NeuroImage 297 (2024) 120695

Contents lists available at ScienceDirect

NeuroImage

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

Predicting changes in brain metabolism and progression from mild cognitive impairment to dementia using multitask Deep Learning models and explainable AI

Fernando García-Gutiérrez [a], Laura Hernández-Lorenzo [a], María Nieves Cabrera-Martín [b], Jordi A. Matias-Guiu [c,*], José L. Ayala [a], the Alzheimer's Disease Neuroimaging Initiative [1]

[a] *Department of Computer Architecture and Automation, Universidad Complutense, Madrid, Spain*
[b] *Department of Nuclear Medicine, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Madrid, Spain*
[c] *Department of Neurology, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Madrid, Spain*

ARTICLE INFO

Keywords:
Neurodegenerative diseases
Alzheimer disease
Neuroimaging
Positron-emission tomography
Artificial intelligence
Machine learning
Deep learning
Automated pattern recognition

ABSTRACT

Background: The prediction of Alzheimer's disease (AD) progression from its early stages is a research priority. In this context, the use of Artificial Intelligence (AI) in AD has experienced a notable surge in recent years. However, existing investigations predominantly concentrate on distinguishing clinical phenotypes through cross-sectional approaches. This study aims to investigate the potential of modeling additional dimensions of the disease, such as variations in brain metabolism assessed via [^{18}F]-fluorodeoxyglucose positron emission tomography (FDG-PET), and utilize this information to identify patients with mild cognitive impairment (MCI) who will progress to dementia (pMCI).

Methods: We analyzed data from 1,617 participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI) who had undergone at least one FDG-PET scan. We identified the brain regions with the most significant hypometabolism in AD and used Deep Learning (DL) models to predict future changes in brain metabolism. The best-performing model was then adapted under a multi-task learning framework to identify pMCI individuals. Finally, this model underwent further analysis using eXplainable AI (XAI) techniques.

Results: Our results confirm a strong association between hypometabolism, disease progression, and cognitive decline. Furthermore, we demonstrated that integrating data on changes in brain metabolism during training enhanced the models' ability to detect pMCI individuals (sensitivity=88.4%, specificity=86.9%). Lastly, the application of XAI techniques enabled us to delve into the brain regions with the most significant impact on model predictions, highlighting the importance of the hippocampus, cingulate cortex, and some subcortical structures.

Conclusion: This study introduces a novel dimension to predictive modeling in AD, emphasizing the importance of projecting variations in brain metabolism under a multi-task learning paradigm.

1. Introduction

Alzheimer's disease (AD), responsible for 60%–80% of dementia cases worldwide, remains a pressing concern with far-reaching implications for current healthcare systems (Alzheimer's and Dementia, 2023). In the United States alone, it is estimated that in 2023, AD affects 6.7 million people aged over 65 years (Rajan et al., 2021). In addition, with the progressive aging of the population and the increase in life expectancy, it is projected that the prevalence of the disease will continue to grow (Rajan et al., 2021). All these factors have placed AD as one of the leading health problems of the 21st century.

Clinically, the typical presentation of AD is characterized by memory loss, changes in other cognitive functions and behavior, together with impairment in daily living activities (Peña-Casanova et al., 2012). These clinical symptoms stem from a cascade of neuropathological events within the brain. Such alterations include the extracellular

---

* Corresponding author.
E-mail addresses: fegarc05@ucm.es (F. García-Gutiérrez), laurah11@ucm.es (L. Hernández-Lorenzo), marcab06@ucm.es (M.N. Cabrera-Martín), jordi.matias-guiu@salud.madrid.org (J.A. Matias-Guiu), jayala@ucm.es (J.L. Ayala).

[1] 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: ADNI.

https://doi.org/10.1016/j.neuroimage.2024.120695
Received 13 May 2024; Accepted 18 June 2024
Available online 26 June 2024
1053-8119/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

formation of amyloid-β plaques and the intracellular aggregation of phosphorylated tau protein tangles (DeTure and Dickson, 2019). Simultaneously, the accumulation of these substrates triggers a neuroinflammatory process and disrupts the neurotransmitter systems, resulting in neuronal loss, atrophy, and synaptic failure, which manifests as the observed clinical symptoms (Jack, 2022).

In this context, several studies have demonstrated that the spectrum of neuropathological events that define the AD landscape initiates approximately 20 years, or even earlier, before the onset of clinical symptoms (Thal et al., 2002). Thus, much of the scientific community efforts have focused on developing biomarkers to detect the disease at its early stage or predict how it will evolve (Ansart et al., 2021). These biomarkers are of interest for intervening in modifiable risk factors, enrolling subjects for clinical trials, studying the pathophysiological mechanisms of the disease, and addressing aspects related to the subject's future planning (Alzheimer's and Dementia, 2023).

In this context, the growing relevance of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) for modeling different aspects of AD is well acknowledged. The potential of these techniques derives from their ability to integrate large amounts of heterogeneous data and uncover non-linear patterns. While their deployment in the clinical environment is still limited, AI techniques are promising tools capable of providing valuable information in the medical decision-making process and biomarker development (Ansart et al., 2021; Sharma et al., 2023).

Therefore, based on the observed alterations in the brain along the AD continuum, the neuroimaging domain has become a focal point for many ML studies. The development of predictive models primarily relies on data derived from structural/functional magnetic resonance imaging (sMRI/fMRI) (Ansart et al., 2021; Sharma et al., 2023) or positron emission tomography (PET) with radiotracers like florbetapir, flortaucipir, or [^{18}F]-fluorodeoxyglucose (FDG), which quantify amyloid, tau protein, and glucose uptake, respectively (Nordberg et al., 2010). As many studies have shown, the high dimensionality and complexity of neuroimaging data have encouraged using ML and DL techniques over classical approaches unable to deal with these aspects (Arbabshirani et al., 2017).

Among these modalities, sMRI has been the most extensively explored (Liu et al., 2018; Ghazi et al., 2019; Abrol et al., 2020; Sauty and Durrleman, 2022). Nevertheless, in recent years, there has been an increasing interest in integrating information from FDG-PET into AI models. It stems from its ability to offer a more direct proxy of neuronal activity by quantifying brain metabolism (Sala and Perani, 2019). As a result, some diagnostic models have been developed, achieving accuracies above 90% in identifying subjects with AD (Katako et al., 2018; Ding et al., 2019; Díaz-Álvarez et al., 2022; Duan et al., 2023) and close to 80% for predicting conversion from mild cognitive impairment (MCI) to AD (Hinrichs et al., 2011; Choi et al., 2018; Iaccarino et al., 2019). Furthermore, although less frequently, additional aspects of the disease have been analyzed, encompassing the prediction of cognitive alterations (Teng et al., 2020; El-Sappagh et al., 2020) and the modeling of brain atrophy (Marinescu et al., 2020).

Nonetheless, there is a notable gap in exploring additional dimensions of the disease that are of significant interest in clinical practice. One of these aspects is the variation in cerebral metabolism patterns along the AD spectrum. Considering the observed alterations in brain metabolism during AD, it is valuable to anticipate future fluctuations as they may indicate a worse prognosis for the patient. For example, several studies have found that reductions in brain metabolism correlate with alterations in different cognitive domains (Landau et al., 2011; Arenaza-Urquijo et al., 2019) and imply a higher conversion rate from MCI to dementia (Ou et al., 2019; Alexander et al., 2002). In this context, the most comparative research on modeling fluctuations using FDG-PET, concentrates on delineating disease-level trajectories (Abi Nader et al., 2020).

The primary objective of this study was to model the metabolic dynamics of the brain across the AD continuum. To this end, we used DL techniques based on feed-forward networks (FFNs) and convolutional neural networks (CNNs). A key aspect of our investigation involved the use of multi-task learning strategies to simultaneously predict declines in glucose metabolism and distinguish between different clinical phenotypes (Zhang and Yang, 2021). Through the integration of several objectives into the multi-task model, we endeavored to refine the model's ability to extract patterns from brain images. This approach aimed to improve the accuracy of identifying MCI individuals who are at risk of evolving into dementia. Moreover, the application of eXplainable AI (XAI) methodologies (Arrieta et al., 2020) was a crucial component of our study. We utilized XAI techniques to obtain deeper insights into the models' decision-making processes. It enabled us to identify factors linked to future reductions in cerebral metabolism and disease progression.

2. Methodology

2.1. Study design

This study used FDG-PET images sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, available at (https://adni.loni.usc.edu). Initiated in 2003, ADNI represents a collaborative effort aimed at exploring whether a combination of serial MRI and PET scans, along with other biomarkers and clinical/neuropsychological evaluations, can effectively track the progression of MCI and early AD.

The dataset comprises 3449 observations from 1617 participants, including 438 healthy controls (HC), 858 individuals with MCI, and 321 subjects diagnosed with dementia at their initial evaluation. This study employed these data across various modeling tasks. Specifically, for examining changes in brain metabolism (detailed in Section 2.3), we utilized a subset containing 1166 observations from 674 subjects. This subset featured individuals with at least two FDG-PET scans conducted within a two-year span, fitting within a six-month window.

In analyzing the transition from MCI to dementia (outlined in Section 2.5), we focused on 904 observations from 419 participants initially diagnosed with MCI, with an additional diagnosis available after two years. Among these observations, 33.72% were linked to cases progressing to dementia (pMCI), whereas the remaining 66.28% pertained to cases that remained stable (sMCI). The clinical and sociodemographic details of the participants are summarized in Table 1.

2.2. Neuroimaging data

Images that were preprocessed (co-registered and averaged) from ADNI were downloaded. Comprehensive information on the acquisition and processing protocols can be accessed at the ADNI website. The images were further processed using Statistical Parametric Mapping 12 (SPM12) software implemented in Matlab 2020b (MathWroks Inc.). Images were realigned and normalized to the Montreal Neurological Institute space using the FDG-PET template presented in Della Rosa et al. (2014) to voxels of size [2, 2, 2] using a 7th Degree B-Spline interpolation. Afterwards, spatial smoothing was performed using a Gaussian kernel with 6 mm full width at half maximum (FWHM). Finally, standardized uptake value ratios (SUVRs) were computed, using the whole cerebellum as the reference region (Dukart et al., 2010).

2.3. Analytical approach to brain metabolism decline

A key goal of our investigation was to model the temporal changes in brain metabolism. We approached this by segmenting brain metabolism data into regions of interest (ROIs) using the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). For each ROI, we assigned the average value of its constituent voxels. We then determined the percentage decrease in the SUVRs across

Table 1

Clinical and socio-demographic characteristics of the sample used in the study.

Diagnosis	Subsample	Subjects	Observations	Sex (%)[a]	Education (years)	MMSE[b]	FDG[c]
All sample	All	1617	3449	44.34	16.05	27.14	0.12
	BMV[d]	674	1166	42.88	16.06	27.63	0.25
HC	All	438	966	51.60	16.43	29.00	0.62
	BMV[d]	223	398	43.50	16.44	29.11	0.58
MCI	All	858	1689	42.19	16.07	27.77	0.23
	BMV[d]	356	601	42.42	16.13	27.81	0.28
	pMCI	81	228	38.27	15.78	26.93	−0.36
	sMCI	338	676	42.31	15.91	27.93	0.37
Dementia	All	321	794	40.19	15.48	22.94	−0.89
	BMV[d]	85	167	44.71	14.80	22.93	−0.75

[a] Percentage of females in the sample.

