Computers in Biology and Medicine 136 (2021) 104678

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Multi-model and multi-slice ensemble learning architecture based on 2D convolutional neural networks for Alzheimer's disease diagnosis☆

Wenjie Kang, Lan Lin *, Baiwen Zhang, Xiaoqi Shen, Shuicai Wu, for the Alzheimer's Disease Neuroimaging Initiative

Intelligent Physiological Measurement and Clinical Translation, Beijing International Platform for Scientific and Technological Cooperation, Department of Biomedical Engineering, Faculty of Environment and Life Sciences, Beijing University of Technology, Beijing, 100124, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Mild cognitive impairment
Structural MRI
Convolutional neural networks
Generative adversarial networks
Deep learning

ABSTRACT

Alzheimer's Disease (AD) is a chronic neurodegenerative disease without effective medications or supplemental treatments. Thus, predicting AD progression is crucial for clinical practice and medical research. Due to limited neuroimaging data, two-dimensional convolutional neural networks (2D CNNs) have been commonly adopted to differentiate among cognitively normal subjects (CN), people with mild cognitive impairment (MCI), and AD patients. Therefore, this paper proposes an ensemble learning (EL) architecture based on 2D CNNs, using a multi-model and multi-slice ensemble. First, the top 11 coronal slices of grey matter density maps for AD versus CN classifications were selected. Second, the discriminator of a generative adversarial network, VGG16, and ResNet50 were trained with the selected slices, and the majority voting scheme was used to merge the multi-slice decisions of each model. Afterwards, those three classifiers were used to construct an ensemble model. Multi-slice ensemble learning was designed to obtain spatial features, while multi-model integration reduced the prediction error rate. Finally, transfer learning was used in domain adaptation to refine those CNNs, moving them from working solely with AD versus CN classifications to being applicable to other tasks. This ensemble approach achieved accuracy values of 90.36%, 77.19%, and 72.36% when classifying AD versus CN, AD versus MCI, and MCI versus CN, respectively. Compared with other state-of-the-art 2D studies, the proposed approach provides an effective, accurate, automatic diagnosis along the AD continuum. This technique may enhance AD diagnostics when the sample size is limited.

1. Introduction

Alzheimer's disease (AD) is a progressive, neurodegenerative form of dementia [1]. Globally, there are around 44 million people who have been diagnosed with AD, and this number is projected to grow to 131.5 million by 2050 [2]. Thus, AD has become a major public health problem worldwide. The main clinical manifestations of AD are memory decline, progressive cognitive dysfunction, and impaired daily living abilities [3]. While the pathogenesis and etiology of AD remain unclear, considerable pathological evidence indicates the central features of AD, including the deposition of the β-amyloid (Aβ) peptide in plaques, the hyperphosphorylation of the Tau protein, and the development of neurofibrillary tangles [4,5]. Mild cognitive impairment (MCI) represents a transitional phase between healthy cognitive aging and AD. The annual conversion rate from MCI to probable AD is about 10%–12% [6]. Neuroimaging plays an important role in determining the diagnostic and prognostic biomarkers of AD. Structural magnetic resonance imaging (sMRI) has been widely used to classify patients with AD at different disease stages [7].

With the development of machine learning, especially deep learning, many computational models have been proposed for the computer-aided classification of AD patients. Convolutional neural networks (CNNs) have recently emerged as a powerful deep learning architecture [8,9]. Convolutional neural networks (CNNs) have recently emerged as a powerful deep learning architecture [10]. Compared with other machine learning approaches, CNNs are superior at discerning subtle and

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

* Corresponding author.
E-mail address: lanlin@bjut.edu.cn (L. Lin).

https://doi.org/10.1016/j.compbiomed.2021.104678
Received 13 May 2021; Received in revised form 20 July 2021; Accepted 20 July 2021
Available online 22 July 2021

diffuse anatomical abnormalities in brain imaging. CNN models in neuroimaging studies can be classified into region of interest (ROI)-level [11], two-dimensional (2D) slice-level [12–16], three-dimensional (3D) patch-level [17–19], and 3D subject-level [20–24]. Shmulev et al. [21] have proposed a 3D ResNet to classify MCI subjects into either MCI converters (MCIc) or MCI non-converters (MCInc), and their method has returned an accuracy value of 62%. Another work from Valliani et al. [15] have classified AD versus CN through a slice-level system in which ResNet18 was pretrained on the ImageNet [25] database, obtaining a classification accuracy of 81.3%. Aderghal et al. [16] have proposed a LeNet-like CNN. They collected MRI and diffusion tensor imaging data from the hippocampal regions on three anatomical planes (axial, sagittal, and coronal) at the slice level and used these data as input. The network was pretrained on the MNIST database, and it obtained a classification accuracy of 86.83% for AD versus CN, 71.75% for AD versus MCI, and 69.85% for MCI versus CN. Notably, training a 3D CNN from scratch requires large amounts of MRI data, and dataset limitations are the main cause of overfitting. To eliminate the requirement for a huge amount of imaging data, researchers have adopted transfer learning techniques that use pretrained 2D CNNs [18,22,26].

