International Journal of
Science, Architecture, Technology and Environment

Integrating MRI, Clinical, and Genetic Features Using a Dual-Attention CNN-MLP for Alzheimer's Diagnosis

Dr. V. Maniraj[1*], K. Vaishali[2]

[1]*Associate Professor, Head of the Department, Department of Computer Science, A.V.V.M.Sri Pushpam College(Autonomous), Poondi, Thanjavur(Dt), Affiliated to Bharathidasan University, Thiruchirappalli, Tamilnadu*

[2]*II M.Sc COMPUTER SCIENCE, Department of Computer Science, A.V.V.M.Sri Pushpam College(Autonomous), Poondi,Thanjavur(Dt), Affiliated to Bharathidasan University, Thiruchirappalli, Tamilnadu*

[*]*Corresponding author, manirajv61@gmail.com*

DOI: https://doi.org/10.63680/ijsate0226038.031

Abstract

Alzheimer's disease (AD) is a progressive and irreversible neurodegenerative disorder, and mild cognitive impairment (MCI) is widely recognized as an intermediate stage preceding AD. Early and accurate identification of AD is crucial for timely prevention and intervention. Although numerous computational approaches have been developed for AD diagnosis, many rely solely on neuroimaging data and overlook additional informative modalities such as genetic and clinical data. Moreover, the interpretability of some existing models remains limited.

To address these issues, we propose a novel dual-attention CNN–MLP framework, termed DANMLP, designed for computer-aided AD diagnosis through the integration of structural MRI (sMRI), clinical variables (including demographic and neuropsychological measures), and APOE genetic information. The DANMLP model is composed of four main components:

(1) a Patch-CNN module for extracting local image features from sMRI patches;

(2) a position self-attention mechanism for modeling intra-patch feature dependencies;

(3) a channel self-attention mechanism for capturing inter-patch feature relationships; and

(4) two multilayer perceptrons for clinical feature extraction and final diagnostic classification.

Experimental results on the ADNI dataset using 5-fold cross-validation demonstrate that DANMLP achieves classification accuracies of 93% for AD vs. MCI and 82.4% for MCI vs. NC, outperforming five state-of-the-art methods by 0.2%–15.2% and 3.4%–26.8%, respectively. Furthermore, individualized visualization of salient brain regions enhances the model's interpretability and provides valuable assistance for early clinical assessment. Overall, these findings indicate that DANMLP is an effective and reliable tool for diagnosing AD and MCI.

Keywords: Object segmentation, Diagnostic imaging, Fuzzy inference, Graph-based image segmentation, Hierarchical neural learning, U-shaped segmentation network, MR imaging, Biological data analysis.

1. Introduction

Alzheimer's disease (AD) is the most prevalent irreversible neurodegenerative disorder affecting the central nervous system in middle-aged and older adults. It typically begins with progressive memory decline and cognitive impairment, and is pathologically characterized by the degeneration of specific neuronal populations, along with the accumulation of neuritic plaques and neurofibrillary tangles. Mild cognitive impairment (MCI), considered a transitional stage between normal aging (NC) and dementia, represents a critical window for slowing AD progression. Although no treatment currently exists to halt or reverse the course of AD, accurate and early diagnosis of AD and MCI is essential for enabling timely clinical interventions, delaying further cognitive deterioration, preserving remaining brain function, and reducing the risk of complications.Given that structural magnetic resonance imaging (sMRI) effectively captures morphological brain changes, numerous machine learning approaches have been proposed to distinguish AD, MCI, and NC individuals from sMRI scans. However, conventional machine learning methods depend strongly on the quality of handcrafted features (e.g., cortical thickness, hippocampal volume, gray matter density) and assume these features sufficiently represent all relevant information—potentially missing important discriminative cues embedded in the raw images. Deep learning (DL), particularly convolutional neural networks (CNNs), overcomes this limitation by automatically learning hierarchical and informative representations directly from the data. CNNs have demonstrated strong capability in extracting high-level features from sMRI images, leading to substantial improvements in brain disease classification. Results from the ADNI dataset confirm that our proposed DANMLP framework enhances diagnostic accuracy while also improving model interpretability.

In summary, the major contributions of this study are as follows:

1. We introduce a dual-attention CNN–MLP framework that integrates sMRI, clinical, and genetic (APOE) data to enhance AD diagnostic performance.