[b] MMSE, Mini-Mental State Examination.

[c] Value provided by ADNI consisting of the average standardized glucose uptake ratio value of angular, temporal, and posterior cingulate. Average values for each group have been presented as z-scores.

[d] Subset of data that has information on variations in brain metabolism (BMV, brain metabolism variations).

two measurements separated by two years, as defined by Eq. (1). We adopted the nomenclature defined in Rolls et al. (2020) to label all the ROIs.

$$\Delta^{\mathrm{SUVR}} = \left(\frac{\mathrm{SUVR}^{(\text{two-years})}}{\mathrm{SUVR}^{(\text{baseline})}} - 1 \right) \cdot 100 \tag{1}$$

When modeling changes in brain metabolism, it is important to recognize that not all ROIs equally contribute to AD pathology. Certain regions may show minimal or no metabolic changes due to the disease's limited impact there. Our analysis, therefore, prioritized ROIs with significant associations with AD, focusing on those exhibiting notable metabolic decrements.

To identify these regions, we conducted a voxel-based morphometry analysis using SPM12. We contrasted brain metabolism between individuals with dementia and HC using a two-sample t-test, and adjusting for both age and average cerebellar metabolism. For each ROI, the percentage of significantly hypometabolic voxels in the dementia group was calculated considering a family-wise error-corrected p-value less than 0.05. Then, the top 20% of the ROIs with the highest percentage of hypometabolic voxels in the dementia group were selected for further analysis.

Furthermore, we utilized generalized linear models (GLMs) to explore the impact of metabolic changes in these ROIs on the progression of the disease. Specifically, we examined their influence on diagnostic shifts and variations in the Mini-Mental State Examination (MMSE) scores, adjusting for demographic factors such as age, sex, and education levels. Additional insights into the ROI selection process and their statistical analysis are provided in Appendix A.

2.4. *Approach to modeling metabolic variations*

For the modeling of variations in brain metabolism, our primary goal was to approximate $\boldsymbol{\Delta}^{\mathrm{SUVR}} \in \mathbb{R}^{n \times m}$ using the function $f(\phi(\mathbf{X}); \boldsymbol{\theta})$. Here, $\boldsymbol{\Delta}^{\mathrm{SUVR}}$ represents the change in SUVR values at $t = 2$ (years) for m ROIs associated with n subjects, as described in Section 2.3. The function ϕ denotes a mapping applied to the input data, where $\mathbf{X} \in \mathbb{R}^{79 \times 95 \times 78}$ corresponds to the subject's FDG-PET image at time $t = 0$, and $\boldsymbol{\theta}$ represents the set of trainable parameters associated with model f. In this study, f denoted FFNs or CNNs with linear units at the output layer. For the FFN models, the function ϕ ($\phi^{(\mathrm{FFN})} : \mathbb{R}^{79 \times 95 \times 78} \rightarrow \mathbb{R}^{90}$) aggregated brain metabolism values within ROIs using the AAL atlas, excluding the cerebellum (Tzourio-Mazoyer et al., 2002). In the case of the CNNs, the function ϕ ($\phi^{(\mathrm{CNN})} : \mathbb{R}^{79 \times 95 \times 78} \rightarrow \mathbb{R}^{67 \times 79 \times 64}$) resampled the input image to voxels of size [3, 3, 3] through nearest-neighbor interpolation, reducing the model's computational cost. Consequently, the input data size for FFNs consisted of a vector of 90 variables, while for CNNs, the input corresponded to a tensor of dimensions [67, 79, 64].

The FFN architecture and its hyperparameters were parameterized and subsequently optimized using a Bayesian optimization (BO) framework employing a tree-structured Parzen estimator (TPE) as a surrogate model (Bergstra et al., 2011). The open-source library Optuna (Akiba et al., 2019) was used for hyperparameter optimization (HPO). In the HPO, the mean absolute error (MAE) obtained on the test set of a five-fold cross-validation was minimized. Overall, 5000 different configurations sampled under the BO-TPE framework were evaluated by randomly sampling the first 1000. The optimized FFN hyperparameters included the number of hidden layers and neurons per layer, the batch size, the learning rate, the use of batch normalization, and the dropout rate. For further details on the parameterization of the FFN-based model, the reader is referred to Appendix B.

On the other hand, given the high computational cost of the CNN-based models, their architecture and hyperparameters were initially derived from existing literature and further fine-tuned through empirical adjustments (Springenberg et al., 2014; Wang et al., 2018; Chen et al., 2021). The final architectural configuration of the CNN-based models is illustrated in Fig. 1. Additionally, we evaluated the impact of model complexity, performance, and computational cost of the scaffold depicted in Fig. 1. For this purpose, we introduced two models with differing complexities by varying the size of the kernels and the number of channels. The simplest model (CNN-A1) comprised a total of 89,479 trainable parameters and $10^{9.061}$ multiply-accumulate operations (MACs) for a forward pass, while the more parameterized model (CNN-A2) featured 276,423 parameters and ∼4 times more MACs than CNN-A1.

The weights of all models were adjusted through backpropagation, utilizing the Huber loss with $\delta = 2$ as the loss function (Eq. (2)), and optimizing the model weights by Adam with decoupled weight decay (AdamW) (Loshchilov and Hutter, 2017).

$$\mathcal{L}^{\mathrm{Huber}}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n} \sum_{i=1}^{n} H(Y_{i,j}, \hat{Y}_{i,j}) \tag{2}$$

where $\mathbf{Y} \in \mathbb{R}^{n \times m}$ and $\hat{\mathbf{Y}} \in \mathbb{R}^{n \times m}$ represent the actual and predicted metabolism variations, and H corresponds to the piecewise function:

$$H(y, \hat{y}) = \begin{cases} \frac{1}{2} \cdot (y - \hat{y})^2, & \text{if } |y - \hat{y}| \leq \delta \\ \delta \cdot (|y - \hat{y}| - \frac{1}{2} \cdot \delta), & \text{otherwise} \end{cases} \tag{3}$$

In our experimental setting, the FFN-based models were trained for a maximum of 500 epochs, while the CNNs were trained for 50 epochs. The batch size for the CNNs was fixed to 20 due to the high memory requirements. Additionally, early stopping was implemented for both models finalizing the training when the value of the loss function over the validation set did not decrease during 15 consecutive epochs. Models were evaluated via cross-validation using ten folds, and 15% of

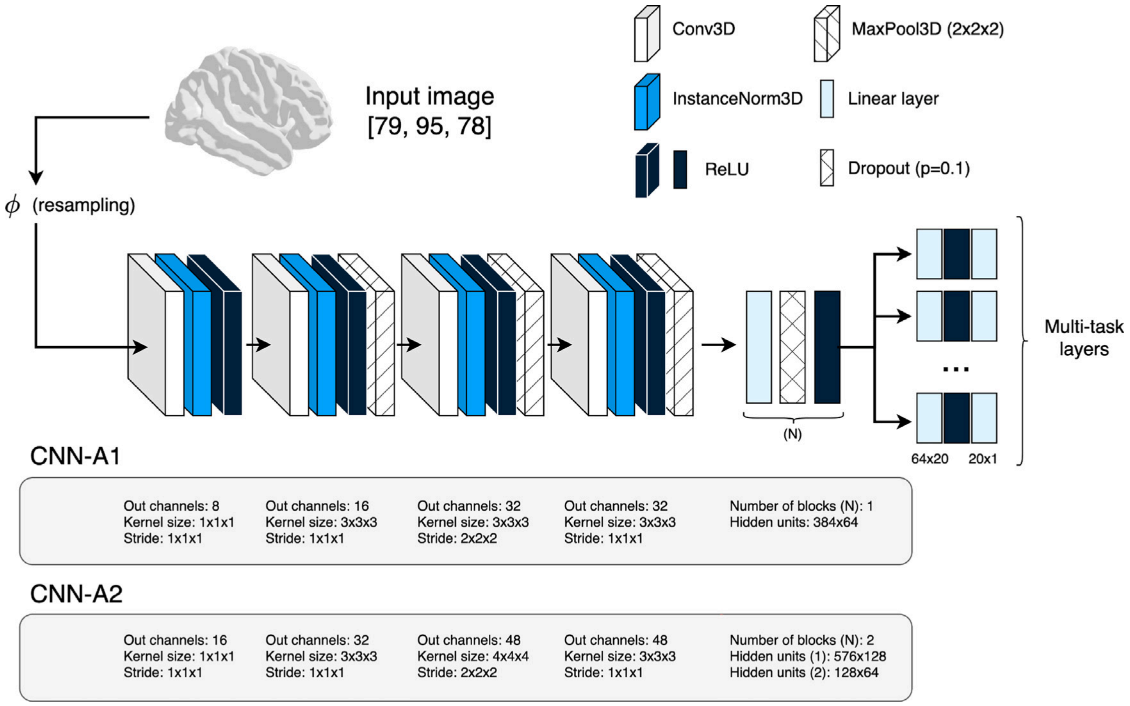


Fig. 1. Overview of the proposed convolutional neural network (CNN) architecture. The process begins with input images of dimensions [79, 95, 78], which are first resampled via a function, ϕ, before passing through several convolutional layers that include instance normalization (InstanceNorm3D) and max pooling (MaxPool3D). The output from these convolutional layers is then flattened and fed into a feed-forward network (FFN) incorporating multi-task projection layers to predict the variation in brain metabolism across specific regions of interest (ROIs). Two distinct architectures, CNN-A1 and CNN-A2, were introduced, each varying in complexity and parameterization, with A1 being the less complex model with fewer parameters, and A2 having a more complex structure with a greater number of parameters.

the training data was used as a validation set. For the FFNs, the data were standardized to z-scores based on the training set statistics. During the separation of data into training, validation, and test sets, splits were performed at the subject level. All the models were developed in Python (v3.11.5) using the PyTorch (v2.1.1) (Paszke et al., 2019) library.

2.5. Approach to modeling MCI-to-dementia conversion

The second aim of this study was to assess the effectiveness of the models introduced in the previous section (Section 2.4) in predicting the conversion from MCI to dementia at two years. To this end, we parameterized a Bernoulli distribution to model $p(\text{pMCI} \mid \phi(\mathbf{X}); \theta)$. For this task, we used the binary cross entropy as a loss function, weighting the positive class to account for class imbalance:

$$\mathcal{L}^{\text{BCE}}(\mathbf{y}, \hat{\mathbf{y}}) = -\frac{1}{n}\sum_{i=1}^{n} w \cdot y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i) \tag{4}$$

where $\mathbf{y} \in \mathbb{R}^n$ and $\hat{\mathbf{y}} \in \mathbb{R}^n$ denote the true and predicted values, and w represents the weighting assigned to the positive class. In our case, we selected $w = 4$ based on the expected conversion ratio reported in the literature (Farias et al., 2009).

