Diagnostic accuracy study of automated stratification of Alzheimer's disease and mild cognitive impairment via deep learning based on MRI

Xiaowen Chen[1]^, Mingyue Tang[2], Aimin Liu[1], Xiaoqin Wei[1]

[1]School of Medical Imaging, North Sichuan Medical College, Nanchong, China; [2]School of Basic Medicine and Forensic Medicine, North Sichuan Medical College, Nanchong, China

Contributions: (I) Conception and design: X Chen; (II) Administrative support: X Wei; (III) Provision of study materials or patients: A Liu; (IV) Collection and assembly of data: M Tang; (V) Data analysis and interpretation: X Chen, X Wei; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiaoqin Wei. School of Medical Imaging, North Sichuan Medical College, No. 234 Fujiang Road, Shunqing District, Nanchong 637000, China. Email: xiaoqin_wei_nsmc@163.com.

Background: Alzheimer's disease (AD) is a widespread neurodegenerative disease that mostly affects the elderly population. Given its prevalence, a precise and efficient stratification system based on AD symptomology that uses functional magnetic resonance imaging (MRI) has great potential in the clinical diagnosis and prognosis estimation of AD patients. It was evident that deep learning methods have performed extremely well in the field of automated stratification of AD based on MRI because of their high predicting accuracy and reliability.

Methods: We proposed a deep convolutional neural network (CNN) and iterated random forest (RF) architecture for MRI image stratification by both anatomical location and image modality using the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. We employed 3 cross-sectional data sets from the ADNI to conduct our binary-stratification [AD and normal controls (NCs), or AD and mild cognitive impairment (MCI)], and multi-stratification (AD, MCI, and NCs) process using MRI. And the accuracy, recall, specificity, area under the curve of receiver operating characteristic curve (AUC), F1 and Matthew's correlation coefficient (MCC) scores to assess accuracy of auxiliary clinical diagnoses.

Results: Compared to other combinations of algorithms, our model obtained remarkable overall stratification accuracies in all different classification sets. In terms of AD *vs.* MCI, the mean training AUC of the 3 runs were 85.1% in 95% confidence intervals (CIs). In terms of AD *vs.* NC, the mean training AUC of the 3 runs was 90.6% in 95% CIs. In terms of the 3 stratifications of AD, MCI, and NC, relative precision, recall, and specificity for each category in the training test (TS) were all near 89%, while the F1 and MCC scores of both sets were 59.9% and 59.5%, respectively.

Conclusions: Using a deep CNN and iterated RF architecture, we showed that brain image stratification is a promising means for evaluating AD, and examining the underlying etiology of the disease, by applying computer and medical images to achieve the early auxiliary diagnosis of AD. However, we still have a long way to go from the discovery of image markers to clinical application.

Keywords: Alzheimer's disease (AD); functional magnetic resonance imaging (functional MRI); deep convolutional neural networks (deep CNN); iterated random forest (iterated RF); Alzheimer's Disease Neuroimaging Initiative (ADNI)

^ ORCID: 0000-0002-0313-6363.

Submitted May 26, 2022. Accepted for publication Jul 06, 2022.
doi: 10.21037/atm-22-2961
View this article at: https://dx.doi.org/10.21037/atm-22-2961

Introduction

Alzheimer's disease (AD) is a primary cause of dementia. In recent years, many studies have employed neuroimaging bio-indicators to stratify AD patients or estimate disease progression (1). In the initial stages of the disease, AD produces learning and memory impairment in the hippocampus. The corresponding symptoms include forgetfulness and confusion (2). As the disease progresses, the patient's visual acuity diminishes, and in combination with serious memory decline, the patient becomes unable to distinguish faces or items (3). This places a heavy burden on patients and society. Thus, improvements in early AD diagnosis, followed by prompt treatment, is critical (4).

In recent years, with rapid advancements in computer and neuroimaging technology, doctors have employed computer and medical images to achieve the early diagnosis of AD. The analysis of AD has become a mainstream trend (5). The current technical difficulties in relation to the analysis of AD primarily include extracting effective classification features from medical images, establishing good robustness, and designing and constructing a simple structure for the classification model (6).

Magnetic resonance imaging (MRI) is a non-invasive research method that measures blood oxygen level-dependent signals in the brain (7). It can accurately determine the amount of patient brain oxygen activity at a given time, and it is widely used in AD diagnostic research (8,9). Thus, it is of the utmost importance to develop objective AD bio-indicators that aid in neuroimaging evaluations for the determination of AD clinical diagnosis and treatment outcomes (10,11).

Nowadays, clinical decision solutions can be interpreted in two different ways by comparing previous knowledge contained in data sets. One is a quick or intuitive approach that uses basic clinical pattern recognition, often used in medical emergencies. But these all have a higher probability of being wrong and providing an incomplete view. The other approach is the slow or rational approach. It is deductive and deliberate, requiring more intelligence, time and cost information. But they make more accurate decisions. Because all of these decisions are based on data collected, analyzed, and stored in complex and heterogeneous forms, it is important to use an algorithmic approach to minimize the computing power required. Machine learning applications are currently making a significant contribution to the global healthcare sector to improve its quality, and will continue to do so (12).

