brain sciences

MDPI

Article

OViTAD: Optimized Vision Transformer to Predict Various Stages of Alzheimer's Disease Using Resting-State fMRI and Structural MRI Data

Saman Sarraf [1,2,*], Arman Sarraf [3], Danielle D. DeSouza [4], John A. E. Anderson [5], Milton Kabia [2] and The Alzheimer's Disease Neuroimaging Initiative [†]

[1] Institute of Electrical and Electronics Engineers, Piscataway, NJ 08854, USA
[2] School of Technology, Northcentral University, San Diego, CA 92123, USA
[3] Department of Electrical and Software Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
[4] Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA 94305, USA
[5] Departments of Cognitive Science and Psychology, Carleton University, Ottawa, ON K1S 5B6, Canada
[*] Correspondence: samansarraf@ieee.org

[†] Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigator within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/howtoapply/ADNIAcknowledgementList.pdf, 15 November 2021.

Abstract: Advances in applied machine learning techniques for neuroimaging have encouraged scientists to implement models to diagnose brain disorders such as Alzheimer's disease at early stages. Predicting the exact stage of Alzheimer's disease is challenging; however, complex deep learning techniques can precisely manage this. While successful, these complex architectures are difficult to interrogate and computationally expensive. Therefore, using novel, simpler architectures with more efficient pattern extraction capabilities, such as transformers, is of interest to neuroscientists. This study introduced an optimized vision transformer architecture to predict the group membership by separating healthy adults, mild cognitive impairment, and Alzheimer's brains within the same age group (>75 years) using resting-state functional (rs-fMRI) and structural magnetic resonance imaging (sMRI) data aggressively preprocessed by our pipeline. Our optimized architecture, known as OViTAD is currently the sole vision transformer-based end-to-end pipeline and outperformed the existing transformer models and most state-of-the-art solutions. Our model achieved F1-scores of 97% ± 0.0 and 99.55% ± 0.39 from the testing sets for the rs-fMRI and sMRI modalities in the triple-class prediction experiments. Furthermore, our model reached these performances using 30% fewer parameters than a vanilla transformer. Furthermore, the model was robust and repeatable, producing similar estimates across three runs with random data splits (we reported the averaged evaluation metrics). Finally, to challenge the model, we observed how it handled increasing noise levels by inserting varying numbers of healthy brains into the two dementia groups. Our findings suggest that optimized vision transformers are a promising and exciting new approach for neuroimaging applications, especially for Alzheimer's disease prediction.

Keywords: Alzheimer's disease; MCI; vision transformer; rs-fMRI; MRI

check for updates

Citation: Sarraf, S.; Sarraf, A.; DeSouza, D.D.; Anderson, J.A.E.; Kabia, M.; The Alzheimer's Disease Neuroimaging Initiative OViTAD: Optimized Vision Transformer to Predict Various Stages of Alzheimer's Disease Using Resting-State fMRI and Structural MRI Data. *Brain Sci.* **2023**, *13*, 260. https://doi.org/10.3390/brainsci13020260

Academic Editor: Michel Dojat

Received: 12 November 2022
Revised: 19 January 2023
Accepted: 1 February 2023
Published: 3 February 2023

1. Introduction

An early diagnosis of Alzheimer's disease (AD) delays the onset of dementia consequences for this life-threatening brain disorder and reduces the mortality rate and billion dollars cost of caring for AD patients [1–4]. The damages that Alzheimer's disease inflict are widespread, mostly targeting memory. Over time, shrinkage of the brain, the atrophy of the posterior cortical brain tissue degradation in the right temporal, parietal, and left frontal lobes and ventricular expansion interfere with patients' language, and memory abilities [5–7]. Researchers consider a transition phase known as mild cognitive impairment (MCI) from normal aging to acute AD, which often takes two to six years. As a result, patients lack focus, exhibit poor decision-making and judgment, experience time and location confusion, and suffer the onset of memory loss [8–10].

Among various biomarkers examinations such as blood and clinical tests, neuroimaging has remained the primary approach for medical practitioners to attempt an early prediction of Alzheimer's disease [11–14]. However, neurologists conduct various neuroimaging tests to diagnose Alzheimer's disease, since the impact of normal aging and early-stage Alzheimer's are barely distinguishable in neuroimaging [15].

Today, artificial intelligence (AI) in neuroimaging is considered an emerging technology where neuroscientists employ and adapt novel and advanced algorithms to analyze medical imaging data [16–18]. Over the past decade, deep learning techniques have enabled medical imaging scientists to predict various stages of Alzheimer's disease [19,20]. Using robust computational resources such as cloud computing, the scientists could implement end-to-end prediction pipelines to preprocess medical imaging data, build complex deep learning models, and post-process results to assist medical doctors in distinguishing early-stage MCI brains from highly correlated normal aging images [21–24].

Convolutional neural networks (CNNs) inspired by the human visual system form the core image classification component of pipelines. CNN-based classifiers consist of sophisticated feature extractors that retrieve hierarchical patterns from brain images and produce highly accurate predictions [25–30]. Although CNN models often require a light preprocessing pipeline, and the models aim to lessen the impact of noise implicitly, many studies have shown that a comprehensive preprocessing pipeline, to prepare neuroimaging data significantly improves prediction performance [31–33].

Advances in CNNs architectures and hybrid CNNs with other architectures, such as recurrent neural networks (RNNs), have significantly improved the performance and multistage AD prediction [34–36]. The central pillar of CNN-based pipelines is the convolutional layer, considered an invariant operator in signal and image processing. The convolutional layer reduces the sensitivity of the image classification pipeline to morphological variations such as shift and rotation [37–39].

Also, multi-dimension filters in CNN models and various combinations of feature map concatenation enhance such models' invariant characteristics [40,41]. However, the high complexity of models with hundreds of millions of trainable parameters, requiring high computations with an enormous amount of data, is considered a disadvantage of such methods [42–44]. Moreover, CNN models incorporate contextual information into training without considering positional embedding offered by transformer block [45,46].

In this study, we explore a novel method to bridge the gap of position-based context extraction through an optimized vision transformer. The literature review shows that using the vision transformer in predicting Alzheimer's disease is in a very early stage, and this study opens a new avenue to employ vision transformers in this domain. We implement two separate end-to-end pipelines to predict the triple class of Alzheimer's disease stages where pre- and postprocessing modules play crucial roles in improving prediction performance. Also, we analyze the impact of merging MCI data with healthy control and Alzheimer's brains to analyze modeling performance in binary classification tasks.

We repeat each model three times using random data splits and assess our pipelines using standard evaluation metrics by averaging across repetitions to ensure the robustness and reproducibility of our models. Finally, we visualize the attention and features maps to demonstrate the global impact of attention mechanisms employed in the architecture.

2. Related Work

Machine learning applications in predicting various stages of Alzheimer's disease have been of interest to numerous researchers, who began employing classical techniques such as support vector machines [47,48]. Researchers extracted features from Alzheimer's imaging data using autoencoders and classical techniques to classify AD and MCI brains. This approach introduced more advanced feature extractors compared to the classical methods, which improved the performance of AD prediction [49,50]. The next generation of predictive models included many CNN architectures to classify mainly AD and HC brains. The successful binary classification motivated imaging and neuroscientists to employ sophisticated techniques to address 3-class prediction tasks of HC vs. MCI vs. AD [51–54].

Besides 2D CNN architectures, 3D convolutional layers enabled scientists to incorporate the volumetric data into the training process. Such approaches produced promising predictions using structural MRI data [55–57]. The 3D models used the signal intensity at the voxel level and applied the convolution operator to 3D filters and previous-layer feature maps. Although 3D models became popular due to producing high accuracy rates, many scientists challenged these techniques, conducting experiments in which 2D models outperformed 3D models [54,58–60].

Some research groups considered using functional MRI 4D data to predict various stages of AD, where they composed the brain images into 2D samples along with depth and time axes. The data decomposition method produced a significant amount of data for training and resulted in a nearly perfect binary classification performance, outperforming most of the models built by structural data. The major challenge in using 4D fMRI data was to establish a preprocessing pipeline to prepare the data for model development [61–64].

Recurrent neural networks (RNNs) and their subsequent architectures, such as long- and short-term memory LSTM models, capture features from a sequence of data that are useful to extract temporal relationships encoded in Alzheimer's imaging data [65,66]. A special use of LSTM models occurs in longitudinal analysis for Alzheimer's disease prediction. In this approach, researchers extract spatial maps from imaging data using various feature extractors, such as multi-layer perceptron (MLP), and train bi-directional LSTMs to address the AD classification problems. This two-step prediction allows neuroscientists to explore the patterns in longitudinal imaging, which are suppressed in cross-sectional methods [67–69]. However, the extra step of explicit feature extraction and the complex impact of sensitive longitudinal analysis remains the major challenge of using such methods [70,71].

The next category of machine learning methods used for Alzheimer's disease prediction is hybrid modeling, where CNNs and RNNs models extract hierarchical and temporal features in a cascade architecture. The CNN component of such networks is considered the central feature extractor, and the RNN-LSTM component extracts position-related features and forms the core of the model [35,71,72].

Multimodal imaging in the same category provides complementary information from each modality, such as fMRI, structural MRI, and PET, that often transfer the predicted labels to a postprocessing or ensemble model. Since the nature of each modality is different, using combined data to build a unique model for AD prediction produces poor performance because the model hardly converges [73–76]. Some researchers considered a hybrid approach, using clinical and imaging data to develop separate models that followed a predictive model. Such a technique offers strong decision-making, since the misprediction by imaging models is compensated for by clinical data [77,78].

Transformers with the various implementations of attention mechanisms stemming from natural language processing (NLP) domains have been of interest to scientists regarding whether such technology is adaptable for Alzheimer's disease prediction [79,80]. For example, a deep neural network with transformer blocks was the core of an Alzheimer's study to assess risks using targeted speech [81].

Transformers' temporal or sequential feature extraction capability allowed researchers to develop end-to-end solutions to predict Alzheimer's through a longitudinal model known as TrasforMesh using structural data [82]. Also, a universal brain encoder based on a transformer with attention mechanisms offered model explainability to analyze 3D MRI data [82]. The transformer technology has motivated scientists to implement predictive models using 3D data in non-Alzheimer's studies, such as defect assessment of knee cartilage [83]. To date, our proposed method of using an optimized vision transformer (OViTAD) to predict various stages of Alzheimer's is considered the first initiative in adopting this technology.

3. Materials and Methods

3.1. Datasets

We used two sets of Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/, 15 July 2021), including fMRI and structural MRI imaging data. We recruited older adults (age group $>$ 75) for both imaging modalities in this study with the aim of suppressing the effect of aging on modeling. Using only older adults in this study enabled us to ensure our models predict the Alzheimer's stages not aging effect. We ensured ground truth quality; we cross-checked the participants' proposed labels by ADNI with their mini-mental state examination (MMSE) scores. The fMRI dataset contained 275 participants scanned for resting-state fMRI (rs-fMRI) studies; we found 52 Alzheimer's (AD), 92 healthy control (HC), and 131 MCI brains in our fMRI dataset. The structural MRI dataset included 1076 participants, where we found 211 AD, 91 HC, and 744 MCI brains. Table 1 shows the participants' demographic details for both modalities categorized into three groups: gender, age, and MMSE scores.

Table 1. The demographic of two sets of ADNI data used in model development shows all the groups are older adults within an age group of $>$75.

Modality	Total	Group	Participant	Female	Age	Male	Age	MMSE
rs-fMRI	284	AD	54	27	80.96 ± 4.64	27	79.0 ± 2.74	22.70 ± 2.10
		HC	99	49	79.78 ± 4.76	50	82.57 ± 3.88	28.82 ± 1.35
		MCI	131	66	79.15 ± 3.09	65	79.72 ± 4.84	26.53 ± 2.51
MRI	1460	AD	577	232	80.98 ± 4.65	345	81.27 ± 4.08	23.07 ± 2.06
		HC	108	51	79.37 ± 3.54	57	80.81 ± 4.42	28.81 ± 1.35
		MCI	775	265	80.28 ± 3.31	510	81.61 ± 4.15	26.53 ± 2.09

3.2. Image Acquisition Protocol

ADNI provided a standard protocol to scientists to acquire imaging data using three Tesla scanners, including General Electric (GE) Healthcare, Philips Medical Systems, and Siemens Medical Solutions machines [84]. We ensured that the two datasets utilized in this study were collected using the same scanning parameters. The protocol stated that the functional scans were performed using an echo-planar imaging (EPI) sequence (150 volumes, repetition time (TR) = 2 second (s), echo to time (TE) = 30 milliseconds (ms), flip angle (FA) = 70 degrees, filed-of-view (FOV) = 20 centimeters (cm)) that produced 64 × 64 matrices with 30 axial slices of 5 millimeters (mm) thickness without a gap. The structural MRI data acquisition employed a 3-dimensional (3D) magnetization prepared rapid acquisition gradient echo sequence known as MPRAGE (TR = 2 s, TE = 2.63 ms, FOV = 25.6 cm) that produced 256 × 256 matrices with 160 slices of 1mm thickness.

3.3. Data Preprocessing

3.3.1. rs-fMRI

We considered an extensive 7-step pipeline to preprocess the rs-fMRI data to preprocess our data from scratch, as the research indicated that enhanced preprocessing rs-fMRI data improved the performance of modeling [85,86]. First, we converted the raw rs-fMRI data, downloaded from ADNI in digital imaging and communications in medicine (DICOM) format, to neuroimaging informatics technology initiative (NIfTI/NII) format using an open-source tool known as the dcm2niix software [87]. We removed skull and neck voxels considered non-brain regions from the structural T1-weighted imaging data corresponding to each fMRI time course using FSL-BET software [88]. Third, using FSL-MCFLIRT [89], we corrected the rs-fMRI data for motion artifact caused by low-frequency drifts, which could negatively impact the time course decomposition. Finally, we applied a standard slice timing correction (STC) method known as Hanning-Windowed Sinc Interpolation (HWSI) to each voxel's time series. According to the ADNI data acquisition protocol, the brain slices were acquired halfway through the relevant volume's TR; therefore, we shifted each time series by a proper fraction relative to the middle point of TR period. We spatially smoothed the rs-fMRI time series using a Gaussian kernel with 5 mm full width half maximum (FWHM). Next, we employed a temporal high-pass filter with a cut-off frequency of 0.01 HZ (Sigma = 90 s) to remove low-frequency noise. We registered the fMRI brains to the corresponding high-resolution structural T1-weighted scans using an affine linear transformation with seven degrees of freedom (7 DOF). Subsequently, we aligned the registered brains to the Montreal Neurological Institute standard brain template (MNI152) using an affine linear transformation with 12 DOF [90]. We resampled the aligned brains by a 4 mm kernel that generated 45 × 54 × 45 brain slices per time course. The rs-fMRI preprocessing pipeline produced 4-dimensional (4D) data, including time series within $T \in [124, 200]$ with the mode of 140 data points per participant; therefore, we obtained 4D NIfTI/NII files of 45 × 54 × 45 × T.

3.3.2. Structural MRI

We preprocessed the structural MRI data from scratch using a 6-step pipeline where we first converted the DICOM raw images to NifTi/NII format using dcm2niix software [87]. Next, we extracted the brain regions by removing the skull and neck tissues from the data [88]. Then, using the FSL-VBM library [91], we segmented the brain images into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). We used the GM images to register to the GM ICBM-152 standard template using a linear affine transformation with 6 DOF. Next, we concatenated the brain images, flipped them along the x-axis, then re-averaged to create a first-pass, study-specific template as a standard approach [88]. Next, we re-registered the structural MRI brains to the template using a non-linear transformation, and then resampled to create a 2 × 2 × 2 mm^3 GM template in the standard space. Per FSL-VBM standard protocol, we applied a modulation technique to the structural MRI data by multiplying each voxel by the Jacobian of the warp field to compensate for the enlargement that occurred via the non-linear component of transformation. Subsequently, we used all the concatenated and averaged 3D GM images (one 3D sample per participant) to create a 4D data stack. Finally, we smoothed the structural MRI data using a range of Gaussian kernels with sigma = 3, 4 (FWHM of 4.6, 7, and 9.3 mm), as the research showed that the smoothing significantly impacted the performance of modeling [92,93]. The structural MRI preprocessing pipeline produced two sets (one set per sigma) of 3D NIfTI/NII files of 91 × 109 × 91.

3.4. Proposed Architecture: Optimized Vision Transformer (OViTAD)

Inspired by a transformer built for natural language processing use cases [94], vision transformers have been adopted for computer vision tasks such as image classification or object detection. The vanilla vision transformer [95] employs a dozen multi-head self-attention (MHSA) layers, considered the transformer blocks building the core of architecture. This algorithm splits an input image into small patches that are passed through positional embedding and transformer encoder layers. During the training process, the positional information is incorporated by attention layers which a similar to better predict farther data points from the current state [94,95]. The vision transformer generated patches from a given set of preprocessed images, converted the 2D arrays into 1D arrays, and decomposed them along the axes for the three channels. The dimension of each patch is calculated by multiplying the number of channels by the height and width of the patches. We prepared the linearly embedded arrays to feed into the next blocks. To address the objective of our multiple-class Alzheimer's prediction, where we used specific imaging data dimensions; we set our transformer's input dimension to 56 × 56 for fMRI and 112 × 112 for structural MRI, which were the closest meaningful dimensions reflecting popular image size. This data-driven approach allowed us to bypass a computationally massive grid search by optimizing the network's hyperparameters. Since we reduced the vision transformer input dimension from 224 × 224 × 3 to 112 × 112 × 3 and 56 × 56 × 3, we reduced the number of heads in MHSA in architecture to optimize the architecture. The core intention was to improve the efficiency of our model while producing the same or better performance compared to the vanilla version with reduced trainable parameters. In the next step, the vision transformer used a positional embedding to feed the arrays to the transformer 8-head self-attention block with six layers in depth, which applied a set of standard steps to the arrays similar to the original architecture [94,95]. To decrease the chance of overfitting, we set our dropout and embedding dropout to 0.1. We used a multi-perceptron layer, known as the fully connected layer of 2048 neurons, to translate the features extracted by the optimal vision transformer to a format usable for the cross-entropy loss function to evaluate classification performance. Figure 1 pictures the architecture of the optimized vision transformer implemented in this study.