However, 3D spatial contextual information, which is required to achieve accurate classifications, cannot be captured by existing 2D CNNs. Ensemble learning (EL) effectively improves classification accuracy and stability by integrating multiple MRI slices or multiple CNNs [12,16,27]. Therefore, the present paper proposes a novel EL CNN system that integrates 11 of the best validation accuracy 2D slice-level models from three CNNs for final classification. In comparison with the CNN classification model based on a single slice, the integration of multiple 2D slices could exploit more effective information contained in 3D MRI data.

The remainder of the paper is organized as follows. Section 2 introduces the preprocessing procedures and proposes an EL method based on deep convolution generative adversarial networks (DCGAN) [28] and other CNN models. Section 3 presents the experimental settings and experimental results of this research. Section 4 discusses the overall results and possibilities for future work, and Section 5 concludes the study.

2. Materials and methods

2.1. Dataset

Data used in the preparation of this article were obtained from the ADNI. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and AD.

The experiment included 1.5T T1-weighted baseline sMRI scans from 798 participants in the ADNI-1 cohort. These subjects were divided into three groups (AD, MCI, and CN) according to their baseline diagnoses. Table 1 summarizes the participants' demographic characteristics, including age, gender, years of education, and Mini-Mental State Examination scores (MMSE) [29]. These subjects were divided into three groups (AD, MCI and CN) by baseline diagnosis. Table 1 summarizes the demographic characteristics of the participants, including age, gender, years of education and mini-mental state examination scores (MMSE). MCI subjects were further divided into two subgroups: MCIc and MCInc, according to 24 months MCI to probable AD conversion status. Some MCI participants were excluded from further analysis due to incomplete follow-up conversion status, or reversion status. Thus, of the 382 MCI subjects, 138 MCIc and 181 MCInc participants remained in the study of the prediction of MCI convert to AD. The sMRI scans were randomly split into training, validation, and test sets at a ratio of 7:1:2.

Table 1
Demographic information of the subjects in ADNI-1.

Characteristic	AD	MCI	CN
Subjects	187	382	229
Age	75.26 ± 7.53	74.71 ± 7.48	75.87 ± 5.02
Gender (Male/Female)	98/89	245/137	119/110
Education	14.66 ± 3.14	15.67 ± 2.90	16.07 ± 2.85
MMSE	23.28 ± 2.04	27.33 ± 1.83	29.11 ± 1.00

The age, education years, and MMSE values are reported as Mean ± Standard deviation (Std).

2.2. Image pre-processing

The MRI scans were first converted from the Digital Imaging and Communications in Medicine (DICOM) format to the Neuroimaging Informatics Technology Initiative (NIfTI) format, using dcm2niigui incorporated in MRIcron software. All subjects were processed and analyzed using the CAT12 toolbox (http://www.neuro.uni-jena.de/cat/) via SPM12 software (http://www.fil.ion.ucl.ac.uk/spm/software/spm). 3D T1-weighted MRI scans were normalized using an affine transformation, followed by non-linear registration, corrected for bias field in homogeneities. Then, the normalized images were segmented into grey matter (GM), white matter, and cerebrospinal fluid and were modulated and normalized into a Montreal Neurological Institute (MNI) template. Before being used for further analysis, the extracted GM density maps (GMDM) were smoothed with a 2.0 mm full width at half maximum (FWHM) Gaussian isotropic kernel. Thus, the preprocessed GMDM had a size of $121 \times 145 \times 121$ voxels and a spatial resolution of $1.5 \times 1.5 \times 1.5$ mm^3.

2.3. 2D slice selection

Slice selection is a critical component of applying 2D CNNs to 3D volumes. Therefore, the 3D volumetric data was sectioned along the coronal direction, and the slices were sampled at 3.0 mm intervals to ensure the diversity of the features. All these slices were used as input images for VGG16, which had been pretrained with the ImageNet dataset. The first four VGG convolution blocks were frozen, while the last convolution block and the dense layers were fine-tuned. Images from the coronal plane were numbered from 1 to 145 and mapped into the standard MNI space from −126 to 90. The 2D slices were resized to 128×128 pixels by cropping and padding. The top 11 2D slices with the highest classification accuracy were selected.

2.4. DCGAN based EL

The concept of generative adversarial networks (GAN) [30] involves training a generator and a discriminator in an adversarial way. Deep convolutional generative adversarial networks (DCGAN) is an upgraded version of the GAN architecture, combining CNN and GAN. In these networks, the generator is composed of a series of transposed convolutional operations. This model assumes a uniform distribution noise vector and generates images. The discriminator is basically a convolutional classifier made up of convolutional layers, max-pooling layers, batch normalization layers, and LeakyReLU activations. Considering the scarcity of neuroimaging data, DCGANs and domain transfer learning are applied to pretrain classification models.

Fig. 1 shows the flowchart of the DCGAN-based method. All 2D slices were used to train the model. Then, the network parameters learned in the DCGAN discriminator were transferred to classify AD versus CN. Majority voting EL was used to combine predictions from multiple selected slices.

