Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Deep learning only by normal brain PET identify unheralded brain anomalies

Hongyoon Choi [a,b,1], Seunggyun Ha [b,1], Hyejin Kang [b], Hyekyoung Lee [b], Dong Soo Lee [a,b,c,*], for the Alzheimer's Disease Neuroimaging Initiative [2]

[a] Department of Nuclear Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[b] Department of Nuclear Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
[c] Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 10 January 2019
Received in revised form 20 March 2019
Accepted 9 April 2019
Available online 16 April 2019

Keywords:
PET
Deep learning
Variational autoencoder
Alzheimer
Anomaly detection

ABSTRACT

Background: Recent deep learning models have shown remarkable accuracy for the diagnostic classification. However, they have limitations in clinical application due to the gap between the training cohorts and real-world data. We aimed to develop a model trained only by normal brain PET data with an unsupervised manner to identify an abnormality in various disorders as imaging data of the clinical routine.

Methods: Using variational autoencoder, a type of unsupervised learning, Abnormality Score was defined as how far a given brain image is from the normal data. The model was applied to FDG PET data of Alzheimer's disease (AD) and mild cognitive impairment (MCI) and clinical routine FDG PET data for assessing behavioral abnormality and seizures. Accuracy was measured by the area under curve (AUC) of receiver-operating-characteristic (ROC) curve. We investigated whether deep learning has additional benefits with experts' visual interpretation to identify abnormal patterns.

Findings: The AUC of the ROC curve for differentiating AD was 0.90. The changes in cognitive scores from baseline to 2-year follow-up were significantly correlated with Abnormality Score at baseline. The AUC of the ROC curve for discriminating patients with various disorders from controls was 0.74. Experts' visual interpretation was helped by the deep learning model to identify abnormal patterns in 60% of cases initially not identified without the model.

Interpretation: We suggest that deep learning model trained only by normal data was applicable for identifying wide-range of abnormalities in brain diseases, even uncommon ones, proposing its possible use for interpreting real-world clinical data.

1. Introduction

Recent advances in deep learning have been rapidly applied to medical fields as it has great advantages in processing high-dimensional data by capturing meaningful discriminative features [1]. Deep learning-based models have been successfully developed for medical image recognition tasks including diagnosis of dermatologic disorders and diabetic retinopathy [2,3]. So far, a few deep learning-based models have been applied to the diagnosis of brain disorders such as Alzheimer's disease (AD), Parkinson's disease and psychiatric disorders [4–7]. Even though these models have recorded high accuracy for discriminating brain disorders from normal controls, their clinical application has not yet been established due to several reasons. One of the critical limitations in the current deep learning-based diagnostic models is that it can be only applied to data similar to a training set as most models rely on supervised learning, while brain images are clinically acquired for patients with various disorders without any prior grouping and its characterization [8]. For example, a deep learning model trained by a cohort composed of AD patients and controls with supervised manner could hardly be applied to patients with a cognitive decline in general population as it includes other types of dementia as well as AD. Moreover, the deep learning model based on supervised training which has successfully differentiated characteristic groups of patients from normal has limitations when applied to the diagnosis of

[*] Corresponding author at: Department of Nuclear Medicine, Seoul National University Hospital, 28 Yongon-Dong, Jongno-Gu, Seoul 110-744, Republic of Korea.
E-mail address: dsl@plaza.snu.ac.kr (D.S. Lee).

[1] These authors equally contribute 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.ebiom.2019.04.022

Research in context

Evidence before this study

A major application of recent deep learning models has been a differential diagnostic classification based on supervised learning. In spite of the remarkable performance, most models based on the supervised learning have limitations in real data as data distribution in a real clinic could be different from the training cohort and include uncommon disorders, which could not be considered for the training. We develop a deep learning model trained only by normal brain PET data to identify the abnormality in various brain disorders. We propose the model that can detect abnormal patterns in a variety of diseases with heterogeneous distribution expecting the application to the clinical practice.

Added value of this study

This is the first study to use unsupervised learning to identify abnormal patterns in brain image data. We suggest the Abnormality Score, which reflects how far a given brain PET data is from the normal data, to define abnormal patterns. The Abnormality Score could differentiate FDG PET of Alzheimer's disease from normal. Moreover, the score at the baseline in mild cognitive impairment patients was correlated with future cognitive decline. We applied this model to the FDG PET data as a clinical routine, which included relatively uncommon disorders. It successfully identified abnormal patterns, some of which had not been initially identified by experts. Moreover, experts' visual interpretation could be aided by this model.

Implications of all the available evidence

Our approach could be used for the identification of metabolic abnormality in brain FDG PET in various types of disorders. The Abnormality Score may be used as a quantitative marker for predicting outcomes and measuring disease severity. The output map of this model could localize abnormal patterns, which can aid experts' visual interpretation as a clinical routine process. Considering heterogeneous patients in various clinical environments, our model based on unsupervised learning maybe clinically applicable for evaluating wide-range of brain diseases, even uncommon ones.

uncommon ones, because it is difficult to collect sufficient data on rare diseases to train the model [9,10].