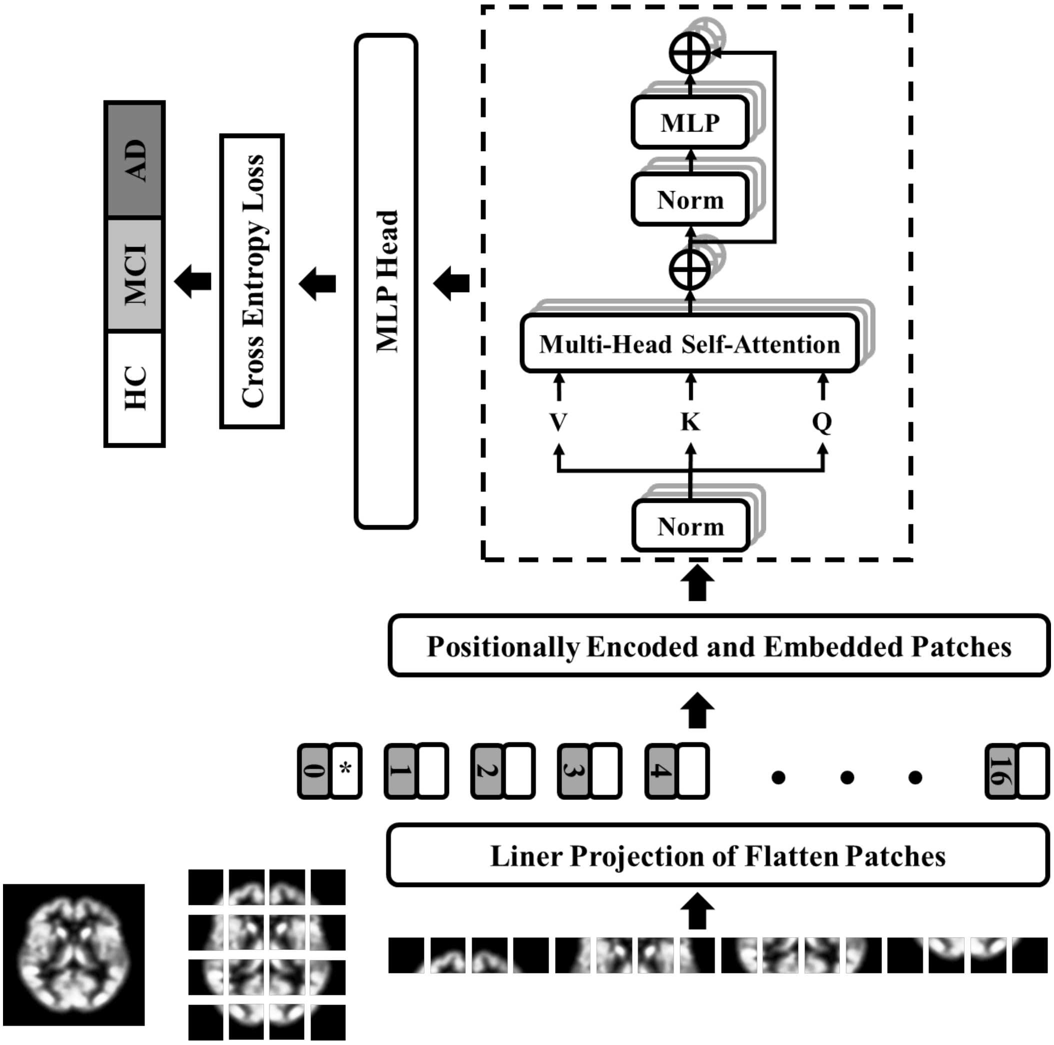

AD
MCI
HC
Cross Entropy Loss
MLP Head
MLP
Norm
Multi-Head Self-Attention
V
K
Q
Norm
Positionally Encoded and Embedded Patches
0
*
1
2
3
4
16
Liner Projection of Flatten Patches

Figure 1. The OViTAD architecture is an optimized ViT shown for structural MRI data composed of a linear projection layer applied to the flattened patches fed into an 8-HSA Transformer. The MLP layer of 2048 parameters translates the features from the transformer encoder to a proper format for the cross-entropy loss function

We used DeepViT, which is a deeper version of a vision transformer, to build our baselines [96]. DeepViT employs a mechanism known as re-attention, instead of MHSA, to reproduce attention maps to increase the diversity of features extracted by the architecture. The re-attention layers benefit from the interaction across various heads to capture further information, which improves the diversity of attention maps through a learnable transformation matrix known as Q. Figure 2 (left) demonstrates the DeepViT transformer block with its re-attention mechanism. To enhance the scope of our benchmarking, we used another vision transformer image classifier known as class-attention in image transformers (CaIT) that introduced a class-attention layer [97]. The CaIT architecture consists of two major components: (a) standard self-attention step which is identical to the ViT transformer, and (b) a class-attention layer step, including a set of operations to convert the positionally embedded patches into class embedding arrays (CLS), followed by a linear classification method. CaIT with the CLS mechanism avoids the saturation of deep vision transformers in the early state and allows the model to further learn across training. Figure 2 (right) shows the CaIT transform block.

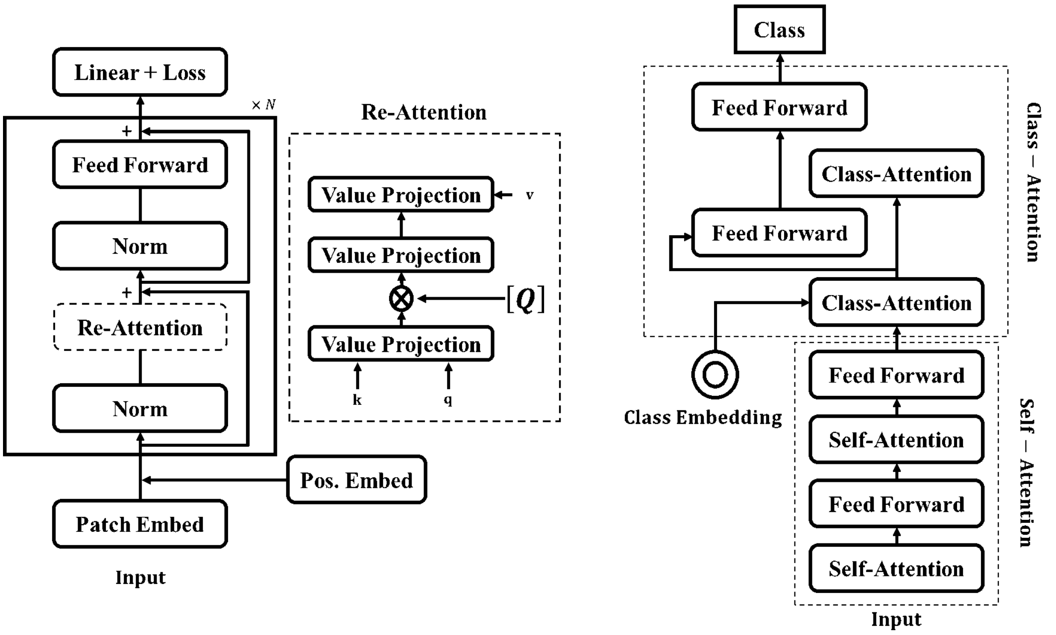


Figure 2. The transformer block in DeepViT architecture includes a re-attention module instead of a standard self-attention layer (**Left**). Class-Attention in Image Transformer architecture consists of a class embedding (CLS) and additional class-attention layers preceded by self-attention layers (**Right**).

3.5. *fMRI Pipeline*

We categorized the preprocessed 4D fMRI samples (one NIfTI/NII per participant) into AD, HC, and MCI classes. In the next step, we used a stratified split of 80%–10%–10% and randomly shuffled data at class level to generate three training, validation, and testing sets. Therefore, the sets included 226, 27, and 31 participants for training, validation, and testing. The main objective of this study was to perform a multiclass prediction; however, we expanded our modeling approach to explore the impact of merging MCI data with two other classes and generated samples for AD + MCI vs. HC and AD vs. HC + MCI experiments. Using our optimized model, we also built AD vs. HC and HC vs. MCI models for a consistent comparison with the literature. To perform a consistent comparison, we used the identical data splits generated for multiclassification for the two binary classifications, where we only modified the corresponding ground truth according to experiments.

3.5.1. Data Decomposition from 4D to 2D

We decomposed the 4D fMRI data and z and t axes into 2D images using a lossless data conversion method to generate portable network graphics (PNG) samples for model development. We first loaded the NIfTI files into memory using the Nibabel package available at https://nipy.org/nibabel/, 20 August 2021 and employed the Python OpenCV library available at OpenCV.org to store the decomposed 2D images in the server. Next, we removed the last ten brain slices and empty brain images to improve data quality. To find the empty slices, we measured the sum of pixel intensity in a given brain image and only stored images with non-zero-sum. Equation (1) shows the details of fMRI data decomposition.

$$\begin{aligned}
&for\ \forall z = 1\ to\ Z - 10 \\
&for\ \forall t = 1\ to\ T \\
&if\ SI_{z,t}(BS_{z,t}(x,y)) = \sum_{X}^{x=1}\sum_{Y}^{y=1} BS(x,y) \neq 0: \\
&BS_{z,t}(x,y) \rightarrow PNG(BS_{z,t}(x,y)) \\
&otherwise: \\
&Ignore\ BS_{z,t}(x,y)
\end{aligned} \tag{1}$$

where X, Y, and Z represent the spatial dimensions of fMRI data (45, 54, 45), and T refers to 140 data points of a given fMRI time course. $SI(z,t)$ represents the sum of voxel intensity in a given brain slice, $BS_{z,t}(x,y)$ represents and PNG denotes the lossless data conversion function. The decomposition module produced 1,433,880 images consisting of 1,141,280, 138,600, and 154,000 samples for training, validation, and testing purposes.

3.5.2. Modeling

The central objective of our fMRI pipeline was to address the multiclass prediction of AD, HC, and MCI using our designed optimal vision transformer. Furthermore, we considered two additional binary classification experiments mentioned earlier: (a) AD + MCI against HC and (b) HC + MCI against AD, to explore the clinical impact of merging MCI with the other classes. We built our optimized vision transformer (OViTAD) and three other baselines—CaIT, DeepViT, and vanilla vision transformer—and used the Amazon Web Services' (AWS) SageMaker infrastructure as our development environment. We spun up a p3.8xlarge instance with 32 virtual central processing units (vCPUs) and 244 gigabyte (GB) memory. The instance included four NVIDIA TESLA V100-SXM2-16GB graphical processing units (GPUs) and 10 GB per second (Gbps) network performance. We trained all the models for 40 epochs and a batch size 64 using the Adam optimization method with a learning rate lr = 3 × 10^{-5}, gamma = 0.7, and stepsize = 1. We monitored modeling performance across the epochs using accuracy rates and loss scores for training and validation sets. We used the accuracy rate of validation sets as the criteria for selecting the best model. We implemented the prediction module to load the stored best models into the memory and predict validation and testing sets with their probability scores at slice level. We evaluated the performance of the models using a standard classification report by calculating precision, recall, F1-score, and accuracy rates. Table A1 in the Appendix A demonstrates the models' performance at slice level for validation and test datasets and three repetitions (random data splits) of fMRI experiments.

3.5.3. Subject-Level Evaluation

We designed our modeling based on the decomposition of brain image into a 2D image; therefore, the performance obtained from the prediction module demonstrated the slice-level performance. To calculate the performance of our models at the subject level (see Table A1 Appendix A), we applied a vote for majority method to the predicted labels by aggregating the results based on subjects' identifiers (IDs). Next, we calculated the probability of each class per subject and then voted for the class with the highest probability. Finally, we used our standard classification report to measure the performance of our models at the subject level. Table A2 in the Appendix A shows the models' performance for validation and test datasets and three repetitions (random data splits) of experiments for fMRI data. First, we calculated the macro average (macro-ave) and weighted average (weighted-avg) for precision, recall, and F1-score evaluation metrics. Next, we analyzed at model level to explore classification performance across the experiments. We used the weighted average scores of the aforementioned four metrics and calculated each experiment's average and standard deviation against three repetitions (random data splits). Table 2 shows the performance of models for validation and test sets with the averaged metrics and the corresponding standard deviation values. We summarized the results of this table in Figure A5 comparing the performance of fMRI models using averaged F1-score for three testing sets.

Table 2. To evaluate the performance of experiments referring to model-level results, we used the weighted average scores of subject-level results and calculated each experiment's average and standard deviation across three repetitions for validation and test sets (random data splits).

Model	Dataset	Precision	Recall	F1-Score	Accuracy
CaIT_ADMCI-HC	Val Test	0.53 ± 0.14 0.54 ± 0.21	0.68 ± 0.02 0.66 ± 0.02	0.57 ± 0.07 0.53 ± 0.04	0.68 ± 0.02 0.66 ± 0.02
CaIT_AD-HCMCI	Val Test	0.66 ± 0 0.65 ± 0	0.81 ± 0 0.81 ± 0	0.73 ± 0 0.72 ± 0	0.81 ± 0 0.81 ± 0
CaIT_AD-HC-MCI	Val Test	0.4 ± 0.14 0.37 ± 0.01	0.54 ± 0.06 0.46 ± 0.02	0.43 ± 0.11 0.37 ± 0.01	0.54 ± 0.06 0.46 ± 0.02
DeepViT_AD_HC_MCI	Val Test	1 ± 0 0.96 ± 0.02	1 ± 0 0.96 ± 0.02	1 ± 0 0.96 ± 0.02	1 ± 0 0.96 ± 0.02
DeepViT_AD_HCMCI	Val Test	0.99 ± 0.02 0.99 ± 0.02	0.99 ± 0.02 0.99 ± 0.02	0.99 ± 0.02 0.99 ± 0.02	0.99 ± 0.02 0.99 ± 0.02
DeepViT_ADMCI_HC	Val Test	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02
ViT_224_8_AD_HC_MCI	Val Test	0.99 ± 0.02 0.97 ± 0.03	0.99 ± 0.02 0.97 ± 0.03	0.99 ± 0.02 0.97 ± 0.03	0.99 ± 0.02 0.97 ± 0.03
ViT_224_8_AD_HCMCI	Val Test	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02
ViT_224_8_ADMCI_HC	Val Test	1 ± 0 0.97 ± 0	1 ± 0 0.97 ± 0	1 ± 0 0.97 ± 0	1 ± 0 0.97 ± 0
ViT_vanilla_AD_HC_MCI	Val Test	0.99 ± 0.02 0.97 ± 0	0.99 ± 0.02 0.97 ± 0	0.99 ± 0.02 0.97 ± 0	0.99 ± 0.02 0.97 ± 0
ViT_vanilla_AD_HCMCI	Val Test	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02
ViT_vanilla_ADMCI_HC	Val Test	1 ± 0 0.98 ± 0.02	1 ± 0 0.98 ± 0.02	1 ± 0 0.98 ± 0.02	1 ± 0 0.98 ± 0.02
OViTAD_AD_HC_MCI	Val Test	0.99 ± 0.02 0.97 ± 0	0.99 ± 0.02 0.97 ± 0	0.99 ± 0.02 0.97 ± 0	0.99 ± 0.02 0.97 ± 0
OViTAD_AD_HCMCI	Val Test	0.99 ± 0.02 0.99 ± 0.02	0.99 ± 0.02 0.99 ± 0.02	0.99 ± 0.02 0.99 ± 0.02	0.99 ± 0.02 0.99 ± 0.02
OViTAD_ADMCI_HC	Val Test	1 ± 0 0.98 ± 0.02	1 ± 0 0.98 ± 0.02	1 ± 0 0.98 ± 0.02	1 ± 0 0.98 ± 0.02
OViTAD_AD_HC	Val Test	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02	1 ± 0 0.99 ± 0.02
OViTAD_HC_MCI	Val Test	1 ± 0 0.97 ± 0.03	1 ± 0 0.97 ± 0.03	1 ± 0 0.97 ± 0.03	1 ± 0 0.97 ± 0.03

3.6. Structural Pipeline

3.6.1. Data Split

We categorized the preprocessed 3D structural MRI samples (one NIfTI/NII per participant) into AD, HC, and MCI classes. In the next step, we used a stratified split of 80%–10%–10% and randomly shuffled data at class-level to generate three training, validation, and testing sets for two sets of preprocessed data S3 (sigma = 3 mm) and S4 (sigma = 4 mm). Therefore, the sets included 1167, 144, and 149 participants for training, validation, and testing, respectively. Similar to the fMRI pipeline, we explored the impact of merging MCI data with AD and HC. We used the identical data splits generated for multiclass prediction to address the binary classification experiments by updating the corresponding ground truth; this strategy allowed us to perform a consistent comparison across experiments and two sigma variations.

3.6.2. Data Decomposition 3D to 2D

We employed the same technique explained in Equation 1 to decompose 3D MRI data into 2D PNG images. As the structural MRI data are constructed without temporal information, we set the time parameter in the equation to T = 1. The structural MRI decomposition module produced 111,899 images per set containing 89,446, 11,040, and 11,413 samples for training, validating, and testing our models.

3.6.3. Modeling

The main objective of the structural MRI pipeline was to conduct a multiclass prediction of AD, HC, and MCI classes using two sets of preprocessed data (sigma = 3, 4) and to evaluate our proposed optimal vision transformer architecture. Also, we used four other models as baselines similar to the fMRI pipeline to investigate the performance of optimal architecture. Furthermore, we considered combining MCI data with AD and HC to classify (a) AD + MCI against HC and (b) HC + MCI against AD. Similar to the fMRI pipeline, we utilized AWS SageMaker as the development environment on a p3.8xlarge instance equipped with NVIDIA GPUs. We trained all the models for 40 epochs and a batch size 64 using the Adam optimization method with a learning rate $lr = 3 \times 10^{-5}$, gamma = 0.7, and step_size = 1. Using loss scores and accuracy rates of training and validation sets, we evaluated the training process and selected the best model based upon the highest accuracy rate obtained from the validation sets. Since we designed our vision transformers to use 2D images, we developed a prediction module to output validation and test sets' labels at slice level. We employed our standard classification report module to generate a macro and weighted average of precision, recall, F1-scores, and accuracy rates. We show the slice-level performance of structural MRI models in Tables A3 and A4 in Appendix A and for sigma = 3, 4.

3.6.4. Subject-Level

We used the predicted labels for brain slices and aggregated the results by the subject IDs to calculate the models' performance at subject level; the slice-level performance is shown in Table A4 Appendix. Then, using the postprocessing module based on the voting for majority concept, we counted the number of each class prediction in an experiment and measured each class probability. In the next step, we assigned the corresponding label of the highest probability to a given subject. Finally, we employed our standard classification reports as described earlier, and generated the evaluation scores at the subject level. Tables A5 and A6 in the Appendix demonstrate the subject-level performance of structural MRI models for preprocessed data with spatial smoothing sigma = 3, 4, respectively. To measure the performance of experiments at the model level, we used the weighted average evaluation scores and calculated the average and standard deviation of the scores for both structural MRI datasets, shown in Table 3. We summarized the results of this table in Figure A6 comparing the performance of sMRI models using averaged F1-score for three testing sets.

Table 3. The models' performance of two sets for structural MRI experiments evaluated by standard evaluation metrics.

Model	Dataset	Precision	Recall	F1-Score	Accuracy
CaIT_S3_AD-HC-MCI	Val	0.74 ± 0.02	0.8 ± 0.02	0.77 ± 0.02	0.8 ± 0.02
	Test	0.71 ± 0.02	0.77 ± 0.03	0.73 ± 0.03	0.77 ± 0.03
CaIT_S3_AD-HCMCI	Val	0.72 ± 0.02	0.72 ± 0.03	0.7 ± 0.03	0.72 ± 0.03
	Test	0.71 ± 0.02	0.7 ± 0.01	0.69 ± 0.01	0.7 ± 0.01
CaIT_S3_ADMCI-HC	Val	0.87 ± 0	0.93 ± 0	0.9 ± 0	0.93 ± 0
	Test	0.85 ± 0	0.92 ± 0	0.88 ± 0	0.92 ± 0
DeepViT_S3_AD-HC-MCI	Val	0.81 ± 0.06	0.85 ± 0.03	0.82 ± 0.04	0.85 ± 0.03
	Test	0.77 ± 0.02	0.84 ± 0.02	0.81 ± 0.02	0.84 ± 0.02
DeepViT_S3_AD-HCMCI	Val	0.84 ± 0.04	0.84 ± 0.05	0.84 ± 0.05	0.84 ± 0.05
	Test	0.83 ± 0.04	0.83 ± 0.04	0.83 ± 0.04	0.83 ± 0.04
DeepViT_S3_ADMCI-HC	Val	0.87 ± 0	0.93 ± 0	0.9 ± 0	0.93 ± 0
	Test	0.85 ± 0	0.92 ± 0	0.88 ± 0	0.92 ± 0
ResNet50_S3_AD-HC-MCI	Val	0.85 ± 0.09	0.86 ± 0.05	0.84 ± 0.06	0.86 ± 0.05
	Test	0.83 ± 0.06	0.85 ± 0.02	0.82 ± 0.02	0.85 ± 0.02
ResNet50_S3_AD-HCMCI	Val	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01
	Test	0.84 ± 0.04	0.84 ± 0.04	0.84 ± 0.04	0.84 ± 0.04
ResNet50_S3_ADMCI-HC	Val	0.87 ± 0	0.93 ± 0	0.9 ± 0	0.93 ± 0
	Test	0.85 ± 0	0.92 ± 0	0.88 ± 0	0.92 ± 0
ViT_S3_AD-HC-MCI	Val	0.84 ± 0.05	0.88 ± 0.02	0.85 ± 0.03	0.88 ± 0.02
	Test	0.84 ± 0.08	0.85 ± 0.03	0.83 ± 0.04	0.85 ± 0.03
ViT_S3_AD-HCMCI	Val	0.84 ± 0.01	0.84 ± 0.01	0.83 ± 0.02	0.84 ± 0.01
	Test	0.84 ± 0.03	0.83 ± 0.02	0.83 ± 0.02	0.83 ± 0.02
ViT_S3_ADMCI-HC	Val	0.87 ± 0	0.93 ± 0	0.9 ± 0	0.93 ± 0
	Test	0.85 ± 0	0.92 ± 0	0.88 ± 0	0.92 ± 0
OViTAD_S3_AD-HC-MCI	Val	0.78 ± 0.02	0.84 ± 0.02	0.81 ± 0.02	0.84 ± 0.02
	Test	0.75 ± 0.03	0.82 ± 0.03	0.79 ± 0.03	0.82 ± 0.03
OViTAD_S3_AD-HCMCI	Val	0.79 ± 0.03	0.77 ± 0.05	0.75 ± 0.07	0.77 ± 0.05
	Test	0.79 ± 0.02	0.77 ± 0.04	0.75 ± 0.06	0.77 ± 0.04
OViTAD_S3_ADMCI-HC	Val	0.87 ± 0	0.93 ± 0	0.9 ± 0	0.93 ± 0
	Test	0.85 ± 0	0.92 ± 0	0.88 ± 0	0.92 ± 0
OViTAD_S3_AD-HC	Val	1 ± 0	1 ± 0	1 ± 0	1 ± 0
	Test	1 ± 0	1 ± 0	1 ± 0	1 ± 0
OViTAD_S3_HC-MCI	Val	1 ± 0	1 ± 0	1 ± 0	1 ± 0
	Test	1 ± 0	1 ± 0	1 ± 0	1 ± 0
CaIT_S4_AD-HC-MCI	Val	0.84 ± 0.03	0.9 ± 0.03	0.87 ± 0.03	0.9 ± 0.03
	Test	0.81 ± 0.01	0.88 ± 0.01	0.84 ± 0.01	0.88 ± 0.01
CaIT_S4_AD-HCMCI	Val	0.87 ± 0.02	0.87 ± 0.02	0.87 ± 0.02	0.87 ± 0.02
	Test	0.86 ± 0.02	0.86 ± 0.02	0.86 ± 0.02	0.86 ± 0.02
CaIT_S4_ADMCI-HC	Val	0.87 ± 0	0.93 ± 0	0.9 ± 0	0.93 ± 0
	Test	0.85 ± 0	0.92 ± 0	0.88 ± 0	0.92 ± 0
DeepViT_S4_AD-HC-MCI	Val	0.85 ± 0.01	0.91 ± 0.01	0.88 ± 0.01	0.91 ± 0.01
	Test	0.82 ± 0.01	0.88 ± 0.01	0.85 ± 0.01	0.88 ± 0.01
DeepViT_S4_AD-HCMCI	Val	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01
	Test	0.89 ± 0.02	0.89 ± 0.02	0.89 ± 0.02	0.89 ± 0.02
DeepViT_S4_ADMCI-HC	Val	0.87 ± 0	0.93 ± 0	0.9 ± 0	0.93 ± 0
	Test	0.85 ± 0	0.92 ± 0	0.88 ± 0	0.92 ± 0