2.5. Multi-model EL

Fig. 2 illustrates the workflow of this algorithm, which contained two

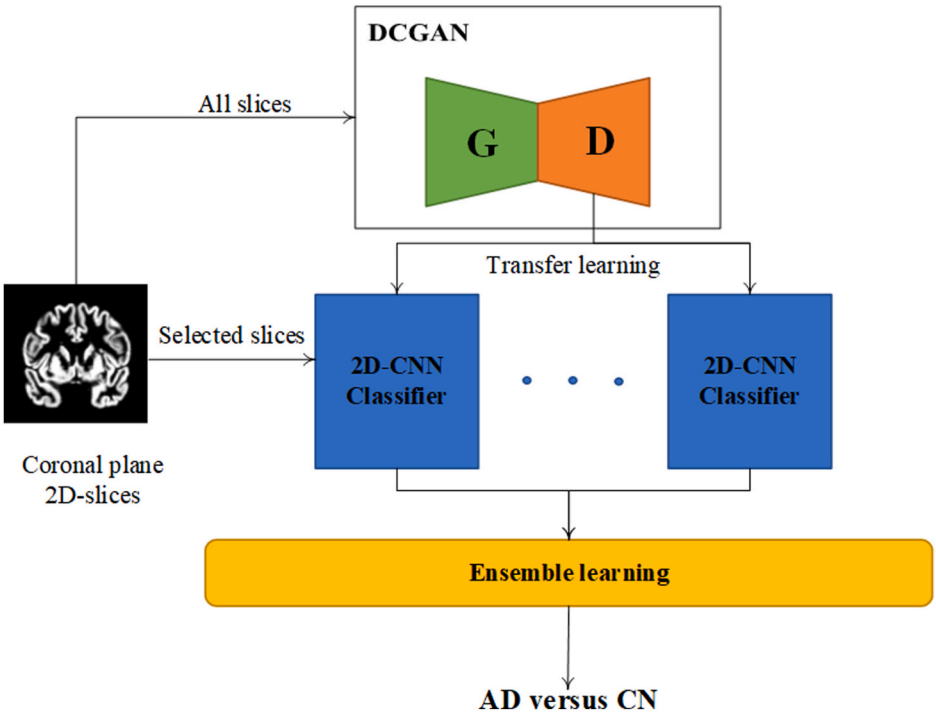

Fig. 1. Flowchart of the DCGAN-based method in AD versus CN classification.

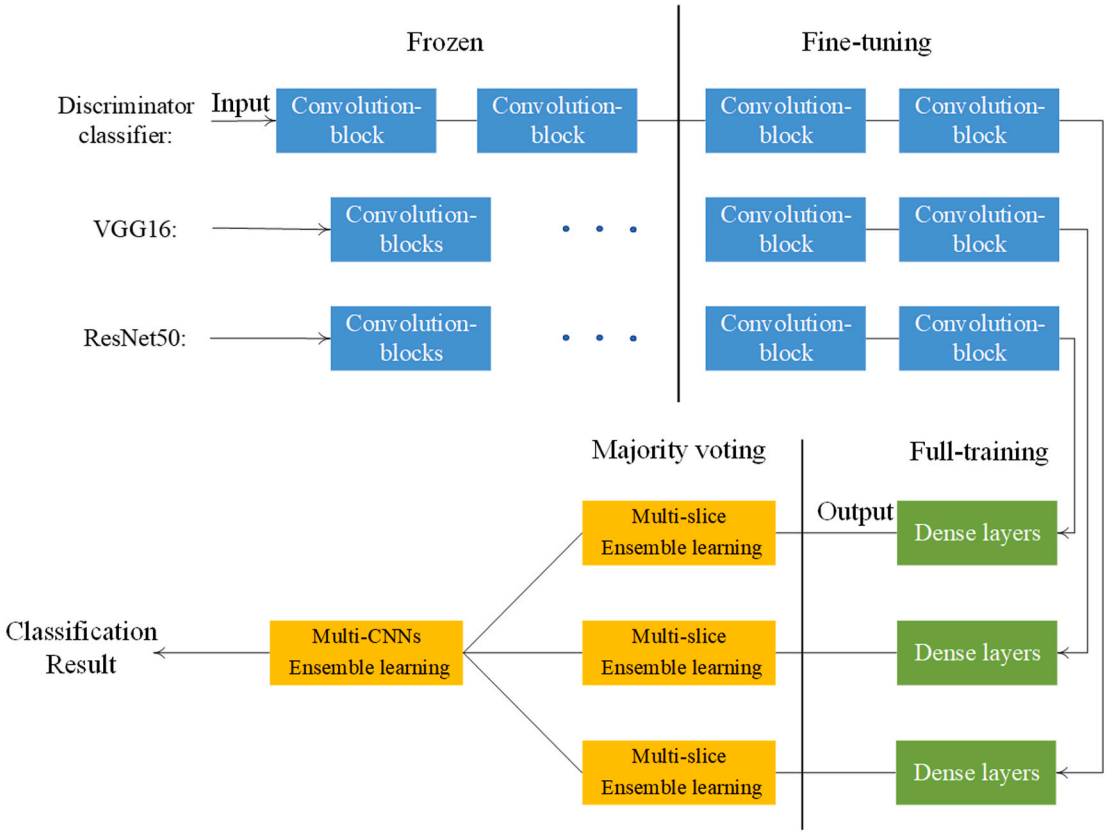

Fig. 2. Flowchart of our proposed EL method.

stages. The first stage involved multi-slice EL. The discriminators, VGG16, and ResNet50 were trained with 11 top-ranked slices from the training set. VGG16 and ResNet50 had also been pretrained with the ImageNet dataset, and transfer learning was applied for classifying AD versus CN. The last two convolution blocks of the classic CNNs were fine-tuned, and the remainder of the convolution blocks were frozen. A majority voting ensemble was used to combine predictions from different slices in each classifier. For the second stage, majority voting

was applied to three classifiers to achieve the final classification.