2. A Patch-CNN module is developed to extract discriminative features from local 3D sMRI patches.

3. A 3D dual-attention block is incorporated to model both positional and channel-wise dependencies, enabling more effective extraction of spatial structural information and improving interpretability at the brain-region level.

The remainder of this paper is organized as follows. Section 2 reviews related work and describes the sMRI, clinical, and APOE genetic data used in this study, along with the proposed DANMLP model. Section 3 presents the experimental setup, and Section 4 reports the results. Section 5 discusses the findings and concludes the paper.

2. Related work

Most existing studies on Alzheimer's disease (AD) diagnosis primarily concentrate on neuroimaging modalities, with positron emission tomography (PET) and magnetic resonance imaging (MRI) receiving particular attention. For instance, Abramova et al. introduced a multi-view separable pyramid network (MiSePyNet) that extracts feature representations from axial, coronal, and sagittal PET views. Although PET effectively reflects cerebral metabolic activity and can detect glucose metabolism abnormalities prior to structural deterioration, it is costly and requires the administration of radioactive tracers.

In contrast, non-invasive imaging techniques such as functional MRI (fMRI), diffusion tensor imaging (DTI), and structural MRI (sMRI) use magnetic fields and radiofrequency pulses to visualize internal brain

structures and are widely applied to in vivo studies of AD-related pathological changes. fMRI captures blood-oxygen-level–dependent signals and enables analysis of brain activity under different conditions. DTI characterizes white matter fiber pathways by measuring water diffusion properties, thus supporting detailed visualization of nerve conduction tracts. sMRI is particularly sensitive to anatomical and morphological changes associated with brain atrophy and can detect structural degeneration.

Several studies have leveraged these imaging modalities for AD classification. For example, Gan et al. proposed a functional connectivity network (FCN) framework based on fMRI to uncover disease-related brain network alterations. De and Chowdhury used three VoxCNNs to separately process 3D Echo Planar Imaging (EPI), Fractional Anisotropy (FA), and Mean Diffusivity (MD) maps from DTI scans, followed by a random forest classifier using region-averaged FA and MD values. Their approach combined CNN and RF outputs through a modulated rank averaging fusion strategy for AD classification. Folego et al. developed ADNet, a method combining 3D CNNs and domain adaptation on whole sMRI scans to identify AD biomarkers; however, its computational complexity increases the risk of overfitting.

Given the heterogeneous and multifactorial nature of AD, single-modality data are typically insufficient for reliable diagnosis. Multi-modality approaches, which integrate complementary information from sources such as sMRI, fMRI, and DTI, have shown improved diagnostic performance. For instance, Gupta et al. extracted features from sMRI, fMRI, and DTI using multiple neuroimaging toolboxes (e.g., DPARSF, FSL) and applied multiple kernel learning (MKL) for classification, though the method's success heavily relies on feature extraction quality. Huang et al. used both sMRI and PET images of the hippocampus in 3D VGG-based CNNs, then fused the learned features through a fully connected network for AD diagnosis. Kang et al. combined sMRI and DTI as inputs to a VGG-16 network for feature extraction via transfer learning, applied LASSO to reduce feature redundancy, and used a support vector machine (SVM) to distinguish early MCI from NC.

Although multimodal neuroimaging approaches achieve promising results, most rely exclusively on combinations of sMRI, fMRI, and PET and often use simplistic fusion strategies. These methods may introduce substantial redundancy, as multimodal neuroimaging data contain overlapping and noisy information. Merely increasing the number of imaging modalities does not necessarily enhance diagnostic performance and may lead to inefficient model training. In contrast, sMRI is widely adopted in clinical practice due to its non-invasive nature, cost-effectiveness, and sensitivity to structural changes. Furthermore, clinical and genetic data—especially APOE genotype—provide complementary information that is not available from imaging alone.

Motivated by these considerations, we propose a dual-attention CNN and MLP framework that separately extracts discriminative AD-related features from sMRI images, clinical measurements, and genetic data. These learned features are then fused using an additional MLP to achieve robust and accurate AD diagnosis.

Attentional mechanisms in medical imaging

The concept of the attention mechanism originates from the visual processing system of animals, which selectively focuses on salient regions when confronted with abundant visual information. In deep learning, attention mechanisms not only enhance interpretability by highlighting influential features, but also improve performance by enabling models to emphasize relevant information while suppressing irrelevant cues. In medical imaging, attention mechanisms are generally categorized into three types: hard attention, soft attention, and self-attention.