Table 3. *Cont.*

Model	Dataset	Precision	Recall	F1-Score	Accuracy
ResNet50_S4_AD-HC-MCI	Val Test	0.93 ± 0.01 0.89 ± 0.06	0.93 ± 0.01 0.91 ± 0.02	0.91 ± 0.01 0.89 ± 0.03	0.93 ± 0.01 0.91 ± 0.02
ResNet50_S4_AD-HCMCI	Val Test	0.93 ± 0 0.91 ± 0.01	0.93 ± 0 0.91 ± 0.01	0.93 ± 0 0.91 ± 0.01	0.93 ± 0 0.91 ± 0.01
ResNet50_S4_ADMCI-HC	Val Test	0.87 ± 0 0.87 ± 0.05	0.93 ± 0 0.92 ± 0	0.9 ± 0 0.89 ± 0.01	0.93 ± 0 0.92 ± 0
ViT_S4_AD-HC-MCI	Val Test	0.9 ± 0.06 0.88 ± 0.06	0.91 ± 0.02 0.9 ± 0.02	0.89 ± 0.02 0.87 ± 0.02	0.91 ± 0.02 0.9 ± 0.02
ViT_S4_AD-HCMCI	Val Test	0.91 ± 0 0.9 ± 0.03	0.91 ± 0 0.9 ± 0.03	0.91 ± 0 0.9 ± 0.03	0.91 ± 0 0.9 ± 0.03
ViT_S4_ADMCI-HC	Val Test	0.87 ± 0 0.85 ± 0	0.93 ± 0 0.92 ± 0	0.9 ± 0 0.88 ± 0	0.93 ± 0 0.92 ± 0
OViTAD_S4_AD-HC-MCI	Val Test	0.86 ± 0.06 0.81 ± 0.01	0.9 ± 0.02 0.87 ± 0.01	0.87 ± 0.03 0.84 ± 0.01	0.9 ± 0.02 0.87 ± 0.01
OViTAD_S4_AD-HCMCI	Val Test	0.9 ± 0 0.89 ± 0.02	0.89 ± 0 0.88 ± 0.02	0.89 ± 0 0.88 ± 0.02	0.89 ± 0 0.88 ± 0.02
OViTAD_S4_ADMCI-HC	Val Test	0.87 ± 0 0.85 ± 0	0.93 ± 0 0.92 ± 0	0.9 ± 0 0.88 ± 0	0.93 ± 0 0.92 ± 0
OViTAD_S4_AD-HC	Val Test	1 ± 0 1 ± 0	1 ± 0 1 ± 0	1 ± 0 1 ± 0	1 ± 0 1 ± 0
OViTAD_S4_HC-MCI	Val Test	1 ± 0 1 ± 0	1 ± 0 1 ± 0	1 ± 0 1 ± 0	1 ± 0 1 ± 0

3.7. Discussion

3.7.1. Technical/Architecture Design

We designed an optimized vision transformer architecture to predict multiple stages of Alzheimer's disease using fMRI and MRI data. Our end-to-end pipeline for two modalities was built on four major components: (a) aggressive preprocessing of fMRI and MRI data, (b) data decomposition from higher dimensions to 2D, (c) vision transformer model development, and (d) postprocessing. The core concept of this study was to explore the capability of vision transformers to predict Alzheimer's stages. We exhaustively trained models to conduct a comprehensive evaluation of our proposed architecture. We investigated the performance of our baselines and our proposed architecture against fMRI and two sets of structural MRI data to address the 3-class AD vs. HC vs. MCI, AD vs. HC + MCI, and AD + MCI vs. HC classifications. To demonstrate the robustness of our modeling approach, we repeated each experiment with random data splits three times. More random data splits, such as five to ten runs, could be explored in future work. We reported the performances at the slice level and subject level, which led us to compare our models across all experiments (model level). We proposed an optimized vision transformer architecture as the core of our end-to-end prediction pipeline. Our optimization approach is based on the scientific fact of using an image input size of architecture that has the closest and most meaningful input dimensions of preprocessed fMRI data. Therefore, we set the architecture input dimension to 56 × 56 and resample our data (45 × 54) to fit our optimal architecture, where the originality of data content remains through minimal upsampling. Next, we consider reducing the number of heads in the multi-head attention layer to decrease the complexity and trainable parameters of the network. We showed in Table 4 that we decreased the input image size and trainable parameters in the optimized network by 75% and 28% compared to the vanilla vision transformer, while improving the models' performance in the fMRI experiments and producing a similar performance to other models in the structural MRI experiments. Unlike grid search-based optimization, which requires massive model development to achieve an optimal architecture and topology, our fact- and data-driven optimization method, which stems from the impact of input size, produced faster converging modeling. This allowed us to explore a broader set of model development and clinical analysis.

Table 4. The number of trainable parameters reduced by 28% compared to vanilla vision transformer and DeepViT while producing higher performance in fMRI and similar performance to other models in structural MRI data.

Model	Input (Channel,x,y)	Params
CaIT	3,224,224	120,707,075
DeepViT	3,224,224	53,532,867
ViT-vanilla	3,224,224	53,532,675
ViT-224-8	3,224,224	40,949,763
OViTAD	3,56,56	38,406,147

We consider the fMRI testing datasets as our gold standard to compare the performance of our models. Unlike training and validation datasets, the testing datasets are unseen and never used in the training processes. The models' performance at the subject. level using fMRI data, shown in Table 3, reveals that OViTAD, DeepViT, ViT-vanilla, and ViT-224-8 in AD-HCMCI classification outperforms other models with an F1-score of 0.99 ± 0.02. Also, among the models trained for the 3-class AD vs. HC vs. MCI prediction, our optimized OViTAD model is on par with the ViT-vanilla, and ViT-224-8 outperforms other models with an F1-score of 0.97 ± 0.02; our optimized models contain much fewer trainable parameters than other models. Also, we investigated the impact of the post-processing step developed based on voting for the majority algorithm, and the results indicated that the models' performance at the subject level (after postprocessing) with an averaged F1-score of 0.89 ± 0.02 across all experiments (testing datasets) are higher (with 3% improvement) than slice-level ones with an averaged F1-score of 0.86 ± 0.02. This finding aligned with the literature [54,58,64] that shows that postprocessing plays a crucial role in improving the performance of modeling and proves that decomposition of data from higher dimensions to 2D and back-transforming the slice-level predictions to the subject level improve the quality of prediction significantly.

Similar to the above approach, we consider the structural MRI (sigma = 3) testing datasets as the golden standard to investigate the best-performing model. The results shown in Table 3 reveal that our OViTAD is on par with DeepViT, ViT-vanilla, and ViT-224-8 in the $ADMCI - HC$ S3 and S4 experiments in terms of F1-scores at the subject level. To explore the central objective of this study, we reviewed the performance of models for 3-class AD vs. HC vs. MCI prediction. The results indicated that ViT-vanilla and ViT-224-8 competed with our OViTAD and produced an F1-score of 0.99 ± 0.01 (0.004 is negligibly higher than OViTAD) using MRI S3. After preprocessing, the original MRI dimension was 91X109, and we downsampled the structural MRI data to 112 × 112, causing a loss in contextual information. Similarly, we analyzed the behavior of our models trained and evaluated by the preprocessed MRI with sigma = 4 testing datasets. Our OViTAD model using MRI S4 was on par with other architectures, producing the best performance with an F1-score of 0.99 ± 0.01.

The results suggest that the input size and number of patches in the attention layers greatly impact the performance of the structural MRI models. In a similar observation to fMRI testing datasets, the models' performance at the subject level (after postprocessing with voting for a majority) increased by 7% compared to the slice-level models across the experiments for sigma = 3, 4. Our analysis indicated that spatial smoothing with a Gaussian kernel of sigma = 3 mm resulted in slightly higher evaluation scores across the study (an average increase of 0.43% in sigma = 3 compared to the sigma = 4 dataset) which aligns with the previous research; however, the improvement is negligible [54,58]. Spatial smoothing is important in preprocessing MRI data that removes random noise in a given voxel's neighborhoods [98,99].

This finding implies that the nature of features extracted by attention layers in the vision transformer should differ from the features extracted by convolutional layers since the impact of sigma = 3, 4 in the previous studies was negligible [54,58]. Next, we calculated the confusion matrix of testing samples normalized per group for the best-performing OViTAD fMRI (test set 2), MRI-S3, and MRI-S4 (test sets 3) models in the multiple classification experiment to predict AD vs. HC vs. MCI, illustrated in Figure 3. The performance of the best-performing OViTAD models for the same test sets across 40 epochs is shown in Figure A4, Appendix A.

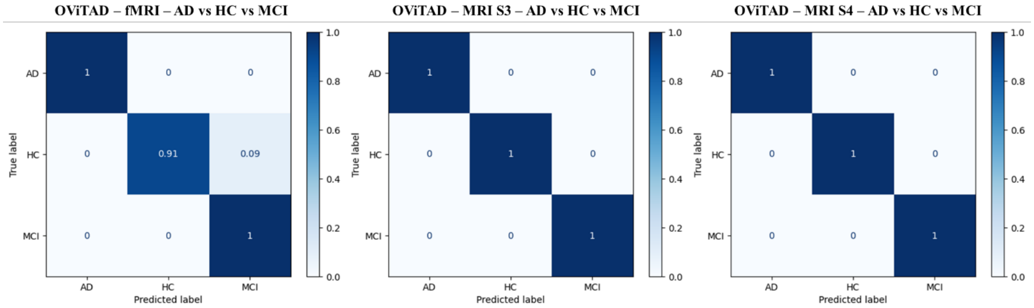


Figure 3. The normalized confusion matrices for the best-performing fMRI (**left**), MRI-S3 (**middle**), and MRI-S4 (**right**) OViTAD models to classify AD vs. HC vs. MCI at subject-level.

Moreover, we comprehensively compared our findings and the recent literature reviews in which the ADNI dataset was used for Alzheimer's disease classification. We carefully selected the most current, highly referenced studies and offered novel techniques where the performance of models was highly competitive. Table 5 compares the performance achieved by OViTAD in the two modalities with the most highly referenced recent literature. Our finding shows that this study offers a broader range of classifications where the optimized vision transformer outperforms the state-of-the-art models.

3.7.2. Clinical Observation

We considered combining the health control brains with Alzheimer's and mild cognitive impairment brains to generate new sets from the ADNI dataset to perform two binary classification tasks using all the models. The fMRI models revealed a consistent pattern in which the AD vs. HC + MCI models outperformed AD + MCI vs. HC by 4.64% with respect to averaged F1-scores across all experiments shown in Table 2. This finding revealed some level of similarity between HC and MCI functional data. Also, the results showed that our predictive models could differentiate HC data from non-HC data, which revealed that our models properly addressed the aging effect in this study. Furthermore, we analyzed the binary models trained by structural MRI data for AD + MCI vs. HC and AD vs. HC + MCI experiments for the two sigma = 3,4. The results indicated that our HC vs. AD + MCI models outperformed AD vs. HC + MCI by 2.82%, respecting the averaged F1-scores across all experiments for the two sigma values shown in Table 3.

Table 5. Comparison between recent studies of Alzheimer's classification using ADNI and our OViTAD. The analysis shows that our study addresses a broader classification aspect with novel vision transformer technology, and our model performance outperformed the literature. Further details is found at Table A7.

Reference	Modality	AD vs. HC vs. MCI	AD + MCI vs. HC	AD vs. MCI+HC	AD vs. HC	MCI vs. HC
Lin et al. 2018 [100]	MRI	-	-	-	88.79%	-
Dimitriadis et al. 2018 [101]	MRI	61.90%	-	-	-	-
Kruthika et al. 2019 [102]	MRI	90.47%	-	-	-	-
Spasov et al. 2019 [103]	MRI + Clinical	-	-	-	-	86%
Basaia et al. 2019 [104]	MRI	-	-	-	98%	87%
Abrol et al. 2020 [105]	MRI	83.01%	-	-	-	-
Shao et al. 2020 [106]	MRI+PET	-	-	-	92.51%	82.53%
Alinsaif et al. 2021 [107]	MRI	-	70.50%	-	62.22%	-
Alinsaif et al. 2021 [107]	MRI	-	91.61%	-	92.78%	-
Ramzan et al. 2019 [63]	rs-fMRI	97.92%	-	-	-	-
Hojjati et al. 2018 [108]	MRI + rs-fMRI	-	-	93%	-	-
Cui et al. 2019 [109]	MRI	-	-	-	91.33%	-
Amoroso et al. 2018 [110]	MRI	38.80%	-	-	-	-
Buvaneswari et al. 2021 [111]	rs-fMRI	-	-	-	-	79.15%
Duc et al. 2019 [61]	rs-fMRI+Clinical	-	-	-	85.27%	-
OViTAD—fMRI	rs-fMRI	0.97 ± 0	0.98 ± 0.02	0.99 ± 0.02	0.99 ± 0.02	0.97 ± 0.03
OViTAD—MRI (Sigma = 3)	MRI	0.9955% ± 0.0039	1 ± 0	0.9955 ± 0.0039	1 ± 0	1 ± 0
OViTAD—MRI (Sigma = 4)	MRI	0.9955% ± 0.0039	1 ± 0	0.9955 ± 0.0039	1 ± 0	1 ± 0

3.7.3. Local and Global Attention Visualization

We extracted the attention weights and produced post-SoftMax for eight self-attention heads with a depth of six. Then, using a random AD fMRI brain slice, we generated the self-attention maps based on OViTAD for AD vs. HC vs. MCI classification as shown in Figure A1, Appendix A. The attention maps in each column represent one self-attention head, whereas the maps in each row represent the depth of attention layers. Also, we explored the impact of attention mechanisms at the global level. We utilized the last feature vector of OViTAD—the fMRI AD vs. HC vs. MCI classification, which is a fully connected layer (FC)—and considered it the global attention feature. The FC layer represents the features produced by the self-attention layers; therefore, it contains the information of global attention. We employed an element-wise operator to obtain the sum of multiplication between each pixel and all the elements in the FC vector. Next, we generated the normalized global attention feature maps for a set of AD fMRI slices in the testing set as shown in Equation (2) and visualized the maps using the CIVIDIS color map, illustrated in Figure 4.

$$\begin{aligned}
&image_{resize} = Resize(image_{original} \rightarrow 56 \times 56)\\
&GlobalAttentionFeatureMap(GAFM) = \sum image_{resize} \times FC_{vector}\\
&GAFM_{normalized} = (GAFM - min(GAFM)) * \frac{255}{max(GAFM) - min(GAFM)}
\end{aligned} \tag{2}$$

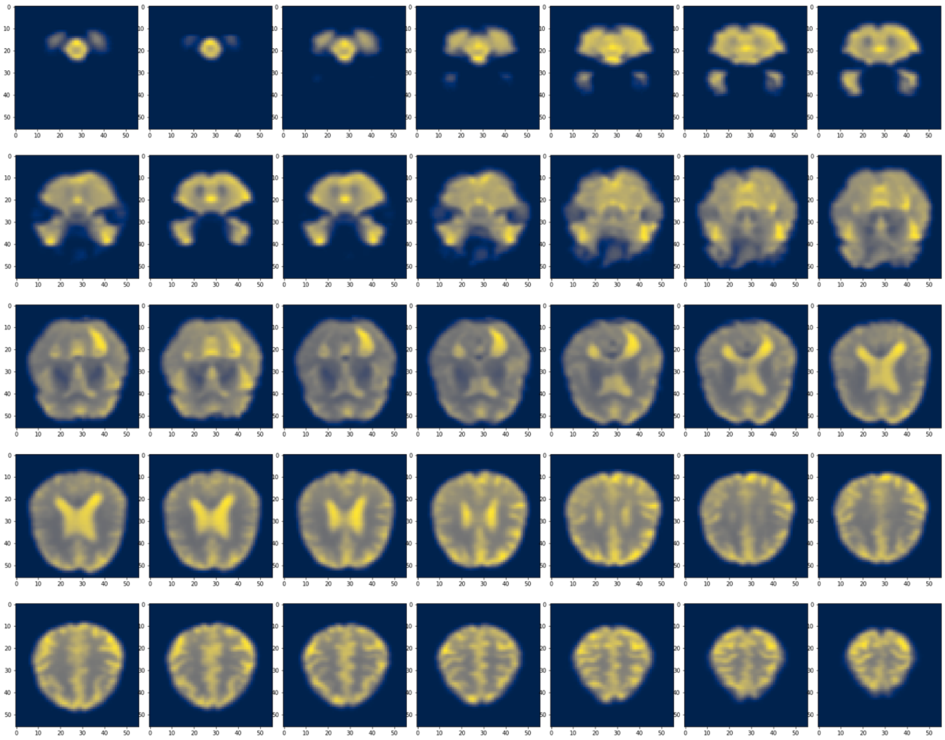


Figure 4. The global attention feature map was obtained by multiplying the FC layer vector by each pixel in the fMRI brain slices and measuring the sum of the multiplication per pixel. Next, we normalized the feature map to (0, 255) and visualized the maps using the CIVIDIS color map. Finally, we selected the first slice of each time-course to demonstrate various brain morphology across the fMRI data acquisition.

3.7.4. Limitations

The number of repetitions for model development is considered a limitation in this research study. Although we utilized a large dataset and generated three random data splits for modeling, it is highly recommended to repeat this exercise up to 10 times with randomly shuffled data to ensure the robustness of OViTAD. Also, we included voting for a majority technique as postprocessing to stabilize models' performance; however, this step would add an extra layer of computation to our pipeline, increasing the modeling cost. Future work could address such a limitation using upper-dimension models, including 3D vision transformers. Training of vision transformer models is costly; therefore, reducing the image input size discussed in this research decreases training time and inspection latency. Finally, this research study outlined an end-to-end machine learning pipeline to predict Alzheimer's disease stages using the ADNI dataset so that the models' performance reflects the accuracy of the pipeline for this dataset. Since the early prediction of this brain disorder is crucial in clinical studies, a variety of existing Alzheimer's datasets should be explored along with ADNI to examine OViTAD performance in future work.