2.6. Transfer learning to different tasks

Compared to AD versus CN classification, several other classification tasks (AD versus MCI, MCI versus CN, and MCIc versus MCInc) are more challenging, because structural changes in MCI are relatively subtle. However, the supplementary knowledge learned from AD versus CN classification can be adopted to enrich the information available when attempting more difficult tasks. For such tasks, this research considered the last two convolutional blocks of each classifier for fine-tuning.

3. Experiments and results

3.1. Model training

All models in this work were deployed in Python 3.7.9 and Tensorflow 2.4 packages on a workstation with Intel Xeon W-2223 CPU with 16 GB of RAM, and a NVIDIA GeForce RTX 3090 GPU 24 GB.

In the DCGAN model, all losses were computed with binary cross entropy. In the training phase, the Adam optimizer was used, with initial learning rates of 2×10^{-3} and 2×10^{-4} for the generator and discriminator, respectively. As training epochs increased, the losses from both the discriminator and the generator were converged to certain constant numbers, and the discriminator's accuracy approached 50%, indicating that the DCGAN had finally reached Nash equilibrium (Fig. 3).

VGG16, ResNet50, and the DCGAN discriminator were used as 2D slice classifiers. To train those CNNs, the Adam [29] optimization algorithm was chosen to iteratively tune the weights. The initial learning rate of the discriminator classifier was set to 1×10^{-3}, and the initial learning rates of VGG16 and ResNet50 were 1×10^{-4}. An early stopping method was implemented, in which the classifier stopped training when the loss function did not improve over 30 epochs with the validation dataset. Uninformative slice classifiers, with an accuracy of less than 60%, were removed from the majority voting process.

3.2. Classification results

Fig. 4 shows the selected slice numbers, the corresponding MNI space coordinates from the Bioimage Suite Web(https://bioimagesuiteweb.github.io/webapp/mni2tal.html) and the corresponding T1 images in MNI space.

To evaluate the diagnostic performance of those classifiers, four widely used performance metrics were applied: accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC). The numbers representing each of these metrics were defined as true positive (TP), true negative (TN), false positive (FP), or false negative (FN), and the first three metrics were calculated as follows: $Accuracy = \frac{TP+TN}{TP+TN+FN+FP}$, $Sensitivity = \frac{TP}{TP+FN}$, and $Specificity = \frac{TN}{TN+FP}$. AUC is calculated based on the area under the receiver operating characteristic curve. Table 2 shows the multi-slice EL's classification results for the AD versus CN task. The classification accuracy of the discriminator classifier was 4.16% and 7.28% higher than the maximum and average accuracies of single-slice classification, respectively. The discriminator classifier also performed better than classic CNN models. Table 3 presents the experimental results obtained via the three EL models across four binary classification tasks (AD versus CN, AD versus MCI, MCI versus CN, and MCIc versus MCInc).

We have observed 90.36% accuracy in classifying AD versus CN. The sensitivity and specificity for this task are 93.94% and 83.78%, respectively. The AUC is noted to be 89.72%. The classification accuracy of AD versus MCI is 77.19%, the AUC for this ensemble model is derived as 71.18%. Classifying the MCI scans from CN scans has an accuracy of 72.36%, with sensitivity and specificity of 74.71% and 84.42%, respectively. Ensemble model yields the accuracy of 63.49% for MCIc versus MCInc classification.

4. Discussion

For slice-level AD versus CN classification, CNNs pretrained using the ImageNet dataset were fine-tuned with GMDM. The discriminator was transferred from classifying fake from real to classifying AD versus CN. Table 4 lists the classification accuracies of the slides selected for AD versus CN classification, further illustrating the accuracy and stability of the single-slice classifier. The top slice covered the following regions: Hippocampus, ParaHippocampal, Temporal_Inf, Temporal_Mid, and Thalamus (all coordinating with the Automated Anatomical Labeling [AAL] template within the MNI space, $Y = -22$). The discriminator achieved the highest accuracy among these classifiers, with an average accuracy of 84.23%. It also had the best stability among the three single classifiers, while ResNet50 had the lowest stability. Improvements in transfer learning, obtained with better models, and pretraining datasets with greater domain matches are complementary. Considering that the slice selection was based on VGG16, and since the selected slices matched VGG16 well, VGG16 outperformed ResNet50. The similarities between the features of the source and target domains were highest in the discriminator, possibly because it achieved the best performance during transfer learning. This research hypothesizes that features from different slices are complementary. Integrating multi-slice results would, therefore, outperform a single-slice result. The classification results from 11 slices were integrated to obtain a subject-level classification. For AD versus CN classification, the multi-slice EL values for the discriminator, VGG16, and ResNet50 were, respectively, 7.28%, 5.24%, and 11.13% higher than the average classification accuracy of a single-slice classifier. Thus, multi-slice EL considerably improved the classification performance of 2D CNN models.