We aimed to develop an unsupervised learning-based model which can be used to identify brain abnormality even in the subjects with unknown heterogeneous distribution. The model trained by ^{18}F-fluorodeoxyglucose (FDG) brain PET of cognitively normal aged subjects was applied to identify abnormal patterns of brain metabolism. We used the data only consisting of brain images of normal subjects for training. The output of the model, defined as Abnormality Score in this study, represents how far a given patient's brain's metabolic pattern is from the distribution of the normal subjects. To show the applicability of this model to various disorders in the clinical setting including relatively uncommon diseases, we applied the model to an independent cohort of patients with heterogeneous brain disorders. As this cohort includes brain disorders showing various brain metabolic patterns which were unheralded and sometimes had no abnormality on visual analysis, we investigated whether the model could capture diverse abnormal metabolic patterns therein. Furthermore, as the model could generate a map representing abnormal patterns, we investigated whether the abnormality pattern map could assist experts' visual interpretation to identify brain abnormalities and their localizations.

2. Materials and methods

2.1. Subjects

A part of the image data of this study was collected from Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu) database. 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. ADNI included subjects from over 50 sites across the US and Canada. The main goal of ADNI has been to develop combined biomarkers by investigating whether serial imaging and biological markers and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease. For up-to-date information, see http://www.adni-info.org. Among FDG PET scans of 393 subjects of normal controls, 353 scans were used for training the model. The remained 40 controls were used as a test set for the model to discriminate abnormal from normal brain.

As abnormal brain data, the model was tested in two different cohorts. Firstly, the model was tested for the FDG PET scans of ADNI database. FDG PET scans of Alzheimer's disease ($n = 243$) and mild cognitive impairment (MCI) ($n = 667$) patients were used. As normal subjects, the remaining 40 controls aforementioned were used to evaluate the model. The diagnostic classification was determined by ADNI Clinical Core and used as a ground-truth for further deep learning-based prediction. Brief demographics of ADNI cohorts are summarized in Supplementary Table 1.

As an independent dataset from the training ADNI dataset, FDG brain PET scans of patients routinely obtained in the clinic were retrospectively collected in a single center. FDG PET scans as a baseline workup for evaluating the etiology of seizures and behavioral abnormalities were retrospectively collected. Baseline FDG PET studies of patients who had structural abnormalities including stroke and tumors were excluded. To compare the brain PET scans of these patients, adult normal FDG brain PET scans were also retrospectively collected. The normal brain PET scans were obtained by subjects who underwent health screening and had no history of neurologic disorders (Demographics are summarized in Supplementary Table 2). For ADNI data, the institutional review boards (IRB) of all participating institutions approved imaging studies and all participants signed a written informed consent. The retrospective study for the independent cohort was approved by IRB of our institute, and informed consent was waived due to the retrospective design. 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.

2.2. Study design

We trained a neural network model based on variational autoencoder (VAE) by normal data to identify abnormal brain metabolism using 353 PET data of normal subjects. FDG PET data and age information was encoded into latent features and then a decoder network of VAE reconstruct the latent features to original FDG PET data. Notably, age information was included to identify age-dependent FDG PET patterns [11]. The VAE model was trained to minimize an error between original and reconstructed brain PET images. The model was trained by a gradient descent algorithm which employs iterative updates of model parameters. Since VAE uses images themselves as labels, it can be regarded as a type of self-supervised learning rather than unsupervised learning in terms of precise meaning. The output of the trained neural network was Abnormality Score, which represents a degree of how far a given brain PET is from normal brain PET.

To evaluate the Abnormality Score, the network was first tested to discriminate AD from normal subjects using the ADNI cohort. For this test, normal PET data of ADNI ($n = 40$) different from training data ($n = 353$) were used to compare with PET data of AD. Additionally, the model was applied to PET data of MCI subjects to find the correlation between Abnormality Score and a future decline of cognitive scores, Mini-Mental State Exam (MMSE) and Clinical Dementia Rating Sum of Boxes (CDR-SB). The network was applied to the independent dataset which was collected from routine clinical PET studies, brain PET scans of patients with seizures and behavioral abnormality as clinical routine data and another normal PET. The application to the independent cohort was aimed to test the feasibility of deep learning model for heterogeneous brain disorders as real-world data as well as the validation of the model. The Abnormality Score of the independent cohort was compared according to the diagnosis. Detailed methods for image acquisition, preprocessing and the deep learning model are described in Supplementary Methods.

2.3. Abnormality score calculation and reconstruction error map

The degree of abnormal patterns in brain metabolism was represented by the reconstruction error of VAE. The mean reconstruction error was measured by the average value of mean square errors of brain voxels. Abnormality Score was the mean reconstruction error of each subject divided by the mean value of normal subjects' reconstruction errors.