4. Conclusions

This study introduced an optimized vision transformer called OViTAD to predict healthy, MCI, and AD brains using rs-fMRI and structural MRI (sigma = 3,4 mm) data. The prediction pipeline included two separate preprocessing stages for the two modalities, training and evaluation of slice-level vision transformers and a postprocessing step based on voting for the majority concept. The results showed that our optimized vision transformer outperformed and was on par with the vision transformers-based benchmark. OViTAD 30% reduced the number of trainable parameters compared to the vanilla ViT. The average performance of OViTAD across three repetitions (random data splits) was 97% ± 0.0 and 99.55% ± 0.39 for the two modalities for the multi-class classification experiments, which outperformed most existing deep learning and CNN-based models. Also, we introduced a method of visualizing the attention mechanism's global effect, enabling scientists to explore crucial brain areas. This study showed that the vision transformers could outperform and compete with the state-of-the-art algorithms to predict various stages of Alzheimer's disease with less complex architectures.

Author Contributions: S.S. designed, implemented and executed the end-to-end project including data processing, modeling, and writing. A.S. collaborated in modeling and software development. D.D.D., J.A.E.A. and M.K. reviewed and proof read the project and manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Institutional Review Board Statement: ADNI board of committee approved the use of the dataset in this research study.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. The slice-level models' performance is described in this table for the validation and test datasets and three repetitions (random data splits). The classification report includes the macro and weighted average precision, recall, and F1-score. The report also includes the accuracy rate and the number of unseen slices used for each model evaluation.

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
CaIT_AD-HC-MCI	Val	1	0.4984	0.4116	0.449	0.3991	0.4984	0.3614	0.442	138600
		2	0.4938	0.4335	0.4598	0.3868	0.4938	0.3519	0.4379	133560
		3	0.4877	0.4148	0.4404	0.355	0.4877	0.2777	0.3673	134400
	Test	1	0.5011	0.4701	0.4865	0.402	0.5011	0.3604	0.4396	155820
		2	0.4989	0.4828	0.4895	0.4005	0.4989	0.3577	0.4377	156100
		3	0.4691	0.3507	0.4031	0.3723	0.4691	0.3148	0.3865	155820
CaIT_AD-HCMCI	Val	1	0.8081	0.404	0.653	0.5	0.8081	0.4469	0.7223	138600
		2	0.8166	0.4083	0.6668	0.5	0.8166	0.4495	0.7341	133560
		3	0.8021	0.9011	0.8413	0.5001	0.8021	0.4452	0.714	134400
	Test	1	0.8113	0.4057	0.6582	0.5	0.8113	0.4479	0.7268	155820
		2	0.8117	0.4058	0.6588	0.5	0.8117	0.448	0.7273	156100
		3	0.7979	0.8989	0.8387	0.5	0.7979	0.4439	0.7082	155820
CaIT_ADMCI-HC	Val	1	0.6799	0.6389	0.66	0.6	0.6799	0.6007	0.6546	138600
		2	0.675	0.6496	0.659	0.548	0.675	0.5121	0.6002	133560
		3	0.6716	0.5506	0.5924	0.5004	0.6716	0.4039	0.5412	134400
	Test	1	0.6502	0.6159	0.6291	0.568	0.6502	0.5536	0.6071	155820
		2	0.6572	0.654	0.655	0.5582	0.6572	0.5233	0.5879	156100
		3	0.6397	0.4783	0.5239	0.4998	0.6397	0.3923	0.5013	155820
DeepViT_ADMCI_HC	Val	1	0.9723	0.9696	0.9723	0.9681	0.9724	0.9711	0.9723	138600
		2	0.9904	0.9895	0.9904	0.9903	0.9904	0.9886	0.9904	138600
		3	0.9832	0.9812	0.9831	0.9858	0.9834	0.9771	0.9832	133560
	Test	1	0.9109	0.9038	0.9105	0.9076	0.9106	0.9004	0.9109	155820
		2	0.989	0.9882	0.989	0.9894	0.989	0.9869	0.989	155820
		3	0.991	0.9904	0.991	0.9882	0.9912	0.9928	0.991	156100

Table A1. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
DeepViT_AD_HCMCI	Val	1	0.9618	0.935	0.9607	0.9618	0.9618	0.9131	0.9618	138600
		2	0.9831	0.9719	0.9831	0.9695	0.9832	0.9743	0.9831	133560
		3	0.9875	0.9804	0.9875	0.9779	0.9876	0.983	0.9875	134400
	Test	1	0.9807	0.9683	0.9806	0.9698	0.9806	0.9668	0.9807	155820
		2	0.9719	0.954	0.9719	0.9542	0.9719	0.9538	0.9719	156100
		3	0.9315	0.8818	0.9275	0.9381	0.9325	0.8451	0.9315	155820
DeepViT_AD_HC_MCI	Val	1	0.9387	0.9322	0.9385	0.9358	0.9392	0.9296	0.9387	138600
		2	0.935	0.9352	0.9352	0.9365	0.9361	0.9347	0.935	133560
		3	0.9299	0.9323	0.9299	0.9282	0.9303	0.937	0.9299	134400
	Test	1	0.8997	0.9049	0.8995	0.9019	0.8996	0.9083	0.8997	155820
		2	0.9069	0.9092	0.9069	0.9123	0.9078	0.9069	0.9069	156100
		3	0.8855	0.8727	0.8837	0.8954	0.8879	0.8589	0.8855	155820
ViT24_8_ADMCI_HC	Val	1	0.9772	0.975	0.9773	0.9723	0.9775	0.978	0.9772	138600
		2	0.9572	0.9527	0.9572	0.9514	0.9573	0.954	0.9572	133560
		3	0.9507	0.9442	0.9507	0.943	0.9508	0.9455	0.9507	134400
	Test	1	0.9181	0.9112	0.9176	0.9166	0.9179	0.9068	0.9181	155820
		2	0.9396	0.9348	0.9393	0.9387	0.9395	0.9314	0.9396	156100
		3	0.9461	0.9412	0.946	0.9438	0.946	0.9388	0.9461	155820
ViT24_8_AD_HCMCI	Val	1	0.9738	0.9564	0.9734	0.9708	0.9736	0.9436	0.9738	138600
		2	0.9849	0.9747	0.9849	0.9761	0.9848	0.9733	0.9849	133560
		3	0.9894	0.9834	0.9894	0.9784	0.9896	0.9886	0.9894	134400
	Test	1	0.9782	0.965	0.9784	0.9571	0.9788	0.9734	0.9782	155820
		2	0.9856	0.9765	0.9856	0.9743	0.9856	0.9787	0.9856	156100
		3	0.932	0.8852	0.9289	0.9274	0.9314	0.8552	0.932	155820
ViT24_8_AD_HC_MCI	Val	1	0.9491	0.943	0.9487	0.9485	0.9497	0.9392	0.9491	138600
		2	0.9318	0.9328	0.932	0.9313	0.9334	0.9354	0.9318	133560
		3	0.922	0.9241	0.9218	0.9183	0.9223	0.9308	0.922	134400
	Test	1	0.9137	0.9174	0.9135	0.9149	0.9137	0.9202	0.9137	155820
		2	0.9207	0.9232	0.9207	0.9226	0.9208	0.9241	0.9207	156100
		3	0.887	0.8705	0.885	0.8871	0.8871	0.8596	0.887	155820

Table A1. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
ViT_vanilla_ADMCI_HC	Val	1	0.9653	0.9622	0.9655	0.9581	0.9661	0.9669	0.9653	138600
		2	0.9376	0.9308	0.9376	0.9312	0.9376	0.9304	0.9376	133560
		3	0.9445	0.9368	0.9444	0.9381	0.9443	0.9357	0.9445	134400
	Test	1	0.9352	0.9298	0.9348	0.9351	0.9352	0.9253	0.9352	155820
		2	0.9198	0.9117	0.9185	0.9278	0.922	0.9011	0.9198	156100
		3	0.9406	0.9355	0.9406	0.9355	0.9406	0.9355	0.9406	155820
ViT_vanilla_AD_HCMCI	Val	1	0.9663	0.9431	0.9655	0.9656	0.9662	0.924	0.9663	138600
		2	0.9777	0.9634	0.9779	0.9568	0.9782	0.9702	0.9777	133560
		3	0.9866	0.9791	0.9867	0.9748	0.9868	0.9836	0.9866	134400
	Test	1	0.9837	0.9735	0.9837	0.9696	0.9839	0.9776	0.9837	155820
		2	0.9706	0.9515	0.9705	0.9559	0.9704	0.9472	0.9706	156100
		3	0.9213	0.8677	0.9179	0.906	0.9195	0.8402	0.9213	155820
ViT_vanilla_AD_HC_MCI	Val	1	0.9359	0.9245	0.935	0.9329	0.9355	0.9183	0.9359	138600
		2	0.9095	0.9092	0.9095	0.9027	0.9112	0.9174	0.9095	133560
		3	0.9255	0.9285	0.9255	0.927	0.9256	0.93	0.9255	134400
	Test	1	0.9158	0.9194	0.9155	0.9248	0.9182	0.9165	0.9158	155820
		2	0.9074	0.9071	0.9072	0.9073	0.9088	0.9085	0.9074	156100
		3	0.8876	0.8653	0.8839	0.8947	0.8901	0.85	0.8876	155820
OViTAD_ADMCI_HC	Val	1	0.9636	0.9602	0.9637	0.9576	0.964	0.9629	0.9636	138600
		2	0.9501	0.9445	0.95	0.9459	0.9499	0.9431	0.9501	133560
		3	0.9435	0.9362	0.9436	0.9345	0.9438	0.938	0.9435	134400
	Test	1	0.9146	0.9073	0.914	0.9134	0.9144	0.9024	0.9146	155820
		2	0.9281	0.9211	0.9271	0.9348	0.9297	0.9116	0.9281	156100
		3	0.9221	0.9155	0.9222	0.9152	0.9222	0.9159	0.9221	155820
OViTAD_AD_HCMCI	Val	1	0.9539	0.9218	0.9527	0.9467	0.9534	0.9012	0.9539	138600
		2	0.9754	0.9586	0.9753	0.9621	0.9753	0.9553	0.9754	133560
		3	0.9854	0.9773	0.9855	0.9727	0.9856	0.9819	0.9854	134400
	Test	1	0.9699	0.9511	0.97	0.9482	0.9701	0.9541	0.9699	155820
		2	0.9843	0.9749	0.9845	0.9653	0.985	0.9853	0.9843	156100
		3	0.9205	0.867	0.9173	0.9022	0.9184	0.8412	0.9205	155820

Table A1. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
OViTAD_AD_HC_MCI	Val	1	0.9239	0.914	0.9232	0.925	0.924	0.9055	0.9239	138600
		2	0.9102	0.91	0.9102	0.9097	0.9112	0.9111	0.9102	133560
		3	0.9192	0.9219	0.9193	0.917	0.921	0.9283	0.9192	134400
	Test	1	0.8939	0.9026	0.8936	0.9057	0.8949	0.9009	0.8939	155820
		2	0.9025	0.9053	0.9024	0.9038	0.9024	0.907	0.9025	156100
		3	0.8882	0.8682	0.8852	0.891	0.889	0.8557	0.8882	155820
OViTAD_AD_HC	Val	1	0.9615	0.9639	0.9618	0.9519	0.9615	0.9574	0.9612	74900
		2	0.9877	0.989	0.9878	0.984	0.9877	0.9864	0.9877	70420
		3	0.9815	0.9775	0.982	0.9838	0.9815	0.9804	0.9816	70700
	Test	1	0.9608	0.9501	0.9627	0.9653	0.9608	0.9569	0.9611	87220
		2	0.9545	0.9429	0.9568	0.9589	0.9545	0.95	0.9549	87500
		3	0.8946	0.9004	0.8962	0.8694	0.8946	0.8814	0.8925	87500
OViTAD_HC_MCI	Val	1	0.9596	0.9576	0.9602	0.9608	0.9596	0.959	0.9597	112000
		2	0.9265	0.9231	0.9277	0.928	0.9265	0.9251	0.9267	109060
		3	0.9283	0.9243	0.929	0.9285	0.9283	0.9262	0.9285	107800
	Test	1	0.9038	0.9049	0.9041	0.9013	0.9038	0.9027	0.9036	126420
		2	0.8921	0.8934	0.8926	0.8894	0.8921	0.8909	0.8918	126700
		3	0.9419	0.9429	0.9421	0.9398	0.9419	0.9411	0.9418	124320

Table A2. The fMRI subject-level models' performance is described in this table for the validation and test datasets and three repetitions (random data splits). The classification report includes the macro and weighted average precision, recall, and F1-score. The report also includes the accuracy rate and the number of unseen subjects aggregated by the postprocessing module and used for each model evaluation. In this table, AD-HC-MCI refers to multiclass (3-class) prediction, AD-HCMCI refers to AD vs. HC + MCI, and ADMCI-HC represents AD + MCI vs. HC binary classifications.

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
CaIT_AD-HC-MCI	Val	1	0.5926	0.4127	0.4974	0.4558	0.5926	0.4131	0.5176	27
		2	0.5556	0.3818	0.4626	0.4188	0.5556	0.3714	0.473	27
		3	0.4815	0.1605	0.2318	0.3333	0.4815	0.2167	0.313	27
	Test	1	0.4516	0.3148	0.3781	0.3463	0.4516	0.284	0.359	31
		2	0.4839	0.3	0.3677	0.3701	0.4839	0.3	0.3823	31
		3	0.4516	0.3148	0.3781	0.3463	0.4516	0.284	0.359	31
CaIT_AD-HCMCI	Val	1	0.8148	0.4074	0.6639	0.5	0.8148	0.449	0.7317	27
		2	0.8148	0.4074	0.6639	0.5	0.8148	0.449	0.7317	27
		3	0.8148	0.4074	0.6639	0.5	0.8148	0.449	0.7317	27
	Test	1	0.8065	0.4032	0.6504	0.5	0.8065	0.4464	0.72	31
		2	0.8065	0.4032	0.6504	0.5	0.8065	0.4464	0.72	31
		3	0.8065	0.4032	0.6504	0.5	0.8065	0.4464	0.72	31
CaIT_ADMCI-HC	Val	1	0.7037	0.6875	0.6944	0.5833	0.7037	0.5714	0.6508	27
		2	0.6667	0.3333	0.4444	0.5	0.6667	0.4	0.5333	27
		3	0.6667	0.3333	0.4444	0.5	0.6667	0.4	0.5333	27
	Test	1	0.6452	0.3226	0.4162	0.5	0.6452	0.3922	0.506	31
		2	0.6774	0.8333	0.7849	0.5455	0.6774	0.4833	0.5753	31
		3	0.6452	0.3226	0.4162	0.5	0.6452	0.3922	0.506	31
DeepViT_ADMCI_HC	Val	1	1	1	1	1	1	1	1	27
		2	1	1	1	1	1	1	1	27
		3	1	1	1	1	1	1	1	27
	Test	1	0.9677	0.964	0.9674	0.9762	0.9693	0.9545	0.9677	31
		2	1	1	1	1	1	1	1	31
		3	1	1	1	1	1	1	1	31

Table A2. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
DeepViT_AD_HCMCI	Val	1	0.963	0.9333	0.9613	0.9783	0.9646	0.9	0.963	27
		2	1	1	1	1	1	1	1	27
		3	1	1	1	1	1	1	1	27
	Test	1	1	1	1	1	1	1	1	31
		2	1	1	1	1	1	1	1	31
		3	0.9677	0.9447	0.9666	0.9808	0.969	0.9167	0.9677	31
DeepViT_AD_HC_MCI	Val	1	1	1	1	1	1	1	1	27
		2	1	1	1	1	1	1	1	27
		3	1	1	1	1	1	1	1	27
	Test	1	0.9677	0.9726	0.9675	0.9778	0.9699	0.9697	0.9677	31
		2	0.9677	0.9726	0.9675	0.9778	0.9699	0.9697	0.9677	31
		3	0.9355	0.9316	0.9354	0.9583	0.9435	0.9141	0.9355	31
ViT24_8_ADMCI_HC	Val	1	1	1	1	1	1	1	1	27
		2	1	1	1	1	1	1	1	27
		3	1	1	1	1	1	1	1	27
	Test	1	0.9677	0.964	0.9674	0.9762	0.9693	0.9545	0.9677	31
		2	0.9677	0.964	0.9674	0.9762	0.9693	0.9545	0.9677	31
		3	0.9677	0.964	0.9674	0.9762	0.9693	0.9545	0.9677	31
ViT24_8_AD_HCMCI	Val	1	1	1	1	1	1	1	1	27
		2	1	1	1	1	1	1	1	27
		3	1	1	1	1	1	1	1	27
	Test	1	1	1	1	1	1	1	1	31
		2	1	1	1	1	1	1	1	31
		3	0.9677	0.9447	0.9666	0.9808	0.969	0.9167	0.9677	31
ViT24_8_AD_HC_MCI	Val	1	1	1	1	1	1	1	1	27
		2	1	1	1	1	1	1	1	27
		3	0.963	0.968	0.9626	0.9762	0.9656	0.963	0.963	27
	Test	1	1	1	1	1	1	1	1	31
		2	0.9677	0.9726	0.9675	0.9778	0.9699	0.9697	0.9677	31
		3	0.9355	0.9316	0.9354	0.9583	0.9435	0.9141	0.9355	31

Table A2. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
ViT_vanilla_ADMCI_HC	Val	1	1	1	1	1	1	1	1	27
		2	1	1	1	1	1	1	1	27
		3	1	1	1	1	1	1	1	27
	Test	1	0.9677	0.964	0.9674	0.9762	0.9693	0.9545	0.9677	31
		2	0.9677	0.964	0.9674	0.9762	0.9693	0.9545	0.9677	31
		3	1	1	1	1	1	1	1	31
ViT_vanilla_AD_HCMCI	Val	1	1	1	1	1	1	1	1	27
		2	1	1	1	1	1	1	1	27
		3	1	1	1	1	1	1	1	27
	Test	1	1	1	1	1	1	1	1	31
		2	1	1	1	1	1	1	1	31
		3	0.9677	0.9447	0.9666	0.9808	0.969	0.9167	0.9677	31
ViT_vanilla_AD_HC_MCI	Val	1	1	1	1	1	1	1	1	27
		2	1	1	1	1	1	1	1	27
		3	0.963	0.9691	0.9632	0.9667	0.9667	0.9744	0.963	27
	Test	1	0.9677	0.9726	0.9675	0.9778	0.9699	0.9697	0.9677	31
		2	0.9677	0.9726	0.9675	0.9778	0.9699	0.9697	0.9677	31
		3	0.9677	0.9582	0.9668	0.9778	0.9699	0.9444	0.9677	31
OViTAD_ADMCI_HC	Val	1	1	1	1	1	1	1	1	27
		2	1	1	1	1	1	1	1	27
		3	1	1	1	1	1	1	1	27
	Test	1	0.9677	0.964	0.9674	0.9762	0.9693	0.9545	0.9677	31
		2	0.9677	0.964	0.9674	0.9762	0.9693	0.9545	0.9677	31
		3	1	1	1	1	1	1	1	31
OViTAD_AD_HCMCI	Val	1	0.963	0.9333	0.9613	0.9783	0.9646	0.9	0.963	27
		2	1	1	1	1	1	1	1	27
		3	1	1	1	1	1	1	1	27
	Test	1	1	1	1	1	1	1	1	31
		2	1	1	1	1	1	1	1	31
		3	0.9677	0.9447	0.9666	0.9808	0.969	0.9167	0.9677	31

Table A2. *Cont.*

Model	**Dataset**	**Repetition**	**Accuracy**	**Precision macro_avg**	**Precision weighted_avg**	**Recall macro_avg**	**Recall weighted_avg**	**F1-Score macro_avg**	**F1-Score weighted_avg**	**Subjects**
OViTAD_AD_HC_MCI	Val	1	1	1	1	1	1	1	1	27
		2	0.963	0.9691	0.9632	0.9667	0.9667	0.9744	0.963	27
		3	1	1	1	1	1	1	1	27
	Test	1	0.9677	0.9726	0.9675	0.9778	0.9699	0.9697	0.9677	31
		2	0.9677	0.9726	0.9675	0.9778	0.9699	0.9697	0.9677	31
		3	0.9677	0.9582	0.9668	0.9778	0.9699	0.9444	0.9677	31
OViTAD_AD_HC	Val	1	1	1	1	1	1	1	1	14
		2	1	1	1	1	1	1	1	14
		3	1	1	1	1	1	1	1	14
	Test	1	1	1	1	1	1	1	1	17
		2	1	1	1	1	1	1	1	17
		3	0.9412	0.9583	0.9461	0.9167	0.9412	0.9328	0.9398	17
OViTAD_HC_MCI	Val	1	1	1	1	1	1	1	1	22
		2	1	1	1	1	1	1	1	22
		3	1	1	1	1	1	1	1	22
	Test	1	0.96	0.9667	0.9627	0.9545	0.96	0.9589	0.9597	25
		2	0.96	0.9667	0.9627	0.9545	0.96	0.9589	0.9597	25
		3	1	1	1	1	1	1	1	25

Table A3. The slice-level performance of structural MRI models using the preprocessed data with spatial smoothing sigma = 3 mm (S3). The naming convention for models and classes is as in the previous tables.

