Journal of Neural Engineering

A whole process interpretable and multi-modal deep reinforcement learning for diagnosis and analysis of Alzheimer's disease

A Whole Process Interpretable and Multi-modal Deep Reinforcement Learning for Diagnosis and Analysis of Alzheimer's Disease

Quan Zhang[1,2], Qian Du[1,2], Guohua Liu[1,2,3,*]

1. College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, China
2. Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Nankai University, Tianjin 300350, China
3. Engineering Research Center of thin film optoelectronics technology, Ministry of Education, Nankai University, Tianjin 300350, China

* Corresponding author: Guohua Liu, liugh@nankai.edu.cn.

Abstract

Objective

Alzheimer's disease (AD), a common disease of the elderly with unknown etiology, has been bothering many people, especially with the aging of the population and the younger trend of this disease. Current AI methods based on individual information or magnetic resonance imaging (MRI) can solve the problem of diagnostic sensitivity and specificity, but still face the challenges of interpretability and clinical feasibility. In this study, we propose an interpretable multimodal deep reinforcement learning model for inferring pathological features and diagnosis of Alzheimer's disease.

Approach

First, for better clinical feasibility, the compressed-sensing MRI image is reconstructed by an interpretable deep reinforcement learning model. Then, the reconstructed MRI is input into the full convolution neural network to generate a pixel-level disease probability of risk map (DPM) of the whole brain for Alzheimer's disease. Finally, the DPM of important brain regions and individual information are input into the attention-based fully deep neural network to obtain the diagnosis results and analyze the biomarkers. 1349 multi-center samples were used to construct and test the model.

Main Results

Finally, the model obtained 99.6%±0.2, 97.9%±0.2, and 96.1%±0.3 area under curve (AUC) in ADNI, AIBL, and NACC, respectively. The model also provides an effective analysis of multimodal pathology and predicts the imaging biomarkers on MRI and the weight of each individual information. In this study, a deep reinforcement learning model was designed, which can not only accurately diagnose AD, but also analyze potential biomarkers.

Significance

In this study, a deep reinforcement learning model was designed. The model builds a bridge between clinical practice and artificial intelligence diagnosis and provides a viewpoint for the interpretability of artificial intelligence technology.

Keywords: deep learning, pathological analysis, biomarker prediction, interpretable artificial intelligence, reinforcement learning

Abbreviation

AD=Alzheimer's disease; **PET**=positron emission tomography; **MRI**=magnetic resonance imaging; **MMSE**=mini mental state examination; **AI**=artificial intelligence; **ROI**=region of interest; **ADNI=** Alzheimer's disease neuroimaging initiative dataset; **AIBL**=Australian imaging biomarker and lifestyle flagship study of aging; **NACC**= National Alzheimer's coordinating center; **t-SNE**=t-distributed stochastic neighbor embedding; **CS-MRI**=Compresses sensing MRI; **HQ-MRI**=High-quality MRI; **MDP**=Markov decision process; **DRL**=Deep reinforcement learning; **A2C**=Advantage actor critical algorithm; **DDPG**=Deep deterministic policy gradient algorithm; **FCN**=Full convolution neural network; **MCC**=Matthews correlation coefficient; **NL**=Normal; **AUC**=Area under curve; **ROC**=Operating characteristic curve; **DPM**=Disease probability of risk map; **CNN**=Convolution neural network; **SS**=Sensitivity-specificity curve.

Footnotes

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/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disease with a hidden onset and unknown etiology [1]. Currently, there is an increasing number of AD patients because of the aging population, the younger trend of this disease, and other reasons, so it is urgent to find an efficient way to fight against AD [2,3].

