

Adaptive 3DCNN-Based Interpretable Ensemble Model for Early Diagnosis of Alzheimer's Disease

Dan Pan, Genqiang Luo, An Zeng[✉], Member, IEEE, Chao Zou, Haolin Liang, Jianbin Wang, Tong Zhang[✉], Member, IEEE, Baoyao Yang[✉], and for the Alzheimer's Disease Neuroimaging Initiative

Abstract—An adaptive interpretable ensemble model based on a 3-D convolutional neural network (3DCNN) and genetic algorithm (GA), i.e., 3DCNN+EL+GA, was proposed to differentiate the subjects with Alzheimer's disease (AD) or mild cognitive impairment (MCI) and also identify the discriminative brain regions significantly contributing to the classifications in a data-driven way. The testing results on the datasets from both the AD Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Studies (OASIS) indicated that 3DCNN+EL+GA outperformed other state-of-the-art deep learning algorithms.

Manuscript received 17 February 2022; revised 15 July 2022; accepted 24 October 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61976058 and Grant 61772143; in part by the Science and Technology Planning Project of Guangdong under Grant 2021A1515012300, Grant 2019A050510041, and Grant 2021B0101220006; in part by the Science and Technology Planning Project of Guangzhou under Grant 202103000034, Grant 202206010007, and Grant 202002020090; in part by the Alzheimer's Disease Neuroimaging Initiative (ADNI), National Institutes of Health, under Grant U01 AG024904; in part by the Department of Defense ADNI under Award W81XWH-12-2-0012; in part by the National Institute on Aging, in part by the National Institute of Biomedical Imaging and Bioengineering; in part by AbbVie; in part by the Alzheimer's Association; in part by the Alzheimer's Drug Discovery Foundation; in part by Araclon Biotech; in part by BioClinica, Inc.; in part by Biogen; in part by Bristol-Myers Squibb Company; in part by CereSpir, Inc.; in part by Cogstate; in part by Eisai Inc.; in part by Elan Pharmaceuticals, Inc.; in part by Eli Lilly and Company; in part by EuroImmun; in part by F. Hoffmann-La Roche, Ltd.; Genentech, Inc.; in part by Fujirebio; in part by GE Healthcare; in part by IXICO Ltd.; in part by Janssen Alzheimer Immunotherapy Research and Development, LLC; in part by Johnson and Johnson Pharmaceutical Research and Development, LLC; in part by Lumosity; in part by Lundbeck; in part by Merck and Company, Inc.; in part by Meso Scale Diagnostics, LLC; in part by NeuroRx Research; in part by Neurotrack Technologies; in part by Novartis Pharmaceuticals Corporation; in part by Pfizer, Inc.; in part by Piramal Imaging; in part by Servier; in part by Takeda Pharmaceutical Company; in part by Transition Therapeutics; and in part by the Canadian Institutes of Health Research, Canada. (*Corresponding author: An Zeng*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Alzheimer's Disease Neuroimaging Initiative Data and Publications Committee (ADNI DPC).

Dan Pan and Genqiang Luo are with the School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou 510665, China (e-mail: pandan@gpnu.edu.cn).

An Zeng, Haolin Liang, Jianbin Wang, and Baoyao Yang are with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China (e-mail: zengan@gdut.edu.cn; baoxyao@gdut.edu.cn).

Chao Zou is with the Research and Development Centre of Guangzhou Department, Agricultural Bank of China, Guangzhou 510430, China.

Tong Zhang is with the Guangdong Provincial Key Laboratory of Computational Intelligence and Cyberspace Information, School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China, and also with the Pazhou Laboratory, Guangzhou 510335, China (e-mail: tony@scut.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TCSS.2022.3223999>, provided by the authors.

Digital Object Identifier 10.1109/TCSS.2022.3223999

More importantly, in these identified brain regions, the discriminative brain subregions at a voxel level were further located with a gradient-based attribution method designed for CNN and illustrated intuitively. Besides these, the behavioral domains corresponding to every identified discriminative brain region (e.g., the rostral hippocampus) were analyzed. It was shown that the resultant behavioral domains were consistent with those brain functions (e.g., emotion) impaired early in the AD process. Further research is needed to examine the generalizability of the proposed ideas and methods in identifying discriminative brain regions and subregions for the diagnosis of other brain disorders (especially little-known ones), such as Parkinson's disease, epilepsy, severe depression, autism, Huntington's disease, multiple sclerosis, and amyotrophic lateral sclerosis, using neuroimaging.

Index Terms—Alzheimer's disease (AD), attribution methods, convolutional neural network (CNN), deep learning (DL), ensemble learning (EL), genetic algorithm, interpretability, magnetic resonance imaging.

I. INTRODUCTION

ALZHEIMER'S disease (AD), accounting for 60%–80% of dementia cases, is the most dominant chronic neurodegenerative brain disorder that ultimately leads to irreversible loss of neurons among the elderly. AD causes cognitive dysfunction and loss of daily living independence [1]. The number of people with dementia is estimated to increase from 57.4 (95% uncertainty interval 50.4–65.1) million cases globally in 2019 to 152.8 (130.8–175.9) million cases in 2050 [2]. Patients with mild cognitive impairment (MCI), which is considered a preclinical stage of AD, are at a high risk of suffering from AD within a short period [3]. A research survey [4] reports that the average annual conversion rate from MCI to AD is 5%–10%. Early detection and intervention for AD through screening MCI are essential for efficacious treatment, new drug development, and delaying disease deterioration [5].

Structural magnetic resonance imaging (sMRI) has been widely used in the computer-aided diagnosis of AD to capture profound brain changes in the neurodegenerative process. Furthermore, deep-learning (DL) methods for medical image analysis, especially deep convolutional neural networks (CNNs), have been applied to various medical imaging modalities, including sMRI, X-rays, computed tomography (CT), ultrasound, and endoscopy. DL methods yield a significant improvement in comparison with conventional learning methods using hand-crafted features owing to their superiority of autonomous feature learning [6], [7], [8], [9], [10].

Ensemble learning (EL) has offered significant advantages in the generalizability and robustness via integrating multiple learning systems [11], which has also been applied to MRI. Li et al. [12] propose an EL model, which divides the whole-brain sMRI into 27 regions and groups the patches in each region using the k -means clustering algorithm. The grouped patches are utilized to train a DenseNet model [13]. The learned characteristics of local patches are integrated to classify AD finally. In addition, Pan et al. [14] develop a “2DCNN (2-D CNNs) +EL” model based on an established eight-layer CNN structure [15] and employ a data-driven and homogeneous EL approach for AD classification. Plus, a “2DCNN+EL+GA” model [16] is designed on the basis of the “2DCNN+EL” model and genetic algorithm (GA). A recent study [17] proposes a dual-task learning method based on separable 3-D CNN (3DCNN), which uses sMRI, demographics, neuropsychological, and genetic data as inputs to predict the MCI-to-AD conversion. Zhu et al. [18] advocate a dual-attention multiinstance deep learning network (DA-MIDL) for detecting early AD and identifying discriminative pathological locations.

This study proposes an adaptive interpretable ensemble model that combines GA and 3DCNN (3DCNN+EL+GA for short) to further boost the explainable diagnostic analysis of AD. The proposed method effectively identifies AD-related discriminative brain regions, which contribute the most to AD classification. Here, all the subjects retrieved from the AD Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) are divided into four different groups, i.e.: 1) AD; 2) MCIC (MCI patients who will convert to AD); 3) MCInc (MCI patients who will not convert to AD); and 4) normal control (NC). Three binary classification tasks are involved, i.e.: 1) AD versus NC; 2) MCIC versus NC; and 3) MCIC versus MCInc. Specifically, a whole-brain MRI image is divided into 246 brain regions, i.e., regions of interest (ROIs), according to the Brainnetome Atlas [19] (<http://atlas.brainnetome.org>). We train a corresponding candidate base classifier, i.e., a 3DCNN model, with each brain-region dataset. GA is then implemented to search for the best combination of base classifiers with optimal generalization ability on the validation dataset. Based on it, the whole-brain MRI classifier ensemble is built to detect AD/MCI on the testing datasets. The experiments on the ADNI database strictly follow the stratified fivefold cross-validation procedure. Moreover, a multicenter research trial is conducted for a binary classification task of AD versus NC. In detail, the models trained with ADNI datasets are examined on the dataset from the Open Access Series of Imaging Studies (OASIS) (<http://www.oasis-brains.org>). In addition, owing to the one-to-one correspondence between a base classifier and a brain region, we also identify discriminative brain regions with significant classification capabilities, while the generalization ability of the whole-brain MRI classifier ensemble is maximized in a data-driven manner. And besides, the behavior domains associated with these AD-related discriminative brain regions are analyzed to verify the effectiveness of the proposed method. Plus, the discriminative brain subregions at a voxel level are further located in these acquired brain regions with a

TABLE I
CHARACTERISTICS OF PARTICIPANTS IN THE TRAINING/TESTING DATASET FROM THE ADNI DATABASE

Variable	AD	MCIC	MCInc	NC
N	137	76	134	162
Gender (male:female)	67:70	43:33	84:50	86:76
Age (year; mean, std)	76.0, 7.3	74.8, 7.3	74.5, 7.2	76.3, 5.4
Weight (kg; mean, std)	70.9, 14.0	72.7, 14.3	76.2, 12.9	73.8, 13.6
MMSE (mean, std)	23.2, 2.0	26.47, 1.84	27.19, 1.71	29.18, 0.96
CDR (mean, std)	0.75, 0.25	0.50, 0.00	0.50, 0.00	0.00, 0.00
GDS (mean, std)	1.59, 1.32	1.38, 1.14	1.52, 1.37	0.80, 1.08

gradient-based attribution method designed for CNNs and also illustrated intuitively. For comparison, the previous studies and experimental results are discussed to facilitate a more objective evaluation of this study.

II. MATERIALS AND METHODS

A. Participants and Datasets

Data partially used in this article were obtained from the ADNI database. The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, M.D. The primary goal of ADNI is to test if serial MRI, positron emission tomography (PET), and other clinical and neuropsychological assessments would be combined to measure the progression of MCI and early AD.

Our experiments are conducted in the same MRI datasets as used in [20] to facilitate comparison with previous work. That is, 137 patients with AD, 76 patients with MCIC (who converted to AD within 18 months), 134 patients with MCInc (who did not convert to AD within 18 months), and 162 cognitively normal elderly controls (NC) (i.e., n = 509 subjects: AD = 137, 18 months MCIC = 76 and MCInc = 134, and NC = 162; Table I) are involved in this study. By following the protocol in [20], MCI patients who have been followed for less than 18 months are excluded here.

To facilitate model development and avoid data leakage, an additional validation dataset with 278 subjects (AD = 100, 36 months MCIC = 39 and MCInc = 39, and NC = 100; Table II) is retrieved from the ADNI database. The validation dataset is the same as the corresponding dataset used in [14]. The subjects in the validation dataset are nonoverlapped with those used in training and/or testing the models (i.e., the above-mentioned 509 subjects). The gender, age, weight, mini-mental state examination (MMSE), clinical dementia rating (CDR), and geriatric depressions scale (GDS) of the subjects in the training/testing and validation datasets are presented in rows two to seven in Tables I and II, respectively. The descriptions of the ADNI subject cohorts, image acquisition protocols, and postacquisition preprocessing procedures can be found at <http://www.adni-info.org>.

Moreover, we retrieve 196 sMRI images of 196 subjects (AD = 105 and NC = 91; Table III) from the OASIS database utilized in [21] to further conduct a multicenter research trial. With the OASIS dataset, the 3DCNN+EL+GA model trained on the ADNI dataset is examined and compared with other models (e.g., 2DCNN+EL and 2DCNN+EL+GA) also trained on the ADNI dataset to verify its generalization capability.

TABLE II
CHARACTERISTICS OF PARTICIPANTS IN THE VALIDATION DATASET FROM THE ADNI DATABASE

Variable	AD	MCIc	MCInc	NC
N	100	39	39	100
Gender (male:female)	60:40	23:16	29:10	45:55
Age (year; mean, std)	74.24, 7.82	74.15, 7.10	76.02, 7.00	73.36, 5.70
Weight (kg; mean, std)	76.04, 15.83	73.59, 14.14	78.35, 12.99	76.16, 15.66
MMSE (mean, std)	23.84, 2.08	27.05, 1.59	27.56, 1.83	28.92, 1.25
CDR (mean, std)	0.82, 0.24	0.50, 0.00	0.50, 0.00	0.00, 0.00
GDS (mean, std)	1.81, 1.56	1.92, 1.35	1.79, 1.45	0.83, 1.34

TABLE III
CHARACTERISTICS OF PARTICIPANTS FROM THE OASIS DATABASE

Variable	AD	NC
N	105	91
Gender (male:female)	58:47	48:43
Age (year; mean, std)	75.79, 7.56	73.74, 9.47
MMSE (mean, std)	23.19, 5.19	29.05, 1.20
CDR (mean, std)	0.86, 0.51	0.0, 0.0

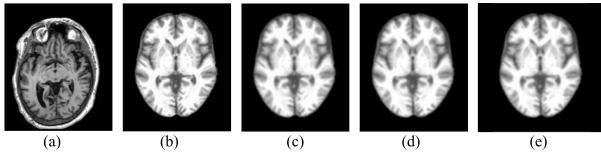


Fig. 1. Preprocessing pipeline. (a) Original image. (b) Skull-stripping and spatial normalization. (c) Smoothing. (d) Gray normalization. (e) Slicing and resizing.

B. MRI Preprocessing

Upon downloading, the T1-weighted MRI data in .nii format are preprocessed using the CAT12 toolkit (<http://www.neuro.uni-jena.de/cat/>) with the default setting. The preprocessing pipeline, which includes skull stripping, registration to the Montreal Neurological Institute (MNI) space, and image smoothing, is shown in Fig. 1. Thus, after preprocessing, all the images have a dimension of $91 \times 109 \times 91$ ($X \times Y \times Z$) with a spatial resolution of $2 \times 2 \times 2$ mm 3 per voxel. The values on the x -, y -, and z -axes are $\{90, 88, 86, \dots, -88, -90\}$, $\{-126, -124, -122, \dots, 88, 90\}$, and $\{-72, -70, -68, \dots, 106, 108\}$, respectively. For example, the coordinates of the two voxels $(0, 0, 0)$ and $(1, 0, 0)$ in the coordinate system of $91 \times 109 \times 91$ are corresponding to the coordinates $(90, -126, -72)$ and $(88, -126, -72)$ in the MNI coordinate system, respectively. The MRI signal intensity value of each voxel is normalized as the original signal intensity value divided by the original maximal signal intensity value of the image, yielding a value between 0 and 1.