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
CaIT_ADMCI-HC_S3	Val	1	0.9317	0.7997	0.9143	0.5188	0.9317	0.5191	0.9025	11040
		2	0.9316	0.7603	0.9098	0.5272	0.9316	0.5345	0.9046	11027
		3	0.931	0.7762	0.9105	0.5191	0.931	0.5197	0.9018	11040
	Test	1	0.921	0.797	0.9021	0.5243	0.921	0.5265	0.8886	11413
		2	0.9198	0.7524	0.8941	0.5177	0.9198	0.5145	0.8861	11427
		3	0.9213	0.797	0.9025	0.5253	0.9213	0.5284	0.8892	11455
CaIT_AD-HCMCI_S3	Val	1	0.6912	0.6796	0.6854	0.657	0.6912	0.6602	0.681	11040
		2	0.6753	0.6604	0.668	0.6412	0.6753	0.6436	0.6651	11027
		3	0.6645	0.6482	0.6559	0.6262	0.6645	0.6274	0.6513	11040
	Test	1	0.6687	0.6526	0.6609	0.6349	0.6687	0.637	0.6586	11413
		2	0.6619	0.6443	0.6532	0.6264	0.6619	0.6279	0.6507	11427
		3	0.672	0.6563	0.6644	0.6378	0.672	0.64	0.6618	11455
CaIT_AD-HC-MCI_S3	Val	1	0.6757	0.6941	0.6764	0.4826	0.6757	0.479	0.6498	11040
		2	0.6637	0.6583	0.6595	0.4656	0.6637	0.4533	0.6354	11027
		3	0.6591	0.616	0.6478	0.4832	0.6591	0.4919	0.6337	11040
	Test	1	0.6496	0.6409	0.6442	0.4701	0.6496	0.464	0.6208	11413
		2	0.6331	0.6646	0.637	0.448	0.6331	0.4311	0.6004	11427
		3	0.6675	0.6157	0.6538	0.4972	0.6675	0.5038	0.6419	11455
DeepViT_ADMCI_HC_S3	Val	1	0.9898	0.958	0.9894	0.9883	0.9897	0.9318	0.9898	11027
		2	0.9901	0.9604	0.9899	0.9796	0.99	0.943	0.9901	11040
		3	0.9829	0.9326	0.9827	0.9444	0.9825	0.9215	0.9829	11040
	Test	1	0.9891	0.9618	0.9889	0.9859	0.9891	0.9405	0.9891	11427
		2	0.9892	0.9618	0.9889	0.9897	0.9892	0.9375	0.9892	11413
		3	0.9869	0.9546	0.9867	0.9691	0.9867	0.9412	0.9869	11455
DeepViT_AD_HCMCI_S3	Val	1	0.9452	0.9421	0.9448	0.9494	0.9462	0.9367	0.9452	11027
		2	0.9505	0.948	0.9503	0.9519	0.9507	0.9448	0.9505	11040
		3	0.9186	0.9137	0.9178	0.9225	0.9196	0.9076	0.9186	11040
	Test	1	0.9404	0.9369	0.9399	0.9454	0.9416	0.9308	0.9404	11427
		2	0.9458	0.943	0.9455	0.9473	0.946	0.9394	0.9458	11413
		3	0.9245	0.9195	0.9236	0.9323	0.927	0.9114	0.9245	11455

Table A3. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
DeepViT_AD_HC_MCI_S3	Val	1	0.9255	0.9152	0.9248	0.9398	0.9284	0.8959	0.9255	11040
		2	0.9176	0.9028	0.917	0.9235	0.9183	0.8855	0.9176	11027
		3	0.9149	0.9058	0.9147	0.9292	0.9155	0.8865	0.9149	11040
	Test	1	0.9202	0.9079	0.9193	0.9388	0.9241	0.8849	0.9202	11413
		2	0.9103	0.9014	0.9099	0.9096	0.911	0.8945	0.9103	11427
		3	0.9208	0.9156	0.9206	0.9364	0.9215	0.8982	0.9208	11455
ViT44_8_ADMCI_HC_S3	Val	1	0.9912	0.965	0.9911	0.9797	0.9911	0.9512	0.9912	11027
		2	0.9913	0.9653	0.9912	0.9817	0.9912	0.9502	0.9913	11040
		3	0.9851	0.9402	0.9847	0.9603	0.9847	0.9221	0.9851	11040
	Test	1	0.9911	0.9693	0.991	0.9821	0.991	0.9573	0.9911	11427
		2	0.9892	0.9624	0.989	0.9827	0.9891	0.9439	0.9892	11413
		3	0.9846	0.9465	0.9844	0.9637	0.9843	0.9306	0.9846	11455
ViT44_8_AD_HCMCI_S3	Val	1	0.9607	0.9587	0.9606	0.963	0.9611	0.9552	0.9607	11027
		2	0.9636	0.9617	0.9634	0.967	0.9642	0.9574	0.9636	11040
		3	0.9372	0.9341	0.937	0.9376	0.9373	0.9311	0.9372	11040
	Test	1	0.9616	0.9595	0.9614	0.9653	0.9623	0.9549	0.9616	11427
		2	0.962	0.9599	0.9618	0.9658	0.9627	0.9553	0.962	11413
		3	0.9362	0.9327	0.9358	0.9383	0.9366	0.9284	0.9362	11455
ViT44_8_AD_HC_MCI_S3	Val	1	0.9359	0.93	0.9358	0.9425	0.9361	0.9189	0.9359	11040
		2	0.9279	0.9158	0.9277	0.9232	0.9279	0.9088	0.9279	11027
		3	0.9287	0.9209	0.9283	0.928	0.93	0.9155	0.9287	11040
	Test	1	0.9339	0.9243	0.9337	0.9407	0.9342	0.9101	0.9339	11413
		2	0.9228	0.9179	0.9227	0.9175	0.9228	0.9185	0.9228	11427
		3	0.9277	0.9219	0.9272	0.9291	0.9295	0.917	0.9277	11455
ViT_vanilla_ADMCI_HC_S3	Val	1	0.9916	0.9658	0.9914	0.9901	0.9915	0.9442	0.9916	11027
		2	0.9915	0.9665	0.9914	0.9765	0.9914	0.9569	0.9915	11040
		3	0.986	0.9433	0.9856	0.9688	0.9857	0.9208	0.986	11040
	Test	1	0.9898	0.9643	0.9896	0.989	0.9898	0.9424	0.9898	11427
		2	0.9892	0.9621	0.989	0.9864	0.9892	0.9405	0.9892	11413
		3	0.9844	0.9455	0.9841	0.9635	0.9841	0.929	0.9844	11455

Table A3. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
ViT_vanilla_AD_HCMCI_S3	Val	1	0.9598	0.9577	0.9596	0.9632	0.9604	0.9533	0.9598	11027
		2	0.9566	0.9541	0.9563	0.9619	0.9578	0.9484	0.9566	11040
		3	0.9262	0.9225	0.9259	0.9257	0.9261	0.9198	0.9262	11040
	Test	1	0.9585	0.9562	0.9582	0.9634	0.9596	0.9507	0.9585	11427
		2	0.9545	0.952	0.9542	0.9596	0.9557	0.9463	0.9545	11413
		3	0.9319	0.928	0.9314	0.9354	0.9327	0.9226	0.9319	11455
ViT_vanilla_AD_HC_MCI_S3	Val	1	0.9439	0.9408	0.9437	0.9497	0.9443	0.9329	0.9439	11040
		2	0.924	0.9109	0.9237	0.9274	0.9241	0.8966	0.924	11027
		3	0.9267	0.911	0.9264	0.9204	0.9268	0.9024	0.9267	11040
	Test	1	0.9385	0.932	0.9382	0.948	0.9393	0.9182	0.9385	11413
		2	0.9214	0.9183	0.9213	0.9267	0.9215	0.9106	0.9214	11427
		3	0.9321	0.9211	0.9316	0.9323	0.9331	0.9118	0.9321	11455
OViTAD_ADMCI_HC_S3	Val	1	0.9886	0.9529	0.9882	0.9862	0.9885	0.9245	0.9886	11027
		2	0.9901	0.9601	0.9899	0.9839	0.99	0.9388	0.9901	11040
		3	0.9828	0.9317	0.9825	0.9474	0.9824	0.9173	0.9828	11040
	Test	1	0.9881	0.9576	0.9877	0.9885	0.9881	0.9311	0.9881	11427
		2	0.9848	0.9462	0.9844	0.9748	0.9846	0.9214	0.9848	11413
		3	0.9791	0.9287	0.979	0.936	0.9788	0.9218	0.9791	11455
OViTAD_AD_HCMCI_S3	Val	1	0.9457	0.9429	0.9455	0.9465	0.9458	0.9399	0.9457	11027
		2	0.9455	0.9427	0.9452	0.9467	0.9457	0.9394	0.9455	11040
		3	0.9167	0.9114	0.9158	0.9221	0.9182	0.9044	0.9167	11040
	Test	1	0.9449	0.9419	0.9446	0.9474	0.9454	0.9375	0.9449	11427
		2	0.9403	0.9371	0.94	0.9428	0.9408	0.9327	0.9403	11413
		3	0.9176	0.9124	0.9167	0.9232	0.9192	0.9053	0.9176	11455
OViTAD_AD_HC_MCI_S3	Val	1	0.931	0.9243	0.9306	0.9385	0.9322	0.9123	0.931	11040
		2	0.9116	0.8931	0.911	0.9078	0.9126	0.8808	0.9116	11027
		3	0.911	0.8936	0.9102	0.9207	0.9126	0.8721	0.911	11040
	Test	1	0.9113	0.8997	0.9107	0.9182	0.9129	0.8847	0.9113	11413
		2	0.9047	0.8983	0.9042	0.9112	0.9058	0.8876	0.9047	11427
		3	0.9109	0.8965	0.9099	0.9219	0.9141	0.8774	0.9109	11455

Table A3. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
OViTAD_AD_HC_S3	Val	1	0.9673	0.9636	0.9671	0.9046	0.9673	0.9311	0.9662	5173
		2	0.9628	0.9456	0.962	0.9034	0.9628	0.9229	0.9619	5164
		3	0.9656	0.9562	0.9651	0.9052	0.9656	0.9285	0.9646	5173
	Test	1	0.9575	0.9607	0.9578	0.8854	0.9575	0.9177	0.9556	5482
		2	0.9653	0.9505	0.9648	0.9237	0.9653	0.9365	0.9648	5480
		3	0.9534	0.9417	0.9526	0.8873	0.9534	0.9116	0.9519	5493
OViTAD_HC_MCI_S3	Val	1	0.9396	0.8412	0.9445	0.8867	0.9396	0.8619	0.9415	6636
		2	0.9183	0.7919	0.9349	0.8841	0.9183	0.8282	0.9238	6630
		3	0.9264	0.8104	0.9362	0.877	0.9264	0.8388	0.9299	6641
	Test	1	0.9344	0.8572	0.9351	0.865	0.9344	0.8611	0.9347	6857
		2	0.9136	0.8045	0.9302	0.8904	0.9136	0.8384	0.9189	6874
		3	0.9144	0.808	0.9232	0.8603	0.9144	0.8308	0.9177	6889

Table A4. The slice-level performance of structural MRI models using the preprocessed data with spatial smoothing sigma = 4 mm (S4). The naming convention for models and classes is as in the previous tables.

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
CaIT_ADMCI-HC_S4	Val	1	0.933	0.859	0.9233	0.5255	0.933	0.5314	0.9048	11040
		2	0.9307	0.724	0.9058	0.5363	0.9307	0.5498	0.9063	11027
		3	0.9329	0.885	0.9265	0.5261	0.9329	0.5324	0.9045	11040
	Test	1	0.9221	0.835	0.9089	0.5293	0.9221	0.5356	0.8905	11413
		2	0.9184	0.6977	0.8862	0.5253	0.9184	0.5288	0.8877	11427
		3	0.9214	0.8152	0.9053	0.5239	0.9214	0.5258	0.8888	11455
CaIT_AD-HCMCI_S4	Val	1	0.7449	0.7367	0.7419	0.7218	0.7449	0.7264	0.7408	11040
		2	0.7387	0.7282	0.7359	0.7196	0.7387	0.7227	0.7362	11027
		3	0.7339	0.7256	0.7305	0.7076	0.7339	0.7124	0.7284	11040
	Test	1	0.7293	0.7191	0.7258	0.7068	0.7293	0.7106	0.7255	11413
		2	0.7142	0.7014	0.7113	0.6955	0.7142	0.6978	0.7121	11427
		3	0.7418	0.733	0.7386	0.7186	0.7418	0.7231	0.7377	11455

Table A4. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
CaIT_AD-HC-MCI_S4	Val	1	0.7468	0.769	0.7496	0.5482	0.7468	0.5548	0.7244	11040
		2	0.7386	0.703	0.7301	0.5375	0.7386	0.5408	0.715	11027
		3	0.7229	0.6716	0.7135	0.5308	0.7229	0.5406	0.6981	11040
	Test	1	0.7188	0.6915	0.711	0.5355	0.7188	0.5397	0.6946	11413
		2	0.7039	0.6711	0.6944	0.5134	0.7039	0.5084	0.6761	11427
		3	0.7308	0.6726	0.7184	0.5474	0.7308	0.5564	0.7056	11455
DeepViT_ADMCI_HC_S4	Val	1	0.9892	0.9564	0.989	0.9788	0.9891	0.9363	0.9892	11027
		2	0.9903	0.9611	0.9901	0.9809	0.9902	0.9431	0.9903	11040
		3	0.9834	0.9327	0.9829	0.9603	0.983	0.9087	0.9834	11040
	Test	1	0.9891	0.9617	0.9888	0.9837	0.989	0.9419	0.9891	11427
		2	0.9871	0.9547	0.9868	0.9786	0.987	0.9334	0.9871	11413
		3	0.985	0.9468	0.9846	0.9729	0.9847	0.924	0.985	11455
DeepViT_AD_HCMCI_S4	Val	1	0.9569	0.9545	0.9566	0.9612	0.9578	0.9494	0.9569	11027
		2	0.9534	0.9508	0.9531	0.9587	0.9547	0.945	0.9534	11040
		3	0.9383	0.935	0.938	0.9404	0.9387	0.9308	0.9383	11040
	Test	1	0.9491	0.9462	0.9487	0.9533	0.95	0.9408	0.9491	11427
		2	0.951	0.9481	0.9506	0.9571	0.9526	0.9418	0.951	11413
		3	0.938	0.9344	0.9375	0.9422	0.939	0.9288	0.938	11455
DeepViT_AD_HC_MCI_S4	Val	1	0.9407	0.9314	0.9405	0.936	0.9406	0.9269	0.9407	11040
		2	0.9348	0.9168	0.9343	0.9401	0.9351	0.8974	0.9348	11027
		3	0.9313	0.9119	0.9309	0.9251	0.9312	0.9001	0.9313	11040
	Test	1	0.9359	0.9217	0.9355	0.9354	0.936	0.9098	0.9359	11413
		2	0.9248	0.915	0.9245	0.9281	0.9252	0.9036	0.9248	11427
		3	0.9351	0.9242	0.9348	0.9409	0.9355	0.9098	0.9351	11455
ViT44_8_ADMCI_HC_S4	Val	1	0.9918	0.9673	0.9917	0.9851	0.9918	0.951	0.9918	11027
		2	0.9933	0.9734	0.9932	0.988	0.9932	0.9597	0.9933	11040
		3	0.985	0.9384	0.9844	0.971	0.9847	0.9107	0.985	11040
	Test	1	0.9935	0.9778	0.9934	0.9897	0.9935	0.9665	0.9935	11427
		2	0.9912	0.969	0.9909	0.9919	0.9912	0.9484	0.9912	11413
		3	0.9846	0.945	0.9841	0.9768	0.9844	0.9179	0.9846	11455

Table A4. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
ViT44_8_AD_HCMCI_S4	Val	1	0.968	0.9663	0.9678	0.9718	0.9687	0.9619	0.968	11027
		2	0.9682	0.9667	0.9681	0.9692	0.9683	0.9644	0.9682	11040
		3	0.95	0.9475	0.9498	0.9508	0.9501	0.9447	0.95	11040
	Test	1	0.9667	0.9648	0.9664	0.9715	0.9677	0.9596	0.9667	11427
		2	0.9671	0.9654	0.967	0.9686	0.9672	0.9627	0.9671	11413
		3	0.9449	0.9421	0.9447	0.9463	0.9451	0.9386	0.9449	11455
ViT44_8_AD_HC_MCI_S4	Val	1	0.9511	0.9451	0.9509	0.9592	0.9517	0.9326	0.9511	11040
		2	0.9468	0.9347	0.9466	0.9454	0.9468	0.9249	0.9468	11027
		3	0.9462	0.9348	0.9461	0.9407	0.9462	0.9292	0.9462	11040
	Test	1	0.9453	0.9331	0.945	0.9552	0.9462	0.9148	0.9453	11413
		2	0.9354	0.9298	0.9352	0.9313	0.9357	0.9289	0.9354	11427
		3	0.9497	0.94	0.9495	0.9473	0.9498	0.9332	0.9497	11455
ViT_vanilla_ADMCI_HC_S4	Val	1	0.9922	0.9685	0.992	0.9911	0.9922	0.9482	0.9922	11027
		2	0.9943	0.9775	0.9942	0.9887	0.9943	0.9669	0.9943	11040
		3	0.988	0.9526	0.9878	0.965	0.9877	0.941	0.988	11040
	Test	1	0.9919	0.9721	0.9918	0.9887	0.9919	0.9568	0.9919	11427
		2	0.989	0.9618	0.9888	0.977	0.9888	0.9477	0.989	11413
		3	0.9858	0.9512	0.9856	0.9608	0.9856	0.9421	0.9858	11455
ViT_vanilla_AD_HCMCI_S4	Val	1	0.9689	0.9675	0.9689	0.9684	0.9689	0.9667	0.9689	11027
		2	0.9621	0.9602	0.962	0.9642	0.9624	0.9569	0.9621	11040
		3	0.9441	0.9411	0.9438	0.9464	0.9445	0.937	0.9441	11040
	Test	1	0.9671	0.9655	0.967	0.9678	0.9672	0.9635	0.9671	11427
		2	0.9609	0.9589	0.9607	0.9639	0.9614	0.9548	0.9609	11413
		3	0.9423	0.9392	0.942	0.9449	0.9428	0.9347	0.9423	11455
ViT_vanilla_AD_HC_MCI_S4	Val	1	0.9472	0.9266	0.947	0.9326	0.947	0.9209	0.9472	11040
		2	0.943	0.9269	0.9427	0.9401	0.9431	0.9151	0.943	11027
		3	0.9366	0.9256	0.9365	0.9265	0.9368	0.9251	0.9366	11040
	Test	1	0.9452	0.9298	0.9449	0.9411	0.945	0.9196	0.9452	11413
		2	0.9397	0.9368	0.9395	0.9467	0.9404	0.9282	0.9397	11427
		3	0.9404	0.9244	0.9402	0.9282	0.9404	0.921	0.9404	11455

Table A4. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Slices
OViTAD_ADMCI_HC_S4	Val	1	0.9901	0.9601	0.9899	0.982	0.99	0.9404	0.9901	11027
		2	0.9889	0.9547	0.9886	0.9811	0.9887	0.9315	0.9889	11040
		3	0.983	0.9299	0.9823	0.9647	0.9825	0.9007	0.983	11040
	Test	1	0.989	0.9613	0.9887	0.9847	0.9889	0.9404	0.989	11427
		2	0.9868	0.9526	0.9863	0.9865	0.9868	0.9239	0.9868	11413
		3	0.9793	0.9256	0.9786	0.9581	0.9787	0.8982	0.9793	11455
OViTAD_AD_HCMCI_S4	Val	1	0.956	0.9539	0.9559	0.9568	0.9561	0.9513	0.956	11027
		2	0.9499	0.9471	0.9495	0.9547	0.951	0.9414	0.9499	11040
		3	0.9355	0.9324	0.9353	0.9349	0.9354	0.9301	0.9355	11040
	Test	1	0.9544	0.952	0.9542	0.9566	0.9548	0.9483	0.9544	11427
		2	0.9495	0.9465	0.9491	0.9559	0.9512	0.94	0.9495	11413
		3	0.939	0.9359	0.9388	0.9392	0.939	0.9332	0.939	11455
OViTAD_AD_HC_MCI_S4	Val	1	0.9403	0.9311	0.9401	0.943	0.9404	0.9203	0.9403	11040
		2	0.9343	0.9163	0.9339	0.9381	0.9346	0.898	0.9343	11027
		3	0.9287	0.9142	0.9282	0.9314	0.9299	0.8999	0.9287	11040
	Test	1	0.9288	0.9132	0.9284	0.9288	0.9288	0.8996	0.9288	11413
		2	0.9209	0.9099	0.9205	0.925	0.9214	0.8969	0.9209	11427
		3	0.9338	0.9167	0.9332	0.9329	0.9349	0.9032	0.9338	11455
OViTAD_AD_HC_S4	Val	1	0.9735	0.9653	0.9732	0.9281	0.9735	0.9455	0.973	5173
		2	0.9562	0.941	0.9553	0.8801	0.9562	0.9072	0.9546	5164
		3	0.9668	0.9513	0.9661	0.915	0.9668	0.932	0.9661	5173
	Test	1	0.9668	0.9637	0.9666	0.9159	0.9668	0.9377	0.9659	5482
		2	0.9578	0.9388	0.957	0.9076	0.9578	0.9223	0.9571	5480
		3	0.955	0.9283	0.9543	0.9085	0.955	0.918	0.9545	5493
OViTAD_HC_MCI_S4	Val	1	0.926	0.807	0.9403	0.8966	0.926	0.843	0.9307	6636
		2	0.9075	0.7718	0.9236	0.8503	0.9075	0.8033	0.9134	6630
		3	0.9368	0.8338	0.9433	0.8891	0.9368	0.8583	0.9392	6641
	Test	1	0.9164	0.8103	0.9294	0.8833	0.9164	0.8405	0.9208	6857
		2	0.9082	0.7955	0.925	0.8773	0.9082	0.8279	0.9138	6874
		3	0.9273	0.8318	0.9351	0.8874	0.9273	0.8562	0.9301	6889

Table A5. The performance of structural MRI models at subject-level for preprocessed data using spatial smoothing with sigma = 3 mm (S3). The naming convention for models and classes is as in the previous tables.