In the current scheme, cerebrospinal fluid biomarkers, PET amyloid, and tau imaging in the detection of AD pathology are difficult to be widely used in actual clinical process [4-7]. In addition, in clinical applications, the imaging biomarkers revealed by MRI still lack specificity in the diagnosis of AD [8-11]. Experienced neuroscientists can only control the sensitivity range of 70.9%–87.3% and the specificity range of 44.3% - 70.8% according to the comprehensive information of patients, such as history, MRI, and bedside Mini Mental State Examination (MMSE) [12]. Therefore, if we can propose an AD auxiliary diagnosis method with the following three functions simultaneously, it will be of positive significance to improve the level of AD diagnosis and further reveal the pathogenesis of AD. (1) According to the comprehensive information of patients, the cognitive state of patients can be accurately and robustly judged. (2) The image biomarkers of AD were automatically analyzed according to a large number of clinical practices. (3) According to clinical practice data, the importance of different clinical information for the diagnosis of AD was automatically analyzed, such as MMSE, age, and gene information. At the same time, the model can also speed up the patient's medical treatment process and reduce the scan time to obtain high-quality MRI.

Artificial intelligence (AI) technology has made a great breakthrough in medical information analysis, but it still faces challenges when applied in clinical practice [13-18]. First, some AI models were trained and tested only in the same dataset. Owing to the lack of additional multi-center datasets, the real performance of the model is difficult to evaluate effectively. Second, the potential impact of the original data distribution on the diagnosis results is rarely considered, and the accuracy of the model prediction may be falsely high. Third, due to the uncertainty of the pathogenesis of AD, multi-modal information including individual differences still needs to be further analyzed. Some studies have focused on the single modal of AD diagnostic methods, such as MRI, PET, and gene information. The individual characteristics of patients are ignored when only the image information is used. The heterogeneity of brain symptoms was ignored when only the biological information was studied. Fourth, the AI model has black box characteristics, and its interpretability is still facing challenges [19, 20]. Fifth, there is a gap between clinical practice and an ideal research environment. The acquisition of MRI data is usually very time-consuming, and the patient needs to stay still during the scan. This not only reduces the patient's sense of experience, but also makes it difficult to ensure that any image obtained is of high quality. Some AI models use data extracted from the zero-filled k-space or original compressed k-space data to obtain reconstructed images. Although this method can improve the efficiency of MRI image acquisition and improve image quality, the process cannot be explained. This makes some pathological features may be subtle changes in the process of reconstruction, which makes the clinical data unreliable. Sixth, the interpretability of the multimodal model has not been solved. When mining medical image features, most models also work around the rule region of interest (ROI), which cannot accurately predict the disease risk of each point in the brain. The interpretability of clinical information is usually ignored, and the impact of each clinical information on the diagnosis results also needs to be studied. These problems prevent AI from being embedded in clinical practice.

To overcome these obstacles, this paper proposes a new AI framework that integrates reinforcement learning and deep learning, as shown in Fig 1. The proposed network has the following functions: (1) According to the multimodal comprehensive information to accurately judge the health condition of patients. (2) This network can explain the process of MRI reconstruction at the pixel level and has higher reliability than the black box model. (3) The disease probability of the risk map (DPM) of AD was generated based on brain MRI. Predicting the disease risk of each pixel in the MRI image, instead of using the ROI set in advance, increases the accuracy of pathological analysis of the model, and shows the evidence of the model judging the health status from the medical image. (4) The degree of attention of the model to different clinical data (such as age, MMSE, etc.), which shows the evidence of the model judging health status from the aspect of physiological parameters. Three datasets were used to train and evaluate the model to evaluate the performance of the model more objectively. The above datasets include the Alzheimer's Disease Neuroimaging Initiative dataset (ADNI) [21], Australian imaging biomarker and lifestyle flagship study of aging (AIBL) [22], and National Alzheimer's Coordinating Center (NACC) [23]. At the same time, t-SNE was used to draw the data distribution map to avoid the potential impact of different original data distributions on the diagnosis results [24, 25].