To date, researchers have developed multiple computer-aided systems to establish a precise disease diagnosis (13). Between the 1970s and 1990s, scientists designed a rule-based expert framework, and post 1990s, they designed supervised models. To train supervised models, features are generally extracted from task-based images; however, such models require human specialists, as well as ample effort, time, and funding (14). This presents an enormous challenge for the continuation of this mode of disease diagnosis (15,16). With the emergence of deep-learning (DL) models, it is feasible to retrieve features directly from imaging data without human intervention. The relative ease of this approach is compelling more research into DL models to enable the precise diagnosis of various diseases (17,18). Compared to the challenges of other image analysis programs [e.g., computed tomography (CT), MRI, X-rays, ultrasounds, and sentiment analysis], DL models have achieved considerable success (16,19). Notably, they have been reported to produce reliable results in terms of disease diagnosis and stratification, especially in the lungs, abdomen, brain, cardiovascular, and retina. However, it is still an enormous challenge for scientists to detect AD using DL models. This is likely due to reduced acquisition and errors in preprocessed medical images, as well as the problematic recognition of the cerebral regions of interest (ROI), disproportional data-set class participants, data-set inaccessibility, and reduced differences between varying classes in different phases of AD (20). To make matters worse, symptoms that are distinct to AD (e.g., hippocampal shrinkage) are also sometimes evident in the healthy aging brain. Further, relative to natural images, medical images are often complex (7,21,22).

To address this problem, this study sought to develop fully automatic CNN models for the multi- or binary-classification of the brain. Using random forest (RF) (23), we demonstrated that data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) can be used to stratify subjects as being either cognitively normal or having dementia (24,25). In this report, we begin with a description of the ADNI data, and then discuss the specific protocols and results employed in the stratification. Next, we examine the statistical information corresponding to the classifiers, report the multiple metrics employed for the stratification assessment, employ the ADNI data set to stratify subjects' disease status using MRI images as predictors, and finally, analyze the results across all classifiers (26,27). We also report the results of a simulation examination of the model's performance using higher resolution images, and finally, we explain the implications of our results, outline future research directions, and address the limitations of our current investigation. We present the following article in accordance with the STARD reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-22-2961/rc).

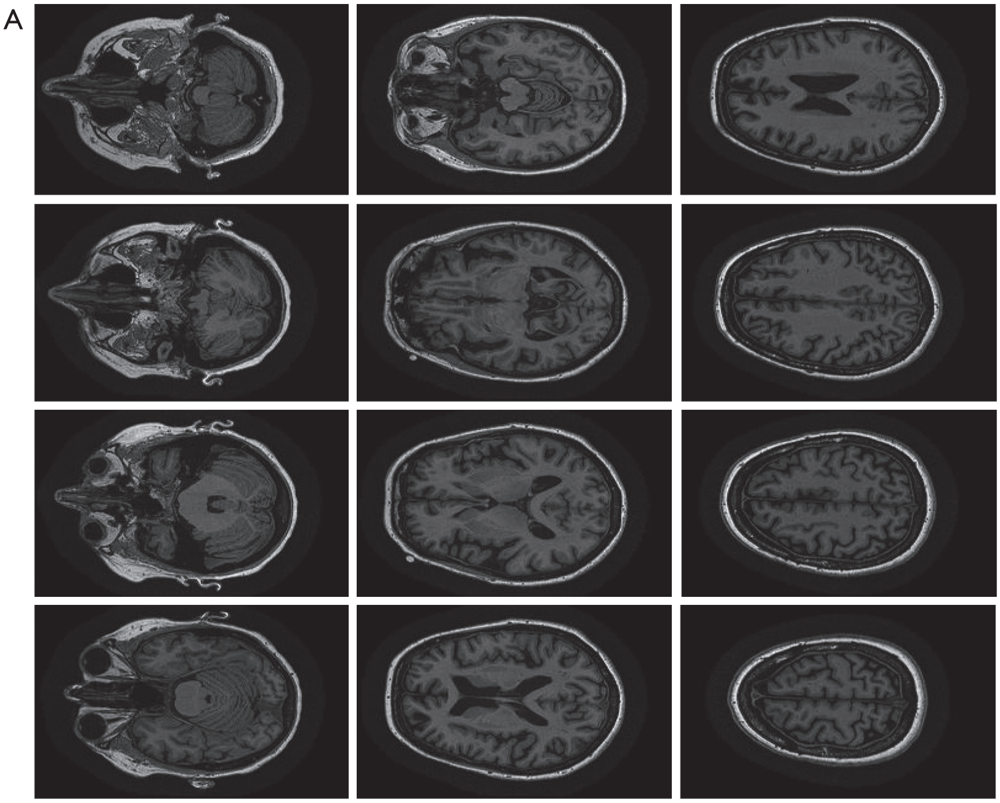
Methods

ADNI details

Data from the ADNI database (adni.loni.usc.edu) were analyzed in this study. ADNI was established in 2003 by Michael W. Weiner, MD, as a public-private partnership (28,29). We used 3 cross-sectional data sets from the ADNI to conduct our binary-stratification [AD or normal controls (NC), or AD and mild cognitive impairment (MCI)] or multi-stratification (AD, MCI, and NC) process using MRI. Overall, we examined 200 subjects, among whom 43 (21.5%) were diagnosed with AD, 97 (48.5%) with MCI, and 60 (30%) as NC (*Figure 1*). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Implementation setup

This section describes the implementation procedure of bio-indicator detection from brain MRI images using deep CNNs and iterated RF. The procedures were executed on Intel(R) Core (TM) i7-7500U, with NVIDIA Tesla V100 32G, and Window 10. Our designed CNNs were trained on brain MRI images, and they predicted and classified brain images as either normal or abnormal. Graphics processing units (GPUs) are known to significantly enhance the training procedure of various models. Intensive computation, matrix multiplication, and other operations