Table 5 shows the classification performance of three CNN models across four tasks. For classifying AD versus CN and predicting MCI conversion to AD, the discriminator classifier achieved the best performance. For classifying AD versus MCI and MCI versus CN, VGG16 performed best. In most tasks, the multi-model EL method, combining multiple classifiers, was more accurate than a single learner alone,

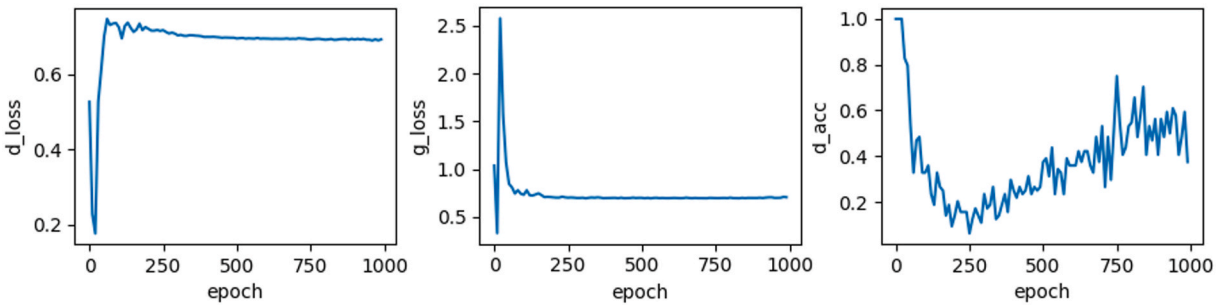

Fig. 3. Training result of DCGAN. d_loss is the loss of discriminator, g_loss is the loss of generator, and d_acc is the output of discriminator.

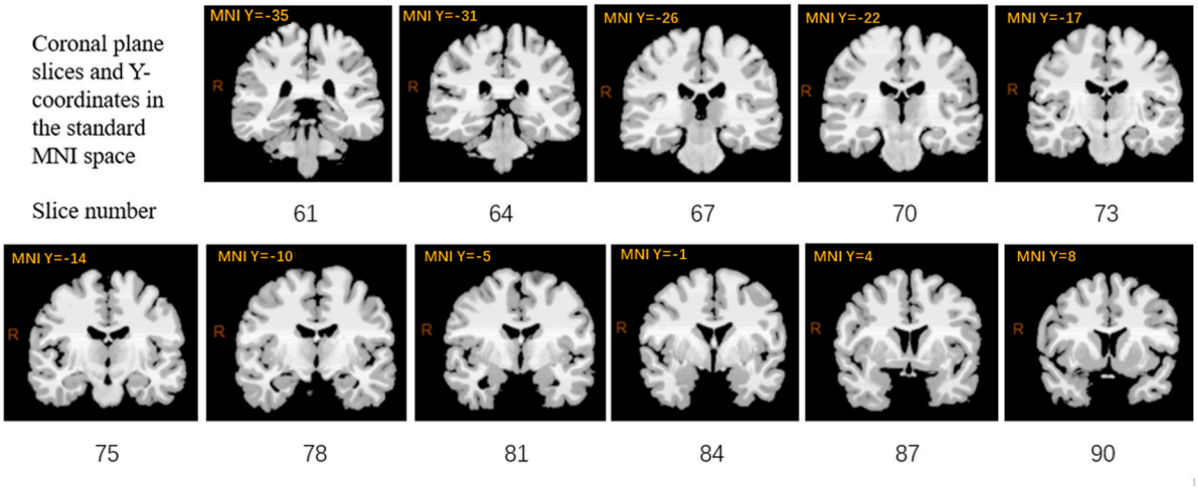

Fig. 4. Location of the selected coronal slices.

Table 2
Classification result of three classifiers for AD versus CN classification.

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
Discriminator	**90.36**	**91.43**	86.49	**89.98**
VGG16	87.95	84.62	**89.19**	88.07
Resnet50	83.13	79.49	83.78	83.20

Table 3
Results of multi-model EL for four binary **classifications.**

Tasks	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
AD versus CN	90.36	93.94	83.78	89.72
AD versus MCI	77.19	68.97	54.06	71.18
MCI versus CN	72.36	74.71	84.42	68.29
MCIc versus MCInc	63.49	57.56	64.29	62.50

Table 4
Single slice classifier performance for AD versus CN classification.

Classifiers	Mean Accuracy(%)	Maximum Accuracy(%)
Discriminator	84.23 ± 2.05	86.75
VGG16	83.57 ± 2.65	86.75
ResNet50	74.70 ± 6.12	83.13

Table 5
Classification performance of three classifiers.