Reconstruction error map of each subject was obtained by subtraction of the output of VAE, reconstructed images, from the input PET images.

2.4. Voxelwise one-sample t-test and distance from a reference brain PET

We compared the Abnormality Score with two conventional voxelwise analyses to identify abnormality, one-sample t-test and distance from a reference brain PET. The brain images of the training cohort were used for estimating voxelwise normal population distribution. Gaussian smoothing with 10 mm kernel was applied to the images. Voxelwise t-scores were calculated for a given brain by using the distribution of the normal cohort. The maximum absolute t-score of each brain was used as the degree of abnormality. As another conventional method, we generated a reference normal brain PET by voxelwise mean of training cohort brains. We then measured Euclidean distances of all brain PET from the reference brain PET to define a conventional measurement for the abnormality.

2.5. Visual interpretation of brain PET

As performed in the clinical setting, brain PET images of patients with various disorders were visually reviewed by more than two experienced nuclear medicine physicians blinded to clinical diagnosis. Abnormal metabolic patterns and their location was annotated. To reveal whether the reconstruction error map could aid visual interpretation, brain PET images combined with corresponding reconstruction error maps were additionally reviewed.

2.6. Statistics

Abnormality Score of two different groups was compared by using Mann Whitney *U* test. Pearson correlation was performed to test the correlation between Abnormality Score and future cognitive decline scores, MMSE and CDR-SB. The decline of cognitive scores was calculated by 2-year follow-up and baseline exams. The performance of discriminating abnormal brain patterns from normal controls was evaluated by ROC curves and AUC was calculated. The optimal cutoff value for Abnormality Score was determined by the point on the ROC curve with minimal distance from 100% sensitivity and 100% specificity. To obtain 95% confidence intervals of ROC AUC, bootstrap resampling was used (1000 iterations).

3. Results

3.1. A model trained only by normal brain PET scans identified AD patterns

A deep neural network model based on VAE was trained by cognitively normal aged subjects. The inputs of the model were FDG PET data and age of subjects. The key concept lies in the fact that the model reconstructs normal brain images similar to training samples with minimized reconstruction error, while the model shall reconstruct brain images with abnormal patterns far from the distribution of normal brains with high reconstruction error. The model was sufficiently trained to reconstruct PET images of normal aged subjects with a minimized loss (Supplementary Fig. 1). We defined 'Abnormality Score' using normalized reconstruction error of images defined by the mean value of mean-squared-errors of brain voxels (Fig. 1a). Abnormality Score of AD patients was significantly higher than normal controls independent of training data (1.93 ± 0.83 vs. 0.99 ± 0.25, $p < 1 \times 10^{-15}$). The area under curve (AUC) of the ROC curve was 0.90 (95% C.I.

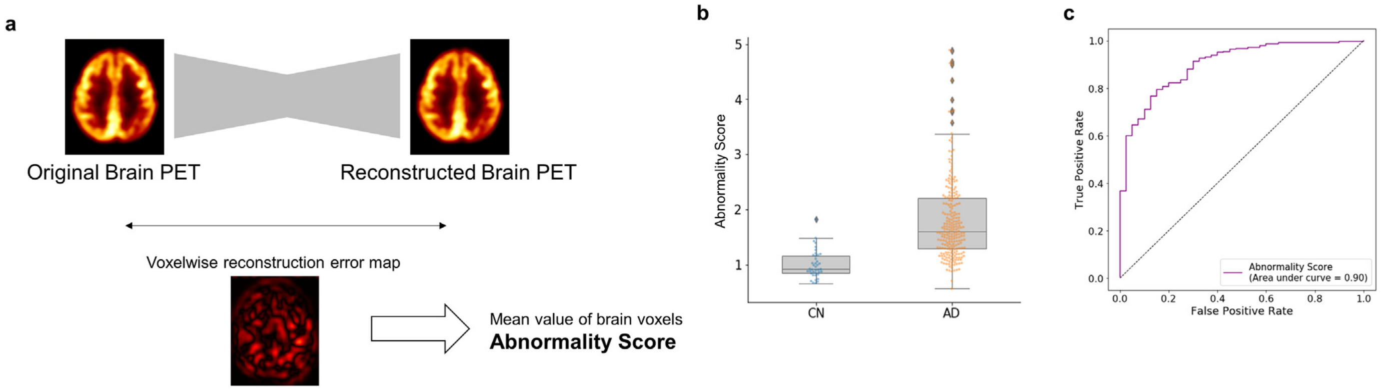

Fig. 1. Abnormality Score of Alzheimer's disease (AD) patients and normal controls. (a) A brief overview of the estimation of abnormality and reconstruction error maps is presented. A variational autoencoder model was trained using brain PET images of cognitively normal subjects. A given brain PET data with abnormal metabolic patterns compared with normal brain distribution shows high reconstruction error. Reconstruction error maps were obtained and mean errors of brain voxels were defined as Abnormality Score. Abnormality Score was measured for AD patients and controls. (b) Abnormality Score of AD patients was significantly higher than that of normal controls (1.93 ± 0.83 vs. 0.99 ± 0.25, $p < 1 \times 10^{-15}$). (c) Receiver-operating-characteristic (ROC) curve analysis was performed to differentiate AD and controls using Abnormality Score. As a performance parameter, the area under curve (AUC) was 0.90.