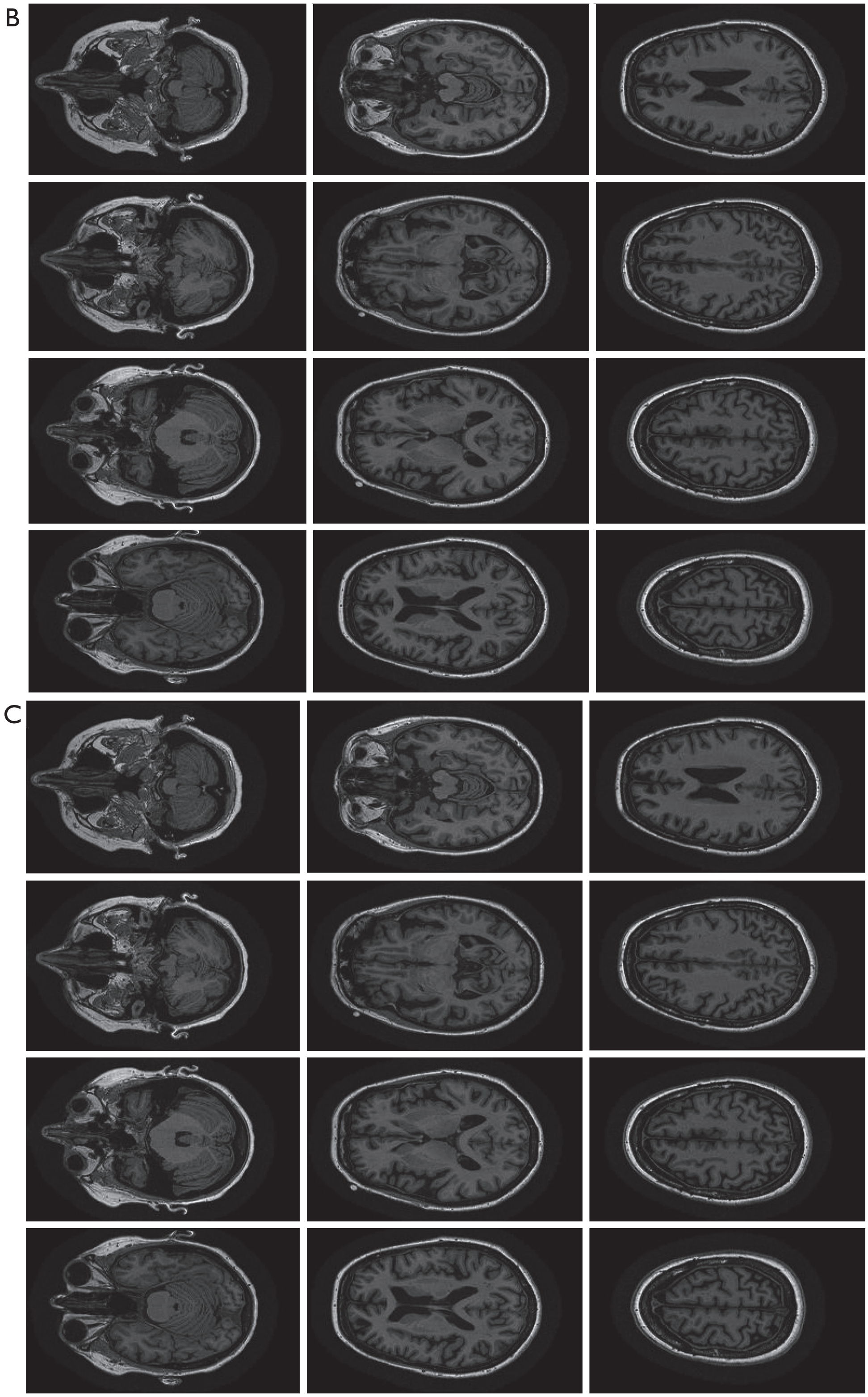


Figure 1 Example image of each modality and anatomical location. (A) AD, (B) MCI, and (C) NC. The original images are obtained from the ADNI's database (https://ida.loni.usc.edu/login.jsp?project=ADNI). AD, Alzheimer's disease; MCI, mild cognitive impairment; NC, normal controls.