To address this problem, we implemented the CNN-A1 architecture (see Fig. 1), which demonstrated superior performance in predicting variations in brain metabolism at two years (see Section 3.2). Furthermore, leveraging insights from previous studies (Zhang and Yang, 2021), we introduced and examined the effect of training the model under a multi-task framework, optimizing several targets simultaneously. In this context, beyond predicting pMCI, we incorporated multiple multi-task layers to predict $\mathbf{\Delta}^{\text{SUVR}}$, as detailed in Section 2.4, and to infer the current patient's diagnosis. For the latter task, we parameterized the diagnosis as a multinomial distribution by adding a softmax function, considering three classes (HC/MCI/dementia), and minimizing the cross-entropy between the model predictions and the ground truth values:

$$\mathcal{L}^{\text{CE}}(\mathbf{Y}, \hat{\mathbf{Y}}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} y_{i,j} \cdot \log(\hat{y}_{i,j}) \tag{5}$$

being $\mathbf{Y} \in \mathbb{R}^{n\times c}$ and $\hat{\mathbf{y}} \in \mathbb{R}^{n\times c}$ the model predictions associated with $c = 3$ possible classes.

Accordingly, the model trained under the multi-task learning paradigm underwent training with the same experimental setup detailed in Section 2.4, increasing batch size to 48, and minimizing the following joint loss function.

$$\mathcal{L}^{\text{Joint}} = \alpha\mathcal{L}^{BCE} + \beta\mathcal{L}^{CE} + \gamma\mathcal{L}^{Huber} \tag{6}$$

where the hyperparameters $\{\alpha, \beta, \gamma\}$ regulate the importance given to each of the tasks. In this study, we select $\alpha = 1, \beta = 2, \gamma = 0.5$, by balancing the magnitude of the terms in each of the individual loss functions. In addition, regarding the multi-task model, it must be considered that not all subjects have all the information (see Section 2.1). Therefore, to handle missing values, we applied a mask when calculating the loss functions for Eqs. (2) and (4). For Eq. (5) this mask was not necessary since all subjects had information about their current diagnosis.

2.6. Model interpretability

The quest for understanding DL models, particularly critical in sectors like healthcare, has underscored the need for robust interpretability techniques (Arrieta et al., 2020). Therefore, to further explore the impact of the input variables on the model predictions, we employed the integrated gradients (IG) method. This algorithm, introduced in Sundararajan et al. (2017), calculates the attributions A_i, of a given model f, and for an input value x_i, based on the straight path from a reference value x_i'. The final attributions for the input variable x_i are determined through Eq. (7):

$$A_i(x_i) = (x_i - x_i') \cdot \int_{\epsilon=0}^{1} \frac{\partial f(x' + \epsilon \cdot (x_i - x_i'))}{\partial x_i} \, d\epsilon \tag{7}$$

The integral's computation is approximated using a Riemann sum over p steps within the interval $[0, 1]$.

In our research, the IG method was leveraged to interpret the multi-task model detailed in Section 2.5. Specifically, attributions were determined for each target variable, employing a black image as the reference point. We focused on attributions for changes in brain metabolism within the designated ROIs, particularly for subjects at the extremes of the Δ^{SUVR} spectrum — those below the 20th and above the 80th percentiles. The goal was to discern model behavior in cases of significant metabolic decline versus relative stability (Oh et al., 2019). This analysis extended to evaluating attributions for predicting pMCI and sMCI phenotypes, as well as for the task of classifying current diagnosis, emphasizing comparisons between HC and dementia cases.

Utilizing the Captum library (Kokhlikyan et al., 2020) for IG implementation, we computed and averaged attributions across the test sets derived from a ten-fold cross-validation, focusing on the absolute values to capture overall influence rather than directional impact. This approach facilitated a nuanced understanding of the model's decision-making process, particularly highlighting the input features' significance across different prediction tasks and clinical conditions.

3. Results

3.1. Identification of the Alzheimer's disease-related brain regions

Our analysis identified fifteen ROIs representing the top 20% of areas exhibiting the most significant hypometabolism in AD. The selected ROIs are presented in Table 2. These ROIs, predominantly localized within the temporal lobe, include the left and right parahippocampal gyri (PHF L/R), the left and right hippocampus (HPO L/R), the left and right inferior temporal gyri (ITG L/R), the left middle temporal gyrus (MTG L), and the left temporal pole of the middle temporal gyrus (TPOmid L). Furthermore, regions in the parietal lobe, such as the left and right angular gyri (ANG L/R), the left and right inferior parietal lobules (IPG L/R), and the left and right posterior cingulate gyri (PCC L/R), along with a subcortical area, the left caudate nucleus (CAU L), were identified for their pronounced metabolic alterations in AD (Rolls et al., 2020).

Significant differences in SUVRs among the selected ROIs were observed across different diagnostic groups at baseline evaluation. Subjects diagnosed as HC typically showed higher SUVRs, indicating relatively healthy brain metabolism. Conversely, individuals diagnosed with MCI displayed a decline in metabolism, with the lowest SUVRs observed in those diagnosed with dementia. Importantly, longitudinal analysis revealed that metabolic changes in these ROIs over a two-year period were statistically associated with an increased risk of MCI progression to dementia, as well as reductions in MMSE scores (detailed in Table 2). For comprehensive insights into the methodology and findings, readers are directed to Appendix A.

3.2. Predictive performance for two-year metabolic changes

This study developed models employing FFNs and CNNs to predict changes in brain metabolism at two years across key ROIs pertinent to AD, as outlined in Section 2.4. The architecture and hyperparameters of the FFNs-based model were optimized using a BO-TPE framework. The optimal FFN architecture comprised a 4-layer network with 250, 174, 100, and 75 neurons respectively, applying dropout rates of 0.15, 0.1, 0.05, and 0.00 for each layer, leveraging the ReLU activation function, and incorporating batch normalization with a batch size of 32. Overall, the FFN-based model had 121,790 trainable parameters.

Comparative analysis revealed that CNN-based models outperformed the optimized FFN model in predicting metabolic changes. Specifically, the CNN-A1 architecture demonstrated superior predictive accuracy, evidenced by an average Pearson correlation coefficient of 0.383 across the test set. This model achieved a MAE of 3.531 and an average explained variance (EV) of 14.388%. In comparison, the FFN model registered a lower correlation coefficient of 0.294, a MAE of 3.681, and an EV of 5.300%. The CNN-A2 model, while offering competitive performance, yielded correlation, MAE, and EV values of 0.365, 3.594, and 12.180%, respectively.

Table 2
Analysis of brain standardized uptake value ratios (SUVRs) of the selected regions of interest (ROIs) relative to patient diagnosis, disease progression, and changes in the Mini-Mental State Examination (MMSE).

ROI	HC vs. Dementia[a]		Disease progression[b]		MMSE changes[c]	
	Z-statistic	p-value	Z-statistic	p-value	Z-statistic	p-value
ITG (L)	−20.291	<0.001*	−3.562	<0.001*	4.836	<0.001*
PHG (L)	−19.203	<0.001*	−3.572	<0.001*	3.305	0.001*
TPOmid (L)	−13.937	<0.001*	−2.220	0.026*	2.547	0.011*
CAU (L)	−11.765	<0.001*	−3.019	0.003*	5.162	<0.001*
PCC (L)	−21.317	<0.001*	−2.712	0.007*	4.232	<0.001*
PHG (R)	−14.888	<0.001*	−2.174	0.030*	2.597	0.009*
ITG (R)	−16.567	<0.001*	−3.102	0.002*	4.861	<0.001*
IPG (L)	−15.050	<0.001*	−2.301	0.021*	3.653	<0.001*
ANG (L)	−21.685	<0.001*	−3.647	<0.001*	4.829	<0.001*
HIP (L)	−16.078	<0.001*	−2.279	0.023*	1.929	0.054
MTG (L)	−16.779	<0.001*	−2.255	0.024*	3.832	<0.001*
HIP (R)	−14.555	<0.001*	−1.971	0.049*	1.368	0.171
ANG (R)	−19.186	<0.001*	−3.452	0.001*	4.165	<0.001*
IPG (R)	−14.099	<0.001*	−1.884	0.060	3.219	0.001*
PCC (R)	−16.667	<0.001*	−2.391	0.017*	2.701	0.007*

* Statistically significant value (p-value < 0.05).

[a] Comparison of baseline SUVRs of healthy controls (HC) and subjects with dementia adjusting for age, sex, and years of formal education.

[b] Comparison of Δ^{SUVR} (see Section 2.3) between individuals with mild cognitive impairment progressing to dementia (pMCI) and individuals who remain stable (sMCI) adjusting for age, sex, and educational level.

[c] Analysis of the association between Δ^{SUVR} (see Section 2.3) and variations in the MMSE defined as $(\mathrm{MMSE}^{(2\text{-years})} - \mathrm{MMSE}^{(\text{baseline})})$. Values adjusted for age, sex, and years of education.

Table 3
Performance of the CNN-A1 model for predicting brain metabolism changes at two years stratified by brain region. Results reflect the average performance obtained on the test from a ten-fold cross-validation.

ROI	Correlation	MAE	RMSE	EV (%)
ITG (L)	0.447	2.994	3.866	19.489
PHG (L)	0.408	2.649	3.558	16.705
TPOmid (L)	0.404	2.768	3.544	16.745
CAU (L)	0.403	3.115	4.086	15.926
PCC (L)	0.397	3.480	4.589	15.053
PHG (R)	0.392	2.772	3.754	15.580
ITG (R)	0.382	3.291	4.262	13.718
IPG (L)	0.380	4.639	6.191	13.908
ANG (L)	0.377	4.822	6.384	13.849
HIP (L)	0.375	2.955	4.065	14.139
MTG (L)	0.375	3.432	4.476	13.752
HIP (R)	0.366	3.073	4.123	13.416
ANG (R)	0.365	4.856	6.354	12.386
IPG (R)	0.340	5.040	6.597	11.073
PCC (R)	0.339	3.090	4.128	10.081
Total	**0.383**	**3.532**	**4.665**	**14.388**

Abbreviations: ROI, region of interest; MAE, mean absolute error; RMSE, root mean squared error; EV, explained variance.

For the model CNN-A1, the regions with the most EV were ITG (L), PHG (L), TPOmid (L), CAU (L), PCC (L), and PHG (R), with values above 15% and a correlation around 0.40. In contrast, the areas with the greatest discrepancy with the model predictions were IPG (R) and PCC (R) with EVs of around 10% and correlations below 0.35. Finally, for ITG (R), IPG (L), ANG (L), HIP (L), MTG (L), HIP (R), and ANG (R), the EVs were in the range of 12%–15% with correlations ranging from 0.365 to 0.382 (see Table 3).