Guan et al. were among the first to apply hard attention in medical image analysis, using an attention-guided CNN (AG-CNN) with three branches to classify thoracic diseases on chest X-ray images. However, hard attention is represented through non-differentiable one-hot encoding, making it incompatible with standard backpropagation and often requiring reinforcement learning for optimization. In contrast, soft attention is fully differentiable and integrates more naturally with deep learning frameworks. For example, Schlemper et al. incorporated attention gates into traditional CNNs to highlight anatomical structures of varying shapes and sizes, while Abramova et al. embedded squeeze-and-excitation (SE) blocks into U-Net to adaptively recalibrate channel-wise feature weights. Although these approaches have shown merit, they primarily focus on the impact of individual features and overlook inter-feature dependencies.

Self-attention mechanisms overcome this limitation by modeling relationships among all input features so that each output representation encodes global contextual dependencies. For instance, Li et al. developed a 3D self-attention CNN for low-dose CT denoising, enabling the model to capture extensive intra- and inter-slice spatial information. Similarly, Shen et al. proposed a multiscale temporal self-attention and dynamic graph convolution hybrid network (MTS-DGCHN) for EEG stereogram recognition, where temporal self-attention was used to learn sequential continuity in EEG signals.

Considering the advantages of self-attention and the fact that Alzheimer's disease is highly complex and heterogeneous—with numerous interactions between brain regions during its progression—we incorporate a 3D dual self-attention mechanism into our model. This approach emphasizes both positional and channel-wise feature correlations, enabling more comprehensive feature representation for effective AD diagnosis.

3. Material and methods

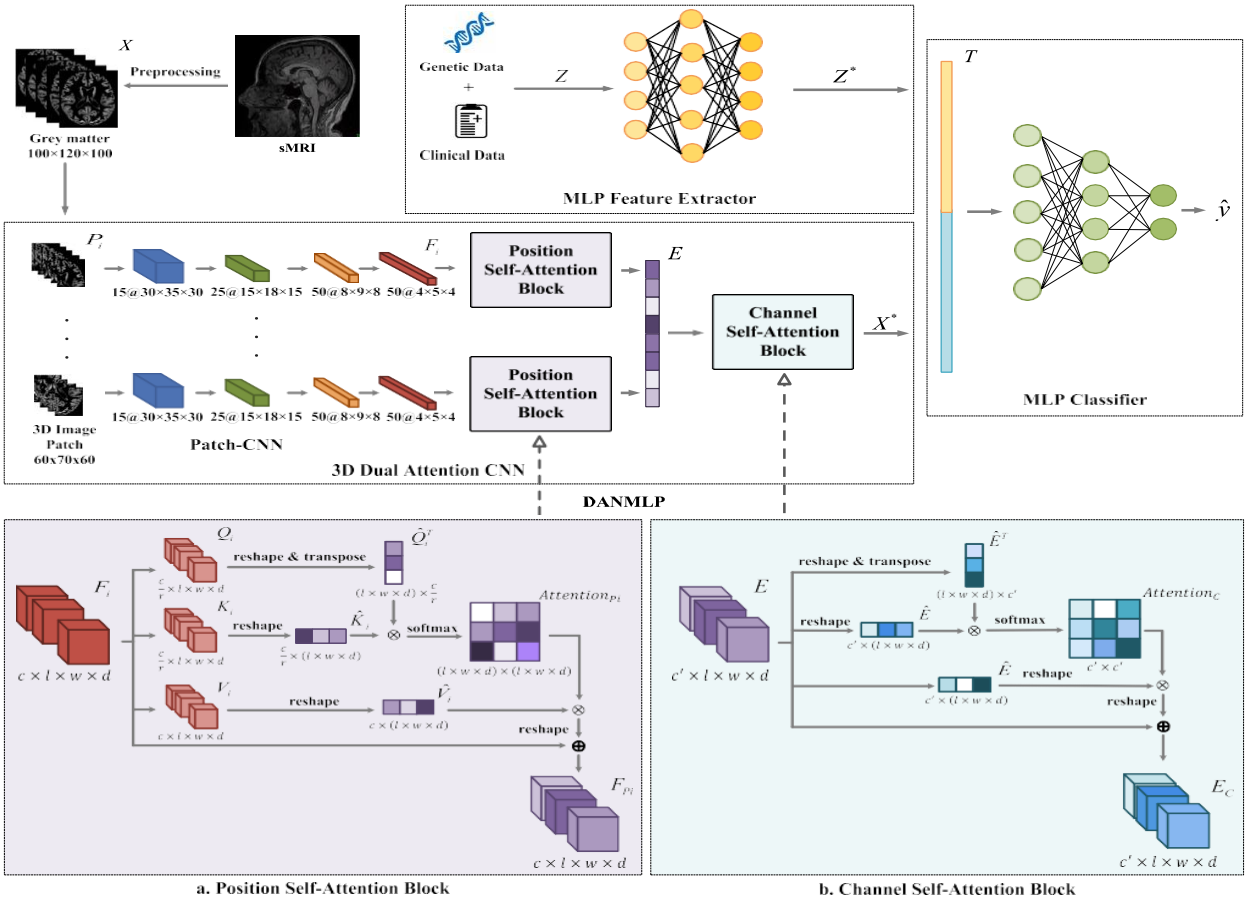
In this section, we first present the data used in this work. Then, we introduce how to preprocess these data. Finally, we show our DANMLP model in detail.

