Journal of Medical Systems (2022) 46:52
https://doi.org/10.1007/s10916-022-01836-w

IMAGE & SIGNAL PROCESSING

Artificial Intelligence on FDG PET Images Identifies Mild Cognitive Impairment Patients with Neurodegenerative Disease

Joan Prats-Climent[1] · Maria Teresa Gandia-Ferrero[2] · Irene Torres-Espallardo[2,3] · Lourdes Álvarez-Sanchez[4] · Begoña Martínez-Sanchis[3] · Consuelo Cháfer-Pericás[4] · Ignacio Gómez-Rico[2] · Leonor Cerdá-Alberich[2] · Fernando Aparici-Robles[5] · Miquel Baquero-Toledo[4] · María José Rodríguez-Álvarez[1] · Luis Martí-Bonmatí[2,5]

Received: 28 February 2022 / Accepted: 16 May 2022

Abstract
The purpose of this project is to develop and validate a Deep Learning (DL) FDG PET imaging algorithm able to identify patients with any neurodegenerative diseases (Alzheimer's Disease (AD), Frontotemporal Degeneration (FTD) or Dementia with Lewy Bodies (DLB)) among patients with Mild Cognitive Impairment (MCI). A 3D Convolutional neural network was trained using images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The ADNI dataset used for the model training and testing consisted of 822 subjects (472 AD and 350 MCI). The validation was performed on an independent dataset from La Fe University and Polytechnic Hospital. This dataset contained 90 subjects with MCI, 71 of them developed a neurodegenerative disease (64 AD, 4 FTD and 3 DLB) while 19 did not associate any neurodegenerative disease. The model had 79% accuracy, 88% sensitivity and 71% specificity in the identification of patients with neurodegenerative diseases tested on the 10% ADNI dataset, achieving an area under the receiver operating characteristic curve (AUC) of 0.90. On the external validation, the model preserved 80% balanced accuracy, 75% sensitivity, 84% specificity and 0.86 AUC. This binary classifier model based on FDG PET images allows the early prediction of neurodegenerative diseases in MCI patients in standard clinical settings with an overall 80% classification balanced accuracy.

Keywords PET · Artificial intelligence · Deep learning · Alzheimer · Mild cognitive impairment · Neurodegenerative diseases

Introduction

Dementia represents a true worldwide epidemic due to the progressive aging of the population. Dementia is a clinical syndrome defined by a persistent deterioration of higher brain functions, such as memory, orientation, calculation, language, and spatial perception. This deterioration entails a

This article is part of the Topical Collection on *Image & Signal Processing*

Joan Prats-Climent and Maria Teresa Gandia-Ferrero equal contributor.

✉ Maria Teresa Gandia-Ferrero
mteresa_gandia@iislafe.es

[1] Instituto de Instrumentación Para Imagen Molecular (I3M), Universitat Politècnica de València (UPV), Camí de Vera, s/n, 46022 Valencia, Spain

[2] Biomedical Imaging Research Group (GIBI230), La Fe Health Research Institute (IIS La Fe), Avenida Fernando Abril Martorell, 46026 Valencia, Spain

[3] Nuclear Medicine Service, La Fe University and Polytechnic Hospital, Avenida Fernando Abril Martorell, 46026 Valencia, Spain

[4] Neurology Service, La Fe University and Polytechnic Hospital, Avenida Fernando Abril Martorell, 46026 Valencia, Spain

[5] Radiology Service, La Fe University and Polytechnic Hospital, Avenida Fernando Abril Martorell, 46026 Valencia, Spain

Published online: 17 June 2022

loss of the patient's autonomy, and a detriment to their social, work and leisure activity. Dementia cases increase exponentially with age and is the main cause of dependency and disability in the elderly, and entails a very significant morbidity, which involves a major economic, social and health cost, that principally falls on family members. Nowadays, it is estimated that around 50 million people suffer from dementia across the world and about 75% of these cases correspond to Alzheimer's disease (AD). The following most common causes of dementia, after AD, are Dementia with Lewy Bodies (DLB) and Frontotemporal Degeneration (FTD). This figure is constantly increasing, and according to the 2016 report by Alzheimer's Disease International, it is expected to reach 131 million people in 2050 if there is no effective treatment [1]. Thus, it is extremely important to provide an early diagnosis in order to ensure the optimal treatment.

Mild cognitive impairment (MCI) is conceptualized as a boundary or transitional state between aging and dementia. Memory deficit is both the usual complaint in MCI and the cardinal feature of AD. The major focus of MCI research has been to distinguish individuals who will progress to AD from those who will not. Interest in MCI has been stimulated by the hope that pharmacologic intervention at this stage may delay or prevent progression to AD. Patients with MCI may stay stable or progress to dementia. The diagnosis of MCI is established by evidence of memory impairment, preservation of general cognitive and functional abilities, and absence of diagnosed dementia. The Clinical Dementia Rating (CDR) a global rating test, was found to distinguish unambiguously among older subjects with a wide range of cognitive function, from healthy to severely impaired. MCI is staged clinically at the 0.5 level on the CDR scale [2–4].

An early recognition of patients with dementia presenting with MCI is clinically challenging, specific diagnosis might be delay before deterioration is evident. Due to the high number of patients with MCI, the ability to diagnose and classify early stage dementia will have an impact on reducing the cost of long-time care by making more knowledgeable decisions regarding clinical interventions and treatment planning [5].

In the last years, developments of diagnostic tests improved the detection of dementia patients at MCI stage. Cerebrospinal fluid (CSF) amyloid-beta 42 ($A\beta_{42}$), $A\beta_{40}$, $A\beta_{42}/A\beta_{40}$ ratio, threonine-181-phosphorylated-tau (p-tau), and total-tau (t-tau) are reliable biomarkers for amyloidosis, tauopathy, and neurodegeneration. These biomarkers allow to distinguish AD at MCI stages. CSF, obtained by lumbar puncture, is an invasive procedure that should not be used in subjects with MCI but a low probability to have a neurodegenerative dementia.

Positron emission tomography (PET) provides a non-invasive quantification of brain metabolism. It has been in the detection and diagnosis of neurodegenerative diseases even in the initial stages when the patient only presents MCI. Although Amyloid PET is recognised as the tracer of choice, its availability and price might be a limiting factor. The most common and easily accessed PET radiopharmaceutical is [^{18}F]FDG [6]. The support of the clinical use of FDG PET in diagnosing prodromal and dementia stages of most neurodegenerative disorders was based on its ability to detect neurodegenerative processes and to inform about the location and extent of neuronal dysfunction at early stages [7]. Although magnetic resonance can also be used in the diagnosis of dementia, this imaging modality is limited in many cases because the brain changes may be too subtle to detect, especially early in the course of the disease [8]. For its overall availability, reduced costs, and its utility in distinguishing distinct patterns of cortical hypometabolism in the initial stages of the neurological disease the FDG PET imaging modality has been the focus of this study.

FDG PET has demonstrated diagnostic and prognostic utility in the evaluation of patients with cognitive impairment and in the distinction between primary neurodegenerative disorders and other cognitive impairment [9–11]. However, sometimes it can be challenging to subjectively differentiate between MCI and neurodegenerative pathologies. To improve the precision and predictivity of FDG PET in this setting, a Deep Neural Network model was trained and externally validate to identify patients with an associated neurodegenerative disease among a group of MCI patients.

Materials and methods

Participants

ADNI dataset

Data for this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (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. ADNI is a global research study that actively supports the investigation and development of treatments that slow or stop the progression of Alzheimer's disease (AD). In this multisite longitudinal study, researchers at 63 sites in the US and Canada track the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) in the human brain with clinical, imaging, genetic and biospecimen biomarkers. For up-to-date information, see www.adni-info.org.