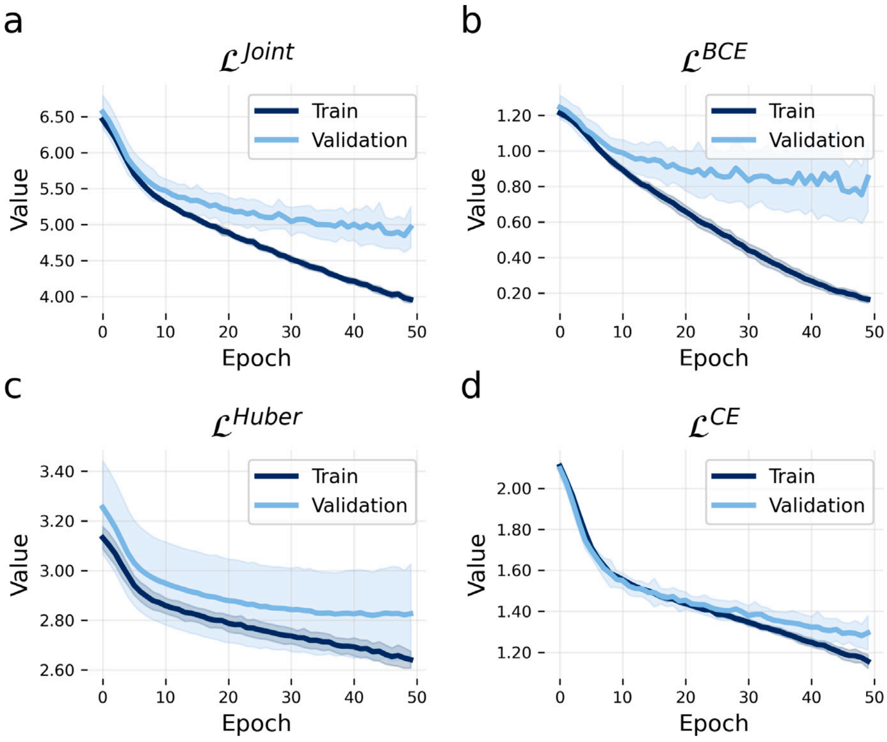


Fig. 2. Convergence of the multi-task model described in Section 2.5. The figure shows the average convergence values during training over the different folds for the (**a**) joint loss, and the components associated with (**b**) the modeling of the progression from mild cognitive impairment to dementia, (**c**) the prediction of two-year changes in brain metabolism, and (**d**) the prediction of the patient's current diagnosis.

3.3. Predicting MCI progression to dementia

As described in Section 2.5, our work successfully adapted the CNN-A1 architecture, initially optimized for modeling changes in brain metabolism, to distinguish between individuals with MCI that remain stable (sMCI) and those progressing to dementia (pMCI) within two years. Additionally, besides training the CNN-A1 model with the existing conversion data from 904 observations, a model was fitted under the multi-task paradigm (CNN-A1 multi-task), minimizing the joint loss function outlined in Eq. (6), by leveraging information from all observations. The convergence of the multi-task model is illustrated in Fig. 2.

The standalone CNN-A1 model yielded an F1-score of 0.628, demonstrating a specificity of 0.799 and a sensitivity of 0.756 for identifying pMCI cases. The introduction of the multi-task learning framework significantly improved these metrics, achieving an F1-score of 0.783, with enhanced specificity (0.869) and sensitivity (0.884). Notably, the multi-task model exhibited reduced variance across different folds of cross-validation, indicating a more robust and consistent performance (as illustrated in Fig. 3).

The evaluation of both models through receiver operating characteristic (ROC) curves, shown in Fig. 4, revealed that the CNN-A1 model achieved an area under the ROC curve (AUC) of 0.83. The multi-task model, however, outperformed this with a higher AUC of 0.92, showcasing its superior predictive capability in identifying MCI patients at risk of progressing to dementia.

For the objectives of the multi-task model, including current diagnosis classification, the model reached an average macro and micro AUC of 0.841 and 0.856, respectively, with corresponding F1-macro/micro scores of 0.676 and 0.670. The model demonstrated a high degree of accuracy in distinguishing between HC and dementia cases, with only minimal misclassification rates between HC and MCI, and between MCI and dementia. Specifically, the model achieved remarkable F1-scores when differentiating HC from dementia (F1-score of 0.956), HC from MCI (F1-score of 0.731), and MCI from dementia (F1-score of 0.782), indicating its effectiveness across various diagnostic distinctions.

In terms of predicting two-year changes in brain metabolism, a slight increase in the MAE was observed, moving from 3.531 (as reported in Section 3.2) to 3.601. This increment, while modest, underscores the challenge of balancing predictive accuracy across multiple tasks within a multi-task learning framework.

3.4. Interpretation of model attributions through integrated gradients

The model attributions for each of the modeled tasks were analyzed using the IG method, as outlined in Section 2.6. In identifying pMCI individuals, the model assigned the highest importance to the PCC (L/R), along with the left putamen (PUT L) and pallidum (PAL L) (see Fig. 5). These same four brain regions also exhibited the highest importance in predicting $\boldsymbol{\Delta}^{\mathrm{SUVR}}$ and inferring the patient's current diagnosis (see Fig. 6).

Additional regions that received high attributions for pMCI classification included the left and right thalamus (THA L/R), PAL (R), left inferior frontal gyrus opercular/orbitalis parts (IFGoper/IFGorb L), left middle cingulate gyrus (MCC L), left cuneus and precuneus (CUN/PCUN L), HIP (L), and left superior occipital gyrus (SOG L).

Consistency in attributions was observed across the tasks. For predicting $\boldsymbol{\Delta}^{\mathrm{SUVR}}$, regions like the THA L, PAL R, CUN L, PCUN L, the left angular gyrus (ANG L), the left lingual gyrus (LIN L), IFGoper L, IFGorb L, and MCC L/R emerged as significant. In the context of diagnosing the patient's current condition, notable regions included the THA L/R, PAL R, CUN L, HIP L, PCUN L, LIN L, MCC L, IFGoper L, and the left amygdala (AMYG L).

These findings underscore the model's reliance on a combination of subcortical structures, limbic regions, and cortical areas involved in memory, executive function, and sensory processing for making predictions related to AD progression. The consistency across tasks highlights

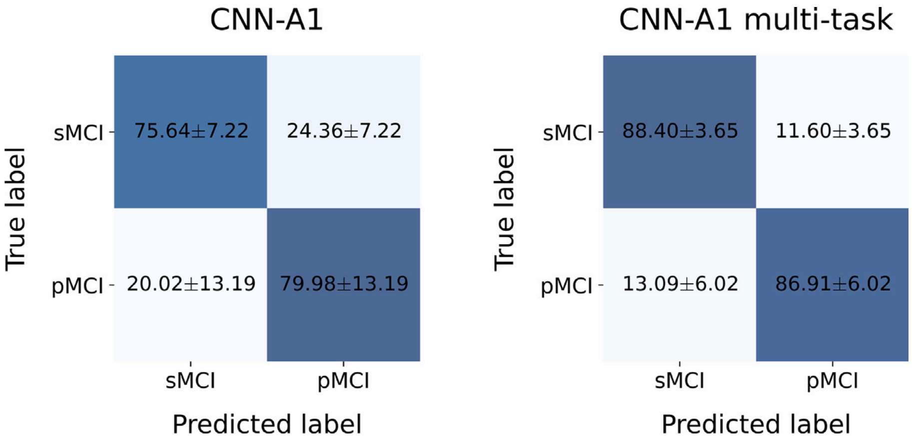


Fig. 3. Confusion matrices associated with the classification of individuals with mild cognitive impairment who will progress to dementia (pMCI) or remain stable (sMCI) over two years. The values denote the average and standard deviation of each of the ten-fold cross-validation folds. These results correspond to the models defined in Section 2.5.

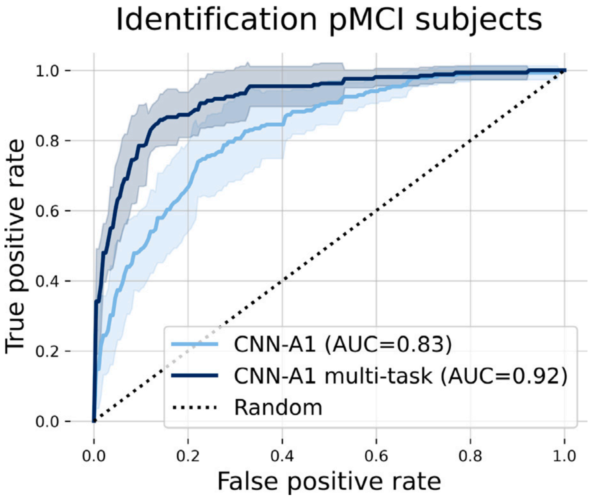


Fig. 4. Receiver operating characteristic (ROC) curves for models developed to identify mild cognitive impairment individuals that will progress to dementia (pMCI) at two years. Average values obtained by ten-fold cross-validation are shown. Abbreviations: AUC, area under the ROC curve.

the interconnectedness of these brain regions in the pathology of AD, reflecting their relevance in both the disease's progression and its impact on brain metabolism.

For detailed insights into the specific attributions across the various tasks and their implications, readers are directed to Appendix C.

4. Discussion

This study introduces a novel framework for the development of predictive models in AD, demonstrating the feasibility of modeling changes in brain metabolism. These models significantly enhance the identification of individuals with MCI progressing to dementia. Additionally, the incorporation of XAI techniques deepened our understanding of the decision-making processes, improving model debugging and reliability. To the best of our knowledge, this is the first study that employs DL techniques to model variations in brain metabolism and leverages this knowledge to identify pMCI individuals.

A primary aim of this study was to predict the evolution of brain metabolism as AD progresses. Notably, within the short timeframe examined, the majority of brain regions exhibit only subtle changes. To streamline the complexity of our analysis and focus on the most relevant data, we selectively targeted several ROIs known for their relevance to AD. This selection process was informed by empirical evidence and focused on areas extensively documented in AD research, including the hippocampus, parahippocampal gyrus, posterior cingulate gyrus (Fjell et al., 2014; Scheef et al., 2012; Leech and Sharp, 2014; Buckner et al., 2005), angular gyrus (Jagust et al., 2006; Smith et al., 2007), medial and inferior temporal regions (McEvoy et al., 2011; Mosconi et al., 2008; McEvoy et al., 2009), inferior parietal gyrus (Crone et al., 2006; Wang et al., 2015), and the left caudate nucleus (Madsen et al., 2010; Jiji et al., 2013). A comprehensive overview of these selected ROIs is provided in Table 2.

Following the identification of key ROIs linked to AD, we employed traditional statistical methodologies to explore the correlation between these ROIs and the progression of AD. Our findings revealed significant associations between declines in FDG-PET, changes in diagnosis, and reductions in the MMSE over two years. These outcomes align with prior research (Landau et al., 2011; Arenaza-Urquijo et al., 2019; Ou et al., 2019; Alexander et al., 2002), confirming the relationship between metabolic decreases and cognitive deterioration. For instance, research by Ou et al. (2019) showed that individuals experiencing a drop in FDG-PET metabolism underwent more pronounced cognitive decline and brain atrophy. Similarly, a study by Alexander et al. (2002) documented significant metabolic slowdowns in the parietal, temporal, and cingulate cortex over a year-long observation. Further studies (Arenaza-Urquijo et al., 2019; Landau et al., 2011) have consistently highlighted the link between metabolic changes and declines in neuropsychological test performances, positing glucose uptake as a predictive marker of AD's cognitive trajectory. Together, these results emphasize the clinical relevance of tracking metabolic shifts within specific ROIs, suggesting that such changes could herald a poorer prognosis. Moreover, this research sheds light on the critical brain regions affected over the course of AD, providing valuable insights into the disease's pathophysiological evolution.