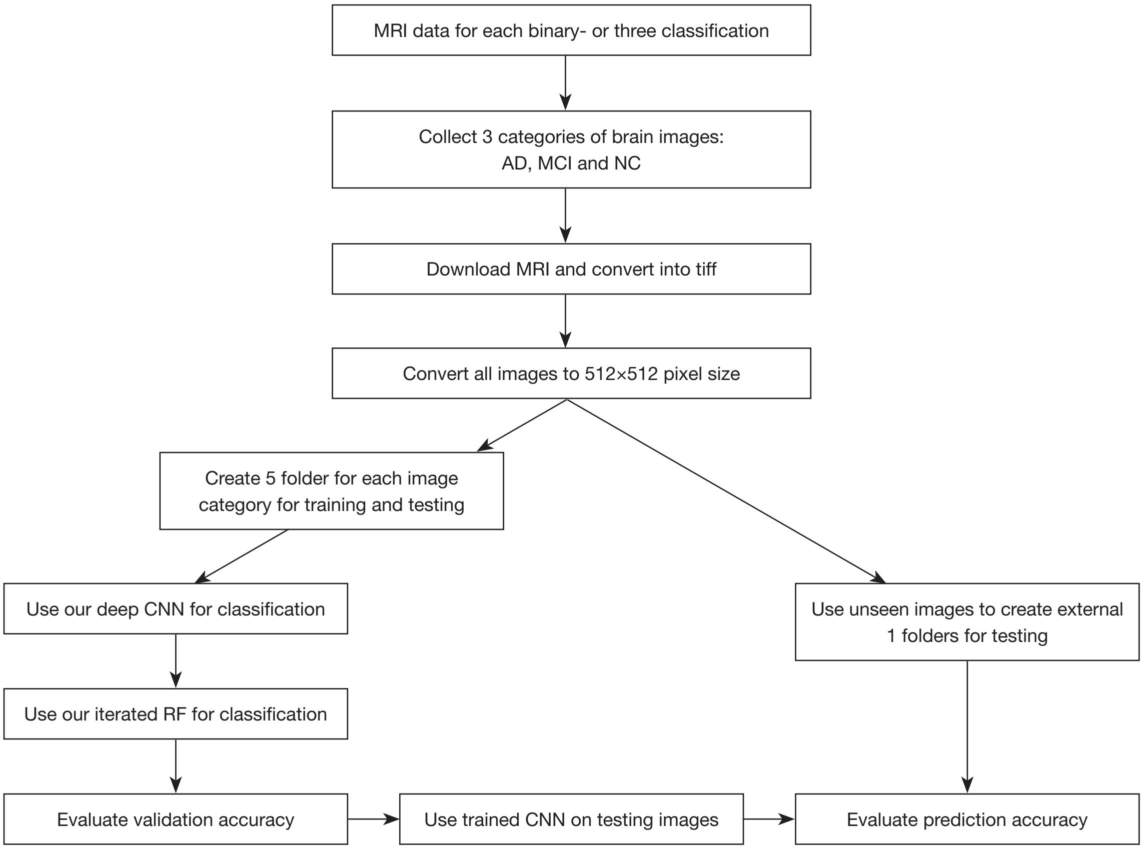

MRI data for each binary- or three classification
Collect 3 categories of brain images: AD, MCI and NC
Download MRI and convert into tiff
Convert all images to 512×512 pixel size
Create 5 folder for each image category for training and testing
Use our deep CNN for classification
Use our iterated RF for classification
Evaluate validation accuracy
Use trained CNN on testing images
Use unseen images to create external 1 folders for testing
Evaluate prediction accuracy

Figure 2 Workflow chart. AD, Alzheimer's disease; MCI, mild cognitive impairment; NC, normal controls; MRI, magnetic resonance imaging; CNN, convolutional neural network.

were included in the training models, such as image stratification. We employed GPUs with machine-learning (ML) frameworks to train our model in this study. We employed several libraries, including Keras, TensorFlow, NumPy, and SciPy, to construct our CNNs. Next, we used an iterated RF to retrieve more specific and relevant characteristics. We also used Python 3.6 to construct certain graphs. The data set we used was composed of T2-weighted MRI brain images in the axial plane, with 512×512 plane resolution. We downloaded the data set from https://adni.loni.usc.edu/. In total, we arbitrarily selected 1,937 images, among which, 621 were AD, 445 were MCI, and 871 were NC. Our study flowchart is provided in *Figure 2*.

Model evaluation

The K-fold cross validation separated participants into *k* independent sets and fitted the model *k* times. At each model fitting, a separate set was used as the test set. The mean prediction error across all the sets provided an estimate prediction for the expected estimation error, and parameter values that reduced the estimate of expected estimation error were employed to fit the classifier/model using all the available data. As most of the data sets had very small sample sizes, they could not be used as separate test sets, and as the *k*-fold cross validation provided an estimation of the expected estimation error, it enabled us to assess model performance while using all the available data to construct the model.

Statistical analysis

To avoid deception, the classifiers must be evaluated in relation to multiple metrics. A considerable challenge in this study was the precision with which a classifier or the probability with which an algorithm could accurately stratify a subject. However, this metric can be misleading, particularly, in conditions in which the sample sizes are

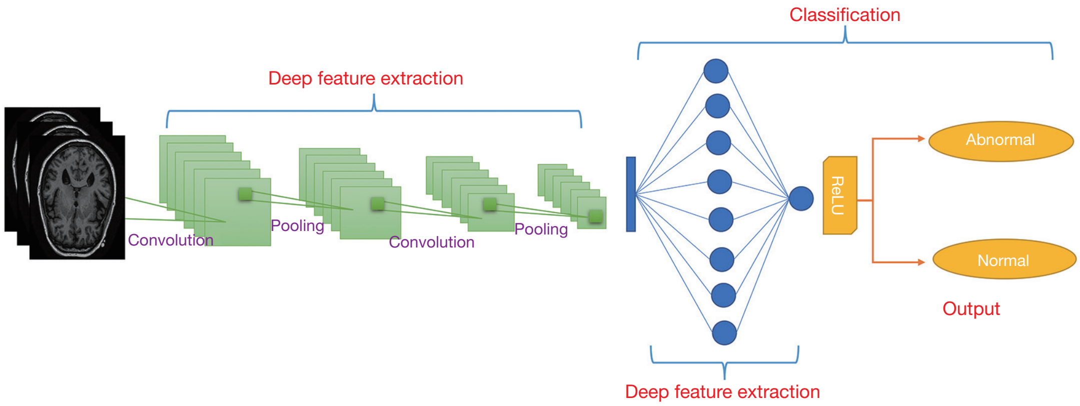


Figure 3 Layered architecture of CNN. CNN, convolutional neural network.

unbalanced. Thus, along with precision, it is critical to predict and consider multiple other metrics while assessing stratification performance (i.e., accuracy, recall, and specificity). Preferably, all metrics must be near 1 if the classifier is performing satisfactorily. Additionally, it may be of benefit to assess stratification performance using measures that assess all 4-confusion matrices [i.e., true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs)]. The Matthew's correlation coefficient (MCC) has advantages over precision value and the F1 score; the form of the metrics used is as follows in Equations 1 and 2:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)\cdot(TP+FN)\cdot(TN+FP)\cdot(TN+FN)}} \quad [1]$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad [2]$$

All the aforementioned metrics must be predicted in a cross-validation process, as our major concern in predicting metrics values is applying the classifier to independent data.