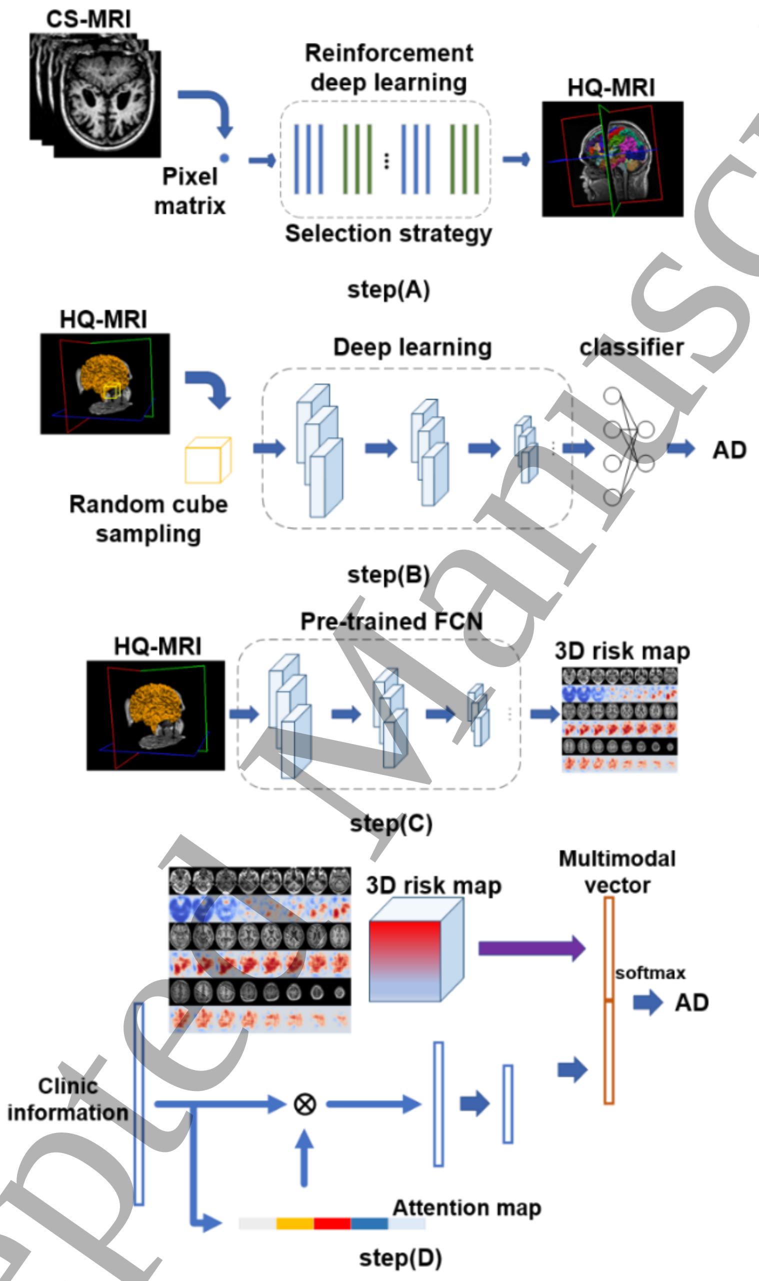

Fig. 1 The schematic diagram of deep learning framework (CS-MRI means compressed sensing MRI. HQ-MRI means high-quality image. AD means Alzheimer's disease. Step (A) designed in this paper is the process of compressed sensing MRI reconstruction. In the deep reinforcement learning model, each pixel of CS-MRI image is regarded as an agent, and the pixel matrix of the whole image is regarded as the model state. DRL model will select a filter for each pixel to improve its own pixel value according to the pathological characteristics, and then use the reward function to evaluate the improvement degree of the image. Finally, the high-quality image can be obtained. Step (B) is the

process of training convolutional classification network. We randomly sampled 47 * 47 * 47 cubes on HQ-MRI, and then these pixels were input into the traditional convolution classification model. Step (C) is the process of generating Alzheimer's disease risk map. The classifier of above convolution classification network is replaced by convolution kernel, which is transformed into FCN. The whole brain pixel level 3D AD risk map can be obtained by inputting HQ-MRI into FCN. Step (D) is a multimodal diagnosis model based on attention mechanism. The 3D risk map and clinical information are input into the model to obtain the diagnosis results. Attention mechanism can analyze biomarkers.)

2. Materials and Methods

2.1 Data Description

This was a retrospective study. The data used in the experiment included the patient's head MRI and clinical parameters (age, gender, MMSE, and ApoE4). All data can be found in ADNI, AIBL and NACC. Demographic information is shown in Table 1.