Building upon the identified ROIs, we utilized DL methodologies to forecast the changes in their SUVRs over a two-year timeframe. The evaluation of our models through cross-validation revealed that the CNN-based model performed optimally, achieving a moderate correlation of 0.38 with the actual metabolic changes. This model registered a MAE of 3.53 and an average EV of 14.39%. The range of metabolic variations observed in our dataset spanned from −12.78 to 7.80, with these values representing the 2.5% and 97.5% percentiles, respectively. Such findings denote a moderate level of predictive accuracy, with an error margin of 17.15% covering 95% of the observed changes. In this context, our research stands out in the field, as to date, there exists only one study that endeavored to model variations in brain metabolism (Abi Nader et al., 2020). Abi-Nader et al. introduced a generative model aimed at predicting temporal shifts in brain metabolism through matrix factorization of 3D volumetric data. Nevertheless, their

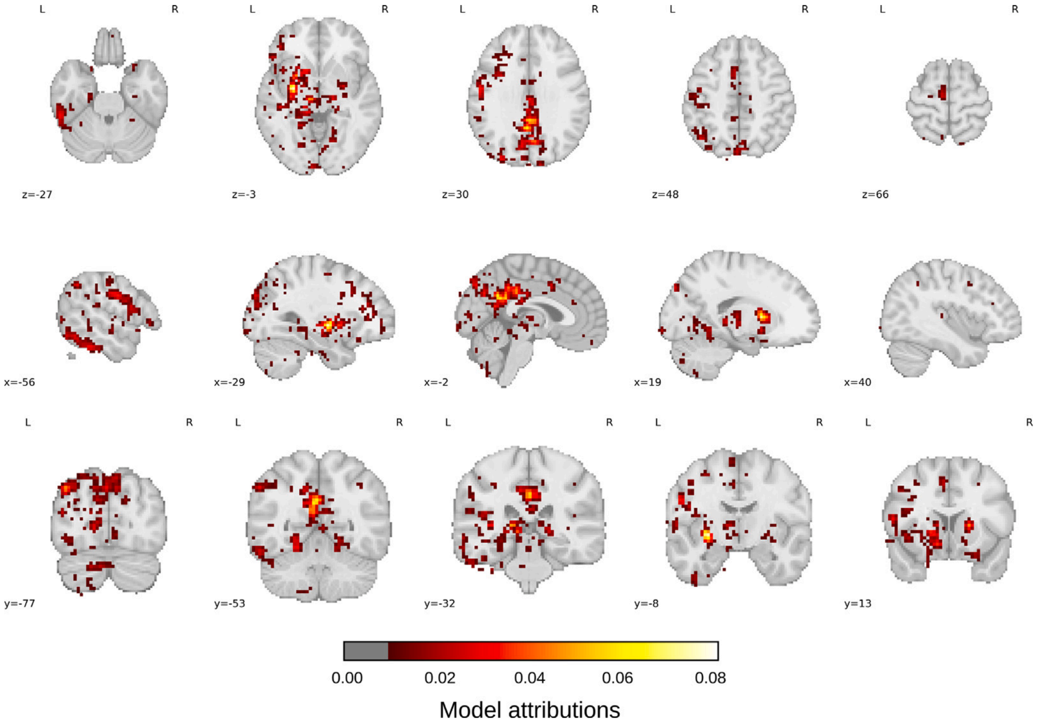
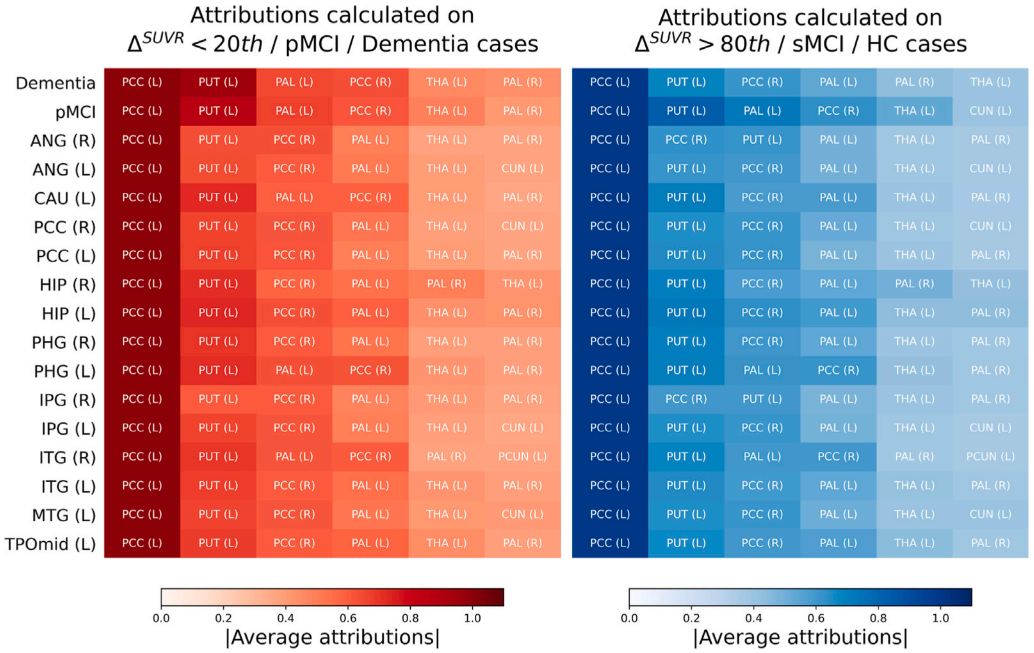


Fig. 5. Attributions of the multi-task model described in Section 2.5 for identifying individuals with mild cognitive impairment (MCI) who progress to dementia within two years (pMCI). A total of 228 images associated with 81 pMCI individuals were used to calculate the attributions. Regions with the highest attributions include the left and right posterior cingulate, left putamen, left and right pallidum, left and right thalamus, and left cuneus and precuneus.



Attributions calculated on $\Delta^{SUVR} < 20th$ / pMCI / Dementia cases

Dementia	PCC (L)	PUT (L)	PAL (L)	PCC (R)	THA (L)	PAL (R)
pMCI	PCC (L)	PUT (L)	PAL (L)	PCC (R)	THA (L)	PAL (R)
ANG (R)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
ANG (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	CUN (L)
CAU (L)	PCC (L)	PUT (L)	PAL (L)	PCC (R)	THA (L)	PAL (R)
PCC (R)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	CUN (L)
PCC (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
HIP (R)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	PAL (R)	THA (L)
HIP (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
PHG (R)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
PHG (L)	PCC (L)	PUT (L)	PAL (L)	PCC (R)	THA (L)	PAL (R)
IPG (R)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
IPG (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	CUN (L)
ITG (R)	PCC (L)	PUT (L)	PAL (L)	PCC (R)	PAL (R)	PCUN (L)
ITG (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
MTG (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	CUN (L)
TPOmid (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)

Attributions calculated on $\Delta^{SUVR} > 80th$ / sMCI / HC cases

Dementia	PCC (L)	PUT (L)	PCC (R)	PAL (L)	PAL (R)	THA (L)
pMCI	PCC (L)	PUT (L)	PAL (L)	PCC (R)	THA (L)	CUN (L)
ANG (R)	PCC (L)	PCC (R)	PUT (L)	PAL (L)	THA (L)	PAL (R)
ANG (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	CUN (L)
CAU (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
PCC (R)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	CUN (L)
PCC (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
HIP (R)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	PAL (R)	THA (L)
HIP (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
PHG (R)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
PHG (L)	PCC (L)	PUT (L)	PAL (L)	PCC (R)	THA (L)	PAL (R)
IPG (R)	PCC (L)	PCC (R)	PUT (L)	PAL (L)	THA (L)	PAL (R)
IPG (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	CUN (L)
ITG (R)	PCC (L)	PUT (L)	PAL (L)	PCC (R)	PAL (R)	PCUN (L)
ITG (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)
MTG (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	CUN (L)
TPOmid (L)	PCC (L)	PUT (L)	PCC (R)	PAL (L)	THA (L)	PAL (R)

Fig. 6. The figure displays the attributions given to each modeled task (Y-axis), by grouping the values into regions of interest (ROIs) using the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). For visualization, the six ROIs with the highest attributions were represented. In the figure on the left, cases exhibiting the most significant decreases in Δ^{SUVR}, progressive mild cognitive impairment (pMCI), or dementia individuals were utilized as input. In figure on the right, attributions were computed for instances with the highest Δ^{SUVR}, stable MCI (sMCI), or healthy controls (HC). Attributions were scaled to the interval [0, 1]. Abbreviations: PCC, posterior cingulate cortex; PUT, putamen; PAL, pallidum; THA, thalamus; CUN, cuneus; PCUN, precuneus.

model suffers from significant constraints, including the assumption of a monotonic disease progression, and it falls short of providing the level of interpretability our DL approach offers.

Furthermore, our results distinctly highlight the advantage of CNN-based models over FFNs in accurately predicting metabolic changes in the brain. This superiority can be ascribed to a number of critical reasons. Firstly, the aggregation of brain metabolism information into ROIs, relying on the average value of the voxels within each region, leads to a significant loss of information, as it ignores the subtle details present in the images. For instance, in a ROI encompassing a large brain volume, metabolism reductions within a specific part of the region may be smoothed out when considering the average value of all its voxels. Secondly, unlike FFNs, CNN models leverage the spatial dimension of the input data, capturing semantic relationships among different brain regions. This spatial awareness empowers CNNs to extract more intricate patterns from the images than a naive FFN (Ren et al., 2019). These aspects are fundamental in elucidating the observed superiority of CNNs in our analysis.

On the other hand, the accurate identification of individuals with pMCI remains a pivotal challenge in AD research (Ansart et al., 2021; Sharma et al., 2023). Given the established links between metabolic changes in the brain and the progression of AD, we delved into the potential of incorporating Δ^{SUVR} information to enhance the prediction of pMCI cases. To achieve this, we crafted multi-task models leveraging the strengths of CNNs, which showed superior capabilities in forecasting Δ^{SUVR}. These models were tasked with simultaneously estimating variations in SUVRs, predicting the current diagnosis, and determining the likelihood of progression from MCI to dementia.

Our findings reveal that adopting a multi-task strategy markedly improves the accuracy in the identification of pMCI subjects. In direct comparison to a singularly focused pMCI prediction model, the multi-task model's specificity and sensitivity reached 88.40% and 86.91%, respectively, from the initial 75.64% and 79.98%. Moreover, this approach significantly stabilized the variability across different folds, indicating a robust enhancement in model reliability. This level of performance mirrors advancements made by other research endeavors. For example, Abrol et al. (2020) crafted a multi-task model that utilized multi-modal data to predict pMCI, achieving an AUC of 0.925, comparable to that obtained in this work. Lee et al. (2019) also reported a notable AUC of 0.86, thanks to their integration of multi-modal data including MRI scans, cerebrospinal fluid data, and cognitive assessments. These outcomes are consistent with findings from additional studies within the field, showcasing the effectiveness of such models (Moradi et al., 2015; El-Sappagh et al., 2020). Our results not only corroborate the prevailing trends in AD research but also underscore the value of multi-task learning frameworks in developing more nuanced and accurate predictive models (Zhang et al., 2012; El-Sappagh et al., 2020; Abrol et al., 2020; Zhang and Yang, 2021).