CNN architecture

CNNs belong to a group of deep neural networks that use convolutional layers (CLs) to filter inputs, and computer neuronal output that is related to certain regions within the input (30). In this way, a CNN is able to extract both spatial and temporal characteristics from an image (28). A weight-sharing method is generally employed in CNN-based CLs to minimize the total quantity of parameters. CNNs are composed of the following 3 building blocks: (I) a CL that examines the spatial and temporal profiles; (II) a subsampling (max-pooling) layer that minimizes or down-samples the dimensionality of an image; (III) a fully connected layer that classifies the input image into different categories. The CNN architecture is shown in *Figure 3*.

We generated our model using CNNs. We modified the original CNN architecture to enable it to read an image 512-by-512-by-1 in size, with 512 as the pixel image size, and 1 as the grayscale image. The collected information was then arbitrarily separated into the training set (TS) and validation set (VS); 80% of the images were used for the TS, and 20% were used for the VS. The algorithm was trained with the TS, and the hyperparameters were tuned with the VS. After several attempts, we made the following modifications: the CNNs were made to possess 5 sets of CLs, with subsequent batch normalization, ReLU, and maximum-pooling layers. The 5 CLs had a filter size of 8-by-8l; however, we slowly increased the filter quantity such that we had 16 filters in the 1st CL, 16 in the 2nd CL, 32 in the 3rd CL, 48 in the 4th CL, and 64 in the 5th CL. We also included 1 default padding and 1 stride in each CL. The entire connected layer was then adjusted to either binary or 3 as we attempted to classify either the binary or 3 categories. Subsequently, we used the TS to predict the labels of the VS and computed the prediction precision (i.e., the fraction of labels that was predicted accurately). The network employed stochastic gradient descent with momentum, and had an initial learning rate of 0.01. After multiple attempts to obtain an optimal result, the quantity of epochs was set to a maximum of 16. To further evaluate the prediction precision of the TS, we generated a test set with the binary or 3 categories of images. Each category contained 30 images that were new and unseen to the algorithm. We computed the prediction precision of these unseen data using the same procedure as

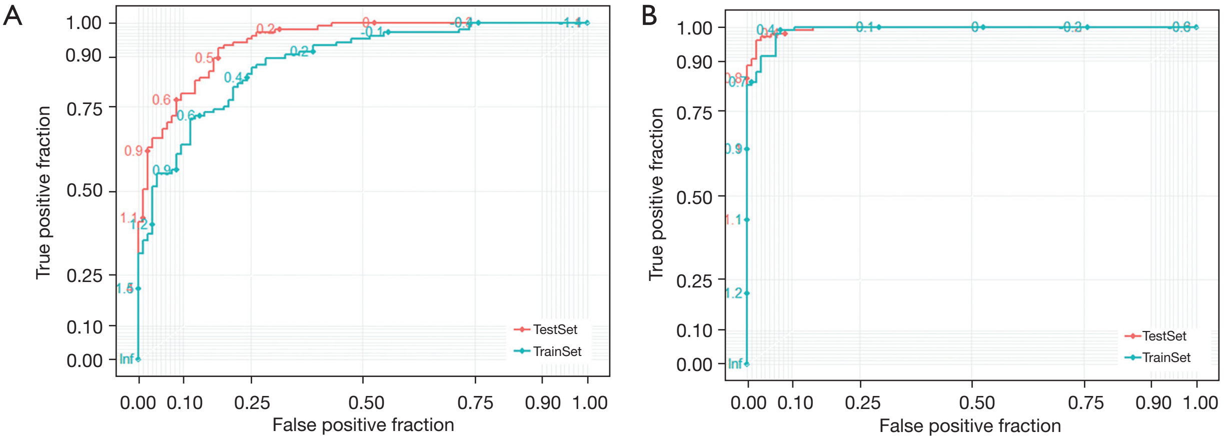


Figure 4 The results of the proposed model's AUC values during the training and testing stages (A) AD *vs.* MCI, and (B) AD *vs.* NC. AUC, area under the curve of receiver operating characteristic curve; AD, Alzheimer's disease; MCI, mild cognitive impairment; NC, normal controls.

that used for the VS.

RF

Using different ML classifiers [i.e., support vector machine (SVM), *k*-nearest neighbor (k-NN), and RF], we assessed each deep characteristic retrieved from the pre-trained CNNs. RF, which was introduced by Breiman, is an ensemble learning algorithm that generates several decision trees using the bagging technique to stratify novel data instance (a deep characteristic of a brain MRI image) to a category target with 3 categories (i.e., AD, MCI, and NC) for 2 MRI data sets. We used the RF algorithm to stratify our input images. We used the Gini index to determine the gain in class "purity" in 1,000 CART trees. We then iteratively repeated these steps until no further improvements could be made. Our arbitrary selection of characteristics reduced the association among different trees and lowered the ensemble error rates. We next fed this observation into all RF stratification trees to predict a category target of new incoming data instance. The RF records the estimation quantity for each category and chooses the category with the most votes as the category label for the new data instance. In our study, the characteristic quantity was adjusted to the square root of the total characteristic quantities to achieve the optimal split. Additionally, we adjusted the quantity of trees from 1 to 150 and chose the 1 with the best precision.