The Brainnetome Atlas containing 246 brain regions of the bilateral hemispheres is employed to extract the brain regions (ROIs) in the preprocessed sMRI images. A specific ROI image (e.g., the left caudal hippocampus, L.cHipp for short) is extracted with its corresponding brain region mask (e.g., No. 217 brain region in the Brainnetome Atlas [19]). Here, the prefix capital letters L and R of a brain region label (e.g., L.cHipp) refer to the left and right cerebral hemispheres, respectively. As an example, Fig. 2 illustrates the slice images of three brain regions, i.e., left caudal hippocampus, right area

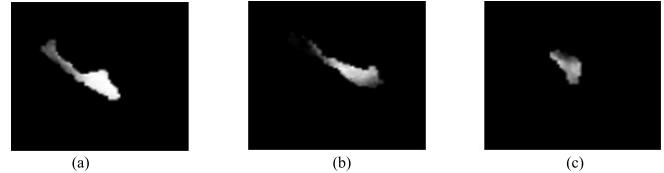


Fig. 2. Slice images of three brain regions. (a) L.cHipp. (b) R.A28/34. (c) L.mAmyg.

28/34 (R.A28/34 for short, No. 116), and left medial amygdala (L.mAmyg for short, No. 211).

C. Convolutional Neural Network

In recent years, CNN-based image recognition has attracted extensive research attention. Hubel and Wiesel [22] depicted the receptive fields, binocular interaction, and the functional structure of the primary visual cortex in cats. Later on, Fukushima and Miyake [23] advocated an artificial neural network named “neocognitron,” which was structurally similar to the hierarchy model of the visual nervous system proposed by Hubel and Wiesel [22]. This unique network architecture decreases the complexity of the neural networks, becoming a characteristic of CNN models. Each input image goes through a series of convolutional layers: filtering layers (kernels), pooling layers, and fully connected layers (FCs). The outputs of FC layers are computed as probability with which the input image is classified as a subject category (e.g., AD) using a softmax function. The weights in CNN models are learned on a training dataset, making CNN appropriate for automatically learning image representations [24].

Recent research [25] has indicated that the effectiveness of DL networks is improved with the increase of network depth while keeping the receptive field unchanged. Inspired by this finding, we design a 3DCNN model, as shown in Fig. 3. The 3DCNN model comprises three convolutional modules, i.e., Conv1, Conv2, and Conv3, and three FC layers, i.e., FC1, FC2, and FC3. Each convolutional module consists of two or three convolutional layers (e.g., Conv1_1 and Conv1_2 in module Conv1) and a pooling layer (e.g., Maxpool1 in module Conv1). All kernel sizes in the convolutional modules are the same. The probabilities for the predicted labels are achieved using three FC layers followed by a softmax layer. The corresponding hyperparameters are detailed in Table IV.

D. GA-Based Ensemble Learning

EL algorithms, such as random forest [26], bagging [27], and boosting [28], are mainly applied to solving a learning

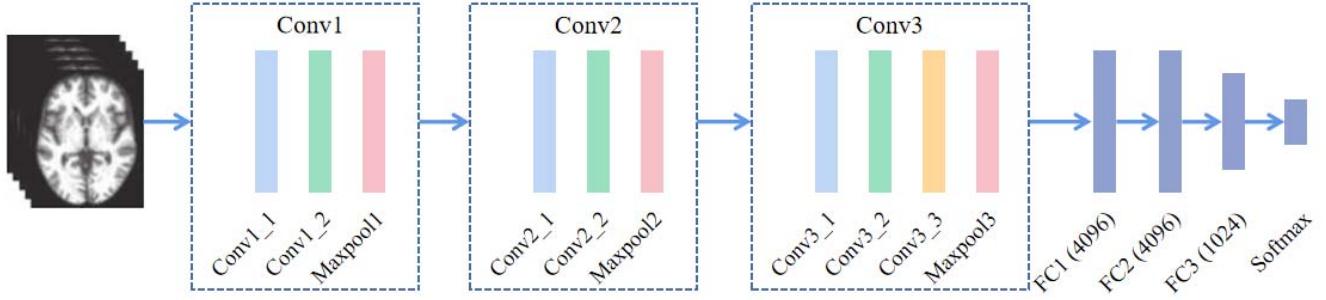


Fig. 3. Structure of proposed 3DCNN model.

TABLE IV
DETAILED HYPERPARAMETERS OF 3DCNN MODEL

Layer Name	Sub-layer Name	Kernel Size	Strides	Filters	Output Size
MRI_images	N/A	N/A	N/A	N/A	91x109x91
Conv1	Conv1_1	3x3x3	1	8	91x109x91
Conv1	Conv1_2	3x3x3	1	8	91x109x91
Conv1	Maxpool1	3x3x3	2	N/A	45x54x45
Conv2	Conv2_1	3x3x3	1	8	45x54x45
Conv2	Conv2_2	3x3x3	1	8	45x54x45
Conv2	Maxpool2	3x3x3	2	N/A	22x27x22
Conv3	Conv3_1	3x3x3	1	8	22x27x22
Conv3	Conv3_2	3x3x3	1	8	22x27x22
Conv3	Conv3_3	3x3x3	1	8	22x27x22
Conv3	Maxpool3	3x3x3	2	N/A	11x13x11
FC1	N/A	N/A	N/A	N/A	4096
FC2	N/A	N/A	N/A	N/A	4096
FC3	N/A	N/A	N/A	N/A	1024
Softmax	N/A	N/A	N/A	N/A	2

task by combining the predictions from multiple base learners. EL can improve the generalizability and robustness of the learning system significantly. According to the ways of generating base classifiers, EL algorithms can be categorized into two groups: homogeneous EL and heterogeneous EL. The homogeneous EL employs the same learning algorithm with different hyperparameters or applies the same learning algorithm to different training data (instances/features) [29]. In contrast, the heterogeneous EL applies different learning algorithms to the same training data (instances/features). This study follows the group of the homogeneous EL, in which the same 3DCNN algorithm with the same hyperparameters is employed to train different base classifiers on different brain-region datasets. The base classifiers with the good generalization performance on the validation dataset are selected and combined to build the classifier ensemble using a simple majority voting scheme. While the classifier ensemble performs the best on the validation dataset, it will be utilized to obtain the classification results for the unseen testing samples.

Hence, building the optimal classifier ensemble is obviously another challenge to be addressed in this study. The GA [30] that relies on biological operations, such as mutation, crossover, and selection, are commonly used to generate high-quality solutions to optimization problems. Researches

have revealed that GA-generated classifier ensembles perform better than the classifiers generated with other techniques [16], [31], [32], [33]. Therefore, in this study, the GA is adopted to search for the best combination of base classifiers in the classifier ensemble using a simple majority voting scheme, adaptively. As conventional GA usually converges slowly, the same extreme fitness value could appear in all offspring and their parents. All offspring are prone to become similar to their parents, which could lead to the termination of evolution, i.e., premature convergence. Thus, we amplify the difference between the offspring and their parents using the following formulas since the difference in fitness values among individuals is often small in this study:

$$\mathbf{D} = \text{FitnessValue} - \min(\text{FitnessValue}) \quad (1)$$

$$p_i = d_i / \sum(\mathbf{D}) \quad (2)$$

where **FitnessValue** is a vector, consisting of the fitness values of all individuals; and d_i is an element in the vector \mathbf{D} . p_i is the probability of selecting the i th individual.

E. Classification Experiment

In this study, three binary classification experiments, i.e., AD versus NC, MCIc versus NC, and MCIc versus MCInc, are

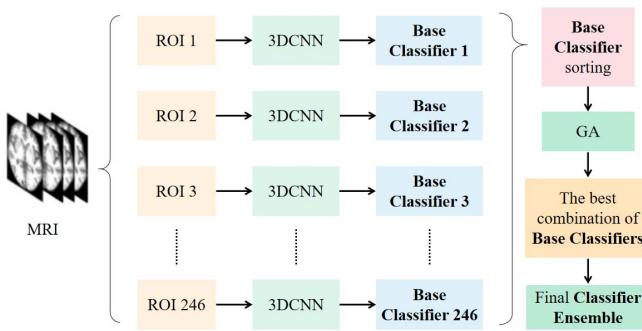


Fig. 4. 3DCNN+EL+GA: adaptive interpretable ensemble model which combines GA and 3DCNN.

conducted. The MRI images of 787 subjects from the ADNI database are partitioned into three datasets: training, testing, and validation datasets. The training/testing dataset ($n = 509$; Table I) is used to establish the base classifiers and examine the performance of the classifier ensemble. The validation dataset ($n = 278$; Table II) is used to evaluate and obtain the best base classifier combination. In addition, the OASIS dataset ($n = 196$; Table III) is employed to test the models trained on the ADNI datasets in the binary classification experiment of AD versus NC. A stratified fivefold cross-validation method is employed for all three binary experiments. That is, each binary classification task is repeatedly performed in five folds. Thus, no image in the testing dataset is applied to building the final classifier ensemble to avoid data leakage in each binary classification task.

Following the stratified fivefold cross-validation method, the training/testing MR images with labels are partitioned into training and testing sets in a ratio of 4:1 for each binary classification task. As described in Section II-B, 246 brain regions are extracted from each MR image with the help of the Brainnetome Atlas. 246 base classifiers (3DCNN models) are trained using the MRI images corresponding to 246 brain regions, respectively.

As shown in Fig. 4, the overall architecture of the proposed 3DCNN+EL+GA model mainly includes a set of 3DCNNs as candidate base classifiers and a GA component. The same 3DCNN algorithm with the same hyperparameters shown in Table IV is applied to different brain-region images to generate respective corresponding candidate base classifiers. In fact, every (candidate) **Base Classifier** (e.g., **Base Classifier 1**) in Fig. 4 is composed of the corresponding five base classifiers trained in the stratified fivefold cross-validation experiments in this study. In essence, a trained **Base Classifier** is corresponding to a brain region. Here, defined as the harmonic mean of precision and recall, the F1-score is calculated to evaluate the performance of each base classifier. Thus, the performance of a trained **Base Classifier** is assessed with its average fivefold F1 scores on the validation dataset, the mean value of the five F1 scores of the five trained base classifiers in a trained **Base Classifier** which is corresponding to a specific brain region.

We rank the trained **Base Classifiers** (i.e., brain regions) by their average fivefold F1 scores on the validation dataset in descending order.

And then, two 3DCNN+EL models, i.e., *vanilla* 3DCNN+EL and 3DCNN+EL+GA, are implemented. In *vanilla* 3DCNN+EL model, the top 30 trained **Base Classifiers** (i.e., brain regions) are selected after all the 246 trained **Base Classifiers** are ranked, using the method similar to the one employed to select the top five sagittal, coronal, and transverse slices in [14]. All of the top 30 trained **Base Classifiers** are ensembled to build the final **Classifier Ensemble**. In detail, for each fold, all of the 30 trained base classifiers corresponding to the top 30 brain regions are combined to construct the final classifier ensemble with a simple majority voting scheme. As a matter of fact, the **Classifier Ensemble** corresponding to the top 30 brain regions consists of the corresponding five classifier ensembles trained in the stratified fivefold cross-validation experiments in this study. Subsequently, the final classifier ensemble fuses the prediction results of multiple base classifiers using the corresponding brain-region image from an unlabeled MRI image as their respective inputs to jointly determine the label of the unlabeled MRI image. Finally, in order to evaluate the performance of the final **Classifier Ensemble**, the corresponding five classifier ensembles achieved in fivefold cross-validation are evaluated on the respective testing datasets, and the average fivefold performance indicators (e.g., F1 score) of the final **Classifier Ensemble**, i.e., the mean values of the respective performance indicators of the five classifier ensembles in the final **Classifier Ensemble**, are calculated.

While in the 3DCNN+EL+GA model, the GA component, just as exhibited in Fig. 4, is adopted to select the **Base Classifiers** (i.e., the corresponding brain regions) employed to form the **Classifier Ensemble**, i.e., the five majority-voting classifier ensembles in the fivefold cross validation, to enable the average fivefold prediction performance of the **Classifier Ensemble** to be the highest on the validation dataset. That is, the GA is employed to adaptively select **Base Classifiers** to maximize the performance (i.e., average fivefold F1 score) of the resultant **Classifier Ensemble** on the validation dataset. The process of GA optimization is shown in Fig. 5.

In detail, the 246 trained **Base Classifiers** are first sorted according to their average fivefold F1 scores on the validation dataset in descending order. The top 30 **Base Classifiers** are selected to perform chromosome encoding. A chromosome is a 0/1 string with a length of 30, that is, a gene in a chromosome denotes a **Base Classifier**. The **Base Classifier** corresponding to the gene that appears to be "1" is set as the selected while that corresponding to the gene appearing to be "0" is set as the nonselected. In fact, while a **Base Classifier** is selected, it means all the five base classifiers corresponding to the **Base Classifier** in the stratified fivefold cross-validation experiments are selected. After iterative optimization with the GA, the optimal combination of selected **Base Classifiers** is achieved to generate the final **Classifier Ensemble** and resultantly, to make a prediction on an unseen MRI image.

For the GA, we set the population number to 30. The algorithm is run for 6000 iterations at most. The crossover probability and mutation probability are set to 0.75 and 0.05, respectively. The F1-score is acted as the fitness value.