In relation to the very competitive results of the multi-task models, we hypothesize that this improvement was due to several reasons. First and foremost, the dataset did not uniformly cover all the targeted variables for every subject. The multi-task learning framework, by accommodating a broader spectrum of training examples, significantly enhanced the capacity for generating rich, meaningful hidden representations within the models. Furthermore, the delineation between sMCI and pMCI is demarcated by a rigid boundary, which may be swayed by a variety of factors, including cognitive reserve, demographic traits, or diagnostic practices. This nuance suggests that a patient's classification as either MCI or dementia – and subsequently as sMCI or pMCI – can sometimes hinge on subtle distinctions in their cognitive assessment, situating them on a fine line between these diagnostic categories. Traditional models that categorize MCI status in a binary fashion might not capture the depth of a patient's cognitive state. In contrast, our approach of concurrently modeling multiple outcomes offers a comprehensive snapshot of a patient's condition, thereby potentially elevating the model's predictive performance. Lastly, an integral advantage of multi-task learning is its inherent function as a regularization technique. This property compels the model to develop more universally applicable internal representations, which, in turn, minimizes the likelihood of overfitting (Zhang and Yang, 2021). Collectively, these factors underscore the multi-task approach's efficacy in refining model predictions.

Finally, recent research has evidenced the effectiveness of XAI techniques in enhancing the understanding of model decisions and boosting their reliability (Arrieta et al., 2020). Within the realm of AD neuroimaging, XAI methodologies have increasingly become a staple, facilitating deeper insights and broader acceptance of model outputs (Oh et al., 2019; Ding et al., 2019; Pohl et al., 2021). Embracing this trend, we integrated the XAI framework into our study to probe deeper into the factors most critically impacting our model's predictions. To achieve this, we employed the IG method to compute attributions within our multi-task modeling approach.

We found that the IG method assigned more importance to typical AD-related regions such as the cingulate cortex and hippocampus, some subcortical structures, parts of the parietal cortex, and, to a lesser extent, some occipital and frontal areas (see Section 3.4). This alignment with well-documented AD research underscores the method's utility in interpreting complex model behaviors. For example, previous research has highlighted the elevated connectivity and metabolic activity of the cingulate cortex, along with its alterations in AD (Leech and Sharp, 2014). Similarly, reduced metabolic activity in the precuneus has been recognized as an early indicator of AD (Scheef et al., 2012), and other research groups have also discussed alterations in the hypothalamus and thalamus (Wang et al., 2006; Fjell et al., 2014; Dai et al., 2015). Moreover, hypometabolism in the occipital and frontal cortex has been documented in more advanced stages of the disease, which potentially explains their inclusion among regions with high attribution (Fjell et al., 2014). Nevertheless, the association with AD was less evident for other areas, such as the putamen or pallidum, which have a minor impact on the disease. In this case, we hypothesize that the model may be capturing nonlinear interactions of these regions with other areas, which were difficult to explain. Nonetheless, we observed the emergence of regions directly implicated in the disease, providing insights into the dynamics of the model.

Collectively, this research offers evidence underscoring the significance of modeling variations in brain metabolism during AD. The findings indicate the feasibility of projecting brain metabolism variations along the AD continuum using DL techniques. So far, most work has focused on differentiating clinical phenotypes (Ghazi et al., 2019; Liu et al., 2018; Abrol et al., 2020; Katako et al., 2018; Ding et al., 2019), predicting cognitive decline (Choi et al., 2018; Iaccarino et al., 2019; Lei et al., 2020; Bhagwat et al., 2018), or forecasting brain atrophy progression (Bône et al., 2018; Marinescu et al., 2020; Sauty and Durrleman, 2022). In this context, predicting the evolution of brain metabolism introduces a novel dimension for developing models in neurodegenerative diseases. This information holds great interest for a potential translation into clinical practice. For instance, in cases where there is an expected decline in brain metabolism, the subject will likely have a poorer prognosis, as supported by our results and previous studies indicating that such changes serve as a reliable indicator of cognitive decline (Landau et al., 2011; Arenaza-Urquijo et al., 2019; Ou et al., 2019; Alexander et al., 2002). Therefore, all the information that can be estimated for a given subject, including variations in the SUVR values, becomes relevant for investigating the pathophysiological mechanisms of the disease, developing clinical trials, evaluating disease-modifying therapies, and implementing personalized treatments (Mintun et al., 2021).

Furthermore, this study provides evidence that modeling SUVR values improves performance on other tasks, including the identification of sMCI and pMCI individuals. This finding is significant, as numerous researchers have reported that modeling diverse complementary tasks

– such as simultaneously predicting clinical phenotypes and cognitive scores – enhances the overall generalization capacity of the models (El-Sappagh et al., 2020; Abrol et al., 2020). Accordingly, this research provides a further pretext for developing models under the multi-task learning framework, indicating the importance of including the prediction of future SUVR values as one of the modeling objectives. Moreover, by introducing XAI techniques, our study showed that it is possible to analyze model predictions, extracting useful information about brain regions that may be of great interest within the pathology.

While this study presents significant advancements, it is not without its limitations. Primarily, the predictive accuracy for changes in brain metabolism is moderate. Future investigations could benefit from employing more sophisticated models that leverage a broader dataset and multi-modal data to enhance predictive performance (Lee et al., 2019; Abrol et al., 2020; Choi et al., 2018). Additionally, akin to many studies utilizing the ADNI database, our analysis is limited by its reliance on a relatively narrow cohort (Samper-González et al., 2017), highlighting the necessity for further studies to examine the scalability and applicability of our analytical framework across diverse populations. Another notable challenge is the substantial computational demand of the CNNs deployed. Subsequent research should investigate more computationally efficient models that can still effectively harness the spatial data present in brain images. In this vein, graph neural networks emerge as a potentially valuable avenue for exploration. Furthermore, our models predominantly rely on data from a singular time point to forecast biennial changes. The incorporation of longitudinal data, capturing the progression of subjects over time, stands as a crucial area for future research, promising to refine and possibly enhance the predictive capabilities of our models.

5. Conclusion

In conclusion, we modeled the dynamic shifts in brain metabolism in AD using multi-task DL models incorporating XAI techniques. Our results demonstrated accurate predictions of metabolic changes and clinical progression from MCI to dementia. Our findings emphasize the relevance of FDG-PET as a biomarker to monitor disease progression and detect patients at high risk of progression to dementia. Overall, this study provides a foundation for future research focused on exploring this new dimension of the disease.

Funding

Jordi A Matias-Guiu is supported by Instituto de Salud Carlos III, Spain through the project INT20/00079 and INT23/00017 (co-funded by European Regional Development Fund “A way to make Europe”)

Ethics declarations

Ethics approval and Consent to participate The Institutional Research Ethics Committee from Hospital Clinico San Carlos approved the research protocol with the 1964 Helsinki declaration and its later amendments. Written informed consent was obtained from all individual participants included in the study or their caregivers.

CRediT authorship contribution statement

Fernando García-Gutiérrez: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Laura Hernández-Lorenzo:** Data curation, Validation, Writing – review & editing. **María Nieves Cabrera-Martín:** Data curation, Writing – review & editing. **Jordi A. Matias-Guiu:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review & editing. **José L. Ayala:** Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflicts of interest regarding the publication of this research article.

Data availability

The authors do not have permission to share data.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U19 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI by the National Institute on Aging, the National Institute of Biomedical is funded Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; 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 and code availability statement

The datasets generated and/or analyzed during the current study are available at ADNI (https://adni.loni.usc.edu/). The code used for this research is available on GitHub.

Appendix A. Identification and analysis of regions of interest in Alzheimer’s disease dementia

This appendix describes the analyses performed on the selected regions of interest (ROIs) (see Section 2.3) using generalized linear models (GLMs) with a Gaussian link function. The objective was to investigate brain metabolism values across diverse phenotypic groups. Additionally, we assessed the influence of variations in standardized uptake value ratios (SUVRs) relative to disease progression.

Firstly, the statistical associations of the SUVR values aggregated in the selected ROIs were explored using the GLM defined by the following equation:

$$\mathbf{y}_{\mathrm{SUVR}} = \beta_0 + \beta_1 \cdot \mathbf{x}_{\mathrm{age}} + \beta_2 \cdot \mathbf{x}_{\mathrm{sex}} + \beta_3 \cdot \mathbf{x}_{\mathrm{yschool}} + \beta_4 \cdot \mathbf{x}_{\mathrm{dementia}} + \varepsilon \,, \qquad \text{(A.1)}$$

where $\mathbf{y}_{\mathrm{SUVR}} \in \mathbb{R}^n$ correspond to the SUVR values at baseline for a given ROI, being n the number of samples; β_0 is the intercept; $\beta_{\{1|2|3\}}$ represents the coefficients of the model for each of its terms; $\mathbf{x}_{\{age|sex|yschool\}} \in \mathbb{R}^n$ are the covariates of the model (age, sex, and years of formal education); $\mathbf{x}_{dementia} \in \mathbb{R}^n$ is the variable of interest identifying the individuals with a diagnosis of dementia; and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ corresponds to the model error. For this model, all baseline acquisitions belonging to subjects with a diagnosis of healthy control (HC) or dementia were selected.

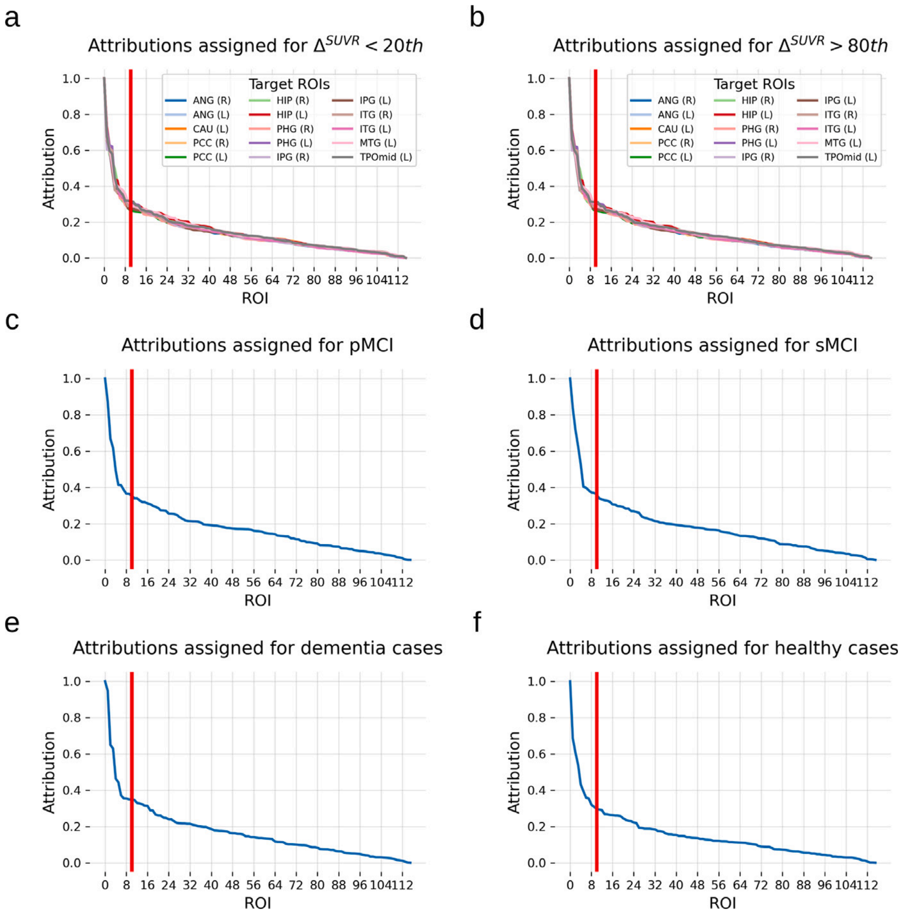


Fig. C.1. Absolute value of the attributions assigned in each of the tasks for each of the regions of interest (ROIs) analyzed. For each task, the attributions allocated to each ROI have been sorted in descending order. The vertical red line shows the cutoff point for the top ten ROIs listed in Table C.1.