Further, we removed the RF usage to train characteristics with a value of 0, and instead used the reserved characteristics for the training. Next, we repeated the same method until the characteristic quantity with a critical value of 0 was <1,000. Subsequently, the characteristics with a critical value <1.00e-7 were removed, and the reserved characteristics were employed for training purposes. The same method was repeated again until the characteristic quantity with a critical value <1.00e-7 was <1,000. Next, the characteristics with a critical value <1.00e-6 were removed, and the reserved characteristics were employed for training purposes. The same method was repeated again until the characteristic quantity with a critical value <1.00e-6 was <1,000. Finally, the iteration was continued until the final 10 most critical characteristics were retrieved.

Results

We trained our model and repeated each run 3 times. The duration of the training and testing was approximately 25–30 minutes. In terms of AD *vs.* MCI, the mean training area under the curve of receiver operating characteristic curve (AUC) of the 3 runs were 85.1%. We next employed our trained network to predict unseen images from our TS. Following 3 runs, the model obtained a mean precision rate of 87.9% (*Figure 4A*). In terms of AD *vs.* NC, the mean training AUC of the 3 runs was 90.6%. We next employed our trained network to predict unseen images from our TS. Following 3 runs, the model obtained a mean precision rate of 92.3% (*Figure 4B*).

In terms of AD *vs.* MCI, relative to the CNNs, the CNNs + SVM, CNNs + *k*-NNs, and CNNs + RF, the

Table 1 Simulation study: average classification performance between AD and MCI

Models	AD *vs.* MCI					
	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 score (%)	MCC (%)
CNN	88.9	87.2	89.1	87.1	58.2	57.5
CNN + SVM	89.1	89.1	88.1	89.1	58.2	58.6
CNN + *k*-NN	89.2	88.2	87.2	89.1	58.1	57.5
CNN + RF	88.1	89.2	89.1	88.1	58.1	56.1
CNN + iterated RF	92.1	92.2	92.4	92.4	62.5	61.5

AD, Alzheimer's disease; MCI, mild cognitive impairment; CNN, convolutional neural network; SVM, support vector machine; *k*-NN, *k*-nearest neighbor; RF, random forest; MCC, Matthew's correlation coefficient.

Table 2 Simulation study: average classification performance between AD and NC

Models	AD *vs.* NC					
	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 score (%)	MCC (%)
CNN	88.9	87.2	89.2	88.3	59.4	59.2
CNN + SVM	88.1	89.1	88.1	89.2	59.3	58.1
CNN + *k*-NN	89.2	89.3	89.1	89.2	59.4	59.1
CNN + RF	89.1	89.2	89.3	89.1	59.4	59.1
CNN + iterated RF	94.6	93.7	94.3	93.2	63.8	61.9

AD, Alzheimer's disease; NC, normal controls; CNN, convolutional neural network; SVM, support vector machine; *k*-NN, *k*-nearest neighbor; RF, random forest; MCC, Matthew's correlation coefficient.

CNNs + iterated RF scores for precision, recall, and specificity for each category in the TS were all near 92%, while the F1 and MCC scores of both sets were 62.5% and 61.5%, respectively (*Table 1*). In terms of AD *vs.* NC, relative to the CNNs, the CNNs + SVM, CNN + *k*-NNs, and CNN + RF, the iterated RF scores for precision, recall, and specificity for each category in the TS were all 93–94%, while the F1 and MCC scores of both sets were 63.8% and 61.9%, respectively (*Table 2*). In terms of the 3 stratifications of AD, MCI, and NC, relative to the CNN, the CNN, CNN + SVM, CNN + *k*-NNs and CNN + RF, the iterated RF scores for precision, recall, and specificity for each category in the TS were all near 89%, while the F1 and MCC scores of both sets were 59.9% and 59.5%, respectively (*Table 3*).

Discussion

Based on our analyses, we found that volumetric models hold great potential in disease stratification, which can aid in determining symptomatic AD diagnosis and possible patient outcomes. DL is an innovative method of discriminative image characteristic extraction (31). The CNN is a frequently employed DL in numerous image analyses and computer vision-based tasks. DL automatically studies visual characteristics from input pixel images via a mechanism of deep-layer receptive field combination and pooling. It is reported to have exceptional performance relative to other traditional ML programs.

MRI is a medical imaging diagnostic program that is safe, non-invasive, non-persistent, and pain-free (32,33). Unlike CT and other imaging, MRI is not associated with radiation (34); rather, it uses a uniform magnetic field and radio-frequency to display the internal system of the human body. Additionally, 2- and 3-dimensional MRI images are usually of high-quality, particularly, in terms of resolution and contrast (35). These digital formats provide enormous medical information regarding internal diseases and soft tissue differentiation that can be used for further analyses and stratification. Further, MRI yields detailed information on soft tissue abnormalities that may not be detected by CT or X-ray radiography (36).

Table 3 Simulation study: average classification performance between AD, MCI and NC

Models	AD *vs.* MCI *vs.* NC					
	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 score (%)	MCC (%)
CNN	88.6	87.2	88.1	87.9	58.1	58.0
CNN + SVM	87.7	88.2	87.9	87.9	59.9	59.6
CNN + *k*-NN	88.4	88.3	88.4	88.1	58.2	58.1
CNN + RF	89.1	89.0	89.2	89.1	59.7	59.1
CNN + iterated RF	89.2	89.1	89.3	89.2	59.9	59.5

AD, Alzheimer's disease; MCI, mild cognitive impairment; NC, normal controls; CNN, convolutional neural network; SVM, support vector machine; *k*-NN, *k*-nearest neighbor; RF, random forest; MCC, Matthew's correlation coefficient.