TABLE V
SELECTED TOP 30 BASE CLASSIFIERS IN AD VERSUS NC

Label of a brain region	Name of a brain region	Label ID	F1-score	Rank
L.mAmyg	medial amygdala	211	0.816413	1
L.cHipp	caudal hippocampus	217	0.787359	2
L.A35/36r	rostral area 35/36	109	0.777667	3
R.cHipp	caudal hippocampus	218	0.769729	4
L.A20rv	rostroventral area 20	103	0.768111	5
L.rHipp	rostral hippocampus	215	0.750821	6
R.lAmyg	lateral amygdala	214	0.749844	7
L.A28/34	area 28/34	115	0.748173	8
R.mAmyg	medial amygdala	212	0.746412	9
R.A35/36r	rostral area 35/36	110	0.730858	10
L.TI	area TI	117	0.726493	11
R.rHipp	rostral hippocampus	216	0.726468	12
R.A28/34	area 28/34	116	0.722993	13
L.A23v	ventral area 23	181	0.721838	14
R.TI	area TI	118	0.715240	15
R.A20rv	rostroventral area 20	104	0.713308	16
L.lAmyg	lateral amygdala	213	0.712451	17
R.A23d	dorsal area 23	176	0.694885	18
L.dId	dorsal dysgranular insula	173	0.685077	19
L.aSTS	anterior superior temporal sulcus	87	0.677440	20
R.rSTS	rostroposterior superior temporal sulcus	122	0.673347	21
R.TE1	TE1.0 and TE1.2	74	0.672133	22
R.A23v	ventral area 23	182	0.669497	23
L.dIa	dorsal agranular insula	167	0.666746	24
R.dId	dorsal dysgranular insula	174	0.661694	25
L.A38m	medial area 38	69	0.660550	26
R.dIg	dorsal granular insula	172	0.656389	27
R.aSTS	anterior superior temporal sulcus	88	0.651160	28
R.A24cd	caudodorsal area 24	184	0.649812	29
L.TE1	TE1.0 and TE1.2	73	0.648118	30

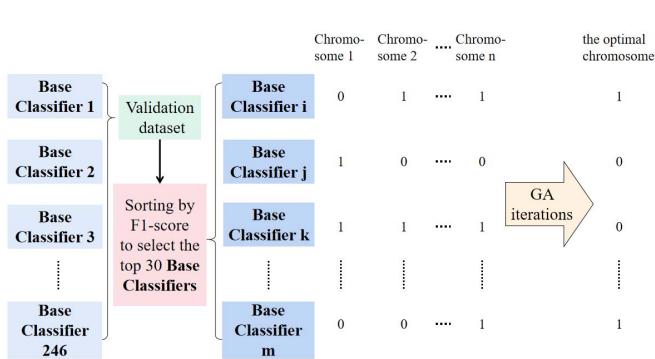


Fig. 5. Process of GA optimization.

All experiments are run on a GPU cluster with five nodes. Each node has two NVIDIA Tesla P100-PCIe-16-GB 250 W cards with 16 GB of video memory and a maximum power consumption of 250 W. On average, it takes 0.78 h to finish training a single-base classifier, and in general, the size of a trained base classifier is 361.83 MB. Hence, it takes about 3.9 h to accomplish a fivefold cross-validation experiment for each base classifier, and subsequently about 959.4 h

(approximately 40 days) to complete the training of all 246 base classifiers in each binary classification task. In this way, it totally takes about 120 days to accomplish all three binary classification tasks.

III. RESULTS

A. Base Classifiers

The trained **Base Classifiers** are first ranked by their average fivefold F1 scores on the validation dataset in descending order. The top 30 **Base Classifiers** and corresponding average fivefold F1 scores (i.e., F1-score in Tables V–VII) calculated in the stratified fivefold cross-validation experiments for the three binary classification tasks (i.e., AD versus NC, MCIc versus NC, and MCIc versus MCInc) are shown in Tables V–VII, respectively. Here, Label ID is the number of a brain region in the Brainnetome Atlas [19], e.g., No. 217 brain region (the left caudal hippocampus).

B. Classification Performance

The experiment is conducted using a stratified fivefold cross-validation procedure. Five performance indicators, namely, classification accuracy, recall, precision, F1-score,

TABLE VI
SELECTED TOP 30 **BASE CLASSIFIERS** IN MCIC VERSUS NC

Label of a brain region	Name of a brain region	Label ID	F1-score	Rank
R.rHipp	rostral hippocampus	216	0.752519	1
R.cHipp	caudal hippocampus	218	0.747415	2
L.rHipp	rostral hippocampus	215	0.744858	3
L.A28/34	area 28/34	115	0.735906	4
R.mAmyg	medial amygdala	212	0.716722	5
L.cHipp	caudal hippocampus	217	0.716470	6
L.mAmyg	medial amygdala	211	0.710065	7
L.TI	area TI	117	0.702743	8
L.rLinG	rostral lingual gyrus	195	0.698902	9
L.A35/36r	rostral area 35/36	109	0.694025	10
L.A13	area 13	49	0.693693	11
R.lAmyg	lateral amygdala	214	0.683823	12
L.A20rv	rostroventral area 20	103	0.680483	13
R.TI	area TI	118	0.672800	14
L.G	hypergranular insula	163	0.665549	15
L.A23v	ventral area 23	181	0.660481	16
R.A20rv	rostroventral area 20	104	0.649841	17
R.A23d	dorsal area 23	176	0.648055	18
L.A23d	dorsal area 23	175	0.645810	19
R.G	hypergranular insula	164	0.641691	20
R.A13	area 13	50	0.639952	21
L.vId/vIg	ventral dysgranular and granular insula	169	0.629449	22
R.A35/36r	rostral area 35/36	110	0.626337	23
R.A38m	medial area 38	70	0.616691	24
R.A31	area 31 (Lc1)	154	0.616233	25
R.A28/34	area 28/34	116	0.615691	26
R.rpSTS	rostroposterior superior temporal sulcus	122	0.614453	27
R.TE1	TE1.0 and TE1.2	74	0.605951	28
R.vmPOS	ventromedial parietooccipital sulcus	198	0.603951	29
L.TE1	TE1.0 and TE1.2	73	0.601435	30

and area under curve (AUC), are applied to evaluate the performance of the achieved final **Classifier Ensemble** with 3DCNN+EL+GA. As shown in Table VIII, its average classification accuracy for all the three classification tasks, i.e., AD versus NC, MCIC versus NC, and MCIC versus MCInc are 89%, 88%, and 71%, respectively. The proposed method outperforms PCA+SVM [20], 2DCNN [15], 2DCNN+EL [14], and 2DCNN+EL+GA [16] in the previous studies and *vanilla* 3DCNN+EL, 3DCNN+EL+Algo.1, 3DCNN+EL+PSO (particle swarm optimization [34]) and 3DCNN+EL+PSO2019 (Phasor PSO [35]) in terms of average accuracy and F1-score for all the three classification tasks. In the 3DCNN+EL+Algo.1, an algorithm for combinatorial optimization of base classifiers in the classifier ensemble (i.e., Algo.1), as shown in Algorithm 1, is advocated to search for the best combination of trained base classifiers, which is performing the best on the validation dataset, to build the classifier ensemble. As well, *vanilla* PSO [34] and Phasor PSO (i.e., PSO2019) [35] are utilized to optimize the **Classifier Ensemble** through adaptively selecting the trained **Base Classifiers** in the 3DCNN+EL+PSO and the 3DCNN+EL+PSO2019, respectively. Moreover, the standard deviations of the results achieved with 3DCNN+EL+GA are 0.03 and 0.02 for the

classification tasks of AD versus NC and MCIC versus NC, respectively, which are approximately a quarter of those obtained by the PCA+SVM [20].

More importantly, when the OASIS dataset, shown in Table III, is utilized as the testing dataset in the binary classification experiment of AD versus NC, the average classification accuracy and the average F1-score of 3DCNN+EL+GA are 76% with the standard deviation of 0.02 and 75% with the standard deviation of 0.02, respectively. As shown in Table IX, these results are superior to those achieved with other models (i.e., 2DCNN+EL, 2DCNN+EL+GA, *vanilla* 3DCNN+EL, 3DCNN+EL+Algo.1, and 3DCNN+EL+PSO) trained on the ADNI datasets and are comparable with those obtained with 3DCNN+EL+PSO2019 on the ADNI datasets. This shows that the proposed 3DCNN+EL+GA still achieves satisfactory generalization performance in the multicenter research trail.

C. Discriminative Brain Regions

Since each base classifier is obtained on each brain-region dataset, a trained base classifier corresponds to a brain region. The brain regions corresponding to the **Base Classifiers** in an achieved final **Classifier Ensemble** are regarded as the discriminative ones that significantly contribute to the

TABLE VII
SELECTED TOP 30 BASE CLASSIFIERS IN MCIC VERSUS MCINC

Label of a brain region	Name of a brain region	Label ID	F1-score	Rank
R.mAmyg	medial amygdala	212	0.664233	1
R.TI	area TI	118	0.626484	2
L.mAmyg	medial amygdala	211	0.622117	3
R.A23d	dorsal area 23	176	0.588003	4
L.cHipp	caudal hippocampus	217	0.578937	5
L.A44v	ventral area 44	39	0.577687	6
L.A4hf	area 4(head and face region)	53	0.575860	7
R.A21c	caudal area 21	82	0.569334	8
L.A31	area 31 (Lc1)	153	0.561922	9
R.A39c	caudal area 39(PGp)	136	0.561899	10
R.A40rv	rostroventral area 40(PFop)	146	0.556160	11
L.A10l	lateral area10	27	0.555881	12
R.IFS	inferior frontal sulcus	32	0.553141	13
R.A1/2/3ulhf	area 1/2/3(upper limb, head and face region)	156	0.552168	14
R.cHipp	caudal hippocampus	218	0.551094	15
L.A20cl	caudolateral of area 20	99	0.550876	16
L.dmPOS	dorsomedial parietooccipital sulcus(PEr)	151	0.549805	17
L.A38m	medial area 38	69	0.549456	18
R.A38m	medial area 38	70	0.549181	19
L.A37mv	medioventral area 37	105	0.547791	20
L.lsOccG	lateral superior occipital gyrus	209	0.547420	21
L.A37lv	lateroventral area 37	107	0.547245	22
L.rHipp	rostral hippocampus	215	0.547137	23
R.A12/47l	lateral area 12/47	52	0.544981	24
L.A6dl	dorsolateral area 6	7	0.543966	25
L.A8vl	ventrolateral area 8	23	0.543307	26
L.rLinG	rostral lingual gyrus	195	0.540670	27
R.A31	area 31 (Lc1)	154	0.540522	28
L.A41/42	area 41/42	71	0.540254	29
R.rHipp	rostral hippocampus	216	0.537359	30

classification performance of the final **Classifier Ensemble**. Thus, in the classification task of AD versus NC, the discriminative brain regions are the rostral hippocampus, caudal hippocampus, medial amygdala, lateral amygdala, rostral area 35/36, anterior superior temporal sulcus, and dorsal area 23. In the classification task of MCIC versus NC, rostral hippocampus, rostral area 35/36 [36], caudal hippocampus, medial amygdala, and area 28/34 [37] are explored as the discriminative brain regions. In the binary classification task of MCIC versus MCInc, GA screens the caudal hippocampus, rostral hippocampus, medial amygdala, medioventral area 37, dorsal area 23, rostroventral area 40, and so on. The details are shown in Tables X–XII.

The discriminative brain regions obtained with 3DCNN+EL+GA are similar to those acquired with 2DCNN+EL [14] to a great extent, though a couple of additional discriminative brain regions are observed with 3DCNN+EL+GA. The additional discriminative brain regions in the binary classification task of AD versus NC include rostral area 35/36, dorsal area 23, and anterior superior temporal sulcus. For the classification task of MCIC versus NC, rostral area 35/36, and area 28/34 are the newly

identified regions. In the MCIC versus MCInc classification experiment, more regions are newly found. They are area TI, dorsal area 23, caudal area 21, rostroventral area 40, inferior frontal sulcus, area 1/2/3 (upper limb, head, and face region), lateral superior occipital gyrus, and dorsolateral area 6. In all three binary classification tasks, the brain regions identified by the proposed 3DCNN+EL+GA contain known AD neuroimaging biomarkers, such as the hippocampus, amygdala, and temporal lobe.

Each discriminative brain region identified by the proposed 3DCNN+EL+GA method here is visualized with the BrainNet Viewer [38], as shown in Fig. 6. In addition, the likelihood ratio [39], which is utilized to measure the significant activation probabilities for each brain region with respect to a given behavioral domain, can be obtained from the hyperlink: <https://atlas.brainnetome.org/>, the official website of the Brainnetome Atlas, to analyze the functions associated with these discriminative brain regions. The bigger the likelihood ratio is, the more associated the behavioral domain is with the brain region. Thus, in each binary classification task (e.g., AD versus NC), for each related discriminative brain region (e.g., L.mAmyg), the top three behavioral domains

Algorithm 1 Algorithm for Combinatorial Optimization of Base Classifiers in the Classifier Ensemble

Require: The validation dataset D , the 246 trained base classifiers corresponding to 246 brain regions (ROIs), $X = \{X_r\}_{r=1,\dots,R}$, $R = 246$, the maximum times T of iterations.

Ensure: The best combination S_{best} of the trained base classifiers among X in the iterations, the prediction accuracy acc_{best} of the classifier ensemble based on the best combination using a simple majority voting scheme on the validation dataset D .