Table 1 Demographic Information (AD means Alzheimer's disease. NL means Normal.)

Dataset	ADNI			AIBL			NACC		
Item	NL (n=226)	AD (n=187)	p	NL (n=395)	AD (n=72)	p	NL (n=269)	AD (n=200)	p
Age(years) [range]	76.0 [60,90]	75.4 [55,91]	0.311	72.3 [60,92]	73.4 [55,93]	0.263	70.4 [55,94]	75.3 [55,95]	<0.0001
Gender (Male,n) [ratio]	117 [51.8%]	97 [51.9%]	0.984	169 [42.8%]	29 [40.3%]	0.693	91 [33.8%]	101 [50.5%]	<0.0001
MMSE(Ave) [range]	29.1 [25,30]	23.4 [18.28]	<0.0001	28.7 [25,30]	20.4 [6,28]	<0.0001	28.9 [20,30]	22.5 [4,30]	<0.0001
APoE4 (positive,n)	58 [25.7%]	119 [63.6%]	<0.0001	30 [7.6%]	19 [26.4%]	0.001	84 [31.2%]	173 [86.5%]	<0.0001

Data from 1,349 patients were used to construct and test the AI model. The inclusion criteria were as follows: age ⩾ 55 years, T1 weighted, 1.5T MRI. The acquisition time of MRI should be within 6 months of diagnosis. In our work, some outlier samples were excluded, including brain tumor, stroke, brain injury, severe depression, Parkinson's disease, epilepsy, Lewy body disease, non-Alzheimer degenerative dementia, mixed dementia, and severe systemic diseases. The exclusion criteria were similar to the ADNI criteria, and the same criteria were used to screen the AIBL and NACC datasets. The samples used in this study had MRI scans, age, sex, MMSE, APoE4, and samples without any information were excluded. Other parameters of MRI are not strictly limited. The serial number of cases used in this paper are shown in the supplementary material Table S1-S3. Through the file name, more accurate parameters of the corresponding MRI can be obtained from the relevant institutions (ADNI, AIBL, NACC). Such as pulse sequence, acquisition matrix, FOV, TR, TE, etc.

The ADNI dataset was divided into three parts: (1) Training sample: 60% of the ADNI data was used to update the global parameters. (2) Tuning samples: 20% of the samples were used to adjust the super parameters. (3) Test sample: The remaining 20% of the samples were used to test the performance of the model. To verify the robustness and generalization ability of the model, two additional datasets (AIBL and NACC) were also used to evaluate the final performance of the model.

2.2 MRI Image Pretreatment

In order to extract the pathological information contained in MRI images more effectively, all MRI images used in this study were preprocessed. The FLIRT toolkit was used to register MRI images. MNI152 was used as the registration template, and the slice thickness was 1 mm. All registered MRI images were manually screened again, and samples with poor registration effects were discarded. It should be noted that all discarded samples were not included in the above demographic statistics. The above statistics only include all samples that meet the requirements.

After obtaining all registered MRI images, we standardized their pixel values. After standardization, the original data are converted into dimensionless data, eliminating the impact of different pixel ranges caused by different measurement conditions on the performance of the model. In other words, the effects of different MRI parameters on model performance are eliminated. The assignment method was used to eliminate singularity. All pixels less than - 1 are assigned - 1, and all pixels greater than 2.5, are assigned 2.5. After the singularity of the MRI image was eliminated, we further eliminated the background of the MRI image, and all the pixels outside the skull were set to − 1.