MRI and ML have greatly benefitted the identification of AD bio-indicators (17,37,38). In most abnormal brain imaging investigations, brain images are classified as either normal or abnormal (39). Upon disease detection, the next steps are typically location identification and personalized treatment design. Previous studies have identified multiple stratification characteristics of AD (3,40). These include the amplitude of low frequency fluctuations or hippocampal association with reduced frequency components, regional homogeneity, functional association with ROI strength, in terms of the automated anatomical labeling (AAL) atlas, whole-brain or selected regional functional correlation connectivity matrices, covariance connectivity matrices, and graph-theoretical measures. For example, to assist AD diagnosis and support the monitoring of disease progression, Dai *et al.* introduced a methodological framework, called the multi-modal imaging and multi-level characteristics with multi-classifiers (M3), to distinguish AD patients from healthy controls (41). Tripoliti *et al.* established a 6-stage procedure based on the characteristics obtained from functional MRI (fMRI) data to stratify AD patients (42). Armananzas *et al.* and Harper *et al.* introduced the direct usage of brain fMRI activation voxels to address the automatic pattern analysis of AD and healthy individuals by applying various ML methods to fMRI data stratification (8,43). However, the ROI-based technique has certain limitations. First, the ROIs examined are usually pre-determined based on previous knowledge and are thus not data-driven or exploratory. Second, early detection accuracy is usually based on an examiner's previous experience. Third, this technique has low efficiency. Fourth, it is difficult to manage mutual data among voxels.

In this study, we introduced a deep CNN and iterated RF architecture for the stratification of brain images using both their anatomical location and image modality. We employed JPEG images and obtained remarkable overall stratification precision in both the TS and VS (>89%). We also obtained markedly elevated F1 and MCC scores (>59%) in each category (of AD, MCI, and NC). Our findings validated the use of deep CNNs and iterated RF in medical image stratification. Our proposed method could potentially decrease the image processing time and save storage space in real-life scenarios. We recommend additional investigations be conducted on other anatomical locations, using other imaging modalities to achieve a fully automated medical image stratification system that can be employed in both clinical and research settings.

Acknowledgments

Funding: This research was partly supported by the City-School Science and Technology Strategic Cooperation Project in Nanchong City (grant No. 19SXHZ0239) and the Project of Sichuan Provincial Primary Health Development Research Center (grant No. SWFZ21-Z-05).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-22-2961/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-2961/coif). All authors report that this research was partly supported by the City-School Science and Technology Strategic Cooperation Project in Nanchong City (grant No.19SXHZ0239) and the

Project of Sichuan Provincial Primary Health Development Research Center (grant No. SWFZ21-Z-05). The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Leach JM, Edwards LJ, Kana R, et al. The spike-and-slab elastic net as a classification tool in Alzheimer's disease. PLoS One 2022;17:e0262367.
2. Weiner MW, Veitch DP, Aisen PS, et al. Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. Alzheimers Dement 2017;13:e1-e85.
3. Archetti D, Ingala S, Venkatraghavan V, et al. Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer's disease. Neuroimage Clin 2019;24:101954.
4. Popuri K, Ma D, Wang L, et al. Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: Independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. Hum Brain Mapp 2020;41:4127-47.
5. Qiu S, Joshi PS, Miller MI, et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain 2020;143:1920-33.
6. Basaia S, Agosta F, Wagner L, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. Neuroimage Clin 2019;21:101645.
7. Mapstone M, Gross TJ, Macciardi F, et al. Metabolic correlates of prevalent mild cognitive impairment and Alzheimer's disease in adults with Down syndrome. Alzheimers Dement (Amst) 2020;12:e12028.
8. Armananzas R, Iglesias M, Morales DA, et al. Voxel-Based Diagnosis of Alzheimer's Disease Using Classifier Ensembles. IEEE J Biomed Health Inform 2017;21:778-84.
9. Shi Y, Zeng W, Deng J, et al. The Identification of Alzheimer's Disease Using Functional Connectivity Between Activity Voxels in Resting-State fMRI Data. IEEE J Transl Eng Health Med 2020;8:1400211.
10. Qiao J, Lv Y, Cao C, et al. Multivariate Deep Learning Classification of Alzheimer's Disease Based on Hierarchical Partner Matching Independent Component Analysis. Front Aging Neurosci 2018;10:417.
11. Weiner MW, Veitch DP, Aisen PS, et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement 2013;9:e111-94.
12. Jayatilake SMDAC, Ganegoda GU. Involvement of Machine Learning Tools in Healthcare Decision Making. J Healthc Eng 2021;2021:6679512.
13. Lima AA, Mridha MF, Das SC, et al. A Comprehensive Survey on the Detection, Classification, and Challenges of Neurological Disorders. Biology (Basel) 2022;11:469.
14. Wang X, Huang W, Su L, et al. Neuroimaging advances regarding subjective cognitive decline in preclinical Alzheimer's disease. Mol Neurodegener 2020;15:55.
15. Lanka P, Rangaprakash D, Dretsch MN, et al. Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets. Brain Imaging Behav 2020;14:2378-416.
16. Ni R. Magnetic Resonance Imaging in Animal Models of Alzheimer's Disease Amyloidosis. Int J Mol Sci 2021;22:12768.
17. Jo T, Nho K, Risacher SL, et al. Deep learning detection of informative features in tau PET for Alzheimer's disease classification. BMC Bioinformatics 2020;21:496.
18. Kothapalli SVVN, Benzinger TL, Aschenbrenner AJ, et al. Quantitative Gradient Echo MRI Identifies Dark Matter as a New Imaging Biomarker of Neurodegeneration that Precedes Tisssue Atrophy in Early Alzheimer's Disease. J Alzheimers Dis 2022;85:905-24.
19. Feng X, Provenzano FA, Small SA, et al. A deep learning MRI approach outperforms other biomarkers of prodromal Alzheimer's disease. Alzheimers Res Ther 2022;14:45.