- 1: Initialize the required variables: the current combination of the trained base classifiers $S = \{\}$, the prediction accuracy $acc \leftarrow 0$ of the current combination (i.e., the classifier ensemble based on the current combination using a simple majority voting scheme) on the validation dataset D , combination change $flag \leftarrow false$, $S_{best} = \{\}$, $S_{temp} = \{\}$, $acc_{best} \leftarrow 0$, $acc_{single} \leftarrow 0$
- 2: **for** $i \leftarrow 1$ to T **do**
- 3: // Traversal of all X_r . Random traversal is recommended here.
- 4: **for** X_r in $(X - S)$ **do**
- 5: Calculate the accuracy acc_{single} of X_r on the validation dataset D
- 6: **if** $acc_{single} < acc - 0.1$ **then**
- 7: // If the accuracy acc_{single} of the X_r much lower than acc_{best} , the X_r is not considered
- 8: continue
- 9: **else if** $acc_{single} > acc$ **then**
- 10: // If the acc_{single} is higher than acc_{best} , add X_r into S unconditionally.
- 11: Add X_r into S , set $flag \leftarrow true$ and break
- 12: **else**
- 13: Copy S to S_{temp} , add X_r into S_{temp}
- 14: Calculate the accuracy acc_{temp} of this combination S_{temp} on the validation dataset D
- 15: **if** $acc_{temp} > acc$ **then**
- 16: Add X_r into S , set $flag \leftarrow true$ and break
- 17: **end if**
- 18: **end if**
- 19: **end for**
- 20:
- 21: // Judge whether the local optimal solution is reached according to whether the current combination S is updated by the above loops.
- 22: **if** $flag == true$ **then**
- 23: $flag \leftarrow false$
- 24: Iterate through the X_r in the combination S , trying to remove one or several X_r to improve the accuracy of the combination
- 25: Calculate the accuracy acc of the updated combination S on the validation dataset D , and update acc_{best} and S_{best}
- 26: **else**
- 27: // If the current combination is not changed after traversing all the X_r , it means that the combination is a local optimal solution.
- 28: Calculate the accuracy acc of the current combination S and update acc_{best} and S_{best}
- 29: // Consider randomly adding a couple of trained base classifiers X_r into S to jump out of the local optimal solution.
- 30: The three trained base classifiers randomly selected among $(X - S)$ are added into S
- 31: Calculate the accuracy acc of the updated combination S on the validation dataset D
- 32: **end if**
- 33: **end for**
- 34: **return** S_{best}, acc_{best}

corresponding to the three biggest likelihood ratios can be obtained, e.g., Interception.Sexuality, 8.96, Emotion.Disgust, 8.89, and Emotion.Happiness, 7.22. And then, for each behavioral domain (e.g., Emotion.Disgust), the sum of all the likelihood ratios corresponding to all related discriminative brain regions acquired in the above-mentioned way is calculated (Fig. 7). In Fig. 7, the vertical and horizontal axes exhibit the relevant behavioral domains and the sum of all the likelihood

ratios of discriminative brain regions associated with a specific related behavioral domain, respectively.

According to the results shown in Fig. 7, the functions related to the brain regions identified by the 3DCNN+EL+GA are mainly concentrated on behavioral domains, such as cognition, emotion, perception, memory, and language. A small number of brain regions are related to internal feelings and activities. At the same time, surveys have shown that memory

TABLE VIII
COMPARISON OF EXPERIMENTAL RESULTS IN ADNI DATASETS

Experiment	Method	Accuracy	AUC	Recall	Precision	F1-score
AD vs. NC	PCA+SVM [20]	0.76±0.11	—	—	—	—
	2DCNN [15]	0.77±0.07	—	—	—	—
	2DCNN+EL [14]	0.84±0.05	0.92±0.03	—	—	—
	2DCNN+EL+GA [16]	0.86±0.05	0.92±0.04	—	—	—
	<i>vanilla</i> 3DCNN+EL	0.86±0.02	0.86±0.02	0.81±0.05	0.80±0.05	0.86±0.02
	3DCNN+EL+Algo.1	0.83±0.02	0.85±0.04	0.84±0.04	0.80±0.03	0.82±0.02
	3DCNN+EL+PSO	0.86±0.03	0.85±0.03	0.80±0.07	0.89±0.08	0.86±0.03
	3DCNN+EL+PSO2019	0.86±0.05	0.85±0.06	0.75±0.10	0.92±0.04	0.85±0.06
MCIc vs. NC	3DCNN+EL+GA	0.89±0.03	0.88±0.04	0.85±0.09	0.90±0.03	0.88±0.03
	PCA+SVM [20]	0.72±0.12	—	—	—	—
	2DCNN [15]	0.74±0.07	—	—	—	—
	2DCNN+EL [14]	0.79±0.04	0.83±0.06	—	—	—
	2DCNN+EL+GA [16]	0.80±0.03	0.82±0.04	—	—	—
	<i>vanilla</i> 3DCNN+EL	0.86±0.05	0.87±0.05	0.88±0.09	0.74±0.07	0.85±0.05
	3DCNN+EL+Algo.1	0.78±0.04	0.73±0.08	0.62±0.11	0.68±0.09	0.63±0.07
	3DCNN+EL+PSO	0.85±0.05	0.83±0.08	0.76±0.11	0.77±0.06	0.83±0.06
MCInc vs. MCInc	3DCNN+EL+PSO2019	0.78±0.04	0.78±0.04	0.75±0.13	0.66±0.10	0.76±0.04
	3DCNN+EL+GA	0.88±0.03	0.87±0.02	0.84±0.05	0.81±0.08	0.87±0.03
	PCA+SVM [20]	0.66±0.16	—	—	—	—
	2DCNN [15]	0.55±0.09	—	—	—	—
	2DCNN+EL [14]	0.62±0.06	0.59±0.07	—	—	—
	2DCNN+EL+GA [16]	0.66±0.04	0.61±0.06	—	—	—
	<i>vanilla</i> 3DCNN+EL	0.68±0.06	0.67±0.05	0.63±0.06	0.56±0.10	0.66±0.06
	3DCNN+EL+Algo.1	0.61±0.04	0.80±0.14	0.74±0.18	0.47±0.03	0.57±0.07
MCInc vs. MCInc	3DCNN+EL+PSO	0.60±0.07	0.62±0.06	0.68±0.09	0.47±0.06	0.59±0.06
	3DCNN+EL+PSO2019	0.64±0.10	0.63±0.12	0.60±0.26	0.49±0.16	0.61±0.13
	3DCNN+EL+GA	0.71±0.08	0.70±0.07	0.65±0.11	0.61±0.12	0.69±0.08

TABLE IX
COMPARISON OF EXPERIMENTAL RESULTS IN OASIS DATASETS

Experiment	Method	Accuracy	AUC	Recall	Precision	F1-score
AD vs. NC	2DCNN+EL [14]	0.73±0.01	0.86±0.01	0.62±0.02	0.92±0.03	0.74±0.01
	2DCNN+EL+GA [16]	0.72±0.02	0.86±0.02	0.58±0.02	0.94±0.02	0.72±0.02
	<i>vanilla</i> 3DCNN+EL	0.74±0.03	0.75±0.02	0.58±0.05	0.89±0.01	0.74±0.03
	3DCNN+EL+Algo.1	0.74±0.005	0.52±0.02	0.56±0.01	0.91±0.01	0.69±0.009
	3DCNN+EL+PSO	0.75±0.01	0.74±0.01	0.64±0.04	0.78±0.04	0.74±0.01
	3DCNN+EL+PSO2019	0.77±0.02	0.76±0.02	0.62±0.05	0.84±0.02	0.76±0.02
	3DCNN+EL+GA	0.76±0.02	0.75±0.02	0.61±0.04	0.83±0.02	0.75±0.02

TABLE X

DETAILS OF DISCRIMINATIVE BRAIN REGIONS IN AD VERSUS NC

Label of a brain region	Name of a brain region	Rank of average five-fold F1-score
L.mAmyg	medial amygdala	1
L.cHipp	caudal hippocampus	2
L.A35/36r	rostral area 35/36	3
R.mAmyg	rostral hippocampus	9
L.lAmyg	lateral amygdala	17
R.A23d	dorsal area 23	18
R.aSTS	anterior superior temporal sulcus	28

loss, cognitive decline, decreased or poor judgment, emotional instability, and language impairment are the most common symptoms of AD patients in the clinical manifestations. Thus, it can be seen that the behavioral domains relevant to the brain regions identified by the proposed 3DCNN+EL+GA method here are basically consistent with the clinical manifestations of AD.

D. Discriminative Brain Subregions

In order to further locate the subregions with more contributions to the classification in the identified discriminative brain

TABLE XI

DETAILS OF DISCRIMINATIVE BRAIN REGIONS IN MCIC VERSUS NC

Label of a brain region	Name of a brain region	Rank of average five-fold F1-score
R.rHipp	rostral hippocampus	1
R.cHipp	caudal hippocampus	2
L.rHipp	rostral hippocampus	3
R.mAmyg	medial amygdala	5
L.mAmyg	medial amygdala	7
L.A35/36r	rostral area 35/36	10
L.A20rv	rostroventral area 20	13
R.A28/34	area 28/34	26

regions, e.g., rostral hippocampus, caudal hippocampus, and medial amygdala, we attempt to visualize the features captured by the base classifiers (i.e., 3DCNN) in the obtained classifier ensemble with gradient-weighted class activation mapping++ (Grad-CAM++) [40], i.e., generalized gradient-based visual explanations for deep convolutional networks. In fact, Grad-CAM++ has been successfully applied to 3D models [41]. Grad-CAM++ considers gradients too, but is based on a different mathematical formulation that enhances the localization of single and multiple instances, in comparison with

TABLE XII
DETAILS OF DISCRIMINATIVE BRAIN REGIONS IN MCIC VERSUS MCINC

Label of a brain region	Name of a brain region	Rank of average five-fold F1-score
R.mAmyg	medial amygdala	1
R.TI	area TI	2
L.mAmyg	medial amygdala	3
R.A23d	dorsal area 23	4
L.A4hf	area 4(head and face region)	7
R.A21c	caudal area 21	8
R.A40rv	rostroventral area 40(PFop)	11
R.IFS	inferior frontal sulcus	13
R.A1/2/3ulhf	area 1/2/3(upper limb, head and face region)	14
R.cHipp	caudal hippocampus	15
L.A20cl	caudolateral of area 20	16
L.A37mv	medioventral area 37	20
L.lsOccG	lateral superior occipital gyrus	21
L.rHipp	rostral hippocampus	23
L.A6dl	dorsolateral area 6	25

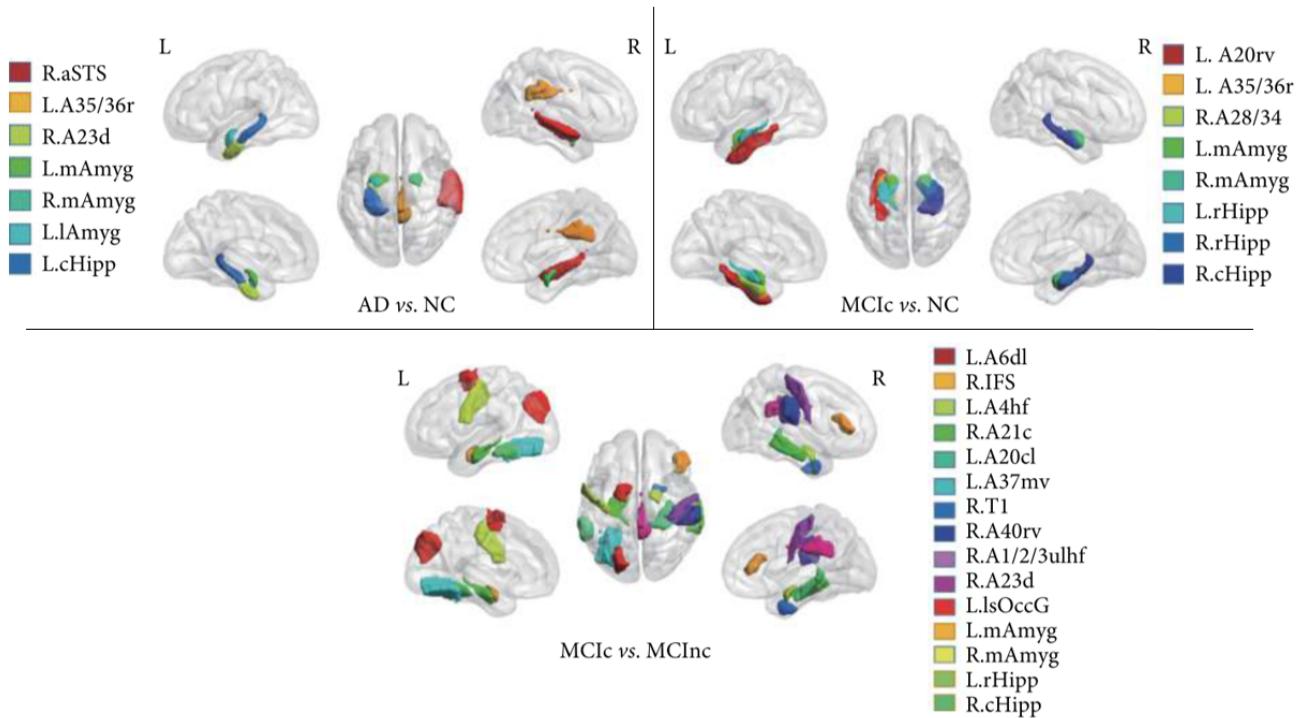


Fig. 6. Identified discriminative brain regions visualized with BrainNet Viewer.

Grad-CAM [42], a popular CAM [43]-based approach that utilizes backpropagation to score the feature maps' locations in a specific layer. First, the 7, 8, and 15 discriminative brain regions are identified with the 3DCNN+EL+GA in the three classification tasks, i.e., AD versus NC, MCIC versus NC, and MCIC versus MCInc, respectively. The trained base-classifiers (3DCNNs) corresponding to these brain regions are saved. For each binary classification task, all available MRI images are input into the five models acquired in the fivefold cross-validation procedure. Among these models, the model with the best generalization performance is utilized to locate the most discriminative subregions with Grad-CAM++.