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
CaIT_ADMCI-HC_S3	Val	1	0.9306	0.4653	0.8659	0.5	0.9306	0.482	0.8971	144
		2	0.9306	0.4653	0.8659	0.5	0.9306	0.482	0.8971	144
		3	0.9306	0.4653	0.8659	0.5	0.9306	0.482	0.8971	144
	Test	1	0.9195	0.4597	0.8454	0.5	0.9195	0.479	0.8809	149
		2	0.9195	0.4597	0.8454	0.5	0.9195	0.479	0.8809	149
		3	0.9195	0.4597	0.8454	0.5	0.9195	0.479	0.8809	149
CaIT_AD-HCMCI_S3	Val	1	0.7431	0.7416	0.7423	0.7087	0.7431	0.7152	0.7338	144
		2	0.7083	0.715	0.7124	0.6588	0.7083	0.6606	0.6871	144
		3	0.6944	0.7055	0.7017	0.6382	0.6944	0.6353	0.6659	144
	Test	1	0.7047	0.704	0.7043	0.6592	0.7047	0.6619	0.6869	149
		2	0.6913	0.6893	0.6901	0.6423	0.6913	0.6426	0.67	149
		3	0.7181	0.7267	0.7233	0.6703	0.7181	0.6737	0.6987	149
CaIT_AD-HC-MCI_S3	Val	1	0.8194	0.5425	0.7603	0.5746	0.8194	0.5561	0.7861	144
		2	0.7986	0.5341	0.7445	0.5556	0.7986	0.5386	0.7625	144
		3	0.7847	0.5228	0.7299	0.5439	0.7847	0.5263	0.7473	144
	Test	1	0.7852	0.5202	0.7195	0.5537	0.7852	0.5318	0.7448	149
		2	0.745	0.4914	0.6807	0.5225	0.745	0.5007	0.7037	149
		3	0.7919	0.5245	0.7255	0.5593	0.7919	0.5374	0.752	149
DeepViT_ADMCI_HC_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149
DeepViT_AD_HCMCI_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9927	0.993	0.9943	0.9931	0.9912	0.9931	144
	Test	1	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149

Table A5. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
DeepViT_AD_HC_MCI_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	0.9933	0.995	0.9933	0.9958	0.9934	0.9944	0.9933	149
		2	0.9866	0.9901	0.9866	0.9901	0.9866	0.9901	0.9866	149
		3	1	1	1	1	1	1	1	149
ViT44_8_ADMCI_HC_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149
ViT44_8_AD_HCMCI_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9927	0.993	0.9943	0.9931	0.9912	0.9931	144
	Test	1	1	1	1	1	1	1	1	149
		2	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		3	1	1	1	1	1	1	1	149
ViT44_8_AD_HC_MCI_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	0.9933	0.995	0.9933	0.9944	0.9934	0.9957	0.9933	149
		3	1	1	1	1	1	1	1	149
ViT_vanilla_ADMCI_HC_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149

Table A5. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
ViT_vanilla_AD_HCMCI_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9927	0.993	0.9943	0.9931	0.9912	0.9931	144
	Test	1	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		2	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		3	1	1	1	1	1	1	1	149
ViT_vanilla_AD_HC_MCI_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	0.9933	0.995	0.9933	0.9944	0.9934	0.9957	0.9933	149
		3	1	1	1	1	1	1	1	149
OViTAD_ADMCI_HC_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149
OViTAD_AD_HCMCI_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9927	0.993	0.9943	0.9931	0.9912	0.9931	144
	Test	1	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		2	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		3	1	1	1	1	1	1	1	149
OViTAD_AD_HC_MCI_S3	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9949	0.993	0.9957	0.9931	0.9942	0.9931	144
	Test	1	0.9933	0.995	0.9933	0.9958	0.9934	0.9944	0.9933	149
		2	0.9866	0.9901	0.9866	0.9901	0.9866	0.9901	0.9866	149
		3	1	1	1	1	1	1	1	149

Table A5. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
OViTAD_AD_HC_S3	Val	1	1	1	1	1	1	1	1	67
		2	1	1	1	1	1	1	1	67
		3	1	1	1	1	1	1	1	67
	Test	1	1	1	1	1	1	1	1	71
		2	1	1	1	1	1	1	1	71
		3	1	1	1	1	1	1	1	71
OViTAD_HC_MCI_S3	Val	1	1	1	1	1	1	1	1	87
		2	1	1	1	1	1	1	1	87
		3	1	1	1	1	1	1	1	87
	Test	1	1	1	1	1	1	1	1	90
		2	1	1	1	1	1	1	1	90
		3	1	1	1	1	1	1	1	90

Table A6. The performance of structural MRI models at subject-level for preprocessed data using spatial smoothing with sigma = 4 mm (S4). The naming convention for models and classes is as in the previous tables.

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
CaIT_S4_ADMCI-HC	Val	1	0.9306	0.4653	0.8659	0.5	0.9306	0.482	0.8971	144
		2	0.9306	0.4653	0.8659	0.5	0.9306	0.482	0.8971	144
		3	0.9306	0.4653	0.8659	0.5	0.9306	0.482	0.8971	144
	Test	1	0.9195	0.4597	0.8454	0.5	0.9195	0.479	0.8809	149
		2	0.9195	0.4597	0.8454	0.5	0.9195	0.479	0.8809	149
		3	0.9195	0.4597	0.8454	0.5	0.9195	0.479	0.8809	149
CaIT_S4_AD-HCMCI	Val	1	0.8681	0.8607	0.8699	0.8666	0.8681	0.8632	0.8686	144
		2	0.8958	0.8889	0.8993	0.8987	0.8958	0.8926	0.8965	144
		3	0.8542	0.8482	0.8538	0.846	0.8542	0.8471	0.8539	144
	Test	1	0.8523	0.8448	0.8535	0.8486	0.8523	0.8465	0.8527	149
		2	0.8389	0.8307	0.8418	0.8375	0.8389	0.8335	0.8397	149
		3	0.8792	0.872	0.8838	0.8825	0.8792	0.8757	0.88	149

Table A6. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
CaIT_S4_AD-HC-MCI	Val	1	0.9097	0.6021	0.8487	0.6491	0.9097	0.6245	0.8778	144
		2	0.9167	0.6069	0.8561	0.655	0.9167	0.6296	0.8848	144
		3	0.8681	0.5743	0.8063	0.614	0.8681	0.5932	0.8356	144
	Test	1	0.8926	0.5907	0.823	0.6441	0.8926	0.616	0.8561	149
		2	0.8725	0.5769	0.8036	0.6285	0.8725	0.6015	0.8365	149
		3	0.8658	0.5722	0.7953	0.6215	0.8658	0.5958	0.829	149
DeepViT_ADMCI_HC_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149
DeepViT_AD_HCMCI_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9927	0.993	0.9943	0.9931	0.9912	0.9931	144
	Test	1	1	1	1	1	1	1	1	149
		2	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		3	1	1	1	1	1	1	1	149
DeepViT_AD_HC_MCI_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9949	0.993	0.9957	0.9931	0.9942	0.9931	144
	Test	1	1	1	1	1	1	1	1	149
		2	0.9866	0.9901	0.9866	0.9901	0.9866	0.9901	0.9866	149
		3	1	1	1	1	1	1	1	149
ViT44_8_ADMCI_HC_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149

Table A6. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
ViT44_8_AD_HCMCI_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149
ViT44_8_AD_HC_MCI_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9949	0.993	0.9957	0.9931	0.9942	0.9931	144
	Test	1	0.9933	0.995	0.9933	0.9958	0.9934	0.9944	0.9933	149
		2	0.9866	0.9901	0.9866	0.9901	0.9866	0.9901	0.9866	149
		3	1	1	1	1	1	1	1	149
ViT_vanilla_ADMCI_HC_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149
ViT_vanilla_AD_HCMCI_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9927	0.993	0.9943	0.9931	0.9912	0.9931	144
	Test	1	1	1	1	1	1	1	1	149
		2	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		3	1	1	1	1	1	1	1	149
ViT_vanilla_AD_HC_MCI_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9949	0.993	0.9957	0.9931	0.9942	0.9931	144
	Test	1	0.9933	0.995	0.9933	0.9958	0.9934	0.9944	0.9933	149
		2	0.9933	0.995	0.9933	0.9958	0.9934	0.9944	0.9933	149
		3	1	1	1	1	1	1	1	149

Table A6. *Cont.*

Model	Dataset	Repetition	Accuracy	Precision macro_avg	Precision weighted_avg	Recall macro_avg	Recall weighted_avg	F1-Score macro_avg	F1-Score weighted_avg	Subjects
OViTAD_ADMCI_HC_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	1	1	1	1	1	1	1	144
	Test	1	1	1	1	1	1	1	1	149
		2	1	1	1	1	1	1	1	149
		3	1	1	1	1	1	1	1	149
OViTAD_AD_HCMCI_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9927	0.993	0.9943	0.9931	0.9912	0.9931	144
	Test	1	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		2	0.9933	0.993	0.9933	0.9945	0.9934	0.9915	0.9933	149
		3	1	1	1	1	1	1	1	149
OViTAD_AD_HC_MCI_S4	Val	1	1	1	1	1	1	1	1	144
		2	1	1	1	1	1	1	1	144
		3	0.9931	0.9949	0.993	0.9957	0.9931	0.9942	0.9931	144
	Test	1	1	1	1	1	1	1	1	149
		2	0.9866	0.9901	0.9866	0.9901	0.9866	0.9901	0.9866	149
		3	1	1	1	1	1	1	1	149
OViTAD_AD_HC_S4	Val	1	1	1	1	1	1	1	1	67
		2	1	1	1	1	1	1	1	67
		3	1	1	1	1	1	1	1	67
	Test	1	1	1	1	1	1	1	1	71
		2	1	1	1	1	1	1	1	71
		3	1	1	1	1	1	1	1	71
OViTAD_HC_MCI_S4	Val	1	1	1	1	1	1	1	1	87
		2	1	1	1	1	1	1	1	87
		3	1	1	1	1	1	1	1	87
	Test	1	1	1	1	1	1	1	1	90
		2	1	1	1	1	1	1	1	90
		3	1	1	1	1	1	1	1	90

Table A7. The summary of highlights for each method.

Reference	Highlights
Lin et al. 2018 [100]	* MCIc vs MCInc 68.68% * FreeSurfer-based Features + 3-layer CNN
Dimitriadis et al. 2018 [101]	* Random Forest Feature Selection + SVM * Model interpretability
Kruthika et al. 2019 [102]	* FreeSurfer-based Features + Multistage Classifier * Further non-ML optimization (PSO) 96.31%
Spasov et al. 2019 [103]	* 3D Images + 3D CNN + Statistical Model * sMCI vs pMCI trained by AD, HC, MCI
Basaia et al. 2019 [104]	* ADNI + non-ADNI data * c-MCI vs s-MCI 75.1%
Abrol et al. 2020 [105]	* 3D Adopted ResNet * Standard 4-way AD, HC, sMCI, pMCI
Shao et al. 2020 [106]	* Hypergraph + Multi-task Feature Selection + SVM
Alinsaif et al. 2021 [107]	* HC + sMCI vs pMCI + AD dataset * 3D Shearlet technique + SVM
Alinsaif et al. 2021 [107]	* HC + sMCI vs pMCI + AD * MobileNet fine-tuned
Ramzan et al. 2019 [63]	* ResNet18 fine-tuned
Hojjati et al. 2018 [108]	* functional connectivity + cortical thickness * SVM
Cui et al. 2019 [109]	* 3D CNN features + RNN
Amoroso et al. 2018 [110]	* Random Forest Feature Selection + Deep Neural Network
Buvaneswari et al. 2021 [111]	* Hippocampal visual features * PCA-SVR
Duc et al. 2019 [61]	* 3D CNN + MMSE Regression
OViTAD - fMRI	* First Vision Transformer for Alzheimer's prediction using rs-fMRI * Aggressive fMRI preprocessing + 4D data decomposition to 2D * postprocessing to retrieve subject-level prediction
OViTAD - MRI (Sigma = 3,4)	* First Vision Transformer for Alzheimer's prediction using MRI * Aggressive fMRI preprocessing + 3D data decomposition to 2D * postprocessing to retrieve subject-level prediction

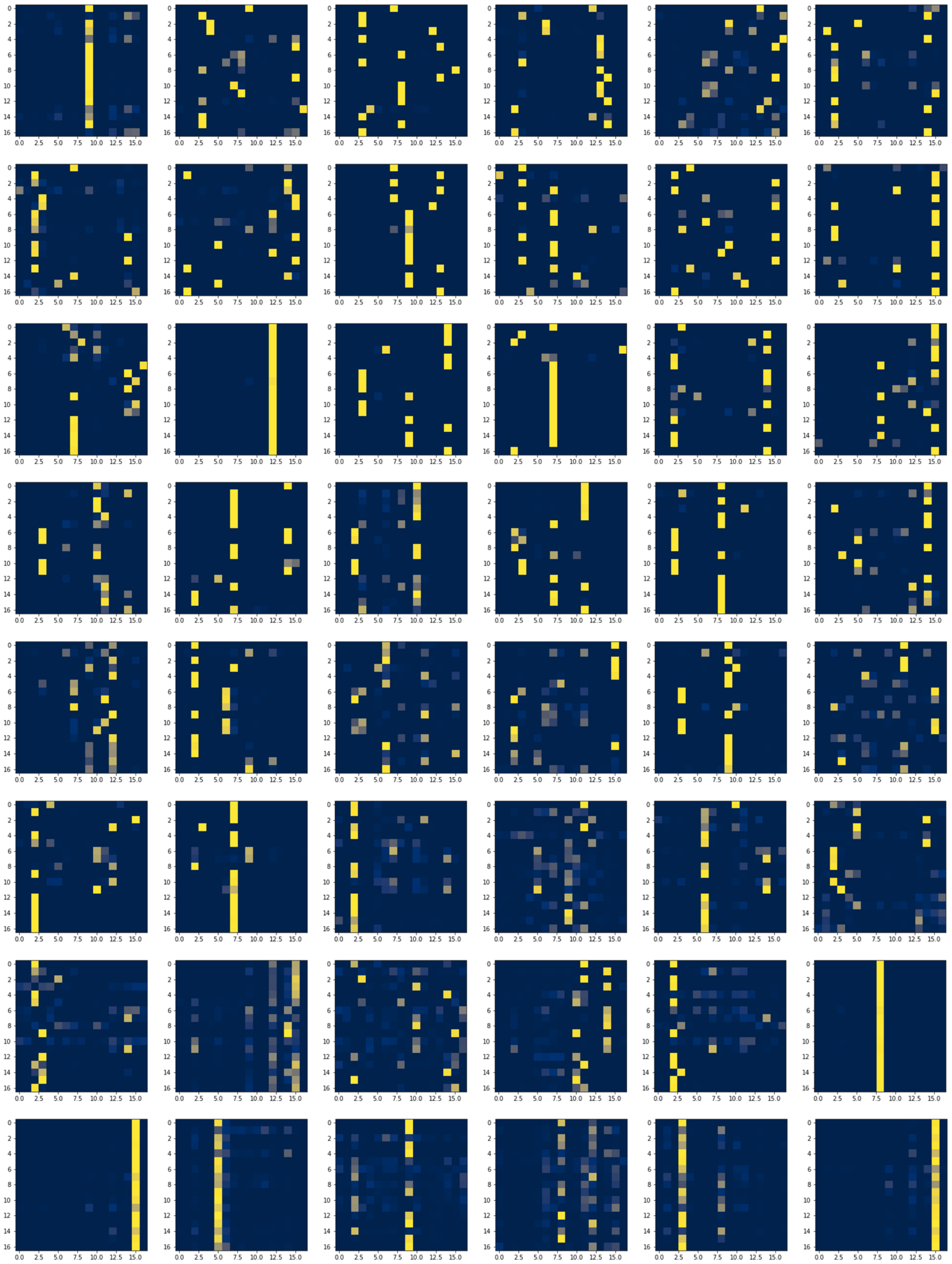


Figure A1. The attention maps for a random AD fMRI slice from the testing set in AD vs. HC vs. MCI in OViTAD with head = 8 and depth = 6, input dimension = 56.

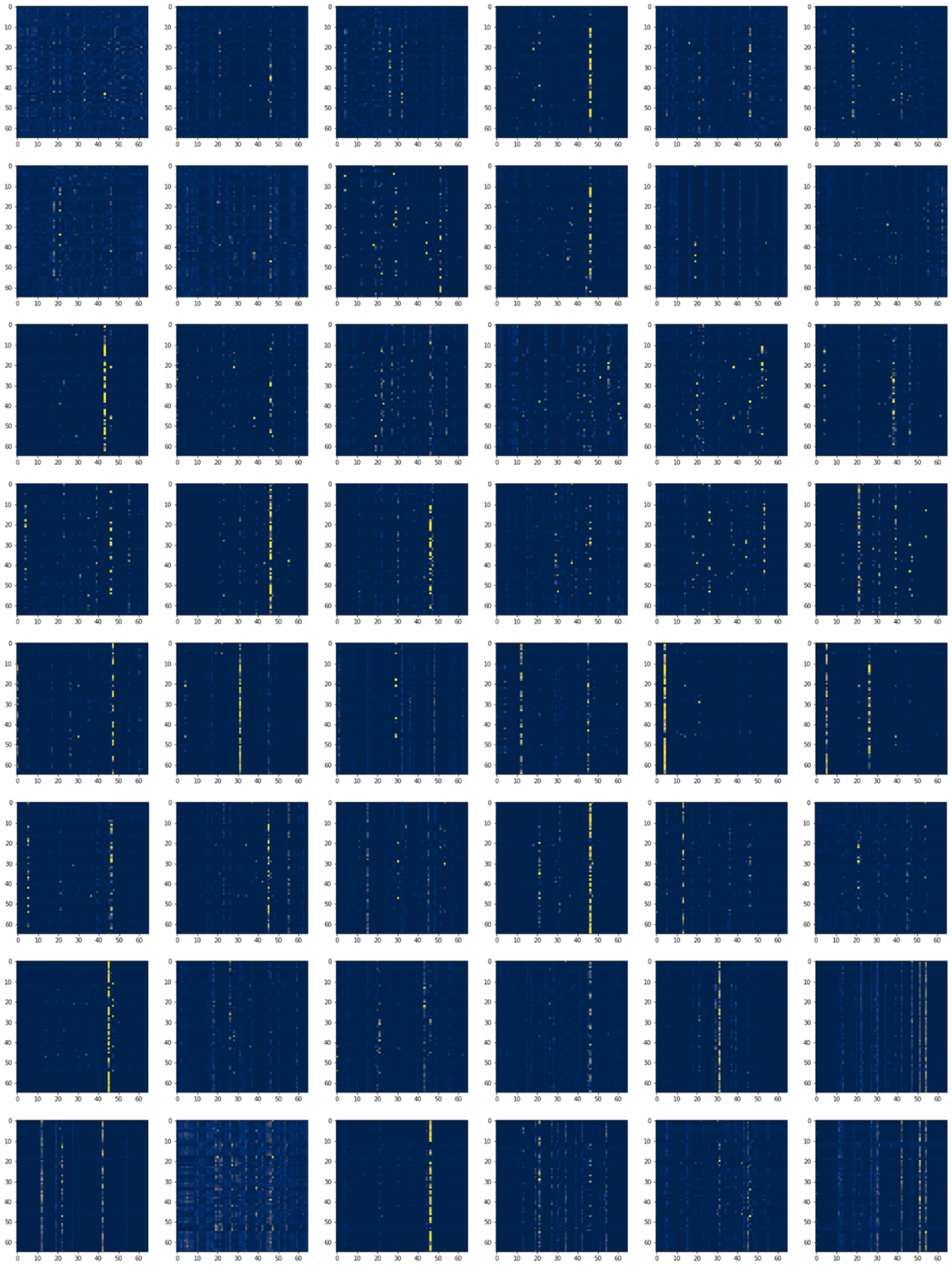


Figure A2. The attention maps for a random AD structural MRI slice from the testing set in AD vs. HC vs. MCI in OViTAD with head = 8 and depth = 6, input dimension = 112.