0.86–0.94) (Fig. 1b, c). As one sample *t*-test for each brain could identify conventionally the degree of abnormality using a voxelwise statistical distribution of normal controls, maximum t-score of each subject was used as a conventional abnormality score. AUC of the ROC curve estimated by the maximum t-score was 0.78 (95% C.I. 0.71–0.85). A distance between a given brain PET from the reference brain was used as another conventional abnormality score. AUC of the ROC curve estimated by the distance from normal reference was 0.80 (95% C.I. 0.73–0.86). The AUCs of conventional abnormality scores were significantly lower than the AUC of Abnormality Score ($p < 0.05$) (Supplementary Fig. 2).

The voxelwise mean-squared-error map drawn for each subject showed variable patterns for AD patients (Fig. 2A). Of note, overall abnormal patterns obtained by mean of reconstruction error maps of AD patients included the posterior cingulate cortex, bilateral parietal cortices and medial frontal cortices (Fig. 2b).

3.2. Correlation between PET abnormality scores and future cognitive decline in MCI patients

We evaluated Abnormality Score for MCI patients to investigate whether the model could capture the characteristic imaging patterns in MCI patients with cognitive decline. Abnormality Scores of MCI-converters were significantly higher than those of MCI-nonconverters (1.22 ± 0.42 vs 1.07 ± 0.33; U = 1.2 × 10 [4], $p < 1 \times 10^{-4}$) (Fig. 3a). MMSE changes from baseline to 2-year follow-up were negatively