Classifiers	Accuracy (%)			
	AD versus CN	AD versus MCI	MCI versus CN	MCIc versus MCInc
Discriminator	**90.36**	74.56	69.11	**66.67**
VGG16	87.95	**77.19**	**69.92**	60.31
ResNet50	83.13	74.56	65.85	61.90

indicating that these classifiers had appropriately learned the different, yet complementary, information. For example, in MCI versus CN classification, the multi-model method improved the classifications by 3.49%. However, when predicting MCI conversion to AD, the multi-model EL method performed relatively poorly as compared to the discriminator, mainly because the pathological changes between MCIc and MCInc are subtle [31], especially in 2D imagery. Limited neuroimaging data and slighter pathological changes between MCIc and MCInc decrease the effectiveness of transfer learning in extracting AD-related pathological features directly from MCI data. This issue can be mitigated by increasing the number of classifiers and changing the data distribution. To accomplish that in this research, three classifiers were trained with training samples of equal weights. Then, the weights of the training samples were updated based on the learning error rate. Hence, subjects with high learning error rates in the subsequent classifier became the focus. This operation was repeated until Classifier-1, Classifier-2, and Classifier-n were obtained. When the number of classifiers increased by n times, weak classifiers could be combined to generate a robust classifier.

To determine how multi-slice EL and multi-model EL contribute to classification performance, an ablation study was conducted (Table 6). A pre-trained VGG16 was employed, using the best slice as the Baseline Model. The best slice was also used in Model 1. Three EL models were adopted for the proposed model and Model 1. The discriminator, which was pre-trained in GAN, was applied in Model 2. For AD versus CN classification, most performance gains were achieved from the multi-slice strategy. For MCI-related classification, multi-model and multi-slice strategies provided basically the same contributions.

The accuracy of MCI conversion predictions was further analyzed. Table 7 shows the various classifiers' accuracy when predicting MCI conversion to AD, including different conversion times. The results show that classification accuracy decreases as conversion time increases, and the discriminator made more accurate predictions for patients who converted from MCI to AD in 18 months. In this study, the period for predicting MCI to AD conversion was set to two years. However, no clear criterion prescribes the exact length of time. According to different scientific research objectives, including a study from the ADNI [32], this time has been said to range from 18 to 36 months. However, regardless of the length of the period, some advantages and disadvantages can be

Table 6
Classification performance of four models.

Model	Multi-slice EL	Multi-modal EL	AD/CN (%)	AD/MCI (%)	MCI/CN (%)
Baseline Model			84.6	68.4	66.7
Model 1		✓	85.6	71.2	65.9
Model 2	✓		**90.4**	74.6	69.1
Proposed Model	✓	✓	**90.4**	**77.2′**	**72.4**

Table 7
True positive rate of MCI convertors in different conversion time.

Model	6 months (n = 9)	12–18 months (n = 13)	24 months (n = 5)
EL	67%	31%	20%
Discriminator	67%	62%	40%
VGG16	67%	46%	40%
ResNet50	56%	23%	60%

observed. The shorter the time between baseline and conversion, the more obvious the pathological changes between MCIc and MCInc, but some cases diagnosed as MCInc might still convert to AD several months later [33]. The longer the time, the harder it is to separate MCIc from MCInc, but the higher the clinical value [22]. This study only used baseline images from the ADNI-1 cohort, and most of these MCIc subjects (100 of 138 subjects) converted to AD within 18 months. In future, longitudinal MCI images should be integrated into the dataset, and the time it takes for MCI to convert to AD should be studied.

Recently, many researchers have devoted their efforts to distinguishing AD from CN using CNNs. For comparison, Table 8 summarizes the classification performances achieved by several state-of-the-art studies. All 2D slice-level algorithms in Table 8 had very similar accuracies in AD versus CN classification, and the approach proposed in this research is ranked as the third best classifier. One reason for this ranking may be that this proposed method fully utilized the multi-layer and multi-model EL to make these AD versus CN classifications. For the three other tasks, the proposed approach was, for several reasons, not that outstanding at first glance. In Aderghal's [16] work, high performance was achieved by using multimodality imaging (sMRI and diffusion tensor imaging). If solely sMRI had been applied, the mean accuracy of the validation and testing sets for AD versus CN, AD versus MCI, and MCI versus CN would have been 88.1%, 76.5%, and 71.3%, respectively. Ahmed's [34] work has shown superior performance compared with other methods. However, it used longitudinal ADNI data, and 351 scans were acquired from 60 subjects. This method presents a data leakage problem. The images from each subject would appear in both the training and test sets, and the classifier would use additional, nonessential information for classification. A 3D CNN can extract 3D spatial information, which normally offers good performance, especially in MCI-related tasks. However, it includes more parameters and requires a large number of images for training [20]. The performance of 3D CNNs will be significantly degraded by limited data [15,35]. Overall, the model proposed in this research has performed comparably in the 2D models presented in previous AD studies.

The subject number denotes the total subject numbers used in training, validation and testing, including AD, MCI and CN. For MCIc versus MCInc classification, discriminator is used.

In general, Table 5 shows that the proposed DCGAN-based model is more advantageous for predicting MCI to AD conversions than classic CNNs. For CNNs, the first several convolutional blocks are locked. Considering the disparity between natural images and GMDM, low-level features learned by classic CNNs may be ineffective for classifying MCIc versus MCInc. DCGAN was originally trained with GMDM images, and the task adaptation transfer for weak supervision tasks was more effective. In the future, multiple discriminators can be integrated into the DCGAN to construct a multi-discriminator DCGAN. This design would guarantee the stability of the training process, and the patterns that are misclassified by different discriminators would change. Replacing CNNs with such discriminators would enhance the final classification performance.

