i.e., AlexNet, VGG16, VGG19, ResNet50, and DenseNet121) were applied on FBB PET images, and the receiver operating characteristic curves and AUC were used to calculate the model's prediction. Among the trained models, the ResNet50 model with a multi-class classification that was trained with axial plane PET images showed good prediction accuracy in identifying AD patients.[13] This showed that the axial plane PET images could build better models and support our observations. With this developed model in the current study, axial plane PET images alone could accurately identify Aβ negative and Aβ positive subjects.

The present study has some limitations. First, the dataset was small, particularly in BAPL2; multi-classification could not be performed; and only binary classifications could be performed. Second, the model could be biased in classifying, in which it needed more data for training to not mismatch the subject's diagnosis group. Third, augmented data in the training dataset could not be the same as the original dataset obtained by nuclear medicine specialists, and this could have affected the trained machine learning model classification accuracy on the test dataset. Although there were not many cases of BAPL2, the presence of minor amyloid loads, such as BAPL2, could have reduced the performance of Aβ positive classification, because it is not easy to classify as Aβ positive or negative.

Therefore, more research is needed on this current study to improve the model's performance with an increased dataset, especially for BAPL2.

In conclusion, the purpose of this study was to design and evaluate an algorithm that could classify Aβ state (positive, negative) in batches of axial plane PET images per subject. Our results showed that the CNN algorithm achieved good performance in classifying Aβ status, and can be used as a reference when clinicians interpret amyloid PET images.

ACKNOWLEDGEMENTS

We thank the Department of Nuclear medicine and Neurology at Chung-Ang University Hospital for providing the necessary tools to make this research successful.

AVAILABILITY OF DATA AND MATERIAL

The datasets used and analyzed during the current study are available from the GitHub repository at https://github.com/chanda1993/A-Status-Prediction-from-PET-Scans.

REFERENCES

1. Porsteinsson AP, Isaacson RS, Knox S, Sabbagh MN, Rubino I. Diagnosis of early Alzheimer's disease: clinical practice in 2021. J Prev Alzheimers Dis 2021;8:371-386.
PUBMED | CROSSREF
2. Shi M, Caudle WM, Zhang J. Biomarker discovery in neurodegenerative diseases: a proteomic approach. Neurobiol Dis 2009;35:157-164.
PUBMED | CROSSREF
3. Mantzavinos V, Alexiou A. Biomarkers for Alzheimer's disease diagnosis. Curr Alzheimer Res 2017;14:1149-1154.
PUBMED | CROSSREF
4. Lee JC, Kim SJ, Hong S, Kim Y. Diagnosis of Alzheimer's disease utilizing amyloid and tau as fluid biomarkers. Exp Mol Med 2019;51:1-10.
PUBMED | CROSSREF
5. Cho K, Kim WG, Kang H, Yang GS, Kim HW, Jeong JE, et al. Classification of 18F-florbetaben amyloid brain PET images using PCA-SVM. Biomed Sci Lett 2019;25:99-106.
CROSSREF
6. Thepa N, Jeong YJ, Kang H, Choi GE, Yoon HJ, Kang DY. A comparative study of [F-18] florbetaben (FBB) PET imaging, pathology, and cognition between normal and Alzheimer transgenic mice. Biomed Sci Lett 2019;25:7-14.
CROSSREF
7. Barthel H, Gertz HJ, Dresel S, Peters O, Bartenstein P, Buerger K, et al. Cerebral amyloid-β PET with florbetaben (18F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. Lancet Neurol 2011;10:424-435.
PUBMED | CROSSREF
8. Tsubaki Y, Akamatsu G, Shimokawa N, Katsube S, Takashima A, Sasaki M, et al. Development and evaluation of an automated quantification tool for amyloid PET images. EJNMMI Phys 2020;7:59.
PUBMED | CROSSREF
9. Cho SH, Choe YS, Kim YJ, Lee B, Kim HJ, Jang H, et al. Concordance in detecting amyloid positivity between 18F-florbetaben and 18F-flutemetamol amyloid PET using quantitative and qualitative assessments. Sci Rep 2020;10:19576.
PUBMED | CROSSREF
10. Kang H, Kim WG, Yang GS, Kim HW, Jeong JE, Yoon HJ, et al. VGG-based BAPL score classification of 18F-florbetaben amyloid brain PET. Biomed Sci Lett 2018;24:418-425.
CROSSREF
11. Kim JY, Oh D, Sung K, Choi H, Paeng JC, Cheon GJ, et al. Visual interpretation of [18F]Florbetaben PET supported by deep learning-based estimation of amyloid burden. Eur J Nucl Med Mol Imaging 2021;48:1116-1123.
PUBMED | CROSSREF
12. Lin W, Tong T, Gao Q, Guo D, Du X, Yang Y, et al. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Front Neurosci 2018;12:777.
PUBMED | CROSSREF
13. Sato R, Lwamoto Y, Cho K, Kang DY, Chen YW. Comparison of CNN models with different plane images and their combinations for classification of Alzheimer's disease using PET images. In: Chen YW, Zimmermann A, Howlett R, Jain L, editors. Smart Innovation, Systems and Technologies book series. Vol 145. Innovation in Medicine and Healthcare Systems, and Multimedia. Singapore: Springer, 2019;169-177.
CROSSREF
14. Ting KM. Confusion matrix. In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning. Boston: Springer, 2011;209.
CROSSREF

15. Simfukwe C, An SS, Youn YC. Comparison of RCF scoring system to clinical decision for the rey complex figure using machine-learning algorithm. Dement Neurocogn Disord 2021;20:70-79.
PUBMED | CROSSREF

16. Sato R, Iwamoto Y, Cho K, Kang DY, Chen YW. Accurate BAPL score classification of brain PET images based on convolutional neural networks with a joint discriminative loss function. Appl Sci (Basel) 2020;10:965.
CROSSREF