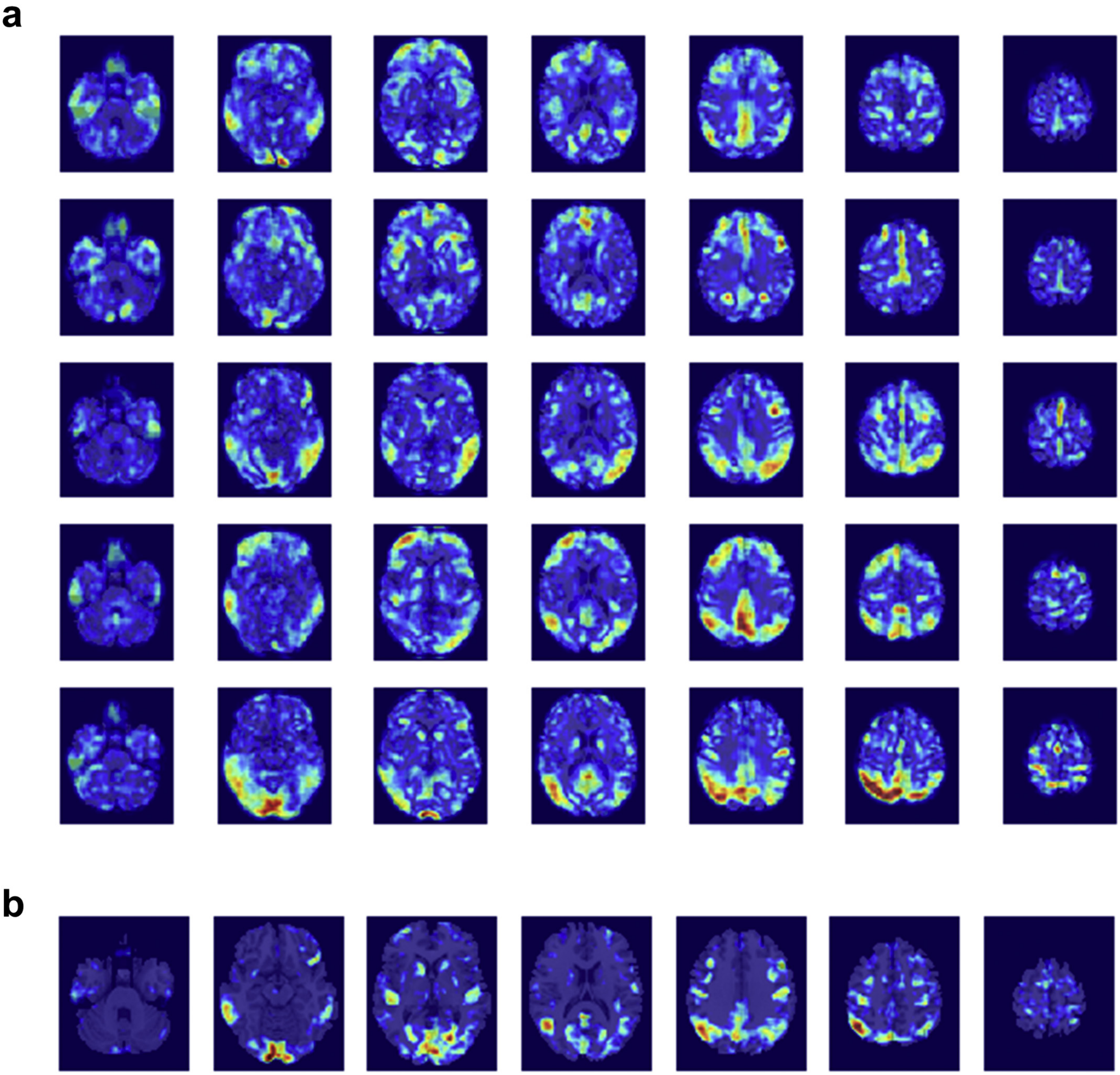

Fig. 2. Reconstruction error maps of AD patients. The output of the model, variational autoencoder, is the reconstructed PET images. The reconstruction error map is obtained by voxelwise mean-squared-error. Voxels with high mean-squared-error represent the location of abnormal patterns as voxelwise reconstruction errors represent contributions of Abnormality Score of the whole brain. (a) The reconstruction error maps were drawn for AD patients. Patients showed different patterns of abnormality, which commonly included the posterior cingulate, bilateral parietal cortices and medial frontal cortices. (b) Overall abnormal patterns of AD patients were generated by the mean image of the reconstruction error maps of AD patients.

correlated with Abnormality Score at baseline ($r = -0.19$, $p < 1 \times 10^{-4}$). CDR-SB changes during 2-years were also positively correlated with Abnormality Score ($r = 0.19$, $p < 1 \times 10^{-4}$) (Fig. 3b–c).

3.3. Feasibility of application to identifying the PET abnormality in patients with heterogeneous brain disorders

We applied the model to brain disorders different from dementia: Baseline brain PET studies for abnormal behavioral symptoms and seizures. These subjects were sampled from the routine cohort of brain FDG PET examinations requested based on various clinical impressions. This cohort consisted of cryptogenic epilepsies of heterogeneous etiology and also included rare disorders such as autoimmune encephalitis (Supplementary Table 2). FDG PET images were performed to evaluate the cause of symptoms and localize the abnormalities. Using our model trained by normal subjects of the different cohort (i.e. ADNI cohort), the Abnormality Scores of PET images were calculated for this new cohort. The Abnormality Scores of these patients were significantly higher than those of healthy controls (U = 93.0, $p = 0.01$). The AUC of the ROC curve for discriminating symptomatic patients from healthy controls was 0.74 (95% C.I. 0.58–0.87) (Supplementary Fig. 3). Sensitivity and specificity were 68.8% and 72.7% when a specific Abnormality Score value, 0.92, was used as a threshold. The threshold value of Abnormality Score was defined as the point on the ROC curve with minimal distance from 100% sensitivity and specificity. A conventional one sample *t*-test using the normal cohort of ADNI as controls could not differentiate symptomatic patients from normal controls. T-scores of patients were not significantly different from those of healthy controls (U = 143.0, $p = 0.18$) and the AUC of the ROC curve was 0.59, which were not significantly higher than 0.5 (95% C.I. 0.44–0.73) (Supplementary Fig. 4).

We compared the deep learning model and visual interpretation by experts for identifying brain metabolic abnormality. According to the visual interpretation performed by experts, 12 of 32 PET scans showed abnormal metabolic patterns, while abnormal patterns could not be visually identified in 20 scans. The Abnormality Score of scans visually abnormal and normal was not significantly different (1.38 ± 1.32 vs. 1.27 ± 0.71, p = n.s) (Table 1). Among 32 scans, the Abnormality Score of 21 scans were higher than the threshold 0.92 (Supplementary Table 3).

We further assessed whether the visual interpretation could be aided by reconstruction error maps. As voxelwise reconstruction errors represented the location of abnormal patterns, the visual interpretation could be aided by them. The visually normal PET scans ($n = 20$) were visually reassessed with corresponding reconstruction error maps to find abnormal metabolic patterns. We found that the visual analysis was helped by the reconstruction error maps to have identified locally decreased metabolism in 12 PET scans (60%, 12/20) (Table 1). In particular, the epileptogenic zones were localized well with abnormal metabolic patterns in neocortical epilepsy patients and in patients with autoimmune encephalitis by the reconstruction error maps (Fig. 4), which had not been visually identified without the reconstruction maps. Of note, 8 abnormal metabolic patterns further identified by the reconstruction error maps showed localized abnormality which corresponded to clinical symptoms of seizures and final diagnosis of neocortical epilepsies, autoimmune encephalitis and degenerative disorders (Supplementary Table 4).