Given an input image I and its output of the base classifier, i.e., the class confidence score (probability) for class c $S^c(I)$, the exponential operation is performed on $S^c(I)$, i.e.,

$Y^c(I) = \exp(S^c(I))$. The gradient weights of voxels in the l th feature map can be calculated as follows:

$$\alpha_{ijk}^{l_c}(I) = \frac{\left(\frac{\partial S^c(I)}{\partial A_{ijk}^l(I)}\right)^2}{2\left(\frac{\partial S^c(I)}{\partial A_{ijk}^l(I)}\right)^2 + \sum_a \sum_b \sum_c A_{abc}^l(I)\left(\frac{\partial S^c(I)}{\partial A_{ijk}^l(I)}\right)^3}. \quad (3)$$

In the equation earlier, $A_{ijk}^l(I)$ refers to the value of the voxel located on $(i, j, k) \in \mathbb{R}^{u,v,w}$ in the l th feature map calculated in the last convolutional layer, while image I is the input and $\{u, v, w\}$ denote the dimensions of the feature maps ($Z = u \times v \times w$). Plus, $(a, b, c) \in \mathbb{R}^{u,v,w}$. In this study, the 3DCNN model has 8 feature maps and $\{u, v, w\}$ equals

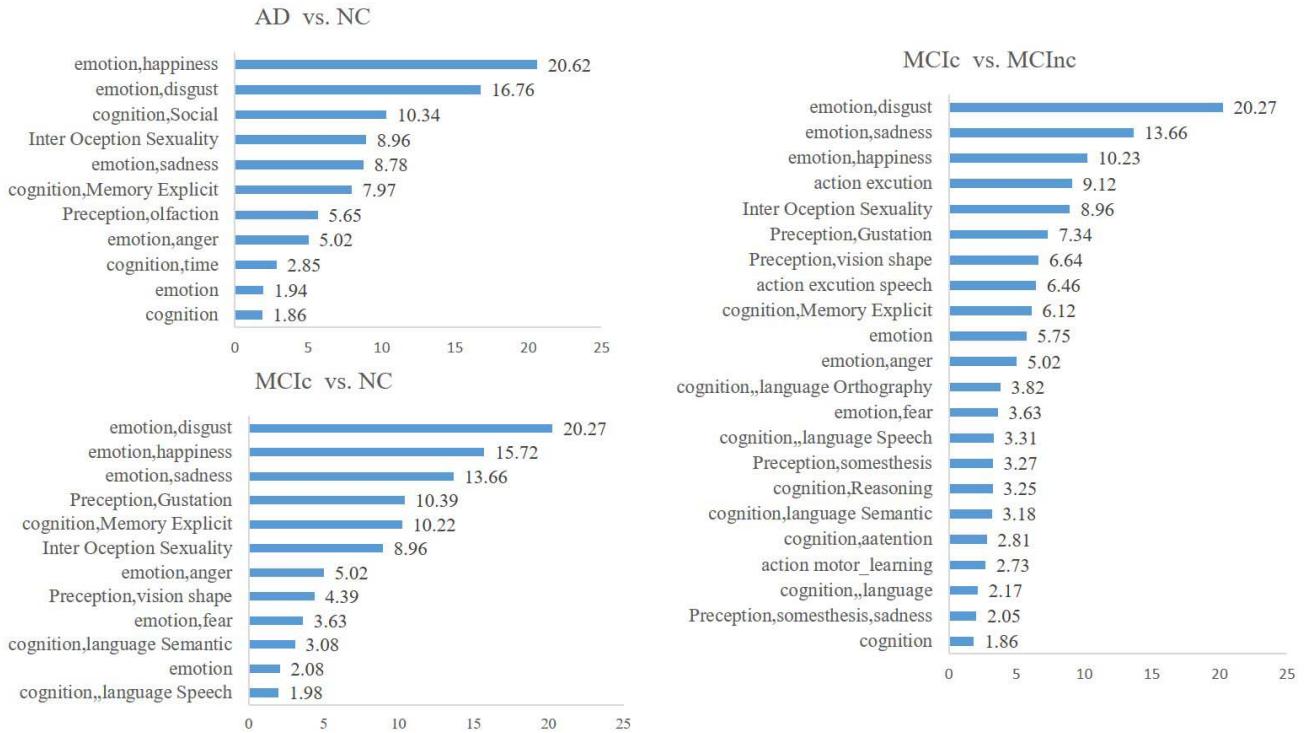


Fig. 7. Distribution of behavioral domains of discriminative brain regions in each binary classification task.

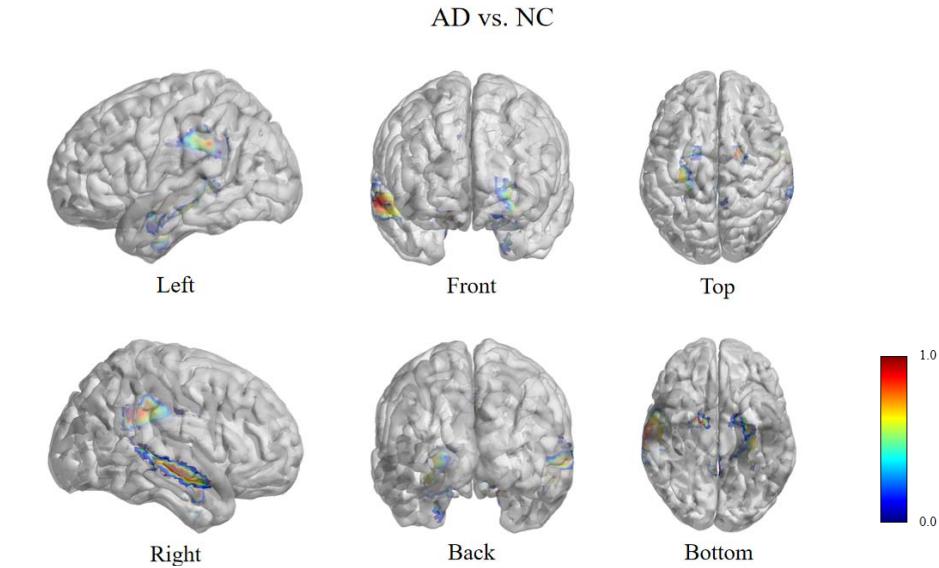


Fig. 8. Whole-brain heat map, including the seven discriminative brain regions, acquired with the 3DCNN+EL+GA in AD versus NC.

to $\{11, 13, 11\}$. And then, the gradient weights of the l th feature map for class C is defined as

$$w^l_c(I) = \sum_i \sum_j \sum_k \alpha_{ijk}^l(I) \cdot \text{relu}\left(\frac{\partial Y^c(I)}{\partial A_{ijk}^l(I)}\right) \quad (4)$$

which leads to

$$L_{ijk}^c(I) = \text{relu}\left(\sum_l w^l_c(I) \cdot A_{ijk}^l(I)\right). \quad (5)$$

Subsequently

$$L_{ijk}^c = \sum_n L_{ijk}^c(I). \quad (6)$$

Here, n is the number of MRI images in class c for a binary classification task. After that

$$L_{ijk}^{c,\text{normalized}} = L_{ijk}^c / \max_{(i,j,k) \in \mathbb{R}^{u,v,w}} (L_{ijk}^c). \quad (7)$$

The dimensions of $L_{ijk}^{c,\text{normalized}}$ are the same as those of the feature maps, which are usually smaller than the input image. Hence, the final Grad-CAM++ explanation map is

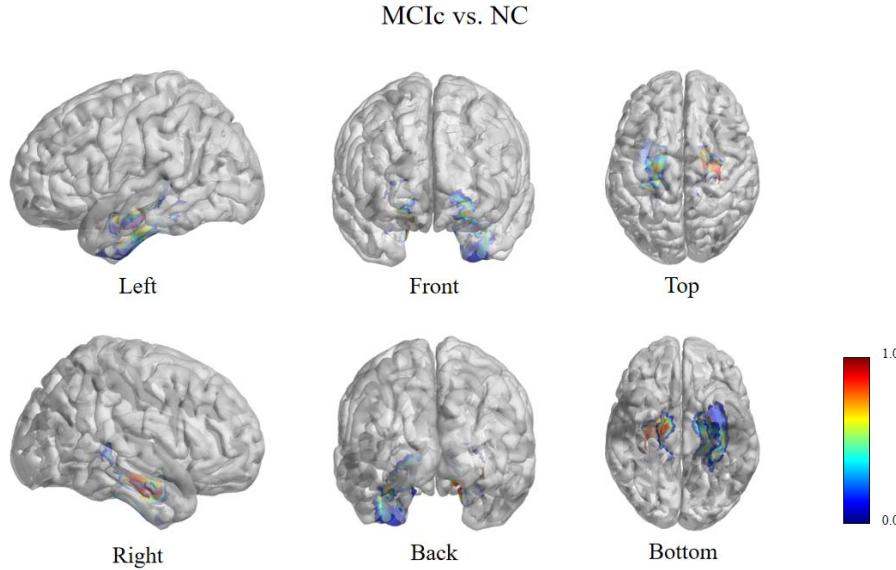


Fig. 9. Whole-brain heat map, including the eight discriminative brain regions, acquired with the 3DCNN+EL+GA in MCIc versus NC.

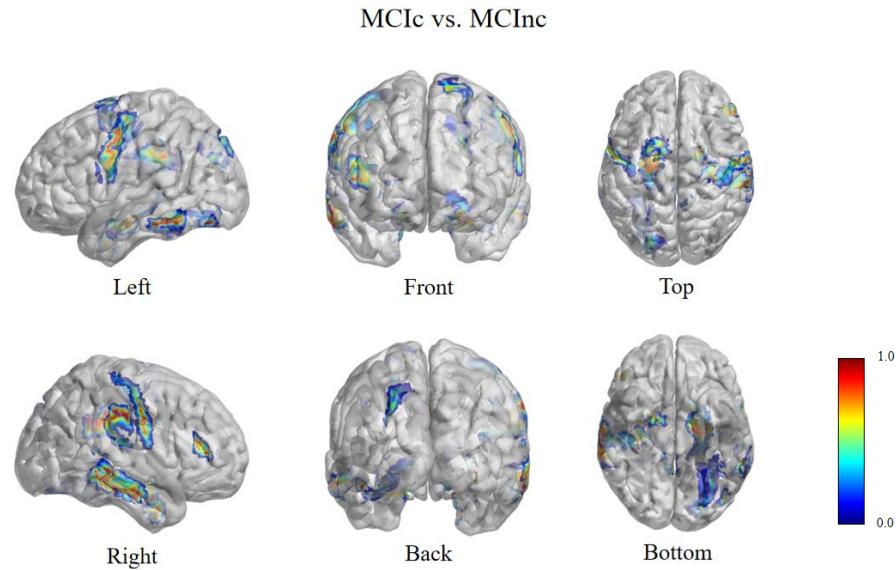


Fig. 10. Whole-brain heat map, including the 15 discriminative brain regions, acquired with the 3DCNN+EL+GA in MCIc versus MCInc.

obtained by upsampling L_{ijk}^c normalized to the size of image I through bilinear interpolation. Thus, the weight of each voxel in image I for class C , i.e., L_{xyz}^c , is acquired, where $(x, y, z) \in \mathbb{R}^{(X, Y, Z)}$ and $\{X, Y, Z\}$ denote the dimensions of an input image ($I = X \times Y \times Z$). In this study, $\{X, Y, Z\}$ equals to $\{91, 109, 91\}$. Based on it, a 3-D heatmap with the same size as the input image shows the subregions contributing the most to the classification for each base classifier, which is corresponding to a brain region. Here, while the value of L_{xyz}^c is closer to 0, the voxel located on (x, y, z) contributes less to the classification and its color is closer to blue. In contrast, while the value of L_{xyz}^c is closer to 1, the voxel located on (x, y, z) contributes more to the classification and its color is closer to red. That is, changes in color from blue to red on the pseudocolor scale represent weights of voxels ranging from 0 to 1. From the perspective of the

visualization of a whole-brain heatmap, the heat maps of the 7, 8, and 15 discriminative brain regions acquired in the binary classification tasks of AD versus NC, MCIc versus MCInc, and MCIc versus MCInc, are shown in Figs. 8–10, respectively, and the heat maps of these obtained brain regions in the three binary classification tasks are individually visualized in Figs. 11–13. As examples, the heat maps of the left and right medial amygdala are visualized in Fig. 14 and the weights of voxels in these two brain regions identified in the classification task of AD versus NC are shown in Tables XIII and XIV, respectively.

IV. CONCLUSION

In this study, we propose an adaptive interpretable ensemble model which combines 3DCNN and GA, i.e., 3DCNN+EL+GA, to achieve two goals: 1) more accurate

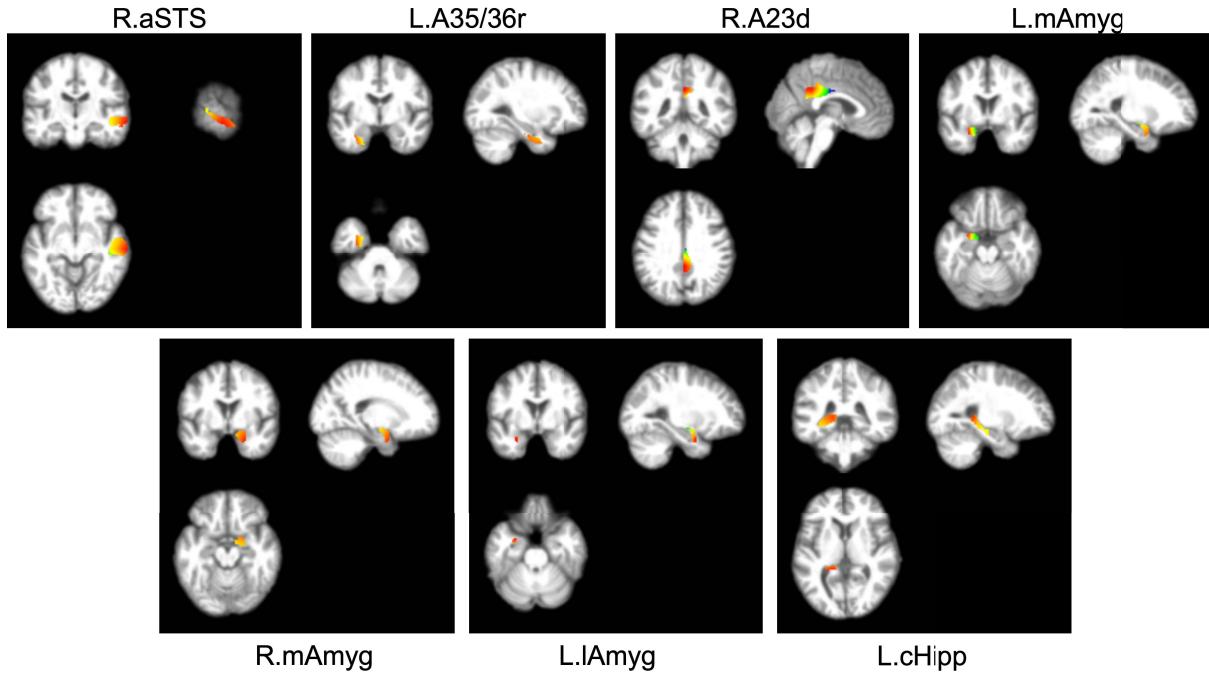


Fig. 11. Individual heat maps of the seven discriminative brain regions acquired with the 3DCNN+EL+GA in AD versus NC.