FDG PET brain images from 822 patients (472 AD, 350 MCI) acquired at the baseline visit of each subject were downloaded from the ADNI open database. Final ADNI diagnosis at follow-up period for each patient was recorded and used as ground truth, which was determined

from neuropsychological testing, imaging, and fluid sample. This provided the dataset for training and internal validation of the proposed model. Demographical and neuropsychological characteristics of the ADNI dataset including gender, age, education years, Clinical Dementia Rating (CDR), Mini-Mental State Examination (MMSE) score and Functional Activities Questionnaire (FAQ) score are depicted in Table 1. CSF values are not considered since it was not collected from everyone in ADNI study, just from those that volunteered and consented to have a lumbar puncture.

La Fe dataset

For the external dataset from our hospital, patients with MCI in the neuropsychological evaluation in absence of functional impairment in daily living activities were recruited. MCI was defined by a regulated neuropsychological evaluation with a global CDR scale equal to 0.5 (questionable dementia). All cases were retrospectively recruited from the Cognitive Disorders Consultation of the Neurology Service from 2013 through 2020. Patient's data were obtained from the hospital electronic clinical records and images from the Picture Archiving and Communication System (PACS). Patients who presented cognitive functions deterioration, mainly memory, in the absence of functional impairment for activities of daily living, were initially included in the database. Those patients without CDR = 0.5 (MCI condition), without final diagnosis or without an FDG PET imaging were excluded. Figure 1 shows inclusion and exclusion criteria. Approval from the La Fe University and Polytechnic Hospital Ethics Committee was obtained for the study and the patients information consent was waived due to the retrospective non-interventional observational research on data design of the study.

Finally, 90 subjects who met the eligibility criteria were included. The final diagnosis was established after considering physical and neuropsychological examination, FDG PET imaging, structural MR imaging, and CSF molecular analysis from the lumbar punction. This final diagnosis that considers all the results of the mentioned examinations is taken within a maximum period of one year. Furthermore, once the patients were retrospectively chosen meeting the inclusion criteria, the database was revised for greater precision in the clinical diagnosis. Patients were stratified into two groups: 71 MCI cases with associated neurodegenerative disease (64 with AD, 4 FTD and 3 DLB patients), and 19 patients with MCI but without an associated neurodegenerative disease (Fig. 1). Overall, there were 52 women and 38 men, aged between 46 and 78 years, 14 having university studies, 21 secondary studies, 54 primary studies and 1 with no studies. The demographic and neuropsychological information is collected in Table 2.

Table 1 Demographical and neuropsychological characteristics of the ADNI dataset

Groups	MCI	AD
N of subjects	350	472
Female/Male	144/206	190/282
Age*	73 ± 7	74 ± 8
Education* (years)	16 ± 3	16 ± 3
CDR*	0.5	0.66 ± 0.24
MMSE*	28 ± 2	25 ± 3
FAQ*	2.03 ± 3.08	9.49 ± 7.29

MCI Mild Cognitive Impairment, *AD* Alzheimer Disease, *MMSE* Mini-Mental State Examination, *FAQ* Functional Activities Questionnaire

*Values are presented as mean ± standard deviation (SD)

FDG-PET

The FDG PET images were retrieved from the PACS. All patients were acquired in a Philips Gemini TF 64 PET/CT scanner. Acquisition protocol followed the European Association of Nuclear Medicine procedure guidelines [12]. All patients were positioned comfortably in a quiet, dimly lit room about 20–30 min before [^{18}F]FDG administration and during the uptake phase of 135–165 MBq [^{18}F]FDG. Images were acquired between 20 and 35 min post-injection, for a total scan length between 10 and 20 min. 3D images were obtained by Ordered Subset Expectation Maximization (OSEM) image reconstruction algorithm using time of flight. All images were validated by a nuclear medicine physician for quality.

Images were downloaded in the DICOM format with $128 \times 128 \times 90$ matrix resolution, and $2 \times 2 \times 2$ mm voxel size. Images were co-registered to a Dementia-Specific FDG PET template from the Montreal Neurological Institute [13] and smoothed with a 4 mm FWHM filter. The software Statistical Parametric Mapping 12 (SPM12) in MATLAB R2014b was used for this task [14]. All scans were resized to a $110 \times 92 \times 92$ resolution with $2 \times 2 \times 2$ mm voxel size as to match with the template resolution and voxel size, adding two extra voxels in the borders of the z-axis with zero value.

Neural network

A 3D CNN was implemented to separate MCI cases with from without associated neurodegenerative disease. Convolutional neural networks are a state-of-the-art algorithm in computer vision problems, since they have greatly benefited from the recent development and availability of high-performance computing systems together with large-scale data repositories [15] They are particularly useful in image classification, where they manage to learn