2.3 Deep Reinforcement Learning (DRL)

Compressed sensing technology can speed up the efficiency of MRI image acquisition and reduce the scanning time for patients. Due to frequency domain down-sampling, compressed sensing MRI (CS-MRI) is usually blurry. Therefore, CS-MRI needs to be reconstructed to obtain high-quality MRI images (HQ-MRI) for diagnosis. HQ-MRI is used to generate DPM. Finally, DPM and patient clinical and genetic information are used to diagnose AD. Simulate the acquisition of CS-MRI by down-sampling the raw T1 image. Here, Cartesian mask with 50% sampling ratio is used to simulate down-sampling columns in k-space. In order to reconstruct MRI, classical methods usually use sparsity constraints, discrete wavelet transform, dictionary learning, etc. Although these classic algorithms have achieved good performance, they still face challenges. One of them is the personalized processing of image different features. It is difficult to design different processing methods for many different image features at the same time. For example, the method of designing filters is usually difficult to automatically use different filters for different pixels simultaneously. Deep learning methods show the potential to solve the above problems. Existing deep learning methods usually input low quality data into the model and obtain high-quality reconstructed images. Although the reconstruction algorithm based on deep learning can achieve better performance, its interpretability still faces challenges. This black box feature may cause potential risks. For example, some small anatomy has changed and it is difficult to be found. And the process of AI model improving image quality is also unknown. In order to solve the above problems, a model based on a reinforced deep learning algorithm is designed. In this study, a deep reinforcement learning framework was designed to reconstruct CS-MRI to make the reconstruction process interpretable, speed up the acquisition of MRI in clinical practice, and improve the quality of MRI images. This principle is shown in Fig 1 (A). The image reconstruction process can be regarded as a Markov decision process (MDP). In this MDP model, each pixel corresponds to an agent that is used to model multi-agent problems. The input image of each time step is regarded as the state of the DRL model. After acquiring state s^t at time step *t*, each agent selects an action from the action space to change its own pixel value, and receives a reward value r^t, which is used to evaluate the degree of improvement of the image. Compared with the traditional methods, this study designs a reward function that considers both the MRI content and the improvement of MRI features, as shown in the Equation 1.

$$\begin{aligned} R = 0.1\left(\left|MRI^{(0)} - MRI_{(\text{targert})}\right| - \left|MRI^{(T)} - MRI_{(\text{targert})}\right|\right) + \left|MRIF^{(0)} - MRIF_{(\text{targert})}\right| \\ - \left|MRIF^{(T)} - MRIF_{(\text{targert})}\right| \end{aligned} \tag{1}$$

Where, $MRI^{(0)}$ means the MRI image should be reconstructed at the initial time step. $MRI_{(target)}$ means target MRI image. $MRI^{(\mathrm{T})}$ represents the reconstructed MRI image at time step T. $MRIF^{(0)}$ means the features of MRI image should be reconstructed at the initial time step. $MRIF_{(target)}$ means the features of target MRI image. $MRIF^{(\mathrm{T})}$ represents the features of the reconstructed MRI image at time step T. Here, the VGG16 model pre-trained by Google is used to extract MRI features. Remove the classification layer of the VGG16 model and input the MRI image to obtain image features.

In this study, CS-MRI is generated using Cartesian sampling in k-space. Cartesian sampling is a Gaussian distribution with a sampling rate of 50%. Cartesian mask with 50% sampling ratio is used to simulate down-sampling columns in k-space. The DRL model is divided into three parts: the feature extraction module, policy selection module, and parameter optimization module. The overall structure of the model is shown in the Supplementary Material Figure S1. The feature extraction module is realized by the convolution network, which is used to extract the potential features of the image. In the above process, the spatial resolution of the image remained unchanged. The core of the policy selection module is the advantage actor critical algorithm (A2C), which aims to generate state-to-action mapping [26]. The output of the feature extraction module is sent to the strategy selection module, and the strategy selection module outputs the probability distribution of actions. The probability distribution generated was evaluated using the value calculation part. The core of the parameter optimization module is the deep deterministic policy gradient (DDPG) algorithm [27]. After the policy selection module selects the action for each agent, a parameter optimization module is used to optimize the parameters of the filter. Finally, the value calculation part reevaluates the parameter optimization module. The loss function of each module is shown in Equation 2-6.

$$L_{SS} = L_\pi + 0.25L_{VC} + 0.1L_E \tag{2}$$

The loss function of the DRL model is L_{SS}. L_π is shown by Equation 3. L_{VC} is shown by Equation 4. L_E is shown by Equation 5.

$$L_\pi = -\log \pi\left(a|s;\theta_s,\theta_f\right)\left(R - V\left(s;\theta_v;\theta_f\right)\right) \tag{3}$$