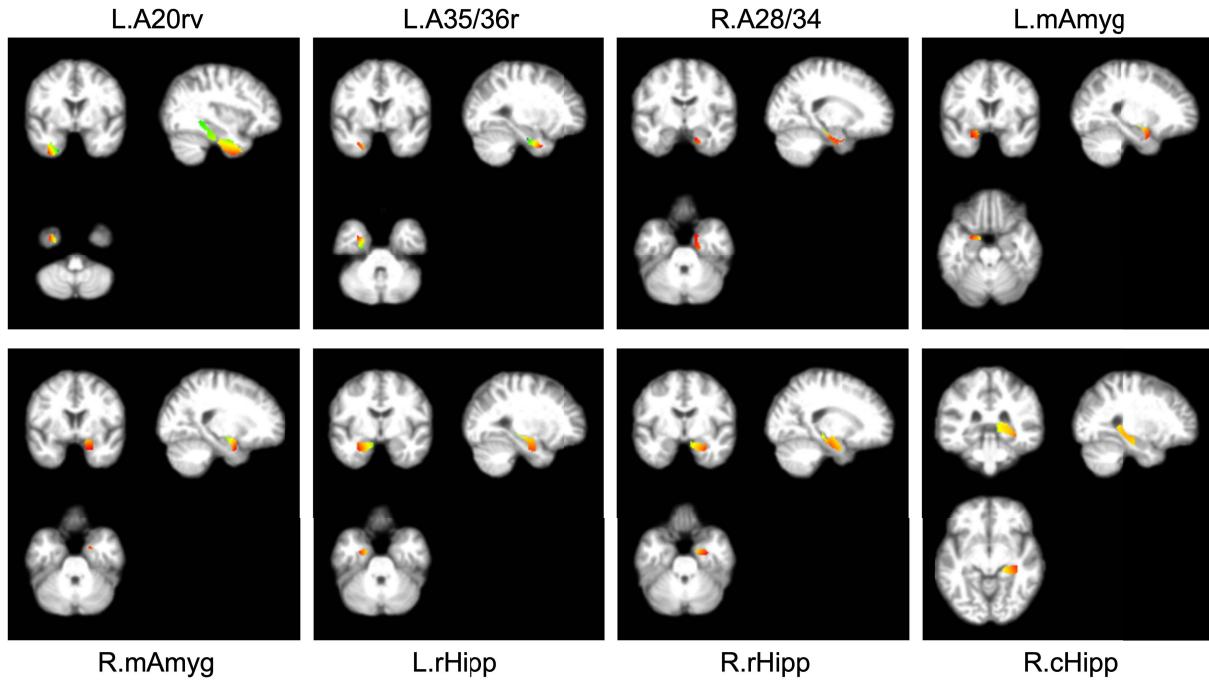


Fig. 12. Individual heat maps of the eight discriminative brain regions acquired with the 3DCNN+EL+GA in MCIc versus NC.

disease classification (AD versus NC, MCIc versus NC, and MCIc versus MCInc) and 2) the identification of discriminative brain regions that could be the biomarkers for early detection in AD progression. The 246 base classifiers (3DCNN models) are trained on the datasets from 246 brain regions, respectively, and are sorted by F1-score on the validation dataset. The GA is then employed to select the best combination of base classifiers

in a classifier ensemble using a simple majority voting scheme. The proposed method breaks through the limitation of manual selection of brain regions by automatically extracting the discriminative features from the MRI images in a data-driven way, which is more general and flexible than the previous CNN model [44] that merely selects hippocampus images according to prior knowledge for the model training.

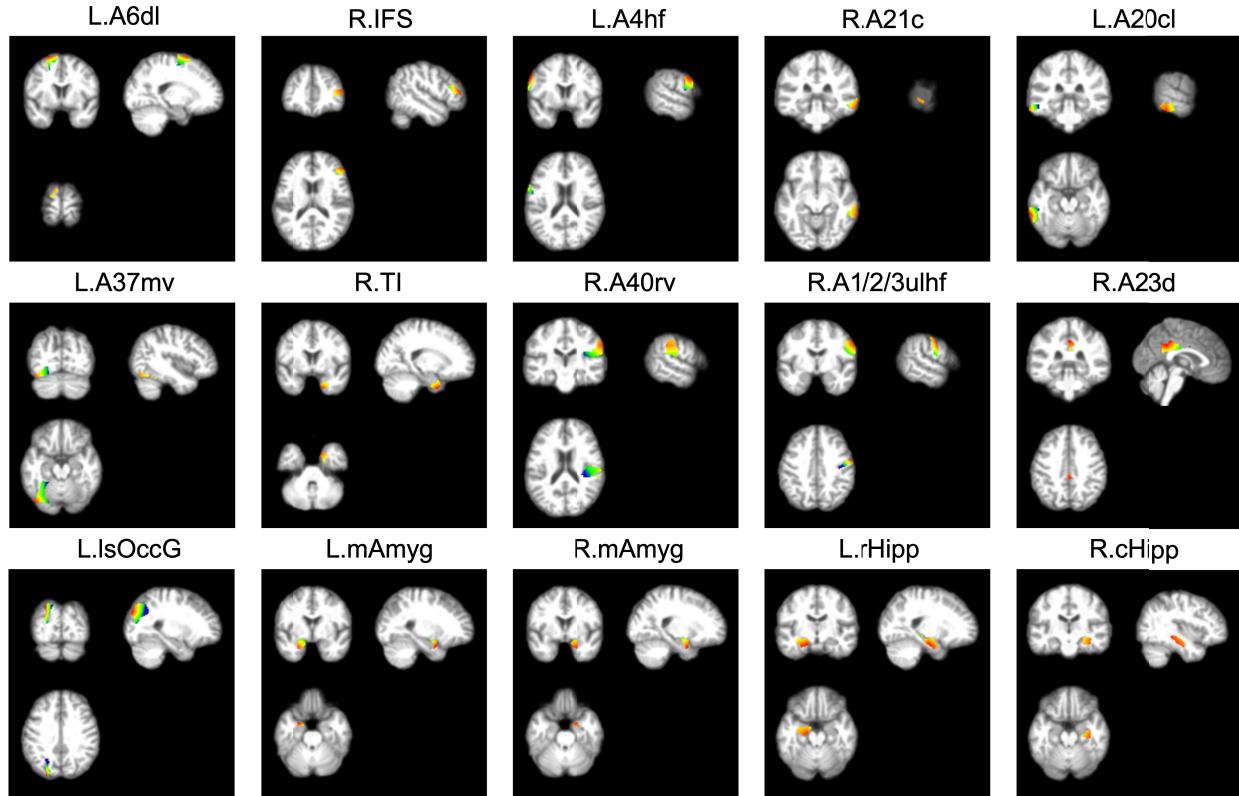


Fig. 13. Individual heat maps of the fifteen discriminative brain regions acquired with the 3DCNN+EL+GA in MCIc versus MCInc.

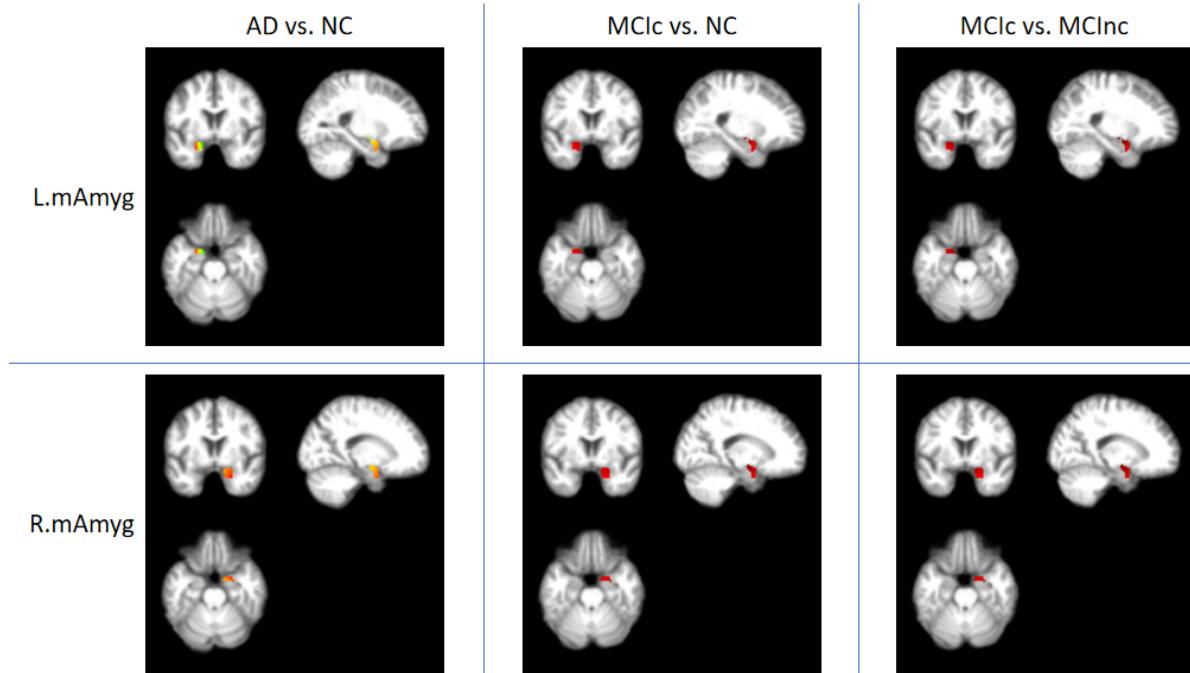


Fig. 14. Heat maps of the left medial amygdala (L.mAmyg) and the right medial amygdala (R.mAmyg) acquired with the 3DCNN+EL+GA in AD versus NC, MCIc versus NC, and MCIc versus MCInc, respectively.

Compared with previous methods, the advocated 3DCNN+EL+GA model has the following merits.

- 1) Under the premise of ROI extraction and EL, the proposed 3DCNN+EL+GA method achieves higher

accuracies than the other methods, e.g., PCA+SVM. Brain-region extraction effectively reduces feature dimensionality. Meanwhile, using a simple majority voting scheme, a classifier ensemble built upon the selected