Data and pre processing

sMRI, clinical and genetic data used in this work were obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.l oni.usc.edu/). In the ADNI database, the subjects are divided into three categories: Alzheimer'sdisease(AD),mildcognitive impairment(MCI)and normal control(NC)by the standard cliniccrite- ria,suchasmini-mental state examination(MMSE)scores,clinicalde- mentia rating (CDR) scores, neuropsychiatric inventory-questionnaire (NPI-Q) scores, and geriatric depression scale (GDS) scores, func- tional assessment questionnaire(FAQ)scores.Weselected750subjects from the ADNI database, including 250 AD, 250 MCI, and 250 NC subjects,who have all the magnetization-prepared rapid gradient-echo(MPRAGE)T1-weightedimage(sMRI),clinical(i.e.,age,gender, MMSE, CDR, etc.) and APOE genotyping data. Each subject's sMRI, clinicalandAPOEgenotypingdataweretakenwithin±6months. Thedemographic,neuropsychologicalandcognitiveassessmentof750 subjects from the ADNI database is shown in Table S1.For subsequent better feature learning and AD diagnosis, we adoptedthetypicalproceduresofAnteriorCommissure(AC)–Posterior Commissure (PC) correction that can eliminate noise introduced by subjectmovementduringthesMRIscanformoreaccuratelocalization and comparison of brain structures in Statistical Parametric Mapping12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/), and skull- stripping,andcerebellumremovalinComputationalAnatomyToolbox to preprocess the original sMRI images downloaded from ADNI. Then, the corrected sMRI images were segmented into Gray Matter (GM), White Matter (WM), and Cerebrospinal Fluid (CSF) according to the tissue probability map (TPM) template. The GM, WM and CSF images [were] normalized to the Montreal Neurological Institute standard space (MNI) using affine linear registration to generate images with 121 ×

145 × 121 voxels. Finally, we removed the border area without information to obtain the images with 100 × 120 × 100 voxels. Only GM images were used in this work.For the genetic data, the APOE locus contains three alleles, $\epsilon 2$, $\epsilon 3$ and$\epsilon 4$,whichcangeneratethreepureheterozygotes(i.e.,$\epsilon 2/2$,$\epsilon 3/3$,

$\epsilon 4/4$)andthreeheterozygotes(i.e.,$\epsilon 2/3$,$\epsilon 2/4$,$\epsilon 3/4$)foratotalofsix common phenotypes. ADNI recorded the phenotypes of the subjects. We counted the number (i.e., 0, 1, 2) of APOE $\epsilon 2$ and APOE $\epsilon 4$ contained in each subject to indicate the APOE genotype.



Schematic of DANMLP model framework. DANMLP comprises three components: 3D dual attention CNN (3D-DACNN), MLP feature extractor and MLP classifier. 3D-DACNNisapatch-levelnetworkusedtolearnthefeaturesofsMRIimages,whichconsistsofaPatch-CNNwithICNNbranches,Ipositionself-attentionblocksandachannelself-attentionblock.MLPfeatureextractorisusedtolearnthefeaturesfromclinicalandgeneticdata.MLPclassifierisadoptedtofusethefeatureslearnedfromimages,clinicalandgeneticdataforNC/MCI/ADclassification.(a)Thearchitectureofthepositionself-attentionblock.Thisblockisusedtoaugmenttheposition-dependentfeatureswithinthepatch.(b)Thearchitectureofthechannelself-attentionblock.Thisblockisusedtoemphasizeinterdependentfeaturemapsbetweenpatches.