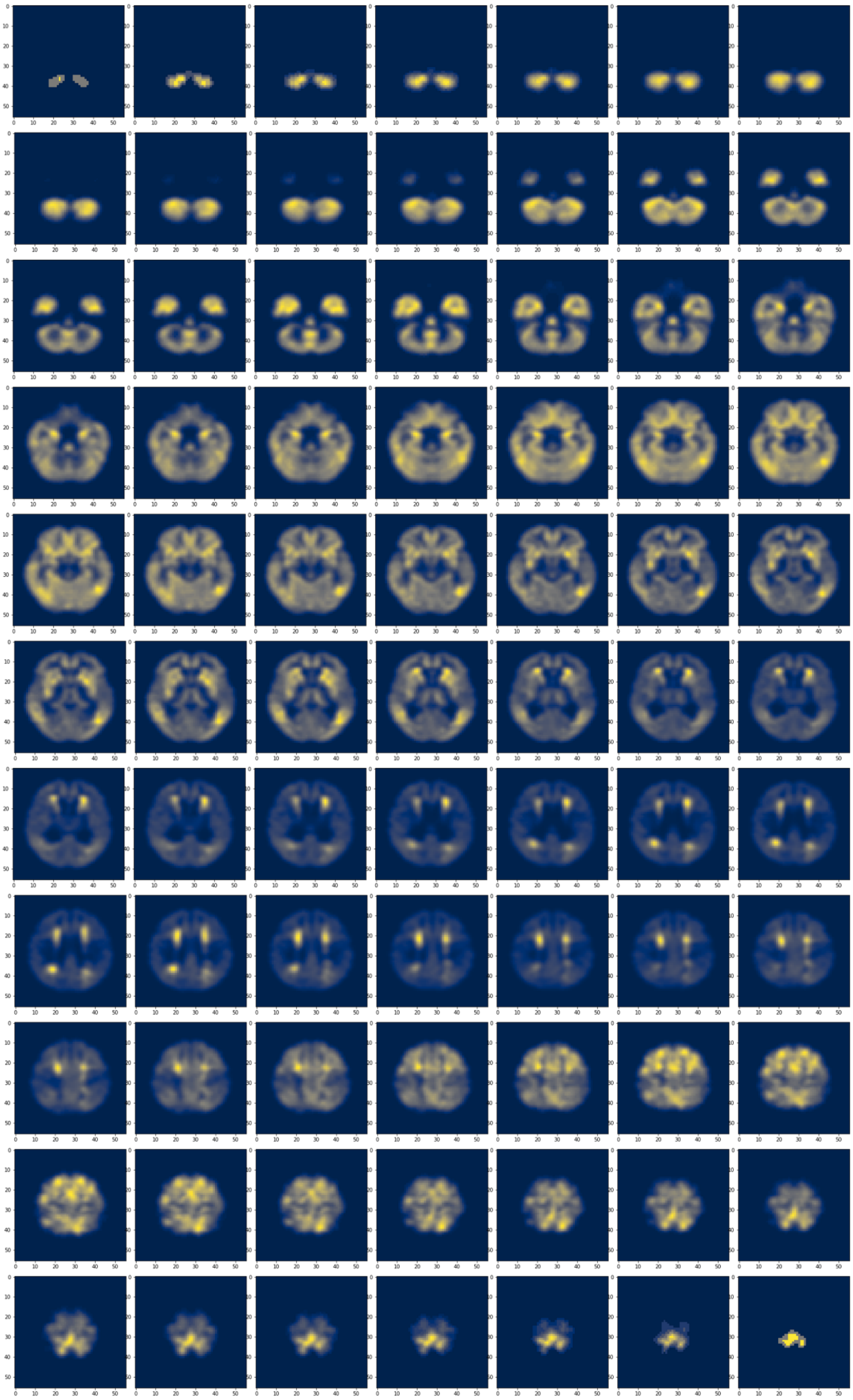
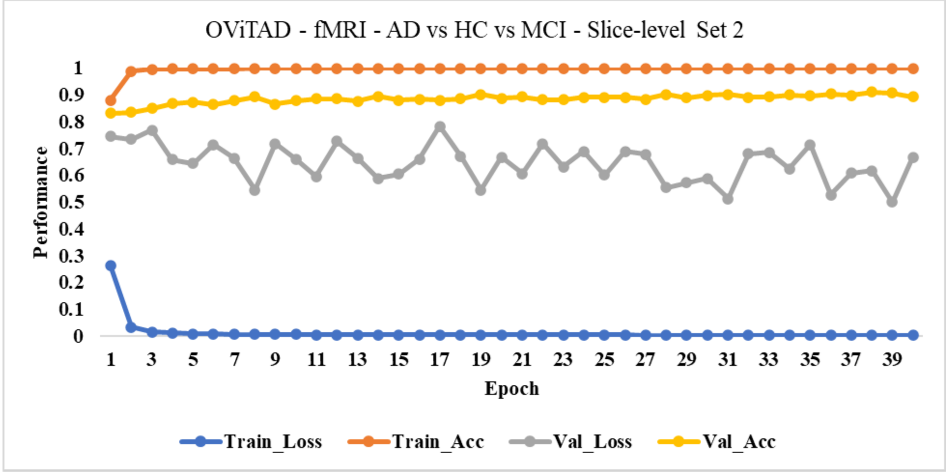
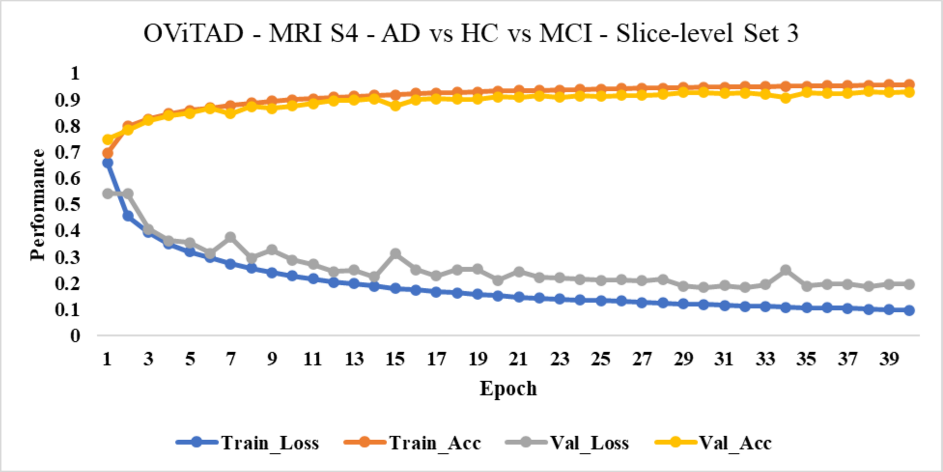


Figure A3. The global attention feature map was obtained by multiplying the FC layer vector by each pixel in the structural MRI brain slices and measuring the sum of the multiplication per pixel. Next, we normalized the feature map to (0,255) and visualized the maps using the CIVIDIS color map. Finally, we selected the first slice of each time-course to demonstrate various brain morphology across the MRI data acquisition.





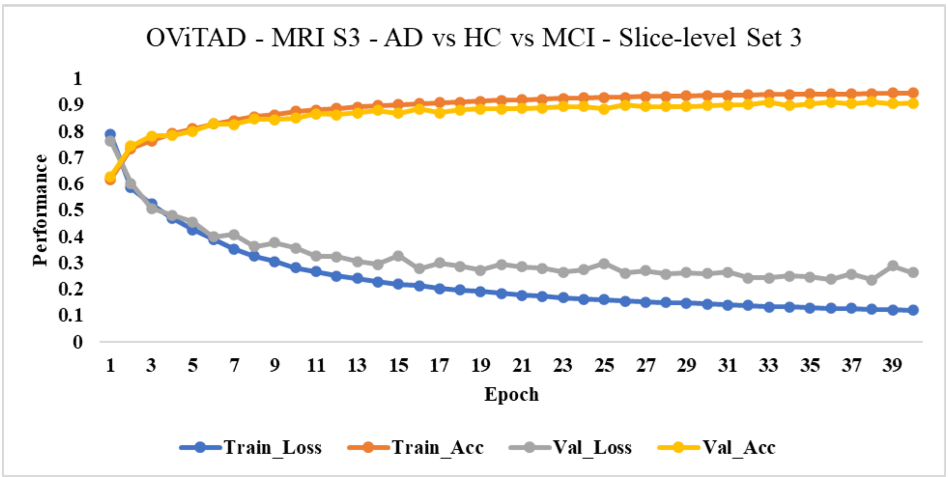


Figure A4. The performance of best-performing models for fMRI, MRI-S3, and MRI-S4 in a multiple classification experiment to predict AD vs. HC vs. MCI includes the training loss and accuracy rates and loss scores for training and validations sets. The modeling was conducted using 2D images, and the metrics shown across represent the slice-level performance used to extract the subject-level metrics.

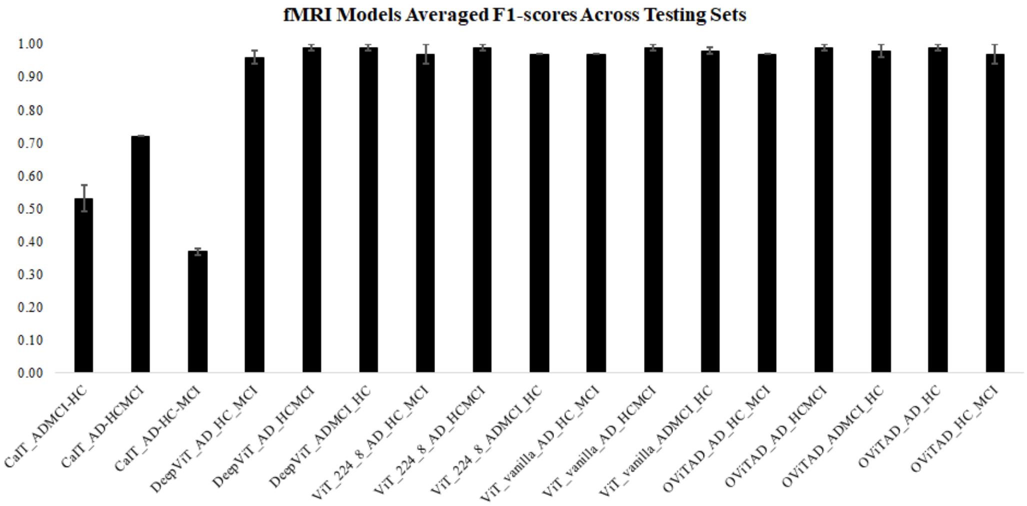


Figure A5. The summary of fMRI models' performance using averaged F1-scores for three testing sets.

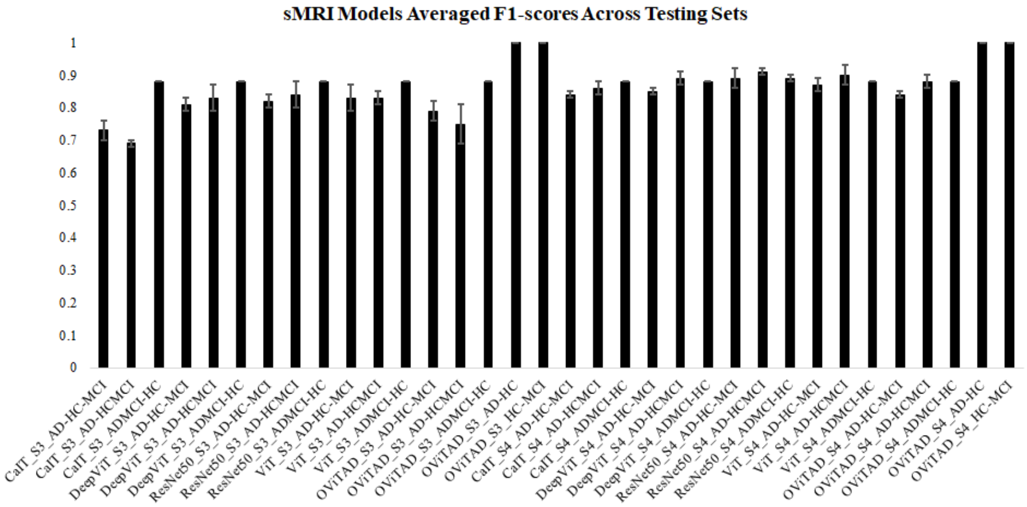


Figure A6. The summary of sMRI models' performance (S3,S4) using averaged F1-scores for three testing sets.

References

1. Lin, P.J.; D'Cruz, B.; Leech, A.A.; Neumann, P.J.; Sanon Aigbogun, M.; Oberdhan, D.; Lavelle, T.A. Family and caregiver spillover effects in cost-utility analyses of Alzheimer's disease interventions. *Pharmacoeconomics* **2019**, *37*, 597–608. [CrossRef] [PubMed]
2. Alzheimer's Association. 2018 Alzheimer's disease facts and figures. *Alzheimer's Dement.* **2018**, *14*, 367–429. [CrossRef]
3. Frisoni, G.B.; Boccardi, M.; Barkhof, F.; Blennow, K.; Cappa, S.; Chiotis, K.; Démonet, J.F.; Garibotto, V.; Giannakopoulos, P.; Gietl, A.; et al. Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *Lancet Neurol.* **2017**, *16*, 661–676. [CrossRef]
4. Rasmussen, J.; Langerman, H. Alzheimer's disease–why we need early diagnosis. *Degener. Neurol. Neuromuscul. Dis.* **2019**, *9*, 123. [CrossRef]
5. Fitzpatrick, A.W.; Falcon, B.; He, S.; Murzin, A.G.; Murshudov, G.; Garringer, H.J.; Crowther, R.A.; Ghetti, B.; Goedert, M.; Scheres, S.H. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017**, *547*, 185–190. [CrossRef]
6. Mazure, C.M.; Swendsen, J. Sex differences in Alzheimer's disease and other dementias. *Lancet Neurol.* **2016**, *15*, 451. [CrossRef]
7. Murphy, M.C.; Jones, D.T.; Jack, C.R., Jr.; Glaser, K.J.; Senjem, M.L.; Manduca, A.; Felmlee, J.P.; Carter, R.E.; Ehman, R.L.; Huston, J., III. Regional brain stiffness changes across the Alzheimer's disease spectrum. *Neuroimage Clin.* **2016**, *10*, 283–290. [CrossRef]
8. Gillis, C.; Mirzaei, F.; Potashman, M.; Ikram, M.A.; Maserejian, N. The incidence of mild cognitive impairment: A systematic review and data synthesis. *Alzheimer's Dementia: Diagn. Assess. Dis. Monit.* **2019**, *11*, 248–256. [CrossRef] [PubMed]
9. Cabeza, R.; Albert, M.; Belleville, S.; Craik, F.I.; Duarte, A.; Grady, C.L.; Lindenberger, U.; Nyberg, L.; Park, D.C.; Reuter-Lorenz, P.A.; et al. Maintenance, reserve and compensation: The cognitive neuroscience of healthy ageing. *Nat. Rev. Neurosci.* **2018**, *19*, 701–710. [CrossRef] [PubMed]

10. Petersen, R.C. Mild cognitive impairment. *Contin. Lifelong Learn. Neurol.* **2016**, *22*, 404.
11. Anthony, M.; Lin, F. A systematic review for functional neuroimaging studies of cognitive reserve across the cognitive aging spectrum. *Arch. Clin. Neuropsychol.* **2018**, *33*, 937–948. [CrossRef] [PubMed]
12. Mateos-Pérez, J.M.; Dadar, M.; Lacalle-Aurioles, M.; Iturria-Medina, Y.; Zeighami, Y.; Evans, A.C. Structural neuroimaging as clinical predictor: A review of machine learning applications. *NeuroImage Clin.* **2018**, *20*, 506–522. [CrossRef] [PubMed]
13. Neale, N.; Padilla, C.; Fonseca, L.M.; Holland, T.; Zaman, S. Neuroimaging and other modalities to assess Alzheimer's disease in Down syndrome. *NeuroImage Clin.* **2018**, *17*, 263–271. [CrossRef] [PubMed]
14. Rathore, S.; Habes, M.; Iftikhar, M.A.; Shacklett, A.; Davatzikos, C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage* **2017**, *155*, 530–548. [CrossRef]
15. Vemuri, P.; Lesnick, T.G.; Przybelski, S.A.; Knopman, D.S.; Lowe, V.J.; Graff-Radford, J.; Roberts, R.O.; Mielke, M.M.; Machulda, M.M.; Petersen, R.C.; et al. Age, vascular health, and Alzheimer disease biomarkers in an elderly sample. *Ann. Neurol.* **2017**, *82*, 706–718. [CrossRef]
16. Lindquist, M. Neuroimaging results altered by varying analysis pipelines, 2020. [CrossRef]
17. Wang, X.; Huang, W.; Su, L.; Xing, Y.; Jessen, F.; Sun, Y.; Shu, N.; Han, Y. Neuroimaging advances regarding subjective cognitive decline in preclinical Alzheimer's disease. *Mol. Neurodegener.* **2020**, *15*, 1–27. [CrossRef]
18. Hainc, N.; Federau, C.; Stieltjes, B.; Blatow, M.; Bink, A.; Stippich, C. The bright, artificial intelligence-augmented future of neuroimaging reading. *Front. Neurol.* **2017**, *8*, 489. [CrossRef] [PubMed]
19. Jo, T.; Nho, K.; Saykin, A.J. Deep learning in Alzheimer's disease: Diagnostic classification and prognostic prediction using neuroimaging data. *Front. Aging Neurosci.* **2019**, 220. [CrossRef] [PubMed]
20. Henschel, L.; Conjeti, S.; Estrada, S.; Diers, K.; Fischl, B.; Reuter, M. Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. *NeuroImage* **2020**, *219*, 117012. [CrossRef]
21. Puranik, M.; Shah, H.; Shah, K.; Bagul, S. Intelligent Alzheimer's detector using deep learning. In Proceedings of the 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 14–15 June 2018; pp. 318–323.
22. Bi, X.; Li, S.; Xiao, B.; Li, Y.; Wang, G.; Ma, X. Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology. *Neurocomputing* **2020**, *392*, 296–304. [CrossRef]
23. Kazemi, Y.; Houghten, S. A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data. In Proceedings of the 2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), St. Louis, MO, USA, 30 May–2 June 2018; pp. 1–8.
24. Tang, Z.; Chuang, K.V.; DeCarli, C.; Jin, L.W.; Beckett, L.; Keiser, M.J.; Dugger, B.N. Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline. *Nat. Commun.* **2019**, *10*, 1–14. [CrossRef] [PubMed]
25. Wen, J.; Thibeau-Sutre, E.; Diaz-Melo, M.; Samper-González, J.; Routier, A.; Bottani, S.; Dormont, D.; Durrleman, S.; Burgos, N.; Colliot, O.; et al. Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. *Med. Image Anal.* **2020**, *63*, 101694. [CrossRef] [PubMed]
26. Liu, M.; Cheng, D.; Wang, K.; Wang, Y. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. *Neuroinformatics* **2018**, *16*, 295–308. [CrossRef]
27. Islam, J.; Zhang, Y. Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Informatics* **2018**, *5*, 1–14. [CrossRef] [PubMed]
28. Song, T.A.; Chowdhury, S.R.; Yang, F.; Jacobs, H.; El Fakhri, G.; Li, Q.; Johnson, K.; Dutta, J. Graph convolutional neural networks for Alzheimer's disease classification. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 414–417.
29. Sarraf, A.; Jalali, A.E.; Ghaffari, J. Recent Applications of Deep Learning Algorithms in Medical Image Analysis. *Am. Acad. Sci. Res. J. Eng. Technol. Sci.* **2020**, *72*, 58–66.
30. Sarraf, A.; Azhdari, M.; Sarraf, S. A comprehensive review of deep learning architectures for computer vision applications. *Am. Acad. Sci. Res. J. Eng. Technol. Sci.* **2021**, *77*, 1–29.
31. Janghel, R.; Rathore, Y. Deep convolution neural network based system for early diagnosis of Alzheimer's disease. *Irbm* **2021**, *42*, 258–267. [CrossRef]
32. Chen, S.; Zhang, J.; Wei, X.; Zhang, Q. Alzheimer's Disease Classification Using Structural MRI Based on Convolutional Neural Networks. In Proceedings of the 2020 2nd International Conference on Big-data Service and Intelligent Computation, Johannesburg, South Africa, 28–30 April 2020; pp. 7–13.
33. Albright, J.; Initiative, A.D.N. Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm. *Alzheimer's Dementia: Transl. Res. Clin. Interv.* **2019**, *5*, 483–491. [CrossRef]
34. Li, F.; Liu, M.; Initiative, A.D.N. A hybrid convolutional and recurrent neural network for hippocampus analysis in Alzheimer's disease. *J. Neurosci. Methods* **2019**, *323*, 108–118. [CrossRef]
35. Feng, C.; Elazab, A.; Yang, P.; Wang, T.; Zhou, F.; Hu, H.; Xiao, X.; Lei, B. Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM. *IEEE Access* **2019**, *7*, 63605–63618. [CrossRef]
36. Dua, M.; Makhija, D.; Manasa, P.; Mishra, P. A CNN–RNN–LSTM based amalgamation for Alzheimer's disease detection. *J. Med. Biol. Eng.* **2020**, *40*, 688–706. [CrossRef]