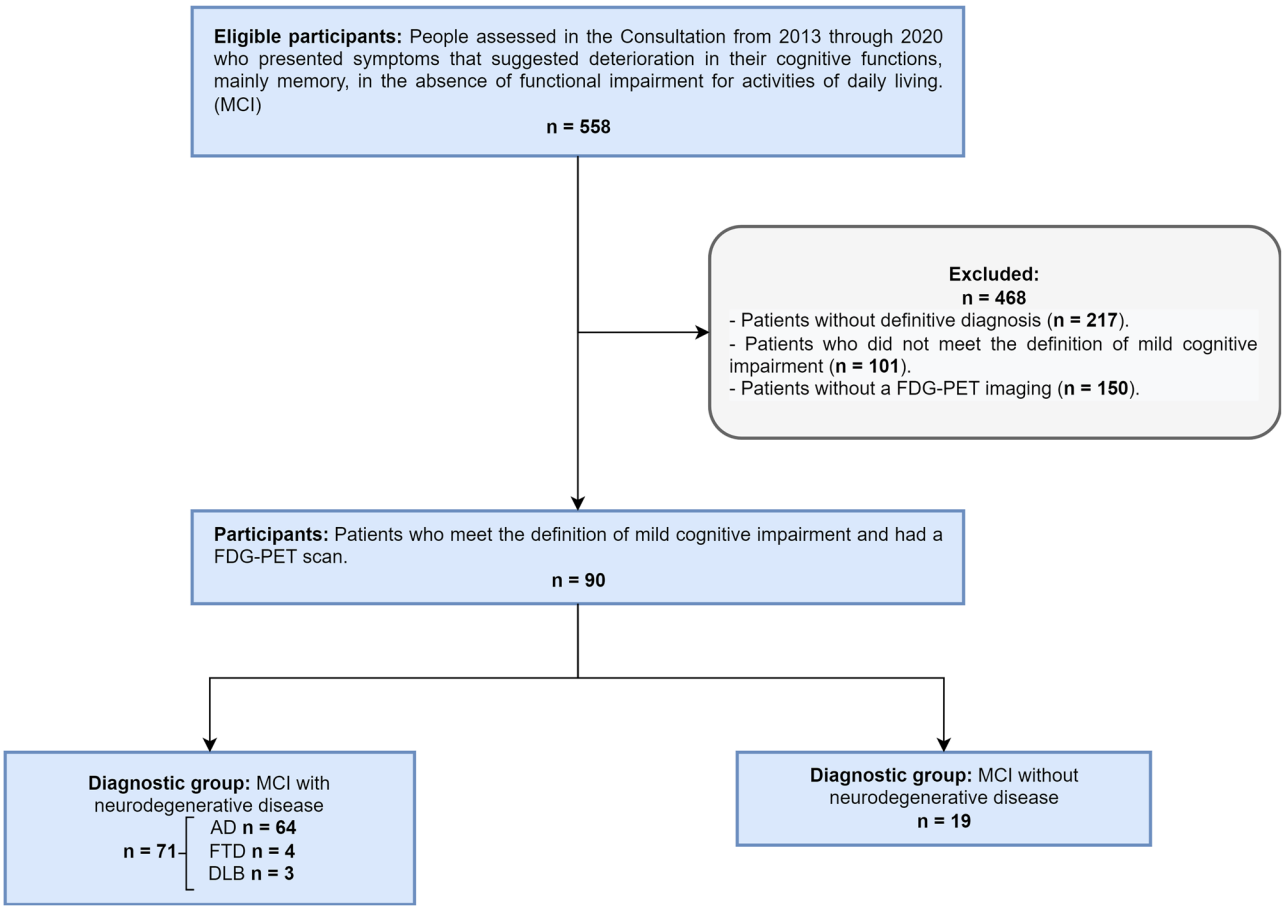

Fig. 1 Eligibility criteria for the participants. From the 558 people assessed in the Consultation 90 participants who met the eligibility criteria were included (71 MCI with neurodegenerative disease and 19 MCI without neurodegenerative disease). All of them met the definition of MCI (Clinical Dementia Rating = 0.5) and had an FDG-PET scan. MCI = Mild Cognitive Impairment, AD = Alzheimer Disease, FTD = Frontotemporal Degeneration, DLB = Dementia with Lewy Bodies

the most important patterns of an input image in the form of an activation map, which is subsequently used in the decision making.

The CNN consists of 3 convolutional and pooling blocks attached to a fully connected layer with rectified linear units as activation function. Convolutional filter size of $3 \times 3x3$, Adam optimizer [16] with $1.5 \cdot 10^{-3}$ learning rate and binary cross-entropy loss function where chosen, with a batch size set to 8. Dropout layers were attached after the convolution blocks and the fully connected layer [17]. This choice of final parameters of the network and batch size was done after some testing starting from a VGG-like convolutional network (Visual Geometry Group Network) [15]which was simplified and adapted for our problem and dataset. A reasonable trade-off between accuracy and computation time was taken into consideration when choosing the architecture and parameter values. The size of the batch was also chosen to avoid computational memory problems. These reduced number of convolutional blocks and filters, together with the addition of dropout layers, were ensured to control overfitting in the training process. A depiction of the network architecture and parameters is shown in Table 3, and the architecture of the neural network is illustrated in Fig. 2.

Python's framework for DL Keras [18] was used for building the model. The 3D CNN was launched on a Linux operating system (Ubuntu 18.04) with 2 Nvidia GeForce RTX 2080 GPUs and Intel i7-7800 × CPU. The 3D CNN was trained and validated on a dataset downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) open database [19]. This dataset consists of 822 subjects (472 AD, 350 MCI). FDG PET brain images acquired at the baseline visit of each subject were used as input for the network. The ADNI diagnosis (baseline diagnosis) considered neuropsychological testing, imaging, and fluid sample. 90% of ADNI dataset was used for training and internal validation of the algorithm and the remaining 10% was used for testing the effectiveness of the algorithm. ADNI image pre-processing included dynamic frame

Table 2 Demographical neuropsychological and cerebrospinal fluid characteristics of the independent sample

Groups	MCI with neurodegenerative disease	MCI without neurodegenerative disease
N of subjects	71	19
Female/Male	41/30	11/8
Age*	69 ± 5	66 ± 7
US/SS/PS/NS	12/16/42/1	2/5/12/0
MMSE*	24 ± 5	25 ± 2
FAQ*	6.94 ± 6.38	5.68 ± 6.16
RBANs*	338 ± 61	357 ± 65
Aβ42 CSF*	559 ± 267	1003 ± 384
p-tau CSF*	110 ± 65	43 ± 17
t-tau CSF*	735 ± 506	286 ± 113

MCI Mild Cognitive Impairment, *US* university studies, *SS* secondary studies, *PS* primary studies, *NS* no studies, *MMSE* Mini-Mental State Examination, *FAQ* Functional Activities Questionnaire, *RBANs* Repeatable Battery for the Assessment of Neuropsychological Status, *CSF* cerebrospinal fluid

*Values are presented as mean ± standard deviation (SD)

co-registration and averaging. Further pre-processing steps were carried out including co-registration to a Dementia-Specific FDG PET template [13] and smoothing with a 4 mm FWHM filter. The software SPM12 [14], was used for this task. With this setup and without taking into consideration the image pre-processing tasks which had to be revised manually, the computation time for our network to load the images, train and test was up to a few hours. Once the network was trained and tested, the model was applied to the hospital independent set of images to perform an external validation of the CNN performance.

Table 3 3D-CNN architecture and hyperparameters

Layer ID	Layer	Kernel number	Kernel size
0	input		
1	Conv3D-1	2	(3,3,3)
2	Max Pooling3D-1		(2,2,2)
3	Dropout (0.4)		
4	Conv3D-2	8	(3,3,3)
5	Max Pooling3D-2		(2,2,2)
6	Dropout (0.4)		
7	Conv3D-3	16	(3,3,3)
8	Max Pooling3D-3		(2,2,2)
9	Dropout (0.4)		
10	Fully Connected	128	
11	Dropout (0.5)		
12	Sigmoid	2	

Results

To evaluate the performance of the developed deep learning model based on ADNI data, several classification metrics such as Sensitivity, Specificity, Balanced Accuracy, and area under the receiver operating characteristic curve (AUC) were considered. The CNN tested on the 10% ADNI dataset (41 MCI and 41 AD subjects) yielded a sensitivity and specificity of 88% (36 out of 41 AD subjects correctly identified) and 71% (29 out of 41 MCI) respectively, with an accuracy of 79% in classifying AD from MCI patients. An AUC of 0.897 was obtained.

This same model was applied to the institution dataset of 90 patients, 19 MCI without an associated neurodegenerative disease and 71 MCI with an associated neurodegenerative disease. The sensitivity and specificity of 75% (53 out of 71 MCI with an associated neurodegenerative disease identified) and 84% (16 out of 19 MCI without an associated neurodegenerative disease identified) values were respectively obtained, with a balanced accuracy of 80% and an AUC of 0.860. In detail, in the MCI with an associated neurodegenerative disease group 48 out of 64 AD, 4 out of 4 FTD and 1 out of 3 DLB were correctly classified by the model. The following table and figures provide an overview of this information: Table 4 includes the comparison between ADNI and La Fe classification metrics. For the ADNI test dataset confusion matrix and receiver operating characteristic curve are depicted, respectively, in Figs. 3 and 4, and for the hospital dataset, in Figs. 5 and 6.

In order to check for the interpretability of the developed model, several visualization tools for the generation of the network attention maps have been investigated, highlighting relevant features of the input image which the learned model uses in the final prediction. Saliency maps [20], SmoothGrad [21] and Grad-CAM [22] visualization algorithms were computed and averaged to both ADNI validation and the institution dataset. Results from the validation set are depicted in Figs. 7, 8, and 9.