Overall architecture of DANMLP model Patch-CNN

The Patch-Net is designed to extract increasingly abstract representations from the originalpatches $P_{n,i}(n =$

$1,2,\dots,N; i = 1,2,\dots,I)$ and to reduce the spatial dimensions of the feature maps. In the DANMLP framework, denotes the total number of subjects, and each subject's sMRI scan is represented as

$$X_n = \{P_{n,i} \mid i = 1,2,\dots,I\},$$

where $P_{n,i} \in \mathbb{R}^{w_1 \times w_2 \times w_3}$ is the i-th 3D patch extracted from the sMRI volume with size $w_1 \times w_2 \times w_3$, and I represents the total number of patches. The clinical and APOE genetic information of subject n is denoted as Z_n, and the corresponding label is $y_n \in Y = \{y_1 = \text{"AD"}, y_2 = \text{"MCI"}, y_3 = \text{"NC"}\}$.

The overall DANMLP architecture (Fig. 1) is composed of three main components:

a 3D dual-attention CNN (3D-DACNN) for learning discriminative features from sMRI patches,an MLP feature extractor for deriving representations from clinical and APOE genotype data, andan MLP classifier that performs AD diagnosis by fusing the outputs of the 3D-DACNN and the MLP feature extractor.

The 3D-DACNN comprises ICNN branches, each responsible for processing the patch located at its corresponding spatial position. Specifically, the i-th CNN branch receives the patch $P_i$as input. Each branch consists of four sequential convolutional blocks with identical architecture. Every block includes a 3D convolution layer, a ReLU activation function, and a 3D max-pooling layer. For the l-th block in the i-th branch, the input feature map $H_l$is passed through the 3D convolutional layer followed by a ReLU nonlinearity. For the initial block, the input is simply the raw patch, i.e.,

$$H^{(1)} = P_i.$$

The value at spatial location (x_1, x_2, x_3)on the j-th feature map of the l-th convolutional block is computed according to the mathematical definitions summarized in Table.

4. Experimental settings and evaluation metrics

We evaluated the proposed DANMLP model on two binary classification tasks: AD vs. MCI and MCI vs. NC, using five-fold cross-validation (5CV). As illustrated in Fig. S1, all subjects were randomly divided into five subsets of roughly equal size. In each fold, one subset was designated as the test set, while the remaining four subsets were split into training and validation sets at a 3:1 ratio. For each binary task, the number of subjects in each class was kept balanced across the training, validation, and test sets to ensure fair comparison. This procedure was repeated five times, with each subset serving once as the test set, and the average performance across the five folds was used as the final evaluation metric.

DANMLP was trained for 40 epochs using the Adam optimizer. The initial learning rate was set to 0.001 for the MLP feature extractor and 0.0001 for the 3D-DACNN. A learning rate annealing strategy was applied, reducing the learning rate by a factor of 0.1 if training loss did not improve for five consecutive epochs. The remaining hyperparameters were configured as follows: batch size = 10, number of patches = 8, patch size = 60 × 70 × 60, reduction ratio = 5, and dropout rate = 0.5. All experiments were conducted on a Linux system equipped with four NVIDIA RTX 3090 GPUs (24 GB each), an Intel Xeon 40 × 2.4 GHz CPU, and 128 GB RAM. The model was implemented using PyTorch 3.7.

Additionally, we compared DANMLP trained with different optimizers (results in Table S3 and Fig. S2). We also examined training and inference time under varying model sizes (Table S4). The results show that Adam consistently outperforms the other three optimization methods in terms of ACC, AUC, and SPE, leading us to adopt it for all experiments. Furthermore, while training time remains relatively constant across model sizes, inference speed decreases as the number of parameters increases. Considering hardware limitations, we selected the DANMLP model with 33

million parameters for this study.