Secondly, we assessed whether changes in brain metabolism at two years, defined as $\boldsymbol{\Delta}^{\mathrm{SUVR}}$ (see Section 2.3), were related to progression from mild cognitive impairmen (MCI) to dementia at two years. For this purpose, we selected all baseline observations of individuals with a diagnosis of MCI, and formulated the following GLM :

$$\mathbf{y}_{\boldsymbol{\Delta}^{\mathrm{SUVR}}} = \beta_0 + \boldsymbol{\beta}_{\mathrm{covar}}^{\intercal} \cdot \mathbf{X}_{\mathrm{covar}} + \beta_{\mathrm{pMCI}} \cdot \mathbf{x}_{\mathrm{pMCI}} + \varepsilon\,, \tag{A.2}$$

where, for simplicity, we have expressed the covariates age, sex and years of education in matrix form (i.e., $\boldsymbol{\beta}_{\mathrm{covar}} \in \mathbb{R}^{3\times 1}$, and $\mathbf{X}_{\mathrm{covar}} \in \mathbb{R}^{3\times n}$).

Finally, we extended the above formulation to explore how variations in MMSE over a two years period (defined as $\Delta_{\mathrm{MMSE}} = (\mathrm{MMSE}^{(\text{2-years})} - \mathrm{MMSE}^{(\text{baseline})})$) were related to $\boldsymbol{\Delta}^{\mathrm{SUVR}}$:

$$\mathbf{y}_{\boldsymbol{\Delta}^{\mathrm{SUVR}}} = \beta_0 + \boldsymbol{\beta}_{\mathrm{covar}}^{\intercal} \cdot \mathbf{X}_{\mathrm{covar}} + \beta_{\mathrm{MMSE}} \cdot \boldsymbol{\Delta}_{\mathrm{MMSE}} + \varepsilon\,, \tag{A.3}$$

Analyses were performed using the statmodels (v0.14.0) (Seabold and Perktold, 2010) library in Python (v3.11.5).

Appendix B. Parameterization of the feed forward networks architecture

This appendix details the hyperparameters of the feed forward network (FFN) models that were subjected to the hyperparameter optimization (HPO) as outlined in Section 2.4. The hyperparameters were parameterized and sampled as detailed below, and optimized by Bayesian optimization using a tree structured parzen estimator (TPE) as a surrogate model.

Firstly, the number of hidden layers was sampled from a uniform distribution with integer values bounded to the interval $[3, 10]$, denoted as $\mathcal{U}^{(int)}(3, 10)$. The number of units per layer l, $n^{(l)}$, was determined through equations:

$$n^{(l)} = \left(\frac{1}{\beta}\right)^{(l)} \cdot \mathrm{init\_layers} \tag{B.4}$$

and

$$n^{(l)} = \mathrm{init\_layers} - \alpha \cdot (l) \tag{B.5}$$

where the utilization of Eq. (B.4) (representing an exponential decay) or Eq. (B.5) (constituting a linear decrease) was determined by a binomial distribution, and $\beta \sim \mathcal{U}(0.8, 5.0)$ and $\alpha \sim \mathcal{U}^{(int)}(25, 250)$ were also considered during the optimization. Furthermore, the value of init_layers was sampled from $\mathcal{U}^{(int)}(250, 1000)$.

Other hyperparameters considered included the batch size, sampled from a categorical distribution over the set $\{16, 32, 48, 64\}$; the learning rate, given by $\mathcal{U}(10^{-7}, 10^{-3})$; the utilization of batch normalization; and the dropout rate, which was determined by applying the same principle as with Eq. (B.5).

Appendix C. Integrated gradient attributions

This appendix provides detailed information on the application of the Integrated Gradients (IG) method as described in Section 2.6. Table C.1 presents the regions to which the model assigned greater importance, while Fig. C.1 illustrates the distribution of importance attributed to each of the regions of interest (ROIs) across the analyzed tasks. These results reveal a skewed distribution in attribution across different regions, indicating that a few areas accumulated the majority of the model's attributions, while others received comparatively lower attributions.

Table C.1
Top ten regions of interest (ROIs) with the highest attribution for each of the modeled tasks using the CNN-A1 multi-task model (see Section 2.4).

$\Delta^{SUVR} < 20$th[a]		$\Delta^{SUVR} > 80$th[a]		pMCI[b]		sMCI[b]		Dementia[c]		Healthy[c]	
ROI	Attr.	ROI	Attr.	ROI	Attr.	ROI	Attr.	ROI	Attr.	ROI	Attr.
PCC (L)	1.00	PCC (L)	1.00	PCC (L)	1.00	PCC (L)	1.00	PCC (L)	1.00	PCC (L)	1.00
PUT (L)	0.68	PUT (L)	0.66	PUT (L)	0.87	PUT (L)	0.84	PUT (L)	0.95	PUT (L)	0.69
PCC (R)	0.60	PCC (R)	0.60	PAL (L)	0.67	PAL (L)	0.72	PAL (L)	0.65	PCC (R)	0.61
PAL (L)	0.54	PAL (L)	0.54	PCC (R)	0.62	PCC (R)	0.63	PCC (R)	0.63	PAL (L)	0.54
THA (L)	0.40	THA (L)	0.40	THA (L)	0.49	THA (L)	0.53	THA (L)	0.46	PAL (R)	0.43
PAL (R)	0.38	PAL (R)	0.38	PAL (R)	0.41	CUN (L)	0.40	PAL (R)	0.44	THA (L)	0.39
CUN (L)	0.36	CUN (L)	0.36	IFGoperc (L)	0.41	MCC (L)	0.40	THA (R)	0.37	CUN (L)	0.36
PCUN (L)	0.35	PCUN (L)	0.35	MCC (L)	0.39	PAL (R)	0.38	CUN (L)	0.36	PCUN (L)	0.35
ANG (L)	0.32	ANG (L)	0.32	IFGorb (L)	0.37	IFGorb (L)	0.37	HIP (L)	0.35	THA (R)	0.32
LING (L)	0.29	LING (L)	0.29	CUN (L)	0.36	IFGoperc (L)	0.37	PCUN (L)	0.35	IFGoperc (L)	0.31

The table shows the attributions scaled to the range [0, 1] for each of the input data and for each of the modeled tasks. The top ten ROIs with the highest attribution in each of the tasks are shown.

[a] Task of predicting changes in brain metabolism, Δ^{SUVR}. In this case, the average attribution of all the modeled ROIs was considered. Information from subjects with greater decrements ($\Delta^{SUVR} < 20$th) or stability ($\Delta^{SUVR} > 80$th) was used as input data to calculate attributions.

[b] Task of identifying individuals with mild cognitive impairment (MCI) who progress to dementia (pMCI) or remain stable (sMCI) at two years. Information from pMCI or sMCI individuals was used as input data for calculating attributions.

[c] Task of identifying the current diagnosis. Information from healthy or dementia individuals was used as input data for calculating attributions.

References

Abi Nader, C., Ayache, N., Robert, P., Lorenzi, M., Initiative, A.D.N., et al., 2020. Monotonic Gaussian process for spatio-temporal disease progression modeling in brain imaging data. Neuroimage 205, 116266.

Abrol, A., Bhattarai, M., Fedorov, A., Du, Y., Plis, S., Calhoun, V., Initiative, A.D.N., et al., 2020. Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease. J. Neurosci. Methods 339, 108701.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2623–2631.

Alexander, G.E., Chen, K., Pietrini, P., Rapoport, S.I., Reiman, E.M., 2002. Longitudinal PET evaluation of cerebral metabolic decline in dementia: a potential outcome measure in Alzheimer's disease treatment studies. Am. J. Psychiatry 159 (5), 738–745.

Alzheimer's, Dementia, 2023. 2023 Alzheimer's disease facts and figures. Alzheimer's Dement. 19 (4), 1598–1695.

Ansart, M., Epelbaum, S., Bassignana, G., Bône, A., Bottani, S., Cattai, T., Couronné, R., Faouzi, J., Koval, I., Louis, M., et al., 2021. Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review. Med. Image Anal. 67, 101848.

Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage 145, 137–165.

Arenaza-Urquijo, E.M., Przybelski, S.A., Lesnick, T.L., Graff-Radford, J., Machulda, M.M., Knopman, D.S., Schwarz, C.G., Lowe, V.J., Mielke, M.M., Petersen, R.C., et al., 2019. The metabolic brain signature of cognitive resilience in the 80+: beyond Alzheimer pathologies. Brain 142 (4), 1134–1147.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115.

Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. Adv. Neural Inf. Process. Syst. 24.

Bhagwat, N., Viviano, J.D., Voineskos, A.N., Chakravarty, M.M., Initiative, A.D.N., et al., 2018. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. PLoS Comput. Biol. 14 (9), e1006376.

Bône, A., Colliot, O., Durrleman, S., 2018. Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 9271–9280.

Buckner, R.L., Snyder, A.Z., Shannon, B.J., LaRossa, G., Sachs, R., Fotenos, A.F., Sheline, Y.I., Klunk, W.E., Mathis, C.A., Morris, J.C., et al., 2005. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. J. Neurosci. 25 (34), 7709–7717.

Chen, X., Li, L., Sharma, A., Dhiman, G., Vimal, S., 2021. The application of convolutional neural network model in diagnosis and nursing of MR imaging in Alzheimer's disease. Interdiscipl. Sci.: Comput. Life Sci. 14, 1–11.

Choi, H., Jin, K.H., Initiative, A.D.N., et al., 2018. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. Behav. Brain Res. 344, 103–109.

Crone, E.A., Wendelken, C., Donohue, S.E., Bunge, S.A., 2006. Neural evidence for dissociable components of task-switching. Cerebral Cortex 16 (4), 475–486.

Dai, Z., Yan, C., Li, K., Wang, Z., Wang, J., Cao, M., Lin, Q., Shu, N., Xia, M., Bi, Y., et al., 2015. Identifying and mapping connectivity patterns of brain network hubs in Alzheimer's disease. Cerebral Cortex 25 (10), 3723–3742.

Della Rosa, P.A., Cerami, C., Gallivanone, F., Prestia, A., Caroli, A., Castiglioni, I., Gilardi, M.C., Frisoni, G., Friston, K., Ashburner, J., et al., 2014. A standardized [18 F]-FDG-PET template for spatial normalization in statistical parametric mapping of dementia. Neuroinformatics 12, 575–593.

DeTure, M.A., Dickson, D.W., 2019. The neuropathological diagnosis of Alzheimer's disease. Mol. Neurodegener. 14 (1), 1–18.

Díaz-Álvarez, J., Matias-Guiu, J.A., Cabrera-Martín, M.N., Pytel, V., Segovia-Ríos, I., García-Gutiérrez, F., Hernández-Lorenzo, L., Matias-Guiu, J., Carreras, J.L., Ayala, J.L., et al., 2022. Genetic algorithms for optimized diagnosis of Alzheimer's disease and Frontotemporal dementia using Fluorodeoxyglucose positron emission tomography imaging. Front. Aging Neurosci. 13, 708932.