Discussion

In this study, a Deep Learning algorithm that classifies patients who present MCI into two diagnostic groups by using FDG PET imaging modality was proposed. The classification allows us to distinguish between those who present a neurodegenerative disease and those who doesn't. Cognitively unimpaired participants were not included because they can be more easily classified by other examinations, and therefore an imaging scan is not

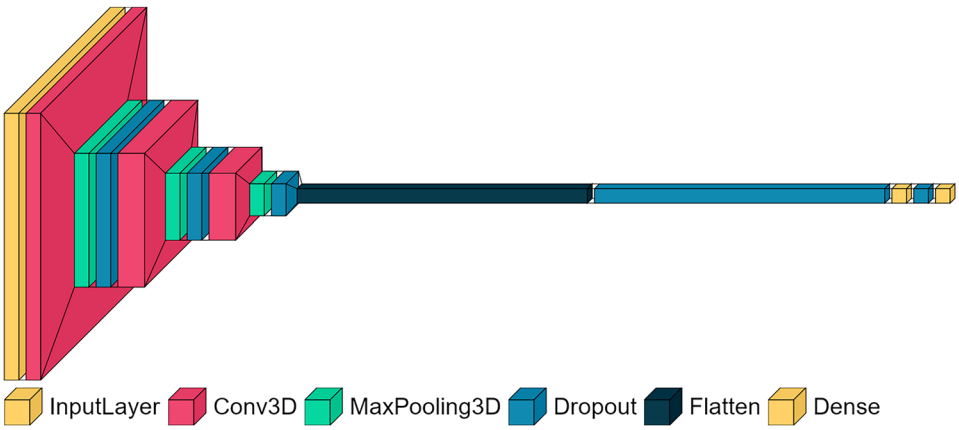

Fig. 2 Neural network architecture. The CNN consists of 3 convolutional and pooling blocks (Conv3D + MaxPooling3D + Dropout) attached to a fully connected layer after a flatten layer

ordered. Likewise, patients with a higher clinical dementia rating were not included either, which present moderate and severe dementia. Although other studies include them, it was considered that the most challenging classification in the clinical routine is to distinguish those patients who have a neurodegenerative disease due to their MCI among a group of patients in their initial stage of cognitive impairment.

The DL neural network was trained and validated first on the ADNI dataset. Then, in order to compare results, the network was validated on a dataset composed of 90 patients whose images were acquired in the hospital La Fe. 71 out of 90 patients were diagnosed as MCI without an associated neurodegenerative disease. The obtained results show that the model trained on the ADNI dataset can be applied on the hospital images preserving 80% of balanced accuracy in the diagnosis classification. The sensitivity and specificity of the validation were 75% and 84% respectively. This validation has demonstrated the reproducibility of the ADNI based trained network in an external dataset. The network exhibits acceptable performance when evaluated on both datasets, enabling to learn the non-complex features in the brain image when performing classification.

A possible source of error can be, on the one hand, the selection bias, because all ADNI subjects had not a CDR = 0.5, which means that the cognitive impairment might not be in the beginnings, while all patients from the validation dataset had CDR = 0.5. Furthermore, the ADNI dataset also included only AD neurodegenerative disease, while in the validation dataset DLB and FTD patients were included in the MCI with an associated neurodegenerative disease group (nevertheless, 100% FTD patients were correctly distinguished from those MCI without an associated neurodegenerative disease, but DLB patients were not as good identified as FTD patients). Even though an algorithm that it has not been trained with those images might not work properly with those cases, this study wanted to bring the results closer to the real hospital situation, where DLB and DFT patients are together with AD patients in the MCI group. On the other hand, unlike magnetic resonance images, there is not a standardized method to process or prepare the FDG PET images, so the considerations may have not be the optimal. Finally, an appropriate partial volume correction (PVC) on FDG PET images might have enhanced the brain network structure analysis and improved classification performance [23]. No PVC was considered due to the lack of MR in part of the 90 patients.

Table 4 3D-CNN classification metrics

Test Set	Balanced Accuracy	Sensitivity	Specificity	AUC
ADNI 10%	0.79	0.88	0.71	0.897
La Fe	0.80	0.75	0.84	0.860

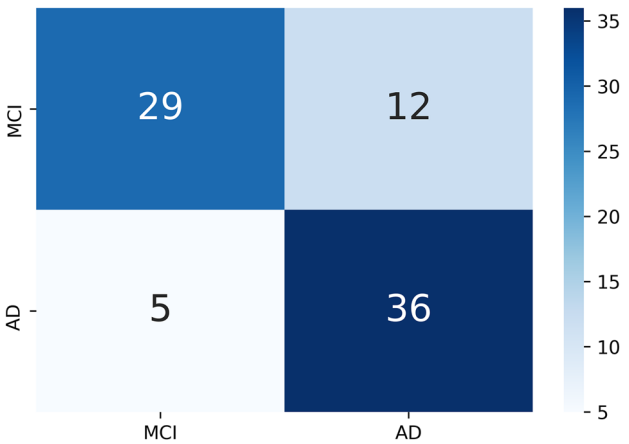

Fig. 3 Confusion matrix for ADNI test dataset. The model performance can be evaluated from this matrix. ADNI test dataset achieved a sensitivity of 88% (36 out of 41 AD subjects correctly identified) and s specificity of 71% (29 out of 41 MCI) with an accuracy of 79% in classifying AD from MCI patients. MCI = Mild Cognitive Impairment, AD = Alzheimer Disease

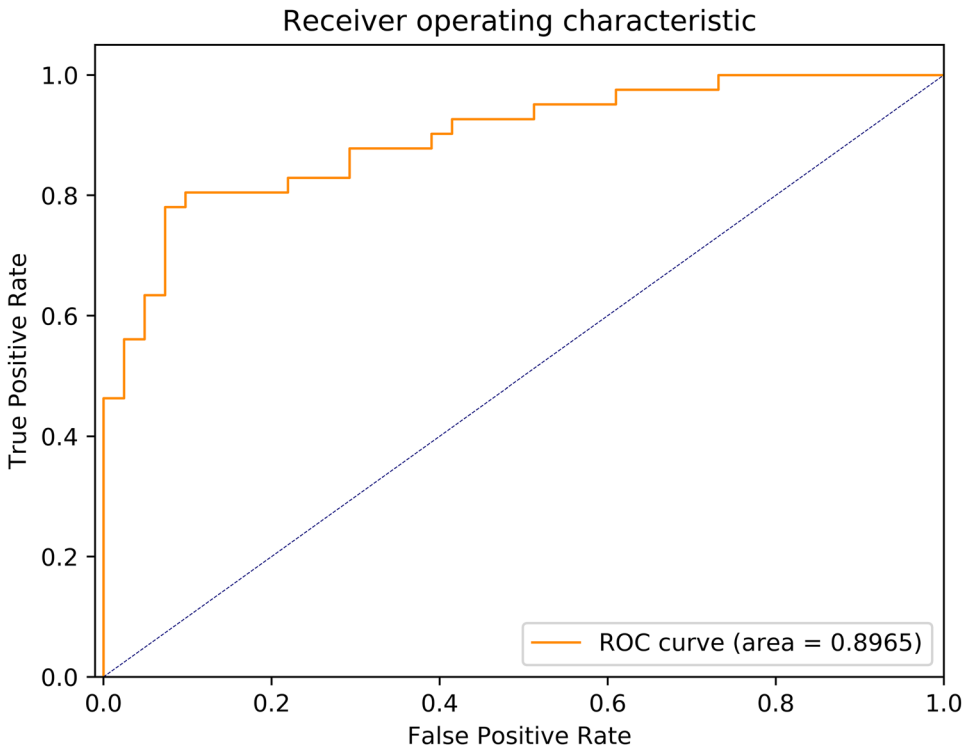

Fig. 4 ROC curves for ADNI test dataset. The model performance can be evaluated from the area under the receiver operating characteristic curve (AUC). For ADNI test dataset the AUC obtained is 0.897