$L_\pi$represents the loss of the strategy selection module part. a represents the selected action (filter). s represents the model state. θ_s represents the parameters of strategy selection module. θ_f represents the parameters of feature extraction module. *R* is reward. *V* is value which is the output of the value calculation section. θ_v represents the parameters of value calculation section.

$$L_{VC} = \left\|R - V(s;\theta_v;\theta_f)\right\|^2 \tag{4}$$

L_{VC}represents the loss of the value calculation section. R is reward. *V* is value which is the output of the value calculation section. θ_v represents the parameters of value calculation section. θ_f represents the parameters of feature extraction module.

$$L_E = \sum_a \pi(a|s;\theta_s,\theta_f)\, log\, \pi\,(a|s;\theta_s,\theta_f) \tag{5}$$

$L_E$represents the negative entropy loss. It is used to encourage action exploration.

$$L_{PO} = -0.5V(s,\theta_p) \tag{6}$$

The action here refers to the filter, and the action space used in this study is shown in Table 2. At time *t*, an image is input as state s^t, and the output image is s^{t+1}. The DRL model used in this study included three-time steps for each episode.

Table 2 Strengthen the Learning Action Space in Deep Reinforcement Learning

Action	size	Action	size
nothing	-	Sobel filter(down)	3*3
Laplace filter	3*3	Box filter	5*5
Unsharp masking	5*5	Bilateral filter	5*5
Sobel filter(left)	3*3	Median filter	5*5
Sobel filter(right)	3*3	Gaussian filter	5*5
Sobel filter(upper)	3*3	Gaussian_s filter	3*3

The number of training iterations was 30,000. When the strategy selection module is trained, the parameter optimization module is frozen, otherwise, strategy selection module is frozen. The model is iterated twice, and the two modules above are alternately. The model optimizer was Adam. The learning rate attenuation method was used in the training process. ReLU was used as an activation function for the entire network. The experimental platform was a workstation with NVIDIA GTX 1080Ti GPU.

2.4 Full Convolution Neural Network (FCN)

This part is used to generate the whole-brain AD DPM. Displaying diagnostic evidence can improve the interpretability of the AI model [28, 29]. The model construction includes two steps: building a convolution classification model and a 3D DPM generation model, as shown in Fig. 1 (B) and (C). The working principle of this part is shown in Fig 2.