4. Discussion

One of the key issues in the clinical application of deep learning is to develop a model that reflect real-world data. Recent deep learning models particularly focusing on the diagnostic classification have resulted in the remarkable accuracy comparable to experts' visual reading [12]. A key drawback of these models was the difficulty in application to patients of the real clinic as some of the disease types and their characteristic patterns of abnormalities would not be included in the recognizable features by the training cohort. A theoretically possible solution is to collect a large dataset which sufficiently covers all types of patients, however, it can result in several practical issues including harmonization, missing data, and class imbalance as well as a huge cost. Our approach adopting the training using only the normal PET data and measuring the degree of abnormality of the PET data of interest could bypass the issues caused by the difference between training datasets and real-world patients [8]. Furthermore, our approach could be an example of a clinically applicable transfer model as it could apply to various types of disorders as a clinical routine.

The Abnormality Score could be used as an imaging biomarker by measuring the degree of abnormality. Considering the correlation between the Abnormality Score and the future cognitive score change in MCI patients, it could be used as a predictive marker for cognitive decline. Although this type of prediction was achieved by supervised learning with higher accuracy [5], the Abnormality Score could not only apply to patients with cognitive decline but various disorders. Since there are various disorders in addition to AD and MCI which show a cognitive problem in the clinical setting, our model has advantages compared with the deep learning models based on supervised learning which show higher accuracy. Furthermore, considering that

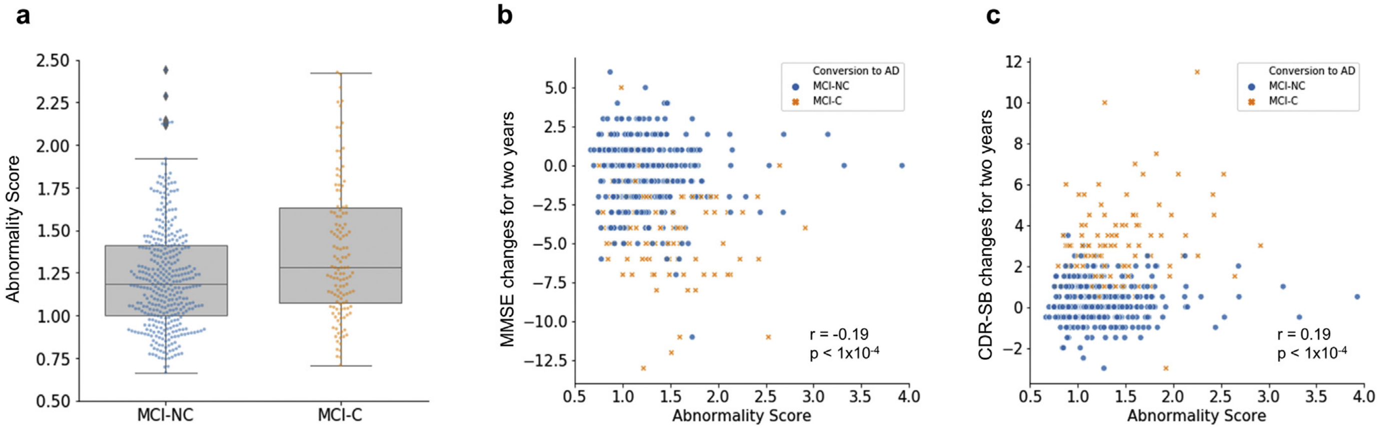

Fig. 3. Abnormality Score as a predictive biomarker for predicting future cognitive decline. We applied our model to brain FDG PET scans of MCI patients. (a) Abnormality Scores of MCI-converters and MCI-nonconverters were significantly different. Those of MCI-converters were significantly higher (1.22 ± 0.42 vs 1.07 ± 0.33; $U = 1.2 \times 10$ [4], $p < 1 \times 10^{-4}$). (b) We evaluated whether Abnormality Score at baseline PET scans predicted the future change of cognitive scores, including Mini-Mental State Exam (MMSE) (B) Clinical Dementia Rating Sum of Boxes (CDR-SB) (c). (b) MMSE changes for 2-years were negatively correlated with Abnormality Score ($r = -0.19$, $p < 1 \times 10^{-4}$). (c) CDR-SB changes for 2-years were also positively correlated with Abnormality Score ($r = 0.19$, $p < 1 \times 10^{-4}$). Note that red dots represent MCI-converters and blue dots represent MCI-nonconverters.

Table 1
Visual interpretation results of brain PET of heterogeneous patients as an initial workup and the results of visual interpretation aided by the reconstruction error map.