Evaluation Metrics

We evaluated classification performance using four common metrics: accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the ROC curve (AUC). These metrics are defined as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, SEN = \frac{TP}{TP + FN}, SPE = \frac{TN}{TN + FP},$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Comparison With State-of-the-Art Methods

To ensure a fair comparison, we reimplemented several representative models using our dataset under the same classification settings, since their original datasets were not publicly available. The compared methods include VoxCNN, ResNet, Xing's method, 3DAN, and Spasov's method. VoxCNN employs a VGG-like architecture with 3D convolutional blocks, whereas ResNet incorporates 21 layers with six VoxRes blocks. Xing's method converts 3D sMRI volumes into 2D representations using approximate rank pooling, followed by a pretrained 2D VGG-11 and an attention-based classifier. 3DAN uses a 3D attention-enhanced CNN for AD classification. Spasov's method integrates dual learning with 3D separable convolutions and incorporates multimodal data including sMRI, local Jacobian determinant images, and demographic, neuropsychological, and APOE ε4 genetic data.Our results demonstrate that DANMLP outperforms these five methods in both classification tasks. For MCI vs. NC, DANMLP achieves an accuracy of 0.824, which is 0.268, 0.210, 0.256, 0.210, and 0.034 higher than VoxCNN, ResNet, Xing's method, 3DAN, and Spasov's method, respectively. For AD vs. MCI, DANMLP reaches an accuracy of 0.930, exceeding the performance of the same methods by 0.106, 0.044, 0.152, 0.060, and 0.002, respectively.Notably, the performance of DANMLP and Spasov's method is substantially better than that of models based solely on sMRI, demonstrating that integrating imaging, clinical, and genetic data can substantially enhance early-stage AD/MCI diagnosis. Although Spasov's method performs comparably to DANMLP, it requires an additional data modality (local Jacobian determinant images), whereas DANMLP achieves superior performance using fewer data sources.

5. Conclusion

In this study, we introduce DANMLP, a novel approach for AD/MCI diagnosis that integrates a dual-attention CNN with an MLP using structural MRI, clinical data, and genetic information. In the DANMLP framework, the Patch-CNN modules are responsible for extracting local image features from sMRI patches. The position self-attention block captures discriminative intra-patch feature relationships, whereas the channel self-attention block models inter-patch dependencies. Additionally, an MLP-based feature extractor learns compact and informative embeddings from the clinical and genetic data. Experimental evaluations show that DANMLP outperforms existing methods in distinguishing AD from MCI and MCI from NC. Moreover, DANMLP effectively identifies AD-relevant brain regions—such as the hippocampus and amygdala—and provides individualized visualizations that can support clinicians in achieving early and accurate AD diagnosis.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article.

Funding

The author received no financial support for the research, authorship and publication of this article.

References

1.Gaugler J, James B, Johnson T, Reimer J, Solis M, Weuve J, et al. Alzheimer'sdiseasefactsandfigures.AlzheimersDementia2022;18(2022):700–89.

2.Mckhann G. Report of the NINCDS-ADRDA work group under the auspices ofdepartment of health and human service task force on Alzheimer's disease.Neurology 1984;34:939–44.

Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drugdevelopment pipeline: 2019. Alzheimer's & Dementia: translational research &clinicalinterventions,5.2019,p.272–93.

3.Colliot O, Chételat G, Chupin M, Desgranges B, Magnin B, Benali H, et al.Discrimination between Alzheimer's disease, mild cognitive impairment andnormal aging by using automated segmentation of the hippocampus. Radiology2008;248:194–201.

4.KlöppelS,StonningtonCM,ChuC,DraganskiB,ScahillRI,RohrerJD,et al. Automatic classification of MR scans in Alzheimer's disease. Brain2008;131:681–9.

5.McDonald C, McEvoy L, Gharapetian L, Fennema-Notestine C, Hagler D, Holl D,et al. Regional rates of neocortical atrophy from normal aging to early Alzheimerdisease. Neurology 2009;73:457–65.

6.Bernal J, Kushibar K, Asfaw DS, Valverde S, Oliver A, Martí R, et al. Deepconvolutional neural networks for brain image analysis on magnetic resonanceimaging: A review. Artif Intell Med 2019;95:64–81.

7.Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medicalimage analysis using convolutional neural networks: A review. J Med Syst2018;42:1–13.

8.Kang W, Lin L, Zhang B, Shen X, Wu S, Initiative AsDN. Multi-model and multi-sliceensemblelearningarchitecturebasedon2Dconvolutionalneuralnetworksfor Alzheimer's disease diagnosis. Comput Biol Med 2021;136:104678.

9.Yagis E, Atnafu SW, García Seco de Herrera A, Marzi C, Scheda R, Giannelli M,et al. Effect of data leakage in brain MRI classification using 2D convolutionalneural networks. Sci Rep 2021;11:1–13.