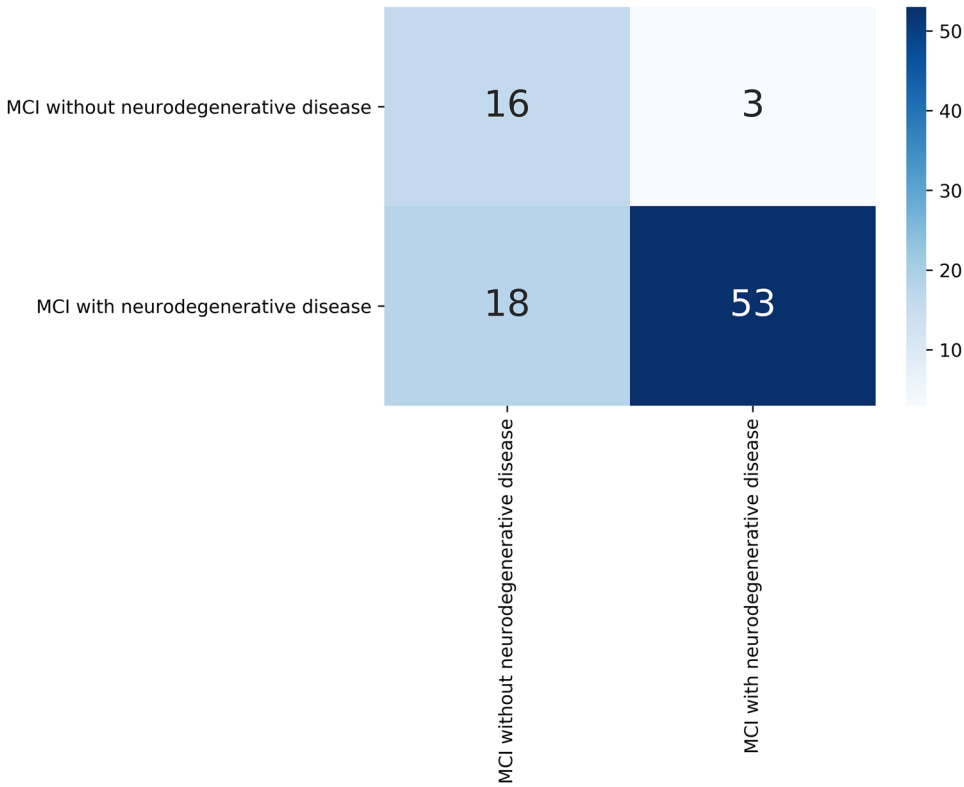

Fig. 5 Confusion matrix for La Fe dataset. The model performance can be evaluated from this matrix. La Fe dataset achieved a sensitivity of 75% (53 out of 71 MCI with an associated neurodegenerative disease identified) and specificity of 84% (16 out of 19 MCI without an associated neurodegenerative disease identified), with a balanced accuracy of 80%. MCI = Mild Cognitive Impairment

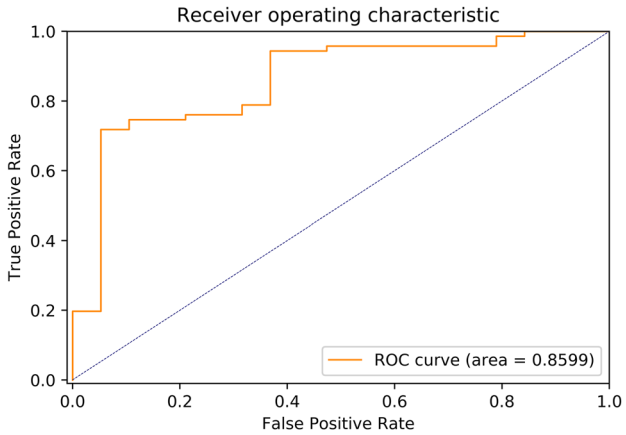

Fig. 6 ROC curves for La Fe dataset. The model performance can be evaluated from the area under the receiver operating characteristic curve (AUC). For La Fe dataset the AUC obtained is 0.897

Another limitation of the study is thought to be the small size of the dataset, compared with other studies [24] which are of the order of thousands of images in the training process. The size of the datasets were restricted by the criteria of our experiment and the availability of data, since only the first scan of each patient at baseline visit was considered. Despite this, a reasonable performance was achieved.

Apart from that, as shown in the network attention maps, Saliency maps and SmoothGrad appear to be highlighting posterior cingulate and superior parietal areas. These brain sections agree with the FDG endophenotype of AD [25]. However, Grad-CAM method show right prefrontal brain area being focused, which is not typical of AD. This leads us to speculate that SmoothGrad is the most reliable method for this study because it is consistent with the FDG endophenotype of AD. Still, these visualization methods for debugging the neural network need to be treated carefully, since they

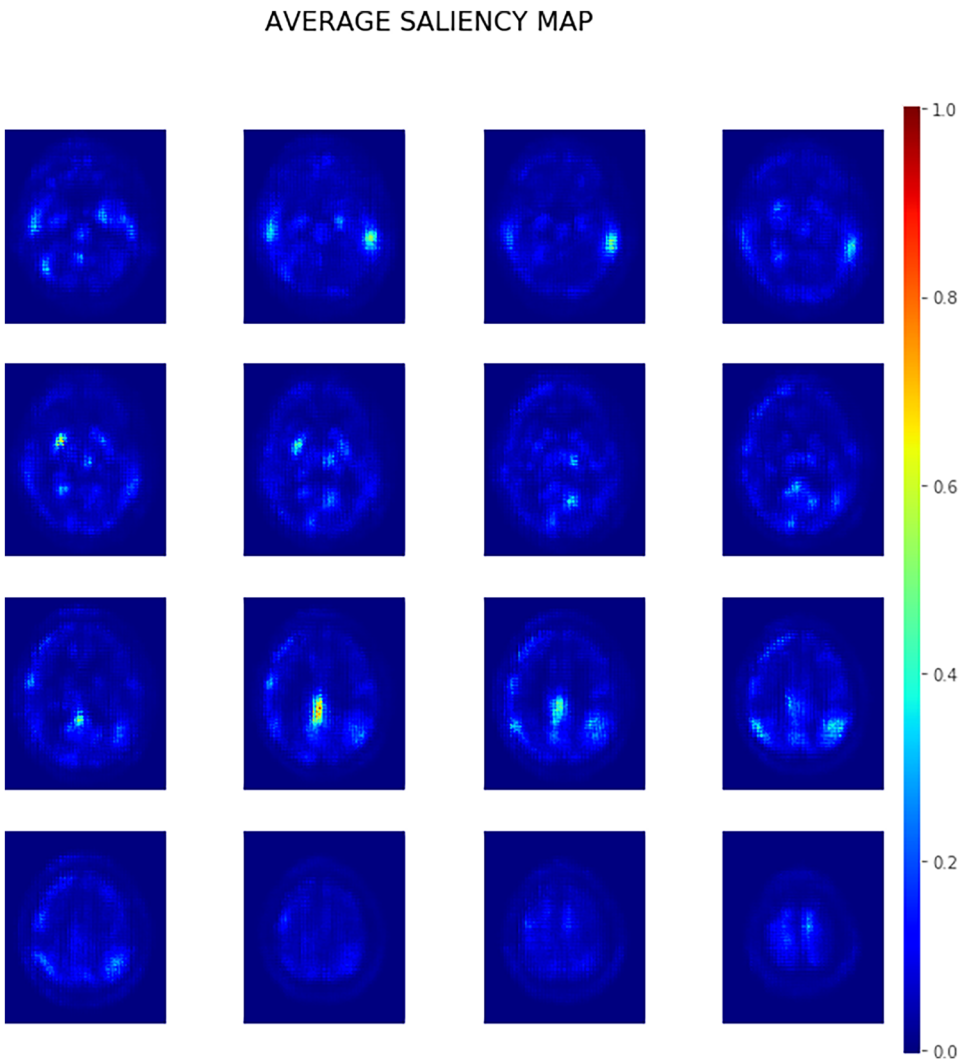

Fig. 7 Saliency maps visualization algorithm computed and averaged on La Fe dataset. Saliency maps are a type of neural network attention maps which lead us to check for the model interpretability. This map appears to be highlighting posterior cingulate and superior parietal areas. These brain sections agree with the FDG endophenotype of Alzheimer Disease