	Visually abnormal	Visually normal
Number of patients	12	20
Diagnosis	8 Temporal lobe epilepsy 3 Neurodegenerative 1 Motor neuron disorder	4 Temporal lobe epilepsy 5 Unknown epilepsy 4 Autoimmune encephalitis 2 Neurodegenerative 2 Frontal lobe epilepsy 1 Parietal lobe epilepsy 2 Normal (Psychogenic seizures)
Abnormality score (Mean ± SD)	1.38 ± 1.32	1.27 ± 0.71
Visual interpretation aided by reconstruction error map		60% (12/20) Further identified suspected abnormality 40% (8/20) Corresponded to final diagnosis

VAE could successfully yield the distribution of variable patterns [11], we thought that the VAE-based model could represent how far a given brain PET is from the normal distribution of brain PET. This idea has been also adapted to anomaly detection for the time-series data [13,14]. Our approach could localize the abnormal patterns, which were expected to be an explainable model by combining clinical routine visual interpretation. As MCI patients show a heterogeneous and clinical manifestation of dementia could be variable, unsupervised learning models including our approach might identify the direction of abnormalities and eventually discover subtypes of patients with clustering algorithms. It could be another future work which might be achieved by unsupervised learning on various imaging modalities as well as other biomarkers. The correlation between the future cognitive decline and the Abnormality Score was weak compared with previous models based on supervised learning [5]. It may be a trade-off to the advantage of the generalized application of our model in heterogeneous patients. As a future work, by combining the advantages, unsupervised learning followed by supervised learning as a model transfer may be an effective method for developing deep learning models with high accuracy for specific tasks with relatively small samples [15].

An important application of our model is an identification of abnormality in an unrelated group of diseases acquired in other institution than the original data even including uncommon disorders. Several natural image classification models based on ImageNet challenge were trained by the same number of image data for each label [16]. However, real-world data, especially in medical fields, include patients' images of many mutually unrelated uncommon disorders. It also obviates the trial to collect statistically sufficient data to be fed to the deep learning algorithm. Furthermore, many diagnostic classifications are based on the complicated clinical presentations including disease progression and treatment response as well as pathologic diagnosis, which result in heterogeneous and uncertain ground-truth labels for medical attributes of the patients or patients' status data. This complexity in diagnostic classification and problems of uncommon disorders eventually demands unsupervised learning.

We showed the feasibility of the application of the VAE-based model for identifying the abnormality in the independent cohort of heterogeneous patients. Though the patients have the variable final diagnosis, Abnormality Score was significantly higher than controls. Moreover, our VAE-based model could identify abnormal patterns for localization of the metabolic abnormality. This localization is clinically important as it can be used as an assistant for experts' reading and to help the requesting clinicians' further planning of surgery or therapy [17,18]. In particular, as the noninvasive localization of the epileptic zone was difficult for cryptogenic epilepsy patients, we suggest that our VAE-based model might be a clinically useful tool. Our model can serve as an assistant system for visual interpretation which will enhance the efficiency and reliability of experts' reading.

As a proof-of-concept study for a clinically applicable deep learning model, we attempted to develop the general model that covers brain images of heterogeneous conditions which correspond to the real-world clinical setting. To prove the model to be applied to various conditions in the clinical setting, the model should be further investigated in other cohorts recruited in other brain disorders. In terms of the model architecture and deep learning methods, there is a room for further modification. A variant type of generative adversarial networks was used for anomaly detection in a previous study [19], which resulted in a good performance for identifying abnormal lesions in optical coherence tomography images. Such a method could be a good candidate for application to brain images as well, even though training and optimization of adversarial networks using 3-dimensional brain image data are still difficult. Further modification and technical improvements in the model only using normal brain data may facilitate the clinical application of deep learning models that can solve real-world problems.