37. Anwar, S.M.; Majid, M.; Qayyum, A.; Awais, M.; Alnowami, M.; Khan, M.K. Medical image analysis using convolutional neural networks: A review. *J. Med. Syst.* **2018**, *42*, 1–13. [CrossRef] [PubMed]
38. Yao, G.; Lei, T.; Zhong, J. A review of convolutional-neural-network-based action recognition. *Pattern Recognit. Lett.* **2019**, *118*, 14–22. [CrossRef]
39. Dhillon, A.; Verma, G.K. Convolutional neural network: A review of models, methodologies and applications to object detection. *Prog. Artif. Intell.* **2020**, *9*, 85–112. [CrossRef]
40. Sornam, M.; Muthusubash, K.; Vanitha, V. A survey on image classification and activity recognition using deep convolutional neural network architecture. In Proceedings of the 2017 ninth international conference on advanced computing (ICoAC), Chennai, India, 14–16 December 2017; pp. 121–126.
41. Sultana, F.; Sufian, A.; Dutta, P. Evolution of image segmentation using deep convolutional neural network: A survey. *Knowl.-Based Syst.* **2020**, *201*, 106062. [CrossRef]
42. Ebrahimighahnavieh, M.A.; Luo, S.; Chiong, R. Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. *Comput. Methods Programs Biomed.* **2020**, *187*, 105242. [CrossRef]
43. Altinkaya, E.; Polat, K.; Barakli, B. Detection of Alzheimer's disease and dementia states based on deep learning from MRI images: A comprehensive review. *J. Inst. Electron. Comput.* **2020**, *1*, 39–53.
44. Murn, L.; Blasi, S.; Smeaton, A.F.; O'Connor, N.E.; Mrak, M. Interpreting CNN for low complexity learned sub-pixel motion compensation in video coding. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 798–802.
45. You, J.; Korhonen, J. Transformer for image quality assessment. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 1389–1393.
46. Li, N.; Liu, S.; Liu, Y.; Zhao, S.; Liu, M. Neural speech synthesis with transformer network. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 6706–6713.
47. Haller, S.; Lovblad, K.O.; Giannakopoulos, P. Principles of classification analyses in mild cognitive impairment (MCI) and Alzheimer disease. *J. Alzheimer's Dis.* **2011**, *26*, 389–394. [CrossRef]
48. Dukart, J.; Mueller, K.; Barthel, H.; Villringer, A.; Sabri, O.; Schroeter, M.L.; Initiative, A.D.N. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. *Psychiatry Res. Neuroimaging* **2013**, *212*, 230–236. [CrossRef]
49. Suk, H.I.; Lee, S.W.; Shen, D.; Initiative, A.D.N. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* **2014**, *101*, 569–582. [CrossRef] [PubMed]
50. Zhu, X.; Suk, H.I.; Lee, S.W.; Shen, D. Canonical feature selection for joint regression and multi-class identification in Alzheimer's disease diagnosis. *Brain Imaging Behav.* **2016**, *10*, 818–828. [CrossRef] [PubMed]
51. Rieke, J.; Eitel, F.; Weygandt, M.; Haynes, J.D.; Ritter, K. Visualizing convolutional networks for MRI-based diagnosis of Alzheimer's disease. In *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*; Springer: Berlin, Germany, 2018; pp. 24–31.
52. Farooq, A.; Anwar, S.; Awais, M.; Rehman, S. A deep CNN based multi-class classification of Alzheimer's disease using MRI. In Proceedings of the 2017 IEEE International Conference on Imaging systems and techniques (IST), Beijing, China, 18–20 October 2017; pp. 1–6.
53. Long, X.; Chen, L.; Jiang, C.; Zhang, L.; Initiative, A.D.N. Prediction and classification of Alzheimer disease based on quantification of MRI deformation. *PLoS ONE* **2017**, *12*, e0173372. [CrossRef] [PubMed]
54. Sarraf, S.; DeSouza, D.D.; Anderson, J.; Tofighi, G. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *BioRxiv* **2017**, 070441.
55. Wang, S.; Wang, H.; Shen, Y.; Wang, X. Automatic recognition of mild cognitive impairment and alzheimers disease using ensemble based 3d densely connected convolutional networks. In Proceedings of the 2018 17th IEEE International conference on machine learning and applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 517–523.
56. Khvostikov, A.; Aderghal, K.; Benois-Pineau, J.; Krylov, A.; Catheline, G. 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies. *arXiv* **2018**, arXiv:1801.05968.
57. Hosseini-Asl, E.; Keynton, R.; El-Baz, A. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In Proceedings of the 2016 IEEE international conference on image processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 126–130.
58. Sarraf, S.; Desouza, D.D.; Anderson, J.A.; Saverino, C. MCADNNet: Recognizing stages of cognitive impairment through efficient convolutional fMRI and MRI neural network topology models. *IEEE Access* **2019**, *7*, 155584–155600. [CrossRef]
59. Soliman, S.A.; Hussein, R.R.; El-Dahshan, E.S.A.; Salem, A.B.M. Intelligent Algorithms for the Diagnosis of Alzheimer's Disease. In *Innovative Smart Healthcare and Bio-Medical Systems*; CRC Press: Boca Raton, FL, USA, 2020; pp. 51–86.
60. Soliman, S.A.; El-Sayed, A.; Salem, A.B.M. Predicting Alzheimer's Disease with 3D Convolutional Neural Networks. *Int. J. Appl. Fuzzy Sets Artif. Intell.* **2020**, *10*, 125–146.

61. Duc, N.T.; Ryu, S.; Qureshi, M.N.I.; Choi, M.; Lee, K.H.; Lee, B. 3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI. *Neuroinformatics* **2020**, *18*, 71–86. [CrossRef]
62. Li, W.; Lin, X.; Chen, X. Detecting Alzheimer's disease Based on 4D fMRI: An exploration under deep learning framework. *Neurocomputing* **2020**, *388*, 280–287. [CrossRef]
63. Ramzan, F.; Khan, M.U.G.; Rehmat, A.; Iqbal, S.; Saba, T.; Rehman, A.; Mehmood, Z. A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks. *J. Med. Syst.* **2020**, *44*, 1–16. [CrossRef]
64. Sarraf, S.; Tofighi, G. Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data. In Proceedings of the 2016 future technologies conference (FTC), San Francisco, CA, USA, 6–7 December 2016; pp. 816–820.
65. Cheng, D.; Liu, M. Combining convolutional and recurrent neural networks for Alzheimer's disease diagnosis using PET images. In Proceedings of the 2017 IEEE International Conference on Imaging Systems and Techniques (IST), Beijing, China, 18–20 October 2017; pp. 1–5.
66. Hong, X.; Lin, R.; Yang, C.; Zeng, N.; Cai, C.; Gou, J.; Yang, J. Predicting Alzheimer's disease using LSTM. *IEEE Access* **2019**, *7*, 80893–80901. [CrossRef]
67. Wang, T.; Qiu, R.G.; Yu, M. Predictive modeling of the progression of Alzheimer's disease with recurrent neural networks. *Sci. Rep.* **2018**, *8*, 1–12. [CrossRef] [PubMed]
68. Sethi, M.; Ahuja, S.; Rani, S.; Bawa, P.; Zaguia, A. Classification of Alzheimer's Disease Using Gaussian-Based Bayesian Parameter Optimization for Deep Convolutional LSTM Network. *Comput. Math. Methods Med.* **2021**, *2021*, 4186666. [CrossRef] [PubMed]
69. Cui, R.; Liu, M.; Li, G. Longitudinal analysis for Alzheimer's disease diagnosis using RNN. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 1398–1401.
70. Bubu, O.M.; Pirraglia, E.; Andrade, A.G.; Sharma, R.A.; Gimenez-Badia, S.; Umasabor-Bubu, O.Q.; Hogan, M.M.; Shim, A.M.; Mukhtar, F.; Sharma, N.; et al. Obstructive sleep apnea and longitudinal Alzheimer's disease biomarker changes. *Sleep* **2019**, *42*, zsz048. [CrossRef]
71. Benoit, J.S.; Chan, W.; Piller, L.; Doody, R. Longitudinal sensitivity of Alzheimer's disease severity staging. *Am. J. Alzheimer's Dis. Other Dementias®* **2020**, *35*, 1533317520918719. [CrossRef]
72. Jabason, E.; Ahmad, M.O.; Swamy, M. Hybrid Feature Fusion Using RNN and Pre-trained CNN for Classification of Alzheimer's Disease (Poster). In Proceedings of the 2019 22th International Conference on Information Fusion (FUSION), Ottawa, ON, Canada, 2–5 July 2019; pp. 1–4.
73. Song, J.; Zheng, J.; Li, P.; Lu, X.; Zhu, G.; Shen, P. An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis. *Front. Digit. Health* **2021**, *3*, 19. [CrossRef]
74. Gupta, Y.; Kim, J.I.; Kim, B.C.; Kwon, G.R. Classification and graphical analysis of Alzheimer's disease and its prodromal stage using multimodal features from structural, diffusion, and functional neuroimaging data and the APOE genotype. *Front. Aging Neurosci.* **2020**, *12*, 238. [CrossRef]
75. Thushara, A.; Amma, C.U.; John, A.; Saju, R. Multimodal MRI Based Classification and Prediction of Alzheimer's Disease Using Random Forest Ensemble. In Proceedings of the 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), Cochin, India, 2–4 July 2020; pp. 249–256.
76. Liu, M.; Cheng, D.; Yan, W.; Initiative, A.D.N. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Front. Neuroinformatics* **2018**, *12*, 35. [CrossRef] [PubMed]
77. Yuen, S.C.; Liang, X.; Zhu, H.; Jia, Y.; Leung, S.W. Prediction of differentially expressed microRNAs in blood as potential biomarkers for Alzheimer's disease by meta-analysis and adaptive boosting ensemble learning. *Alzheimer's Res. Ther.* **2021**, *13*, 1–30. [CrossRef] [PubMed]
78. Kim, J.; Park, Y.; Park, S.; Jang, H.; Kim, H.J.; Na, D.L.; Lee, H.; Seo, S.W. Prediction of tau accumulation in prodromal Alzheimer's disease using an ensemble machine learning approach. *Sci. Rep.* **2021**, *11*, 1–8. [CrossRef]
79. Hu, D. An introductory survey on attention mechanisms in NLP problems. In Proceedings of the Proceedings of SAI Intelligent Systems Conference, London, UK, 5–6 September 2019; pp. 432–448.
80. Letarte, G.; Paradis, F.; Giguère, P.; Laviolette, F. Importance of self-attention for sentiment analysis. In Proceedings of the Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Brussels, Belgium, 1 November 2018; pp. 267–275.
81. Roshanzamir, A.; Aghajan, H.; Soleymani Baghshah, M. Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech. *BMC Med. Informatics Decis. Mak.* **2021**, *21*, 1–14. [CrossRef] [PubMed]
82. Sarasua, I.; Pölsterl, S.; Wachinger, C.; Neuroimaging, A.D. TransforMesh: A Transformer Network for Longitudinal Modeling of Anatomical Meshes. In Proceedings of the International Workshop on Machine Learning in Medical Imaging, Strasbourg, France, 27 September 2021; pp. 209–218.
83. Wang, S.; Zhuang, Z.; Xuan, K.; Qian, D.; Xue, Z.; Xu, J.; Liu, Y.; Chai, Y.; Zhang, L.; Wang, Q.; et al. 3DMeT: 3D Medical Image Transformer for Knee Cartilage Defect Assessment. In Proceedings of the International Workshop on Machine Learning in Medical Imaging, Strasbourg, France, 27 September 2021; pp. 347–355.

84. Jack, C.R., Jr.; Bernstein, M.A.; Fox, N.C.; Thompson, P.; Alexander, G.; Harvey, D.; Borowski, B.; Britson, P.J.; Whitwell, J.L.; Ward, C.; et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med.* **2008**, *27*, 685–691. [CrossRef]
85. Churchill, N.W.; Spring, R.; Afshin-Pour, B.; Dong, F.; Strother, S.C. An automated, adaptive framework for optimizing preprocessing pipelines in task-based functional MRI. *PLoS ONE* **2015**, *10*, e0131520. [CrossRef] [PubMed]
86. Churchill, N.W.; Oder, A.; Abdi, H.; Tam, F.; Lee, W.; Thomas, C.; Ween, J.E.; Graham, S.J.; Strother, S.C. Optimizing preprocessing and analysis pipelines for single-subject fMRI. I. Standard temporal motion and physiological noise correction methods. *Hum. Brain Mapp.* **2012**, *33*, 609–627. [CrossRef] [PubMed]
87. Li, X.; Morgan, P.S.; Ashburner, J.; Smith, J.; Rorden, C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J. Neurosci. Methods* **2016**, *264*, 47–56. [CrossRef]
88. Smith, S.M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **2002**, *17*, 143–155. [CrossRef]
89. Jenkinson, M.; Bannister, P.; Brady, M.; Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **2002**, *17*, 825–841. [CrossRef]
90. Fonov, V.; Evans, A.C.; Botteron, K.; Almli, C.R.; McKinstry, R.C.; Collins, D.L.; the Brain Development Cooperative Group. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* **2011**, *54*, 313–327. [CrossRef]
91. Smith, S.M.; Jenkinson, M.; Woolrich, M.W.; Beckmann, C.F.; Behrens, T.E.; Johansen-Berg, H.; Bannister, P.R.; De Luca, M.; Drobnjak, I.; Flitney, D.E.; et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **2004**, *23*, S208–S219. [CrossRef]
92. Scarpazza, C.; Tognin, S.; Frisciata, S.; Sartori, G.; Mechelli, A. False positive rates in Voxel-based Morphometry studies of the human brain: Should we be worried? *Neurosci. Biobehav. Rev.* **2015**, *52*, 49–55. [CrossRef]
93. Mikl, M.; Mareček, R.; Hluštík, P.; Pavlicová, M.; Drastich, A.; Chlebus, P.; Brázdil, M.; Krupa, P. Effects of spatial smoothing on fMRI group inferences. *Magn. Reson. Imaging* **2008**, *26*, 490–503. [CrossRef]
94. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
95. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
96. Zhou, D.; Kang, B.; Jin, X.; Yang, L.; Lian, X.; Jiang, Z.; Hou, Q.; Feng, J. Deepvit: Towards deeper vision transformer. *arXiv* **2021**, arXiv:2103.11886.
97. Touvron, H.; Cord, M.; Sablayrolles, A.; Synnaeve, G.; Jégou, H. Going deeper with image transformers. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 32–42.
98. Alakörkkö, T.; Saarimäki, H.; Glerean, E.; Saramäki, J.; Korhonen, O. Effects of spatial smoothing on functional brain networks. *Eur. J. Neurosci.* **2017**, *46*, 2471–2480. [CrossRef]
99. Chen, Z.; Calhoun, V. Effect of spatial smoothing on task fMRI ICA and functional connectivity. *Front. Neurosci.* **2018**, *12*, 15. [CrossRef] [PubMed]
100. Lin, W.; Tong, T.; Gao, Q.; Guo, D.; Du, X.; Yang, Y.; Guo, G.; Xiao, M.; Du, M.; Qu, X.; et al. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Front. Neurosci.* **2018**, 777. [CrossRef]
101. Dimitriadis, S.I.; Liparas, D.; Initiative, A.D.N. How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: From Alzheimer's disease neuroimaging initiative (ADNI) database. *Neural Regen. Res.* **2018**, *13*, 962. [CrossRef]
102. Kruthika, K.; Maheshappa, H.; Initiative, A.D.N. Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. *Informatics Med. Unlocked* **2019**, *14*, 34–42. [CrossRef]
103. Spasov, S.; Passamonti, L.; Duggento, A.; Lio, P.; Toschi, N.; Initiative, A.D.N. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage* **2019**, *189*, 276–287. [CrossRef]
104. Basaia, S.; Agosta, F.; Wagner, L.; Canu, E.; Magnani, G.; Santangelo, R.; Filippi, M.; Initiative, A.D.N. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage Clin.* **2019**, *21*, 101645. [CrossRef] [PubMed]
105. Abrol, A.; Bhattarai, M.; Fedorov, A.; Du, Y.; Plis, S.; Calhoun, V.; Initiative, A.D.N. Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease. *J. Neurosci. Methods* **2020**, *339*, 108701. [CrossRef]
106. Shao, W.; Peng, Y.; Zu, C.; Wang, M.; Zhang, D.; Initiative, A.D.N. Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease. *Comput. Med. Imaging Graph.* **2020**, *80*, 101663. [CrossRef]
107. Alinsaif, S.; Lang, J.; Initiative, A.D.N. 3D shearlet-based descriptors combined with deep features for the classification of Alzheimer's disease based on MRI data. *Comput. Biol. Med.* **2021**, *138*, 104879. [CrossRef]
108. Hojjati, S.H.; Ebrahimzadeh, A.; Khazaee, A.; Babajani-Feremi, A.; Initiative, A.D.N. Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI. *Comput. Biol. Med.* **2018**, *102*, 30–39. [CrossRef]
109. Cui, R.; Liu, M.; Initiative, A.D.N. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Comput. Med. Imaging Graph.* **2019**, *73*, 1–10. [CrossRef]

110. Amoroso, N.; Diacono, D.; Fanizzi, A.; La Rocca, M.; Monaco, A.; Lombardi, A.; Guaragnella, C.; Bellotti, R.; Tangaro, S.; Initiative, A.D.N.; et al. Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge. *J. Neurosci. Methods* **2018**, *302*, 3–9. [CrossRef] [PubMed]
111. Buvaneswari, P.; Gayathri, R. Detection and Classification of Alzheimer's disease from cognitive impairment with resting-state fMRI. *Neural Comput. Appl.* **2021**, *1*, 1–16. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.