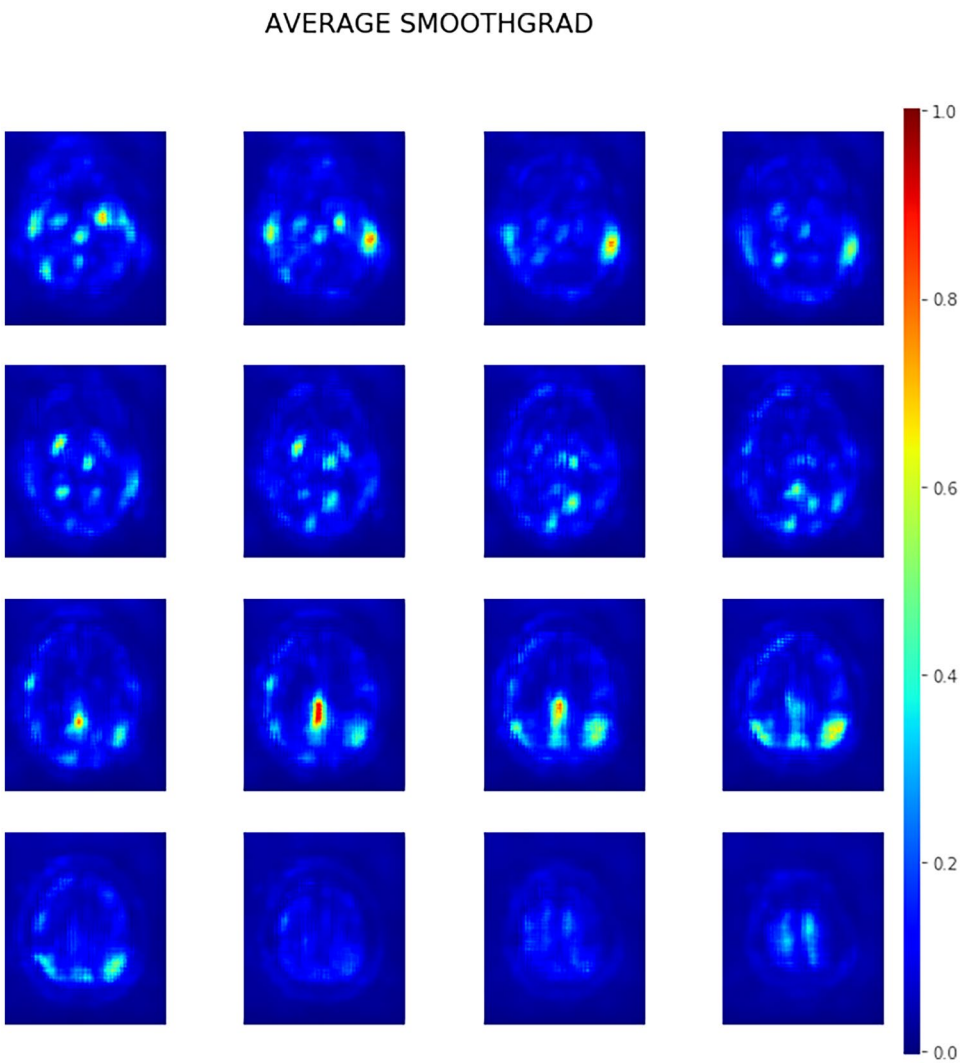

Fig. 8 SmoothGrad visualization algorithm computed and averaged on La Fe dataset. SmoothGrad maps are a type of neural network attention maps which lead us to check for the model interpretability. This map appears to be highlighting posterior cingulate and superior parietal areas. These brain sections agree with the FDG endophenotype of Alzheimer Disease

are very fragile to small perturbations in the input image [26]. Other studies have also reported this fact. Ding et al. [24] found that the Saliency map suggested that their model was considering the whole brain when making a prediction. Further research in this topic is needed in order to achieve more robustness in the interpretation of neural networks.

There are several AI studies which their aim is to classify neurodegenerative diseases [5, 24, 27–31]. A brief summary is depicted in Table5. Although some of them achieve better results for the accuracy, specificity, or sensitivity metrics, there are few studies which only used PET imaging for the classification, most of them also use MR and/or other biomarkers. Additionally, most of the studies do not focus on classifying patients who have CDR = 0.5 (MCI condition), which is the difficult clinical goal. For example, Manhua et al. [31] obtained better balanced accuracy than in this study using FDG PET images for differentiating AD or MCI vs HC, but in their study these subjects can be perfectly classified (accuracy = 1, sensibility = 1 and specificity = 1) by their CDR, because all HC subjects had CDR = 0, AD patients had CDR = 0.8 ± 0.25 and MCI subjects, CDR = 0.5 ± 0.04, while in the validation dataset of the present study all subjects had the MCI condition (CDR = 0.5). Moreover, few studies have done an external validation to verify its reproducibility. Ding et al. [24] also developed a deep learning algorithm using ADNI FDG PET images and validated the model on an independent dataset. Unlike the present study, they also included cognitively unimpaired participants, so they did a 3-class classification. Another difference between both studies is that Ding et al. did not exclude patients with a high clinical dementia rating in their validation, that may not be in the initial stage of cognitive impairment. Although 3-class classification results are not easy to compare with the results of a binary classifier, if you analyse the balanced accuracy Ding et al. obtained 59% on their model validation in the independent dataset while in this study 80% balanced accuracy was achieved. They also compared the algorithm performance