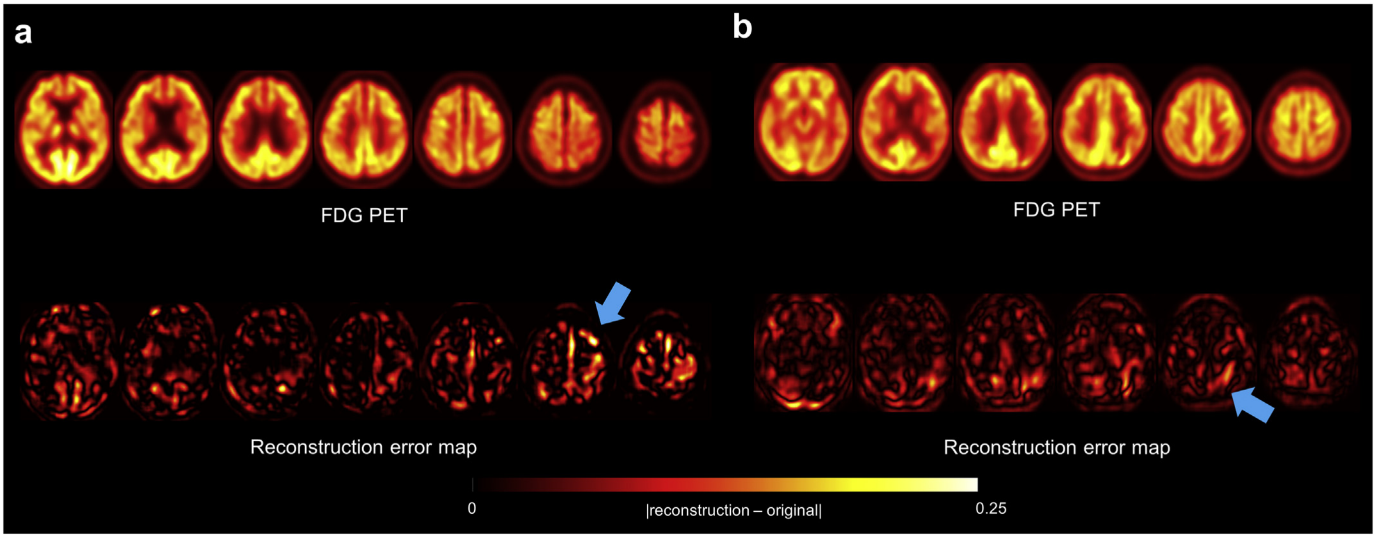

Fig. 4. Identification of abnormal metabolic patterns aided by the reconstruction error map. (a) Brain FDG PET image of a patient with autoimmune encephalitis was initially interpreted as it showed normal brain metabolism pattern according to the experts' reading. The reconstruction error map highlighted the relatively high reconstruction error in the left frontal cortex, which corresponded to the clinical symptom, right side movement abnormality. (b) Identification of abnormality in the left parietal cortex in brain PET images of patients with parietal lobe epilepsy was aided by the reconstruction error map.

5. Conclusions

We introduced a deep learning model trained only by normal brain PET data to identify abnormal patterns. This unsupervised learning-based approach has advantages in the flexible application for heterogeneous patients, even for uncommon disorders. Given that it is difficult to collect a sufficiently large brain image dataset that covers all diseases, especially rare diseases, our approach could be appropriate for real-world heterogeneous clinical data from various disorders. Furthermore, we showed that our model could be combined with experts' image interpretation, the current clinical routine, by assisting the identification and localization of abnormal patterns. We expect that our approach, which has extensibility to various diseases and may have synergy with current clinical routine practice, can facilitate the application of deep learning to nuclear medicine.

Acknowledgment

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (No. 2017M3C7A1048079), and funded by the Korean government (MOE) (No. 2016R1D1A1A02937497), and funded by the Korean Government (MSIT) (No.2017R1A5A1015626), and funded by the Korea government (MSIP) (No. 2015M3C7A1028926).

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.

Author contributions

H.C. and D.S.L. designed the study. S.H. collected imaging and clinical data. H.C. and S.H. analyzed the data. H.K. and H.L. interpreted results and supported the analysis. All authors wrote and edited the manuscript. All authors approved the manuscript.

Role of funding sources

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflict of interests

The authors declare no competing financial interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.04.022.

References

[1] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436.
[2] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115.
[3] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Jama 2016;316(22):2402–10.
[4] Ithapu VK, Singh V, Okonkwo OC, et al. Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment. Alzheimers Dement 2015;11(12):1489–99.
[5] Choi H, Jin KH, Initiative AsDN. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. Behav Brain Res 2018;344:103–9.
[6] Choi H, Ha S, Im HJ, Paek SH, Lee DS. Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. NeuroImage 2017;16:586–94.
[7] Pinaya WH, Gadelha A, Doyle OM, et al. Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. Sci Rep 2016; 6:38897.
[8] Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19:221–48.
[9] Ravı D, Wong C, Deligianni F, et al. Deep learning for health informatics. IEEE J Biomed Health Inform 2017;21(1):4–21.
[10] Choi H. Deep learning in nuclear medicine and molecular imaging: current perspectives and future directions. Nucl Med Mol Imaging 2017:1–10.
[11] Choi H, Kang H, Lee DS. Alzheimer's disease neuroimaging I. predicting aging of brain metabolic topography using variational autoencoder. Front Aging Neurosci 2018;10:212.
[12] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
[13] Suh S, Chae DH, Kang H-G, Choi S. Echo-state conditional variational autoencoder for anomaly detection. Neural networks (IJCNN). 2016 international joint conference on: IEEE; 2016. p. 1015–22.
[14] Xu H, Chen W, Zhao N, et al. Unsupervised anomaly detection via variational autoencoder for seasonal KPIs in web applications. Proceedings of the 2018 world wide web conference on world wide web; 2018: International World Wide Web Conferences Steering Committee; 2018. p. 187–96.
[15] Shin H-C, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35(5):1285–98.
[16] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. Int J Comput Vision 2015;115(3):211–52.
[17] Lee SK, Lee SY, Kim KK, Hong KS, Lee DS, Chung CK. Surgical outcome and prognostic factors of cryptogenic neocortical epilepsy. Ann Neurol 2005;58(4):525–32.
[18] Kim YK, Lee DS, Lee SK, Seok-Ki K. Differential features of metabolic abnormalities between medial and lateral temporal lobe epilepsy: quantitative analysis of (18) F-FDG PET using SPM. J Nucl Med 2003;44(7):1006.
[19] Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. International conference on information processing in medical imaging. Springer; 2017. p. 146–57.