TABLE XIII

WEIGHTS OF VOXELS IN THE LEFT MEDIAL AMYGDALA FROM AD MRI IMAGES IN THE BINARY CLASSIFICATION TASK OF AD VERSUS NC

Location	Weight	Location	Weight	Location	Weight	Location	Weight
(-26, 0, -22)	1	(-22, -2, -20)	0.75007832	(-18, -2, -22)	0.587150065	(-16, 2, -18)	0.442326078
(-26, 0, -20)	0.980988921	(-20, -2, -28)	0.741794141	(-18, 2, -20)	0.57719668	(-14, -2, -22)	0.442053315
(-24, 2, -26)	0.96910727	(-20, 0, -26)	0.740055448	(-18, 0, -20)	0.569534442	(-16, 0, -18)	0.441757438
(-26, 0, -18)	0.961977842	(-22, 0, -16)	0.732517679	(-20, -6, -16)	0.56948727	(-16, -2, -18)	0.441188797
(-24, 2, -24)	0.950094588	(-20, 2, -24)	0.731079964	(-24, -12, -12)	0.565402999	(-16, -4, -18)	0.440620157
(-24, 0, -28)	0.949807667	(-24, -6, -16)	0.7288373	(-16, -2, -28)	0.56323512	(-16, -6, -18)	0.440051517
(-24, 0, -26)	0.938699921	(-20, -2, -26)	0.727901382	(-18, -2, -20)	0.561872204	(-14, 0, -22)	0.435761992
(-24, 2, -22)	0.931081906	(-22, -2, -18)	0.727405434	(-16, 0, -28)	0.556004608	(-18, -6, -12)	0.433076783
(-24, -2, -28)	0.920353162	(-20, -2, -30)	0.725439058	(-18, 2, -18)	0.555008694	(-14, 2, -22)	0.42947067
(-24, 0, -24)	0.918508381	(-22, -4, -20)	0.724894958	(-18, -4, -20)	0.554209966	(-14, -4, -20)	0.420416175
(-24, 2, -20)	0.912069225	(-20, 0, -24)	0.717502986	(-16, 2, -28)	0.548774096	(-18, -8, -12)	0.419234795
(-24, -2, -26)	0.908292573	(-20, 2, -22)	0.709950413	(-16, -2, -26)	0.547510191	(-12, -2, -22)	0.418688843
(-24, -2, -30)	0.904484322	(-22, -2, -16)	0.704732548	(-18, 0, -18)	0.545801519	(-14, -2, -20)	0.415800084
(-24, 0, -22)	0.898316841	(-20, -2, -24)	0.703926009	(-20, -6, -14)	0.542666075	(-16, -2, -16)	0.414608449
(-24, 2, -18)	0.893056543	(-22, -4, -18)	0.700921187	(-16, 0, -26)	0.541410975	(-12, 0, -22)	0.412697596
(-24, -2, -24)	0.886922175	(-20, 0, -22)	0.694950524	(-22, -10, -12)	0.537839173	(-16, -4, -16)	0.412372845
(-24, 0, -20)	0.878125301	(-20, 2, -20)	0.688820862	(-18, -2, -18)	0.536594343	(-14, 0, -20)	0.411183992
(-24, -2, -22)	0.865551776	(-20, -2, -22)	0.679950635	(-16, 2, -26)	0.535311759	(-16, -6, -16)	0.41013724
(-22, 2, -26)	0.860658392	(-22, -4, -16)	0.676947417	(-18, -4, -18)	0.527387167	(-12, -4, -20)	0.398315739
(-24, 0, -18)	0.857933761	(-22, -6, -18)	0.674436941	(-16, -2, -24)	0.520929842	(-14, -6, -18)	0.395428573
(-22, 0, -28)	0.851356902	(-20, 0, -20)	0.672398062	(-18, -6, -18)	0.518179991	(-12, -2, -20)	0.393926733
(-24, -2, -20)	0.844181378	(-20, 2, -18)	0.667691311	(-16, 0, -24)	0.516497591	(-14, -4, -18)	0.392487713
(-22, 2, -24)	0.840587276	(-18, 2, -28)	0.656396115	(-16, 2, -24)	0.512065339	(-14, -2, -18)	0.389546853
(-22, 0, -26)	0.839377685	(-20, -2, -20)	0.655975262	(-18, -2, -16)	0.511316482	(-16, -2, -14)	0.388028101
(-24, 0, -16)	0.837742221	(-18, 0, -28)	0.654455373	(-22, -12, -12)	0.507452272	(-14, 0, -18)	0.386605992
(-22, -2, -28)	0.831073652	(-18, -2, -28)	0.652514631	(-18, -4, -16)	0.500564369	(-16, -4, -14)	0.384125532
(-24, -2, -18)	0.822810979	(-20, 0, -18)	0.6498456	(-14, -2, -26)	0.494559776	(-16, -6, -14)	0.380222964
(-22, 2, -22)	0.820516159	(-22, -6, -16)	0.649162285	(-16, -2, -22)	0.494349494	(-12, -6, -18)	0.374738153
(-22, -2, -26)	0.818096978	(-18, 2, -26)	0.643760637	(-20, -8, -12)	0.493730434	(-12, -4, -18)	0.371951388
(-22, 0, -24)	0.818005684	(-18, 0, -26)	0.640733212	(-16, 0, -22)	0.491584206	(-12, -2, -18)	0.369164623
(-22, -2, -30)	0.81496169	(-20, -4, -20)	0.639552462	(-18, -6, -16)	0.489812255	(-14, -6, -16)	0.365824881
(-24, -4, -20)	0.810237454	(-18, -2, -26)	0.637705786	(-16, 2, -22)	0.488818919	(-14, -4, -16)	0.364559251
(-22, 2, -20)	0.800445043	(-20, -2, -18)	0.631999888	(-18, -2, -14)	0.486038621	(-14, -2, -16)	0.363293622
(-22, 0, -22)	0.796633683	(-22, -6, -14)	0.62388763	(-18, -4, -14)	0.47374157	(-10, -6, -18)	0.354047732
(-22, -2, -24)	0.795424092	(-18, 2, -24)	0.621572651	(-20, -10, -12)	0.471615989	(-16, -6, -12)	0.350308688
(-24, -4, -18)	0.787688198	(-18, 0, -24)	0.617000288	(-16, -4, -20)	0.468867469	(-12, -6, -16)	0.34677156
(-22, 2, -18)	0.780373927	(-20, -4, -18)	0.614154177	(-14, -2, -24)	0.468306546	(-12, -4, -16)	0.345587037
(-22, 0, -20)	0.775261681	(-18, -2, -24)	0.612427925	(-16, -2, -20)	0.467769146	(-14, -2, -14)	0.337040391
(-22, -2, -22)	0.772751206	(-20, -2, -16)	0.608024515	(-16, 0, -20)	0.466670822	(-14, -4, -14)	0.336630789
(-22, 0, -18)	0.75388968	(-18, 2, -22)	0.599384666	(-16, 2, -20)	0.465572499	(-14, -6, -14)	0.336221188
(-20, 0, -28)	0.752906138	(-20, -6, -18)	0.596308466	(-18, -6, -14)	0.461444519	(-12, -6, -14)	0.318804968
(-24, -6, -18)	0.752565416	(-18, 0, -22)	0.593267365	(-14, 0, -24)	0.460339992		
(-20, 2, -26)	0.752209515	(-20, -4, -16)	0.588755893	(-12, -2, -24)	0.443450953		

base classifiers in a data-driven way can achieve better generalization performance, especially in the multicenter testing datasets, i.e., OASIS datasets.

- 2) Since the extracted ROIs are corresponding to the brain regions in the Brainnetome Atlas and each trained base classifier is corresponding to a specific brain region, the resultant majority-voting ensemble is interpretable. While the acquired ensemble achieves satisfactory classification generalizability, it could “tell” us which

brain regions are discriminative and informative in an intuitive way.

- 3) In comparison with other algorithms for the combinatorial optimization (e.g., PSO and Algo.1), the proposed GA-based method can overcome the premature convergence by amplifying the difference between the offspring and the parents to some extent, and thus, has advantages in optimizing the feature combinations by screening base classifiers, which

TABLE XIV

WEIGHTS OF VOXELS IN THE RIGHT MEDIAL AMYGDALA FROM AD MRI IMAGES IN THE BINARY CLASSIFICATION TASK OF AD VERSUS NC

Location	Weight	Location	Weight	Location	Weight	Location	Weight
(24, 2, -28)	1	(20, 0, -24)	0.864962274	(20, -4, -18)	0.814542425	(14, 0, -20)	0.781999254
(24, 2, -26)	0.986647801	(20, 0, -22)	0.861861624	(16, 2, -22)	0.814248812	(16, -4, -20)	0.780240026
(24, 2, -24)	0.971108758	(24, -4, -22)	0.859459942	(20, -4, -16)	0.813930464	(16, -8, -14)	0.780139741
(24, 0, -28)	0.961405372	(20, 0, -20)	0.858760975	(20, -4, -14)	0.813318504	(24, -8, -16)	0.779508856
(24, 2, -22)	0.955569716	(22, -2, -20)	0.858506867	(16, -4, -12)	0.812909134	(20, -8, -12)	0.77905471
(22, 2, -28)	0.950589403	(20, 0, -18)	0.855660326	(24, -6, -18)	0.812739518	(18, -8, -14)	0.778658861
(24, 0, -26)	0.950512646	(18, 2, -16)	0.85333513	(18, -2, -22)	0.81247271	(16, -2, -24)	0.778403566
(26, -2, -26)	0.950302941	(20, -2, -30)	0.852685465	(18, -2, -24)	0.809536928	(14, -2, -18)	0.778320195
(22, 2, -26)	0.940123526	(20, 0, -16)	0.852559676	(16, -2, -16)	0.809315045	(22, -8, -16)	0.777405054
(24, 2, -20)	0.940030673	(18, 2, -18)	0.852083099	(18, -4, -14)	0.809030181	(20, -8, -14)	0.777177981
(24, 0, -24)	0.937022886	(22, -2, -18)	0.851858475	(16, 0, -20)	0.807478586	(22, -8, -14)	0.775697101
(26, -2, -24)	0.934070374	(18, 2, -28)	0.851768209	(16, 2, -24)	0.807399757	(20, -8, -16)	0.775301252
(22, 2, -24)	0.930181508	(18, 2, -20)	0.850831068	(18, -2, -28)	0.807102855	(16, -6, -18)	0.77522743
(22, 0, -30)	0.925040372	(24, -4, -20)	0.850068744	(22, -6, -20)	0.806716262	(14, -4, -16)	0.775209249
(24, 2, -18)	0.924491631	(18, 2, -22)	0.849579038	(18, -2, -26)	0.806601145	(24, -8, -14)	0.774216221
(24, 0, -22)	0.923533125	(18, 2, -24)	0.848327007	(24, -6, -16)	0.805397602	(22, -8, -12)	0.773989148
(24, -2, -28)	0.922810743	(18, 2, -26)	0.847074977	(18, -4, -16)	0.805252522	(14, 0, -22)	0.773954635
(24, -2, -30)	0.922222311	(20, -2, -28)	0.845672151	(16, -4, -14)	0.804741857	(18, -8, -16)	0.773197451
(22, 2, -22)	0.92023949	(22, -2, -16)	0.845210083	(22, -6, -18)	0.803361496	(14, -6, -14)	0.772666415
(26, -2, -22)	0.917837808	(20, -2, -26)	0.842526594	(16, 2, -28)	0.802357612	(16, -2, -26)	0.770675697
(22, 0, -28)	0.917415425	(24, -4, -18)	0.840677546	(14, 2, -18)	0.801767552	(14, -2, -20)	0.76999152
(24, -2, -26)	0.914377492	(20, -2, -24)	0.84067029	(16, -2, -18)	0.801587176	(24, -8, -12)	0.768923586
(22, 2, -20)	0.910297472	(20, -2, -22)	0.838813985	(18, -4, -18)	0.801474864	(16, -2, -28)	0.768533559
(24, 0, -20)	0.910043364	(22, -2, -14)	0.838561691	(16, -6, -12)	0.801047484	(14, -4, -18)	0.766596517
(22, 0, -26)	0.909287785	(22, -4, -22)	0.837613144	(16, 2, -26)	0.800550702	(14, 0, -24)	0.765910016
(26, -4, -26)	0.908868373	(18, 0, -16)	0.837307594	(16, 0, -22)	0.800190124	(14, -6, -16)	0.763769627
(24, -2, -24)	0.902937013	(20, -2, -20)	0.83695768	(22, -6, -16)	0.80000673	(20, -10, -12)	0.762228793
(20, 2, -28)	0.901178806	(18, 0, -18)	0.835213687	(18, -6, -12)	0.798464055	(14, -2, -22)	0.761662844
(22, 0, -24)	0.90099258	(20, -2, -18)	0.835101375	(24, -6, -14)	0.798055686	(20, -10, -14)	0.759107719
(22, 2, -18)	0.900355454	(16, 2, -16)	0.834795977	(18, -4, -20)	0.797697205	(14, -4, -20)	0.757983785
(24, 0, -18)	0.896553603	(26, -6, -20)	0.833446607	(22, -6, -14)	0.796651964	(12, 0, -20)	0.756635538
(26, -4, -24)	0.895087557	(20, -2, -16)	0.83324507	(16, -4, -16)	0.79657458	(12, -2, -16)	0.756339009
(20, 2, -26)	0.893599251	(18, 0, -20)	0.833119781	(20, -6, -12)	0.79580627	(14, -6, -18)	0.754872839
(22, 0, -22)	0.892697374	(22, -4, -20)	0.832611565	(20, -6, -14)	0.795248242	(22, -10, -14)	0.754742238
(24, -2, -22)	0.891496533	(20, -2, -14)	0.831388766	(14, -2, -14)	0.794977546	(22, -10, -12)	0.754681098
(22, 2, -16)	0.890413435	(24, -4, -16)	0.831286349	(20, -6, -16)	0.794615858	(14, -2, -24)	0.753334169
(20, 2, -24)	0.889254258	(18, 0, -22)	0.831025874	(14, 2, -20)	0.794006989	(12, -4, -14)	0.751022424
(22, -2, -30)	0.887453888	(18, 0, -28)	0.829435532	(20, -6, -18)	0.793983474	(12, 0, -22)	0.751005047
(20, 0, -30)	0.885573754	(18, 0, -24)	0.828931968	(16, -2, -20)	0.793859306	(12, -2, -18)	0.7504033
(20, 2, -22)	0.884909264	(16, 2, -18)	0.827946922	(18, -6, -14)	0.793844521	(24, -10, -14)	0.750376756
(22, 0, -20)	0.884402169	(22, -4, -18)	0.827609986	(16, 0, -24)	0.792901662	(24, -10, -12)	0.747133403
(22, -2, -28)	0.884241447	(18, 0, -26)	0.826838061	(16, -6, -14)	0.792440799	(12, -4, -16)	0.744781498
(26, -4, -22)	0.88130674	(18, -2, -14)	0.82421584	(14, 0, -18)	0.790043874	(12, -2, -20)	0.744467592
(20, 2, -20)	0.88056427	(22, -4, -16)	0.822608407	(18, -6, -16)	0.789224986	(12, -6, -14)	0.73977013
(24, -2, -20)	0.880056054	(16, 0, -16)	0.822055511	(16, -8, -12)	0.789185833	(12, -4, -18)	0.738540571
(22, -2, -26)	0.878452043	(18, -2, -16)	0.821280058	(16, -4, -18)	0.788407303	(12, -2, -22)	0.738531883
(20, 2, -18)	0.876219276	(16, 2, -20)	0.821097867	(14, -2, -16)	0.786648871	(22, -12, -12)	0.735373048
(22, 0, -18)	0.876106964	(24, -6, -20)	0.820081434	(14, 2, -22)	0.786246426	(12, -6, -16)	0.733223986
(20, 0, -28)	0.873425479	(18, -2, -18)	0.818344275	(16, -2, -22)	0.786131436	(12, -2, -24)	0.732596174
(20, 2, -16)	0.871874282	(22, -4, -14)	0.817606827	(16, 0, -26)	0.785613199	(12, -4, -20)	0.732299645
(22, -2, -24)	0.871803651	(16, -2, -14)	0.817042915	(16, 0, -28)	0.785445586	(12, -6, -18)	0.726677842
(24, -2, -18)	0.868615575	(20, -4, -22)	0.815766345	(18, -6, -18)	0.784605452	(24, -12, -12)	0.72534322
(20, 0, -26)	0.868062923	(18, -2, -20)	0.815408493	(18, -8, -12)	0.784120271		
(22, 0, -16)	0.867811759	(20, -4, -20)	0.815154385	(16, -6, -16)	0.783834114		
(22, -2, -22)	0.865155259	(16, 0, -18)	0.814767049	(14, -4, -14)	0.783821981		

substantially improves classification performance. With the modified GA, the effective information in MRI images GA can be fully utilized in a data-driven manner.

- 4) The proposed 3DCNN+EL+GA is effective in determining brain regions with high discriminability. We also analyze the behavioral domains corresponding to the discriminative brain regions. The obtained behavioral domains are related to emotion, memory, and language, which are involved with AD patients' clinical manifestations, such as memory loss and cognitive decline.
- 5) For each discriminative brain region, the corresponding subregions which contribute more to the good classification performance are further located and illustrated on the basis of the trained base classifiers using the corresponding brain-region datasets, with the help of the attribution methods.