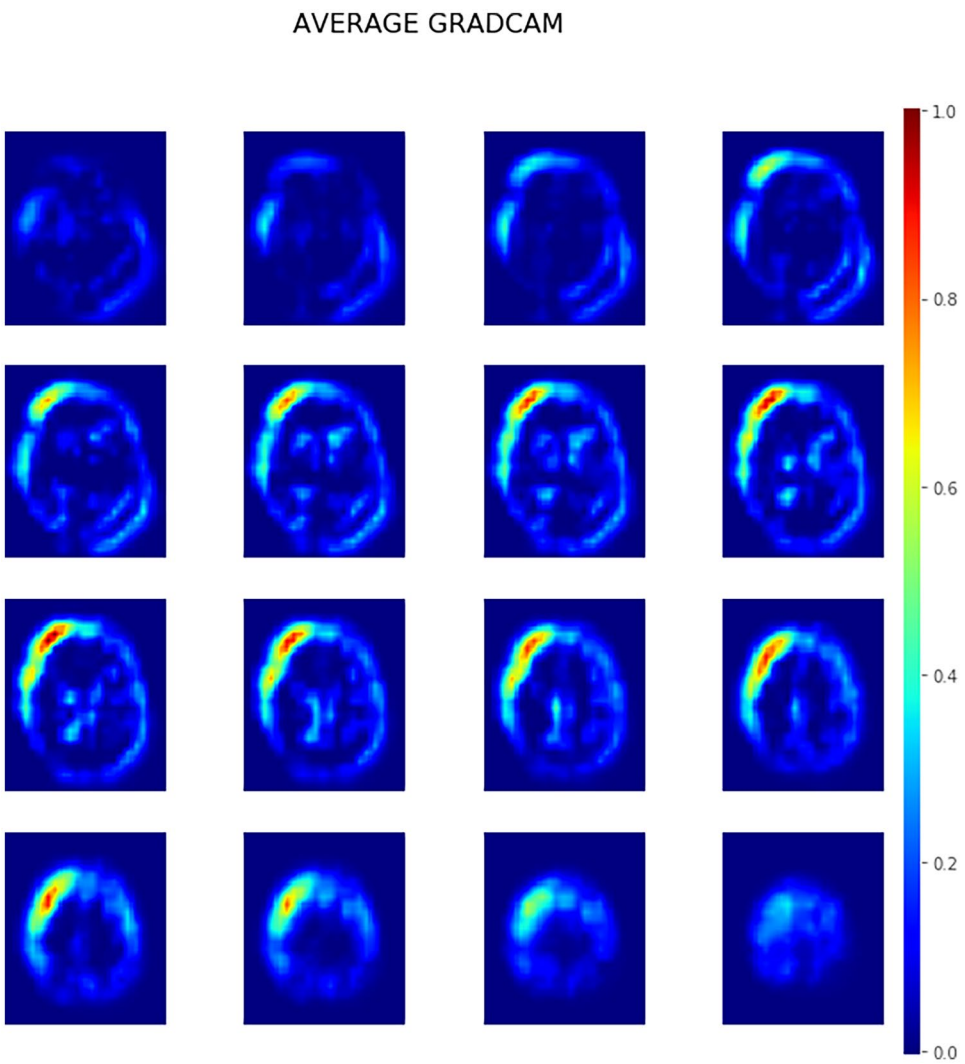

Fig. 9 Grad-CAM visualization algorithm computed and averaged on La Fe dataset. Grad-CAM maps are a type of neural network attention maps which lead us to check for the model interpretability. This map shows right prefrontal brain area being focused, which is not typical of Alzheimer Disease

to that of radiologic readers and they showed that the model outperformed three radiology readers in predicting the final diagnosis of AD. Their result reinforces the evidence that deep learning models applied to the classification of neurodegenerative diseases are a great support for nuclear medicine physicians. However, this study has gone beyond the comparison with radiology readers. In the present study the performance of the Deep Learning algorithm has been tested against the final diagnosis of the patient, which was determined not only by the reports of the radiologists but the neuropsychological examinations and CSF molecular analysis. It has been considered that the comparison against the diagnosis as the ground truth follows a stricter criterion and therefore, it is a strongest goal.

Table 5 State of the art studies where Artificial Intelligence has been applied to classify neurodegenerative diseases

Reference	Data source	Dataset modalities	N subjects	Dataset split ratio	Classification groups	Sensibility	Specificity	Balanced accuracy	AUC	External validation
[3]	ADNI	FDG PET + sMRI + CSF + APOE	N = 158 (38 AD, 36 MCIs, 46 MCIc, 38 HC)	train 75% test 25%	AD vs HC	1	0.96	0.98	0.983	No
					MCIs vs MCIc	1	0.89	0.95	0.936	
					AD vs MCIs	1	0.92	0.96	0.968	
					AD vs MCIc	0.92	0.93	0.92	0.946	
					HC vs MCIc	0.95	1	0.97	0.964	
					HC vs MCIs	1	0.89	0.95	0.952	
		FDG PET			AD vs HC	0.90	0.90	0.90	0.925	
					MCIs vs MCIc	1	0.85	0.93	0.904	
					AD vs MCIs	0.90	0.92	0.91	0.892	
					AD vs MCIc	0.66	0.88	0.77	0.840	
					HC vs MCIc	0.88	1	0.94	0.911	
					HC vs MCIs	0.92	0.80	0.86	0.897	
[22]	ADNI	FDG PET	N = 1002 (236 AD, 406 MCI, 360 Non-AD/MCI)	train 90% test 10%	AD vs	0.81	0.94	0.65	0.920	Yes
					MCI vs	0.54	0.68		0.630	
					Non-AD/MCI	0.59	0.75		0.730	
	Author's institution		N = 40 (7 AD, 7 MCI, 26 Non-AD/MCI)	test 100%	AD	1	0.82	0.59	0.980	
					MCI	0.43	0.58		0.340	
					Non-AD/MCI	0.35	0.93		0.840	
[25]	ADNI	MRI + FDG PET	N = 397 (93 AD, 76 pMCI, 128 sMCI, 100 HC)	tenfold cross-validation	AD vs HC	0.98	0.92	0.95	0.968	No
					pMCI vs HC	0.83	0.89	0.86	0.911	
					sMCI vs HC	0.71	0.60	0.65	0.692	
[26]	ADNI + European DLB Consortium	FDG PET	N = 757 (200 AD, 200 MCI, 157 DLB, 200 HC)	train 90% test 10%	AD vs	0.91	0.92	0.70	0.964	No
					MCI vs	0.17	0.94		0.714	
					DLB vs	0.86	1		0.962	
					HC	0.88	0.90		0.947	
[27]	ADNI	MRI + FDG PET	N = 397 (93 AD, 76 pMCI, 128 sMCI, 100 HC)	-	AD vs HC	0.93	0.94	0.93	0.967	No
					pMCI vs HC	0.81	0.84	0.83	0.884	
					sMCI vs HC	0.63	0.67	0.65	0.670	
		FDG PET			AD vs HC	0.91	0.86	0.88	0.945	
					pMCI vs HC	0.78	0.79	0.78	0.860	
					sMCI vs HC	0.64	0.66	0.64	0.666	
[28]	Chosun University, NRCD	MMSE + rs-fMRI	N = 331 (133 AD, 198 HC)	tenfold cross-validation	AD vs HC	0.98	0.67	0.85	-	No
[29]	ADNI	FDG PET	N = 339 (93 AD, 146 MCI, 100 HC)	tenfold cross-validation	AD vs HC	0.91	0.91	0.91	0.953	No
					MCI vs HC	0.78	0.80	0.79	0.839	

AUC Area Under the receiver operating characteristic Curve, *ADNI* Alzheimer's Disease Neuroimaging Initiative, *FDG PET* Fluorodeoxyglucose Positron Emission Tomography, *sMRI* structural Magnetic Resonance Image, *CSF* Cerebrospinal Fluid, *APOE* Apolipoprotein E genotype, *AD* Alzheimer Disease, *MCIs* Mild Cognitive Impairment stable, *MCIc* Mild Cognitive Impairment converter, *HC* Healthy Controls, *MCIp* progresive Mild Cognitive Impairment, *DLB* Dementia with Lewy Bodies, *rs-fMRI* resting-state functional Magnetic Resonance Image, *NRCD* National Research Center for Dementia

Conclusions

A binary classifier model based on a 3D Convolutional neural network using the FDG PET patient baseline image was successfully trained and validated. This algorithm allows the early prediction of neurodegenerative diseases in MCI patients in standard clinical settings with an overall 80% classification balanced accuracy.

The model can help for the early non-invasive prediction of neurodegenerative diseases in patients with MCI. This Deep Learning model is still not being used in the hospital, but it can be integrated as an experimentation decision support tool for the nuclear medicine physicians in day-to-day care. In order to improve the model for future applications, a new neural network architectures should be further developed to try to improve results by applying transfer learning from the DL neural network trained on the ADNI images to the hospital dataset. Also including other non-image variables as in biometric values which are important in the neurodegenerative disease diagnosis to finally get better classification results. Further routines of the code could be investigated to improve performance, as in a more exhaustive search of the optimal parameters of the network and of the pre-processing image steps.

Acknowledgements Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). 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

Funding This work was financially supported by INBIO 2019 (DEEPBRAIN), INNVA1/2020/83(DEEPPET) funded by Generalitat Valenciana, and PID2019-107790RB-C22 funded by MCIN/AEI/10.13039/501100011033/. 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.; Cogstate; 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 Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Data availability The training and test datasets analysed during the current study are part of the ADNI repository, http://www.loni.ucla.edu/ADNI. The validation dataset analysed during the current study is available from the corresponding author upon reasonable request. The deep learning network algorithm developed during the current study is available from the first author upon reasonable request.

Declarations

Ethics approval Approval was granted by the La Fe University and Polytechnic Hospital Ethics Committee for the study with registration number: 2019–084-1 and the patients information consent was waived due to the retrospective non-interventional observational research on data design of the study.