Ding, Y., Sohn, J.H., Kawczynski, M.G., Trivedi, H., Harnish, R., Jenkins, N.W., Lituiev, D., Copeland, T.P., Aboian, M.S., Mari Aparici, C., et al., 2019. A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain. Radiology 290 (2), 456–464.

Duan, J., Liu, Y., Wu, H., Wang, J., Chen, L., Chen, C., 2023. Broad learning for early diagnosis of Alzheimer's disease using FDG-PET of the brain. Front. Neurosci. 17, 1137567.

Dukart, J., Mueller, K., Horstmann, A., Vogt, B., Frisch, S., Barthel, H., Becker, G., Möller, H.E., Villringer, A., Sabri, O., et al., 2010. Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies. Neuroimage 49 (2), 1490–1495.

El-Sappagh, S., Abuhmed, T., Islam, S.R., Kwak, K.S., 2020. Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. Neurocomputing 412, 197–215.

Farias, S.T., Mungas, D., Reed, B.R., Harvey, D., DeCarli, C., 2009. Progression of mild cognitive impairment to dementia in clinic-vs community-based cohorts. Arch. Neurol. 66 (9), 1151–1157.

Fjell, A.M., McEvoy, L., Holland, D., Dale, A.M., Walhovd, K.B., Initiative, A.D.N., et al., 2014. What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. Prog. Neurobiol. 117, 20–40.

Ghazi, M.M., Nielsen, M., Pai, A., Cardoso, M.J., Modat, M., Ourselin, S., Sørensen, L., Initiative, A.D.N., et al., 2019. Training recurrent neural networks robust to incomplete data: Application to Alzheimer's disease progression modeling. Med. Image Anal. 53, 39–46.

Hinrichs, C., Singh, V., Xu, G., Johnson, S.C., Initiative, A.D.N., et al., 2011. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. Neuroimage 55 (2), 574–589.

Iaccarino, L., Sala, A., Perani, D., Initiative, A.D.N., 2019. Predicting long-term clinical stability in amyloid-positive subjects by FDG-PET. Ann. Clin. Transl. Neurol. 6 (6), 1113–1120.

Jack, C.R., 2022. Advances in Alzheimer's disease research over the past two decades. Lancet Neurol. 21 (10), 866–869.

Jagust, W., Gitcho, A., Sun, F., Kuczynski, B., Mungas, D., Haan, M., 2006. Brain imaging evidence of preclinical Alzheimer's disease in normal aging. Ann. Neurol. 59 (4), 673–681.

Jiji, S., Smitha, K.A., Gupta, A.K., Pillai, V.P.M., Jayasree, R.S., 2013. Segmentation and volumetric analysis of the caudate nucleus in Alzheimer's disease. Eur. J. Radiol. 82 (9), 1525–1530.

Katako, A., Shelton, P., Goertzen, A.L., Levin, D., Bybel, B., Aljuaid, M., Yoon, H.J., Kang, D.Y., Kim, S.M., Lee, C.S., et al., 2018. Machine learning identified an Alzheimer's disease-related FDG-PET pattern which is also expressed in Lewy body dementia and Parkinson's disease dementia. Sci. Rep. 8 (1), 13236.

Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., et al., 2020. Captum: A unified and generic model interpretability library for pytorch. arXiv preprint arXiv:2009.07896.

Landau, S.M., Harvey, D., Madison, C.M., Koeppe, R.A., Reiman, E.M., Foster, N.L., Weiner, M.W., Jagust, W.J., Initiative, A.D.N., et al., 2011. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. Neurobiol. Aging 32 (7), 1207–1218.

Lee, G., Nho, K., Kang, B., Sohn, K.-A., Kim, D., 2019. Predicting Alzheimer's disease progression using multi-modal deep learning approach. Sci. Rep. 9 (1), 1952.

Leech, R., Sharp, D.J., 2014. The role of the posterior cingulate cortex in cognition and disease. Brain 137 (1), 12–32.

Lei, B., Yang, M., Yang, P., Zhou, F., Hou, W., Zou, W., Li, X., Wang, T., Xiao, X., Wang, S., 2020. Deep and joint learning of longitudinal data for Alzheimer's disease prediction. Pattern Recognit. 102, 107247.

Liu, M., Zhang, J., Adeli, E., Shen, D., 2018. Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. 43, 157–168.

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

Madsen, S.K., Ho, A.J., Hua, X., Saharan, P.S., Toga, A.W., Jack, Jr., C.R., Weiner, M.W., Thompson, P.M., Initiative, A.D.N., et al., 2010. 3D maps localize caudate nucleus atrophy in 400 Alzheimer's disease, mild cognitive impairment, and healthy elderly subjects. Neurobiol. Aging 31 (8), 1312–1325.

Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Eshaghi, A., Toni, T., et al., 2020. The alzheimer's disease prediction of longitudinal evolution (TADPOLE) challenge: Results after 1 year follow-up. arXiv preprint arXiv:2002.03419.

McEvoy, L.K., Fennema-Notestine, C., Roddey, J.C., Hagler, Jr., D.J., Holland, D., Karow, D.S., Pung, C.J., Brewer, J.B., Dale, A.M., 2009. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology 251 (1), 195–205.

McEvoy, L.K., Holland, D., Hagler, Jr., D.J., Fennema-Notestine, C., Brewer, J.B., Dale, A.M., 2011. Mild cognitive impairment: baseline and longitudinal structural MR imaging measures improve predictive prognosis. Radiology 259 (3), 834–843.

Mintun, M.A., Lo, A.C., Duggan Evans, C., Wessels, A.M., Ardayfio, P.A., Andersen, S.W., Shcherbinin, S., Sparks, J., Sims, J.R., Brys, M., et al., 2021. Donanemab in early Alzheimer's disease. New Engl. J. Med. 384 (18), 1691–1704.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., Initiative, A.D.N., et al., 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage 104, 398–412.

Mosconi, L., De Santi, S., Brys, M., Tsui, W.H., Pirraglia, E., Glodzik-Sobanska, L., Rich, K.E., Switalski, R., Mehta, P.D., Pratico, D., et al., 2008. Hypometabolism and altered cerebrospinal fluid markers in normal apolipoprotein E E4 carriers with subjective memory complaints. Biol. Psychiatry 63 (6), 609–618.

Nordberg, A., Rinne, J.O., Kadir, A., Långström, B., 2010. The use of PET in Alzheimer disease. Nat. Rev. Neurol. 6 (2), 78–87.

Oh, K., Chung, Y.-C., Kim, K.W., Kim, W.-S., Oh, I.-S., 2019. Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning. Sci. Rep. 9 (1), 1–16.

Ou, Y.-N., Xu, W., Li, J.-Q., Guo, Y., Cui, M., Chen, K.-L., Huang, Y.-Y., Dong, Q., Tan, L., Yu, J.-T., 2019. FDG-PET as an independent biomarker for Alzheimer's biological diagnosis: a longitudinal study. Alzheimer's Res. Ther. 11 (1), 1–11.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.

Peña-Casanova, J., Sánchez-Benavides, G., de Sola, S., Manero-Borrás, R.M., Casals-Coll, M., 2012. Neuropsychology of Alzheimer's disease. Arch. Med. Res. 43 (8), 686–693.

Pohl, T., Jakab, M., Benesova, W., 2021. Interpretability of deep neural networks used for the diagnosis of Alzheimer's disease. Int. J. Imaging Syst. Technol..

Rajan, K.B., Weuve, J., Barnes, L.L., McAninch, E.A., Wilson, R.S., Evans, D.A., 2021. Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020–2060). Alzheimer's Dement. 17 (12), 1966–1975.

Ren, F., Yang, C., Qiu, Q., Zeng, N., Cai, C., Hou, C., Zou, Q., 2019. Exploiting discriminative regions of brain slices based on 2D CNNs for Alzheimer's disease classification. Ieee Access 7, 181423–181433.

Rolls, E.T., Huang, C.-C., Lin, C.-P., Feng, J., Joliot, M., 2020. Automated anatomical labelling atlas 3. Neuroimage 206, 116189.

Sala, A., Perani, D., 2019. Brain molecular connectivity in neurodegenerative diseases: recent advances and new perspectives using positron emission tomography. Front. Neurosci. 13, 617.

Samper-González, J., Burgos, N., Fontanella, S., Bertin, H., Habert, M.-O., Durrleman, S., Evgeniou, T., Colliot, O., Initiative, A.D.N., 2017. Yet another ADNI machine learning paper? Paving the way towards fully-reproducible research on classification of Alzheimer's disease. Mach. Learn. Med. Imaging 10541, 53–60.

Sauty, B., Durrleman, S., 2022. Riemannian metric learning for progression modeling of longitudinal datasets. In: IEEE 19th International Symposium on Biomedical Imaging. ISBI, pp. 1–5.

Scheef, L., Spottke, A., Daerr, M., Joe, A., Striepens, N., Kölsch, H., Popp, J., Daamen, M., Gorris, D., Heneka, M.T., et al., 2012. Glucose metabolism, gray matter structure, and memory decline in subjective memory impairment. Neurology 79 (13), 1332–1339.

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and statistical modeling with python. In: Proceedings of the 9th Python in Science Conference.

Sharma, R., Goel, T., Tanveer, M., Lin, C.T., Murugan, R., 2023. Deep-learning-based diagnosis and prognosis of Alzheimer's disease: A comprehensive review. IEEE Trans. Cogn. Dev. Syst. 15 (3), 1123–1138.

Smith, C., Chebrolu, H., Wekstein, D., Schmitt, F., Jicha, G., Cooper, G., Markesbery, W., 2007. Brain structural alterations before mild cognitive impairment. Neurology 68 (16), 1268–1273.

Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M., 2014. Striving for simplicity: The all convolutional net. arXiv 2014, arXiv preprint arXiv:1412.6806.

Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic attribution for deep networks. In: International Conference on Machine Learning. PMLR, pp. 3319–3328.

Teng, L., Li, Y., Zhao, Y., Hu, T., Zhang, Z., Yao, Z., Hu, B., (ADNI), A.D.N.I., 2020. Predicting MCI progression with FDG-PET and cognitive scores: a longitudinal study. BMC Neurol. 20, 1–10.

Thal, D.R., Rüb, U., Orantes, M., Braak, H., 2002. Phases of Aβ-deposition in the human brain and its relevance for the development of AD. Neurology 58 (12), 1791–1800.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15 (1), 273–289.

Wang, S.-H., Phillips, P., Sui, Y., Liu, B., Yang, M., Cheng, H., 2018. Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. J. Med. Syst. 42, 1–11.

Wang, Z., Xia, M., Dai, Z., Liang, X., Song, H., He, Y., Li, K., 2015. Differentially disrupted functional connectivity of the subregions of the inferior parietal lobule in Alzheimer's disease. Br. Struct. Funct. 220, 745–762.

Wang, L., Zang, Y., He, Y., Liang, M., Zhang, X., Tian, L., Wu, T., Jiang, T., Li, K., 2006. Changes in hippocampal connectivity in the early stages of Alzheimer's disease: evidence from resting state fMRI. Neuroimage 31 (2), 496–504.

Zhang, D., Shen, D., Initiative, A.D.N., et al., 2012. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 59 (2), 895–907.

Zhang, Y., Yang, Q., 2021. A survey on multi-task learning. IEEE Trans. Knowl. Data Eng. 34 (12), 5586–5609.