We conclude this article with a discussion on limitations and future research work. As suggested by [45], AD is heterogeneous with numerous subtypes. Plus, the proposed 3DCNN+EL+GA is a data-driven machine learning method. Hence, the effectiveness of the obtained model is inevitably affected by the bias in data collection. According to the experimental results, the prediction performances in the binary classification task of MCIC versus MCInc are much lower than those in AD versus NC or MCIC versus NC for the nine algorithms. There are probably at least three reasons for it. The first reason could be that the datasets in MCIC versus MCInc are smaller than those in AD versus NC or MCIC versus NC. The second might be related to the heterogeneity in AD [45]. Especially, from the perspective of neuroimaging, the heterogeneity could be more obviously exhibited in the early stage AD than in the late stage of AD, which might lead to the fact that the brain regions identified in MCIC versus MCInc are more than and at the same time, generally less discriminative than those discerned in AD versus NC or in MCIC versus NC. The third could be the MRI differences between MCIC and MCInc might be smaller than those between AD and NC or those between MCIC and NC. All of these mean that the identified brain regions in MCIC versus MCInc might vary from dataset to dataset and be sensitive to the algorithm used to build the model. Besides, the prediction results in MCIC versus MCInc might vary from brain atlas to brain atlas to some extent. Additionally, since some brain regions are too small to effectively train corresponding base classifiers, they could be inevitably excluded by the proposed method here. This could result in the loss of important features for the classification to some degree. The possible solution to this problem warrants further investigation. More importantly, how to let the trained model with satisfactory prediction performance "tell" us more about what it has noticed to help predict in an understandable way is another challenge. Hopefully, the advocated method and ideas might help identify novel candidate neuroimaging biomarkers for other brain disorders (especially little-known ones), such as Parkinson's disease, epilepsy, severe depression, autism, Huntington's disease, multiple sclerosis, and amyotrophic lateral sclerosis, in a data-driven manner. The relevant research is worthwhile.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their constructive comments on this article. Alzheimer's Disease Neuroimaging Initiative (ADNI)-Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org), North Bethesda, MD, USA. The grantee organization is the Northern California Institute for Research and Education, San Francisco, CA, USA, and the study is coordinated by the Alzheimer's Therapeutic Research Institute, University of Southern California, Los Angeles, CA, USA. ADNI data are disseminated by the Laboratory for Neuro Imaging, University of Southern California.

AUTHOR CONTRIBUTIONS STATEMENT

Dan Pan and An Zeng designed and coordinated the study. Dan Pan, Genqiang Luo, An Zeng, Chao Zou, Haolin Liang, Jianbin Wang, Tong Zhang, and Baoyao Yang carried out experiments and data processes. Dan Pan, Genqiang Luo, and An Zeng reviewed the study design and data processing and edited the result presentation and interpretation. All authors drafted and revised this article and approved the submission of the final version of this article.

DECLARATION OF INTEREST

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

- [1] L. V. Hong-Meng, Z. Di, and C. Xue-Bin, "Deep learning for early diagnosis of Alzheimer's disease based on intensive alexnet," *Comput. Sci.*, vol. 44, no. S1, pp. 50–60, 2017.
- [2] E. Nichols and T. Vos, "Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the global burden of disease study 2019," *Lancet Public Health*, vol. 17, no. 2, pp. e105–e125, Dec. 2021.
- [3] M. Silveira and J. Marques, "Boosting Alzheimer disease diagnosis using PET images," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2556–2559.
- [4] A. J. Mitchell and M. Shiri-Feshki, "Rate of progression of mild cognitive impairment to dementia—meta-analysis of 41 robust inception cohort studies," *Acta Psychiatrica Scandinavica*, vol. 119, no. 4, pp. 252–265, 2009.
- [5] J. Liu, J. Zhao, S. Duan, and J.-W. Zhao, "Advances in mild cognitive impairment," *Prog. Modern Biomed.*, vol. 17, no. 11, pp. 2170–2173, 2017.
- [6] X. Chen et al., "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102444.
- [7] D. Pan, C. Zou, H. Rong, and A. Zeng, "Early diagnosis of Alzheimer's disease based on three-dimensional convolutional neural networks ensemble model combined with genetic algorithm," *J. Biomed. Eng.*, vol. 38, no. 1, pp. 47–55, Feb. 2021.

- [8] W. Yin, L. Li, and F.-X. Wu, "Deep learning for brain disorder diagnosis based on fMRI images," *Neurocomputing*, vol. 469, pp. 332–345, Jan. 2022.
- [9] S. Zhao, X. Wu, B. Chen, and S. Li, "Automatic spondylolisthesis grading from MRIs across modalities using faster adversarial recognition network," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101533.
- [10] D. Xu, Z. Tian, R. Lai, X. Kong, Z. Tan, and W. Shi, "Deep learning based emotion analysis of microblog texts," *Inf. Fusion*, vol. 64, pp. 1–11, Dec. 2020.
- [11] C. Luo, Z. Tan, G. Min, J. Gan, W. Shi, and Z. Tian, "A novel web attack detection system for Internet of Things via ensemble classification," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5810–5818, Aug. 2021.
- [12] F. Li and M. Liu, "Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks," *Computerized Med. Imag. Graph.*, vol. 70, pp. 101–110, Dec. 2018.
- [13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [14] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, "Early detection of Alzheimer's disease using magnetic resonance imaging: A novel approach combining convolutional neural networks and ensemble learning," *Frontiers Neurosci.*, vol. 14, p. 259, May 2020.
- [15] S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, "Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling," *J. Med. Syst.*, vol. 42, no. 5, p. 85, May 2018.
- [16] A. Zeng et al., "Discovery of genetic biomarkers for Alzheimer's disease using adaptive convolutional neural networks ensemble and genome-wide association studies," *Interdiscipl. Sci., Comput. Life Sci.*, vol. 13, no. 4, pp. 787–800, 2021.
- [17] S. Spasov, L. Passamonti, A. Duggento, P. Lió, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *NeuroImage*, vol. 189, pp. 276–287, Apr. 2019.
- [18] W. Zhu, L. Sun, J. Huang, L. Han, and D. Zhang, "Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2354–2366, Sep. 2021.
- [19] L. Fan et al., "The human brainnetome atlas: A new brain atlas based on connectional architecture," *Cerebral Cortex*, vol. 26, no. 8, pp. 3508–3526, 2016.
- [20] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, and I. Castiglioni, "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach," *Frontiers Neurosci.*, vol. 9, p. 307, Sep. 2015.
- [21] P. J. LaMontagne et al., "OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease," *medRxiv*, 2019. [Online]. Available: <https://www.medrxiv.org/content/early/2019/12/15/2019.12.13.19014902>, doi: [10.1101/2019.12.13.19014902](https://doi.org/10.1101/2019.12.13.19014902).
- [22] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, p. 106, 1962.
- [23] K. Fukushima and S. Miyake, "Neocognitron: A Self-organizing neural network model for a mechanism of visual pattern recognition," in *Competition and Cooperation in Neural Nets*, S.-I. Amari and M. A. Arbib, Eds. Berlin, Germany: Springer, 1982, pp. 267–285, doi: [10.1007/978-3-642-46466-9_18](https://doi.org/10.1007/978-3-642-46466-9_18).
- [24] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, May 2015.
- [25] E. Hosseini-Asl, R. Keynto, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3D convolutional network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 126–130.
- [26] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [27] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
- [28] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Apr. 1997.
- [29] C.-X. Zhang and J.-S. Zhang, "A survey of selective ensemble learning algorithms," *Chin. J. Comput.*, vol. 34, no. 8, pp. 1399–1410, Oct. 2011.
- [30] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, Feb. 1996.
- [31] Z. Yi, T. Xu, W. Shang, W. Li, and X. Wu, "Genetic algorithm-based ensemble hybrid sparse ELM for grasp stability recognition with multimodal tactile signals," *IEEE Trans. Ind. Electron.*, vol. 70, no. 3, pp. 2790–2799, Mar. 2023.
- [32] T. Mao, A.-S. Mihaita, F. Chen, and H. L. Vu, "Boosted genetic algorithm using machine learning for traffic control optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7112–7141, Jul. 2022.
- [33] W. Yang, Y. Zhang, H. Wang, P. Deng, and T. Li, "Hybrid genetic model for clustering ensemble," *Knowl.-Based Syst.*, vol. 231, Nov. 2021, Art. no. 107457.
- [34] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 4, Mar. 1995, pp. 1942–1948.
- [35] M. Ghasemi, E. Akbari, A. Rahimnejad, S. E. Razavi, S. Ghavidel, and L. Li, "Phasor particle swarm optimization: A simple and efficient variant of PSO," *Soft Comput.*, vol. 23, no. 19, pp. 9701–9718, Oct. 2019.
- [36] J. A. Harasty, G. M. Halliday, J. J. Kril, and C. Code, "Specific temporoparietal gyral atrophy reflects the pattern of language dissolution in Alzheimer's disease," *Brain*, vol. 122, no. 4, pp. 675–686, Apr. 1999.
- [37] A. T. Du, "Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease," *J. Neurol., Neurosurgery Psychiatry*, vol. 71, no. 4, pp. 441–447, Oct. 2001.
- [38] M. Xia, J. Wang, and Y. He, "BrainNet viewer: A network visualization tool for human brain connectomics," *PLoS ONE*, vol. 8, no. 7, Jul. 2013, Art. no. e68910.
- [39] J. Wang et al., "Determination of the posterior boundary of Wernicke's area based on multimodal connectivity profiles," *Human Brain Mapping*, vol. 36, no. 5, pp. 1908–1924, May 2015.
- [40] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 839–847.
- [41] E. S. N. Joshua, M. Chakravarthy, and D. Bhattacharyya, "Lung cancer detection using improvised grad-cam++ with 3D CNN class activation," in *Smart Technologies in Data Science and Communication*, S. K. Saha, P. S. Pang, and D. Bhattacharyya, Eds. Singapore: Springer, 2021, pp. 55–69, doi: [10.1007/978-981-16-1773-7_5](https://doi.org/10.1007/978-981-16-1773-7_5).
- [42] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
- [43] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.
- [44] K. Aderghal, J. Benois-Pineau, K. Afdel, and C. Gwenaelle, "FuseMe: Classification of sMRI images by fusion of deep CNNs in 2D+e projections," in *Proc. 15th Int. Workshop Content-Based Multimedia Indexing*, 2017, pp. 1–7.
- [45] J. W. Vogel and O. Hansson, "Subtypes of Alzheimer's disease: Questions, controversy, and meaning," *Trends Neurosciences*, vol. 45, no. 5, pp. 342–345, May 2022.



Dan Pan received the B.S. degree from Nanchang University, Nanchang, China, in 1995, the M.S. degree and the Ph.D. degree in circuits and systems from the South China University of Technology, Guangzhou, China, in 1998 and 2001, respectively, and the master's degree in business administration from the Sobey School of Business, Saint Mary's University, Halifax, NS, Canada, in 2011.

He is currently an Associate Professor and a Senior Engineer with the School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou. His research interests include artificial intelligence, deep learning, and their applications, especially in the fields of medicine and healthcare.



Genqiang Luo was born in Guangdong, China, in 1996. He received the B.Eng. degree in communication engineering from Hunan University of Science and Engineering, Yongzhou, China, in 2020. He is currently pursuing the M.S. degree in electronic information with the School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou, China.

His research interests include medical image processing and robot operating systems.



Jianbin Wang was born in 1999. He is currently pursuing the M.S. degree in computer science and technology with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China.

His current research interests include deep learning and computer vision.



An Zeng (Member, IEEE) received the B.S. degree from Nanchang Hangkong University, Nanchang, China, in 1999, and the M.S. and Ph.D. degrees in applied computer technology from the South China University of Technology, Guangzhou, China, in 2002 and 2005, respectively.

From 2008 to 2010, she was a Post-Doctoral Fellow with the Faculty of Computer Science, Medical School of Dalhousie University, Halifax, NS, Canada. She is currently a Professor and a Vice Dean with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou. Her research interests include deep learning, artificial intelligence, and their applications in the fields of medicine and healthcare.

Tong Zhang (Member, IEEE) received the B.S. degree in software engineering from Sun Yat-sen University, Guangzhou, China, in 2009, and the M.S. degree in applied mathematics and the Ph.D. degree in software engineering from the University of Macau, Macau, China, in 2011 and 2016, respectively.

He is currently a Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou. His research interests include affective computing, evolutionary computation, neural network, and other machine learning techniques and their applications.



Dr. Zhang has been working on publication matters for many IEEE conferences. He is an Associate Editor of the IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS.



Chao Zou received the M.S. degree in computer science and technology from Guangdong University of Technology, Guangzhou, China, in 2020.

He is currently a Software Engineer with the Research and Development Centre of Guangzhou Department, Agricultural Bank of China, Guangzhou. His research interests include medical image processing, machine learning, and pattern recognition.

Baoyao Yang received the B.Sc. degree in computer science and technology from the South China University of Technology, Guangzhou, China, in 2014, and the Ph.D. degree in computer science from Hong Kong Baptist University, Hong Kong, in 2018.

In 2019, she was on sabbatical leave as a Visiting Researcher with the Machine Intelligence Laboratory, Department of Information Science and Technology, The University of Tokyo, Tokyo, Japan, for two months. She is currently an Associate Professor with the School of Computers, Guangdong University of Technology, Guangzhou. She has authored or coauthored around 20 papers in top-tier journals and conferences, including the IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY (IFS), IEEE Cybernetics, Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence (AAAI), and International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). Her research interests include federated learning, cross-modal adaptation, and machine learning in medical applications.



Haolin Liang was born in Guangdong, China, in 1997. He received the B.Sc. degree in optoelectronic information science and engineering from Guangdong University of Technology, Guangzhou, China, in 2016, where he is currently pursuing the M.S. degree in computer science and technology with the School of Computer Science and Technology.

His research interests include medical image processing and deep learning.