20. Fernández Montenegro JM, Villarini B, Angelopoulou A, et al. A Survey of Alzheimer's Disease Early Diagnosis Methods for Cognitive Assessment. Sensors (Basel) 2020;20:7292.
21. Counts SE, Ikonomovic MD, Mercado N, et al. Biomarkers for the Early Detection and Progression of Alzheimer's Disease. Neurotherapeutics 2017;14:35-53.
22. Hall Z, Chien B, Zhao Y, et al. Tau deposition and structural connectivity demonstrate differential association patterns with neurocognitive tests. Brain Imaging Behav 2022;16:702-14.
23. Lebedev AV, Westman E, Van Westen GJ, et al. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. Neuroimage Clin 2014;6:115-25.
24. Bhagwat N, Pipitone J, Voineskos AN, et al. An artificial neural network model for clinical score prediction in Alzheimer disease using structural neuroimaging measures J Psychiatry Neurosci 2019;44:246-60.
25. Song A, Yan J, Kim S, et al. Network-based analysis of genetic variants associated with hippocampal volume in Alzheimer's disease: a study of ADNI cohorts. BioData Min 2016;9:3.
26. Carro E, Bartolomé F, Bermejo-Pareja F, et al. Early diagnosis of mild cognitive impairment and Alzheimer's disease based on salivary lactoferrin. Alzheimers Dement (Amst) 2017;8:131-8.
27. Amini M, Pedram MM, Moradi A, et al. Single and Combined Neuroimaging Techniques for Alzheimer's Disease Detection. Comput Intell Neurosci 2021;2021:9523039.
28. Jo T, Nho K, Saykin AJ. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. Front Aging Neurosci 2019;11:220.
29. He P, Qu H, Cai M, et al. Structural Alteration of Medial Temporal Lobe Subfield in the Amnestic Mild Cognitive Impairment Stage of Alzheimer's Disease. Neural Plast 2022;2022:8461235.
30. Punjabi A, Martersteck A, Wang Y, et al. Neuroimaging modality fusion in Alzheimer's classification using convolutional neural networks. PLoS One 2019;14:e0225759.
31. Rathore S, Habes M, Iftikhar MA, et al. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. Neuroimage 2017;155:530-48.
32. Jansen MG, Geerligs L, Claassen JAHR, et al. Positive Effects of Education on Cognitive Functioning Depend on Clinical Status and Neuropathological Severity. Front Hum Neurosci 2021;15:723728.
33. Mondragón JD, Marapin R, De Deyn PP, et al. Short- and Long-Term Functional Connectivity Differences Associated with Alzheimer's Disease Progression. Dement Geriatr Cogn Dis Extra 2021;11:235-49.
34. Zhang B, Lin L, Liu L, et al. Concordance of Alzheimer's Disease Subtypes Produced from Different Representative Morphological Measures: A Comparative Study. Brain Sci 2022;12:187.
35. Huang W, Li X, Li H, et al. Accelerated Brain Aging in Amnestic Mild Cognitive Impairment: Relationships with Individual Cognitive Decline, Risk Factors for Alzheimer Disease, and Clinical Progression. Radiol Artif Intell 2021;3:e200171.
36. Dyrba M, Hanzig M, Altenstein S, et al. Improving 3D convolutional neural network comprehensibility via interactive visualization of relevance maps: evaluation in Alzheimer's disease. Alzheimers Res Ther 2021;13:191.
37. Berlyand Y, Weintraub D, Xie SX, et al. An Alzheimer's Disease-Derived Biomarker Signature Identifies Parkinson's Disease Patients with Dementia. PLoS One 2016;11:e0147319.
38. Llano DA, Bundela S, Mudar RA, et al. A multivariate predictive modeling approach reveals a novel CSF peptide signature for both Alzheimer's Disease state classification and for predicting future disease progression. PLoS One 2017;12:e0182098.
39. Goryawala M, Zhou Q, Barker W, et al. Inclusion of Neuropsychological Scores in Atrophy Models Improves Diagnostic Classification of Alzheimer's Disease and Mild Cognitive Impairment. Comput Intell Neurosci 2015;2015:865265.
40. Korolev IO, Symonds LL, Bozoki AC, et al. Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification. PLoS One 2016;11:e0138866.
41. Dai Z, Yan C, Wang Z, et al. Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). Neuroimage 2012;59:2187-95.
42. Tripoliti EE, Fotiadis DI, Argyropoulou M, et al. A six stage approach for the diagnosis of the Alzheimer's disease based on fMRI data. J Biomed Inform 2010;43:307-20.

43. Harper L, Fumagalli GG, Barkhof F, et al. MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. Brain 2016;139:1211-25.

(English Language Editor: L. Huleatt)

Cite this article as: Chen X, Tang M, Liu A, Wei X. Diagnostic accuracy study of automated stratification of Alzheimer's disease and mild cognitive impairment via deep learning based on MRI. Ann Transl Med 2022;10(14):765. doi: 10.21037/atm-22-2961