Competing interests The authors declare that they have no competing interests.

References

1. Karagiannidou, M. P., Comas-Herrera, A., Knapp, M., Guerchet, M. (2016) World Alzheimer Report 2016 Improving healthcare for people living with dementia. Coverage, Quality and costs now and in the future. Alzheimer's Disease International (ADI). https://www.alzint.org/u/WorldAlzheimerReport2016.pdf Accessed 30 May 2022.
2. Galende, A. V., Ortiz, M. E., Velasco, S. L., Luque, M. L., de Miguel, C. LDS., Prieto, C., Jurczynska, CP. (2021) Report by the Spanish Foundation of the Brain on the social impact of Alzheimer disease and other types of dementia. *Neurologia* **36**.
3. Morris, JC., Storandt, M., Miller, JP., McKeel, DW., Price, JL., Rubin, EH., Berg, L. (2001) Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology*.
4. Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., Martin, R. L. (1982) A new clinical scale for the staging of dementia. *British Journal of Psychiatry*.
5. Gupta, Y., Lama, R. K., Kwon, GR. (2019) Prediction and Classification of Alzheimer's Disease Based on Combined Features From Apolipoprotein-E Genotype, Cerebrospinal Fluid, MR, and FDG-PET Imaging Biomarkers. *Frontiers in Computational Neuroscience* **13**.
6. Gamez-Cenzano, C., Robles-Barba, J., Rodriguez-Bel, J. L., Gascon-Bayarri, J., Cortes-Romera, M., Sabate-Llobera, A., Gracia-Sanchez, LM., Romero-Zayas, I., Rocaengronyat, M., Vercher-Conejero, J., Majos-Torro, C., Soriano-Mas, C., Aguilera Grijalvo, C. (2015) Impact of PET brain imaging using F18-FDG and F18-FLORBETAPIR in patients with cognitive impairment. *European Journal of Nuclear Medicine and Molecular Imaging* **42**.
7. Nobili, F., Arbizu, J., Bouwman, F., Drzezga, A., Agosta, F., Nestor, P., Walker, Z., Boccardi, M. (2018) European Association of Nuclear Medicine and European Academy of Neurology recommendations for the use of brain 18 F-fluorodeoxyglucose positron emission tomography in neurodegenerative cognitive impairment and dementia: Delphi consensus. *Eur J Neurol*.
8. Dave, A., Hansen, N., Downey, R., Johnson, C. (2020) FDG-PET Imaging of Dementia and Neurodegenerative Disease. *Seminars in Ultrasound, CT and MRI* **41**.

9. Silverman, DHS., Mosconi, L., Ercoli, L., Chen, W., Small, GW. (2008) PET Scans Obtained for Evaluation of Cognitive Dysfunction. *Seminars in nuclear medicine* **38**.
10. Marcus, C., Mena, E., Subramaniam, RM. (2014) Brain PET in the diagnosis of Alzheimer's disease. *Clinical Nuclear Medicine* **39**.
11. Gámez-Cenzano, C., Rodríguez-Bel, L., Gascón-Bayarri, J., Reñé-Ramírez, R., Campdelacreu-Fumado, J., Turón-Sans, J., Soriano-Mas, C., Vercher-Conejero, J., Gràcia-Sánchez, L., Llinares-Tello, E., Pons-Escoda, A., C., MT. (2016) Role of 18F-FDG-PET and amyloid-PET imaging on patient management in mild cognitive impairment or dementia. *European Journal of Nuclear Medicine and Molecular Imaging* **43**.
12. Varrone, L. A., Asenbaum, S., Vander-Borght, T., Booij, J., Nobili, F., Någren, K., Darcourt, J., Kapucu, O. L., Tatsch, K., Bartenstein, P., Van, K. (2009) EANM procedure guidelines for PET brain imaging using [18F]FDG, version 2. *European Journal of Nuclear Medicine and Molecular Imaging* **36**.
13. Della Rosa, PA., Cerami, C., Gallivanone, F., Prestia, A., Caroli, A., Castiglioni, I., Gilardi, M. C., Frisoni, G., Friston, K., Ashburner, J., Perani, D. (October 2014) A standardized [18F]-FDG-PET template for spatial normalization in statistical parametric mapping of dementia. *Neuroinformatics* **12**.
14. Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W. (2007) *Statistical Parametric Mapping*. Academic Press London.
15. Simonyan, K., A., Z. (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556.
16. Kingma, D. P., Ba, J. (2014) Adam: A Method for Stochastic Optimization..
17. Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014) Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* **15**.
18. Chollet, Francois, others (Accessed 2005) Keras. Available at: https://github.com/fchollet/keras
19. (Accessed ADNI) Alzheimer's Disease Neuroimaging Initiative. Available at: http://adni.loni.usc.edu/
20. Simonyan, K., Vedaldi, A., Zisserman, A. (2014) Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *CoRR*.
21. Smilkov, R., Thorat, N., Kim, B., Viégas, F., Wattenberg, M. (2017) Smoothgrad: removing noise by adding noise. *Workshop on Visualization for Deep Learning, ICML*.
22. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization..
23. Yang, J., Hu, C., Guo, N., Dutta, J., Vaina, LM., Johnson, KA., Sepulcre, J., El-Fakhri, G., Li, Q. (2017) Partial volume correction for PET quantification and its impact on brain network in Alzheimer's disease. *Scientific Reports* **7**.
24. Ding, Y., Sohn, J., Kawczynski, M., Trivedi, H., Harnish, R., Jenkins, N., Lituiev, D., Copeland, T., Aboian, M., Mari Aparici, C., Behr, S., Flavell, R., Huang, S., Zalocusky, K., Nardo, L., Seo, Y., Hawkins, R. (2018) A deep learning model to predict a diagnosis of Alzheimer disease by using 18 F-FDG PET of the brain. *Radiology* **290**.
25. Johnson, K. A., Fox, N. C., Sperling, R. A., Klunk, W. E. (2012) Brain imaging in Alzheimer disease. *Cold Spring Harbor* **2**.
26. Ghorbani, A., Abid, A., Zou, J. (2019) Interpretation of Neural Networks Is Fragile. *Proceedings of the AAAI Conference on Artificial Intelligence* **33**.
27. Feng, C., Elazab, A., Yang, P., Wang, T., Zhou, F., Hu, H., Xiao, X., Lei, B. (2019) Deep Learning Framework for Alzheimer's Disease Diagnosis via 3D-CNN and FSBi-LSTM. *IEEE Access* **7**.
28. Etminani, K., Soliman, A., Davidsson, A., Chang, JR., Martínez-Sanchis, B., Byttner, S., Camacho, V., Bauckneht, M., Stegeran, R., Ressner, M., Agudelo-Cifuentes, M. (2022) A 3D deep learning model to predict the diagnosis of dementia with Lewy bodies, Alzheimer's disease, and mild cognitive impairment using brain 18F-FDG PET. *European Journal of Nuclear Medicine and Molecular Imaging* **49**.
29. Liu, M., Cheng, D., Wang, K., Wang, Y. (2018) Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis. *Neuroinformatics* **16**.
30. Duc, NT., Ryu, S., Qureshi, MNI., Choi, M., Lee, KH., Lee, B. (2020) 3D-Deep Learning Based Automatic Diagnosis of Alzheimer's Disease with Joint MMSE Prediction Using Resting-State fMRI. *Neuroinformatics* **18**.
31. Manhua, L., Cheng, D., Weiwu, Y. (2018) Classification of Alzheimer's Disease by Combination of Convolutional and Recurrent Neural Networks Using FDG-PET Images. *Frontiers in Neuroinformatics* **12**.