5. Conclusion

This paper has proposed a multi-model, multi-slice EL framework based on 2D CNNs and intended to improve AD diagnostic performance. The proposed EL framework includes two stages. In the first stage, a majority voting ensemble is used to combine multi-slice predictions with complementary results. During the second stage, VGG16, ResNet50, and the DCGAN discriminator are used to construct an ensemble classifier to improve the robustness of AD versus CN classification. Transfer learning is a possible solution for handling biomedical classification problems with a small dataset. Two transfer strategies have been implemented: domain transfer (i.e., fine-tuning pretrained classic CNNs with GMDM images) and task adaptation transfer (i.e., using an AD versus CN classifier as a baseline model for an MCI versus CN task). The experimental results have shown that the EL framework achieved good classification results in most tasks, as compared with a single classifier. Comparing the proposed approach with other benchmark approaches indicates that the former performs comparably to existing state-of-the-art approaches, proving its effectiveness. Considering the lack of pretrained 3D CNNs, the proposed 2D approach is cost-effective and well suited for AD studies with small sample sizes. Future perspectives for this research include studying optimized slice selection strategies for each classifier, incremental ensemble size, time of MCI to AD conversion, and the replacement of CNNs with different DCGAN discriminators.

Author contributions

Kang Wenjie: designed and performed the experiments, performed the data analyses and wrote the manuscript.

Lin Lan: contributed to the conception of the study, and manuscript revision.

Zhang Baiwen: contributed to experiment preparation.

Shen Xiaoqi: contributed to data pre-processing.

Wu Shuicai: contributed to the design the study.

Table 8
Comparison of classification performance of state-of-art studies based on baseline sMRI data of ADNI-1.

Study	CNN Model	Subject Numbers	Accuracy			
			AD versus CN (%)	AD versus MCI (%)	MCI versus CN (%)	MCIc versus MCInc (%)
Aderghal et al. 2020 [16]	2D slice-level	1551	92.5	85.0	80.0	
Lin et al. 2018 [12]	2D slice-level	725	88.8			73.0
Pan et al. 2020 [27]	2D slice-level	787	84		79	62
Ahmed et al. 2020 [34]	2D slice-level	351	93.6	85.5	81.7	
Lian et al. 2020 [18]	3D patch-level	1447	90.3			80.9
Liu et al. 2018 [36]	3D patch-level	1526	91.1			76.9
Our approach	2D slice-level	798	90.4	77.2	72.4	66.7

Acknowledgement

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Declaration of competing interest

None declared.

Funding

This research was financially supported by grants from Natural Science Foundation of Beijing Municipality (L182010), National Natural Science Foundation of China (81971683), and the Scientific Research General Project of Beijing Municipal Education Committee (KM201810005033).

References

[1] E.D. Roberson, L. Mucke, 100 years and counting: prospects for defeating Alzheimer's disease, Science 314 (2006) 781–784.
[2] 2020 Alzheimer's Disease Facts and Figures, Alzheimers Dement (2020).
[3] B. Zhang, L. Lin, S. Wu, Z. Al-Masqari, Multiple subtypes of Alzheimer's disease base on brain atrophy pattern, Brain Sci. 11 (2021).
[4] A. Dechamps, L. Fasotti, J. Jungheim, E. Leone, E. Dood, A. Allioux, P.H. Robert, X. Gervais, N. Maubourguet, M.G. Olde Rikkert, R.P. Kessels, Effects of different learning methods for instrumental activities of daily living in patients with Alzheimer's dementia: a pilot study, Am J Alzheimers Dis Other Demen 26 (2011) 273–281.
[5] B. Zhang, L. Lin, S. Wu, A review of brain atrophy subtypes definition and analysis for Alzheimer's disease heterogeneity studies, J. Alzheim. Dis. 80 (2021) 1339–1352.
[6] S. Gauthier, B. Reisberg, M. Zaudig, R.C. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, H. Chertkow, J.L. Cummings, M. de Leon, H. Feldman, M. Ganguli, H. Hampel, P. Scheltens, M.C. Tierney, P. Whitehouse, B. Winblad, Mild cognitive impairment, Lancet 367 (2006) 1262–1270.
[7] J. Ashburner, K.J. Friston, Voxel-based morphometry—the methods, Neuroimage 11 (2000) 805–821.
[8] S. Rathore, M. Habes, M.A. Iftikhar, A. Shacklett, C. Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages, Neuroimage 155 (2017) 530–548.
[9] N. Burgos, O. Colliot, Machine learning for classification and prediction of brain diseases, recent advances and upcoming challenges 33 (2020) 439–450.
[10] Applications of deep learning to MRI images: a survey, Big Data Mining and Analytics 1 (2018) 1–18.
[11] H.I. Suk, S.W. Lee, D. Shen, I. Alzheimer's Disease Neuroimaging, Deep ensemble learning of sparse regression models for brain disease diagnosis, Med. Image Anal. 37 (2017) 101–113.
[12] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, X. Qu, I. Alzheimer's Disease Neuroimaging, Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment, Front. Neurosci. 12 (2018) 777.
[13] M. Liu, D. Cheng, W. Yan, I. Alzheimer's Disease Neuroimaging, Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images, Front. Neuroinf. 12 (2018) 35.
[14] A.B. Tufail, Y.-K. Ma, Q.-N. Zhang, Binary classification of Alzheimer's disease using sMRI imaging modality and deep learning, J. Digit. Imag. 33 (2020) 1073–1090.
[15] A. Valliani, A. Soni, Deep residual nets for improved Alzheimer's diagnosis, Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics (2017) 615, 615.
[16] K. Aderghal, K. Afdel, J. Benois-Pineau, G. Catheline, Improving Alzheimer's stage categorization with Convolutional Neural Network using transfer learning and different magnetic resonance imaging modalities, Heliyon 6 (2020), e05652.
[17] H.I. Suk, S.W. Lee, D. Shen, I. Alzheimer's Disease Neuroimaging, Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis, Neuroimage 101 (2014) 569–582.
[18] C. Lian, M. Liu, J. Zhang, D. Shen, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2020) 880–893.
[19] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, I. Alzheimer's Disease Neuroimaging, L. Shen, M. Xu, A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease, Neuroimage 208 (2020) 116459.
[20] G. He, A. Ping, X. Wang, Y. Zhu, Alzheimer's disease diagnosis model based on three-dimensional full convolutional DenseNet, in: 2019 10th International Conference on Information Technology in Medicine and Education, ITME, 2019, pp. 13–17.
[21] Y. Shmulev, M. Belyaev, Predicting Conversion of Mild Cognitive Impairments to Alzheimer's Disease and Exploring Impact of Neuroimaging, Springer International Publishing, Cham, 2018, pp. 83–91.
[22] J. Bae, J. Stocks, A. Heywood, Y. Jung, L. Jenkins, V. Hill, A. Katsaggelos, K. Popuri, H. Rosen, M.F. Beg, L. Wang, I. Alzheimer's Disease Neuroimaging, Transfer learning for predicting conversion from mild cognitive impairment to dementia of Alzheimer's type based on a three-dimensional convolutional neural network, Neurobiol. Aging 99 (2021) 53–64.
[23] S. Qiu, P.S. Joshi, M.I. Miller, C. Xue, X. Zhou, C. Karjadi, G.H. Chang, A.S. Joshi, B. Dwyer, S. Zhu, M. Kaku, Y. Zhou, Y.J. Alderazi, A. Swaminathan, S. Kedar, M. H. Saint-Hilaire, S.H. Auerbach, J. Yuan, E.A. Sartor, R. Au, V.B. Kolachalama, Development and validation of an interpretable deep learning framework for Alzheimer's disease classification, Brain 143 (2020) 1920–1933.
[24] J. Hu, Z. Qing, R. Liu, X. Zhang, P. Lv, M. Wang, Y. Wang, K. He, Y. Gao, B. Zhang, Deep learning-based classification and voxel-based visualization of frontotemporal dementia and Alzheimer's disease, Front. Neurosci. 14 (2020) 626154.
[25] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems -, vol. 1, Curran Associates Inc., Lake Tahoe, Nevada, 2012, pp. 1097–1105.
[26] B. Cheng, D. Zhang, D. Shen, Domain transfer learning for MCI conversion prediction, Med Image Comput Comput Assist Interv 15 (2012) 82–90.
[27] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, X. Song, Early detection of Alzheimer's disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning, Front. Neurosci. 14 (2020) 259.
[28] A. Radford, L. Metz, S.C.J.C. Science, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, 2015.
[29] P.S. Aisen, Q&A: the Alzheimer's disease neuroimaging initiative, BMC Med. 9 (2011) 101.
[30] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (2020) 139–144.
[31] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, I. Alzheimer's Disease Neuroimaging, Multimodal classification of Alzheimer's disease and mild cognitive impairment, Neuroimage 55 (2011) 856–867.
[32] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-Gonzalez, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, O. Colliot, I. Alzheimer's Disease Neuroimaging, B. Australian Imaging, a. Lifestyle flagship study of, Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation, Med. Image Anal. 63 (2020) 101694.
[33] J. Young, M. Modat, M.J. Cardoso, A. Mendelson, D. Cash, S. Ourselin, I. Alzheimer's Disease Neuroimaging, Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment, Neuroimage Clin 2 (2013) 735–745.
[34] S. Ahmed, B.C. Kim, K.H. Lee, H.Y. Jung, I. Alzheimer's Disease Neuroimaging, Ensemble of ROI-based convolutional neural network classifiers for staging the Alzheimer disease spectrum from magnetic resonance imaging, PloS One 15 (2020), e0242712.
[35] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, M. Filippi, I. Alzheimer's Disease Neuroimaging, Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks, Neuroimage Clin 21 (2019) 101645.
[36] M. Liu, J. Zhang, E. Adeli, D. Shen, Landmark-based deep multi-instance learning for brain disease diagnosis, Med. Image Anal. 43 (2018) 157–168.