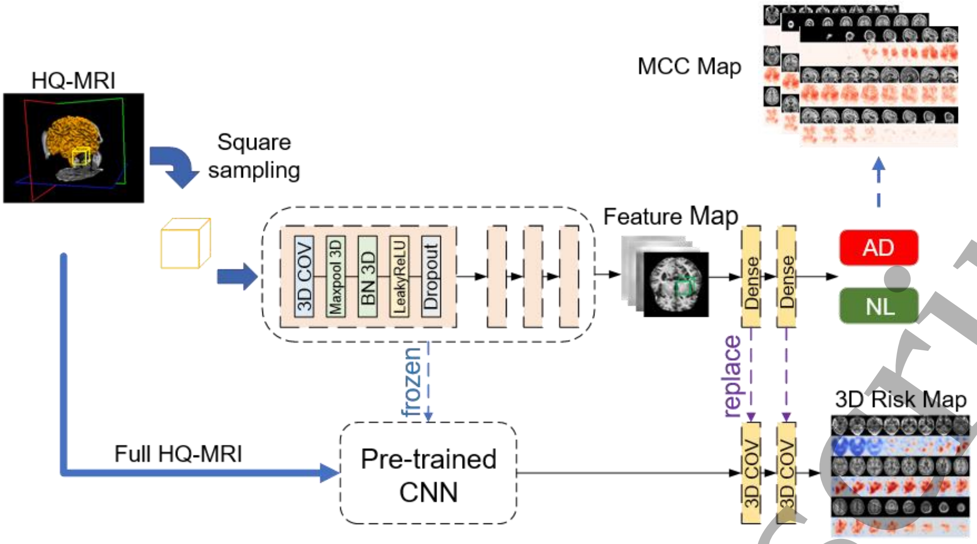

Fig. 2 Schematic diagram of building a fully convolutional neural network to obtain DPM and matthews correlation coefficient distribution maps (The size of each HQ-MRI is 181*217*181. HQ-MRI is randomly sampled by cubes, and each HQ-MRI is randomly sampled 5000 times. The cube pixel block is input into the convolutional neural network for binary classification. According to the classification results, we can calculate the matthews correlation coefficient (MCC) score of each part of the brain. The MCC score reflects the model's attention in different brain regions when diagnosing AD. After obtaining a satisfactory classification model, the convolutional block is frozen, and the dense layers are replaced with convolutional layers. The parameters of the original fully connected layer are filled into the new 3D convolutional layer. The parameter of dropout is 0.1. Here, the FCN construction is completed. The risk map of the whole brain can be obtained)

The convolution classification model needs first to sample 5000 voxel blocks with a size of 47 × 47 × 47 from each original image with a size of 181 * 217 * 181 and then to be trained by the above voxel blocks with the classification function softmax. The output of the classification model is health status. The MCC was used to weigh the degree of model attention on different voxel blocks reflecting the different brain regions while diagnosing AD. Through cube sampling and statistics of the corresponding MCC, the importance of any part of the brain to the diagnosis of AD can be determined. The higher the MCC score of the brain domain, the more important it is to diagnose AD. The principle of MCC is shown in Equation 7.

$$\mathrm{MCC} = \frac{[(\mathrm{TP}\times\mathrm{TN})-(\mathrm{FP}\times\mathrm{FN})]}{\sqrt{(\mathrm{TP}+\mathrm{FP})\times(\mathrm{TP}+\mathrm{FN})\times(\mathrm{TN}+\mathrm{FP})\times(\mathrm{TN}+\mathrm{FN})}} \tag{7}$$

Where, TN means true negative, TP means true positive, FN means false negative and FP means false positive.

The 3D DPM generation model is based on the convolution classification model described above. The output of this model is a convolution block with dimensions of 2 × 1 × 1 × 1. The 1 × 1 × 1 convolution block represents a 3D voxel. The number 2 of the block refers to the two channels of AD and normal (NL). Therefore, the AD DPM of the whole brain region can be obtained by inputting a 3D MRI image. The optimizer of this model is Adam, and the training technique of this model is the learning rate decay method. LeakyReLU is the activation function of the entire network, and dropout technology is used to increase the generalization ability of the model.

2.5 Multi-modal Diagnosis Model

In this part, the DPM of high MCC score brain regions and individual clinical information are used to comprehensively judge the health status of patients using a multilayer perceptron network and explain the diagnosis basis of the model. As shown in Fig 1D, we embed the attention mechanism in the multilayer neural network. Through the attention map, we can obtain the attention degree of the model to different clinical information explaining the parameter weights for AD diagnosis. The individual clinical information used in this study included age, gender, MMSE score, and APoE4. A schematic diagram of the model is shown in Supplementary Material Figure S2.

3 Results

This section describes the performance of the model in diagnosing AD. In addition, the interpretability of each part is presented. This includes an interpretable image reconstruction process, interpretable image diagnosis process, and interpretable biological information diagnosis process.

3.1 Interpretable MRI Reconstruction Process

The reinforcement deep learning model designed in this study can explain the reconstruction process of CS-MRI. Here, we use different colors to represent different filters. If the relevant agent is processed by a certain action, the pixel represented by the above agent is overlaid by the color corresponding to the filter. Traditional AI methods usually input low-quality images into the model, and then output high-quality images. Although these methods can achieve good performance, researchers are not sure how AI reconstructs images. At the same time, the above situation also has hidden dangers, that is, in the process of reconstructing the image, some subtle pathological features may have been changed. The model in this paper can design a reasonable filter for each pixel according to the image characteristics, and visualize the distribution of different filters. Different colors were used to represent different certain actions, so that relevant researchers can intuitively observe how the model reconstructs the MRI. At the same time, it is also convenient to judge whether the above reconstruction process is reasonable and whether the important pathological features have undergone substantial changes, which increases the interpretability and reliability of the model to a certain extent. Compared with the traditional deep learning model with black box characteristics, this model has higher interpretability and reliability, as shown in Fig 3.

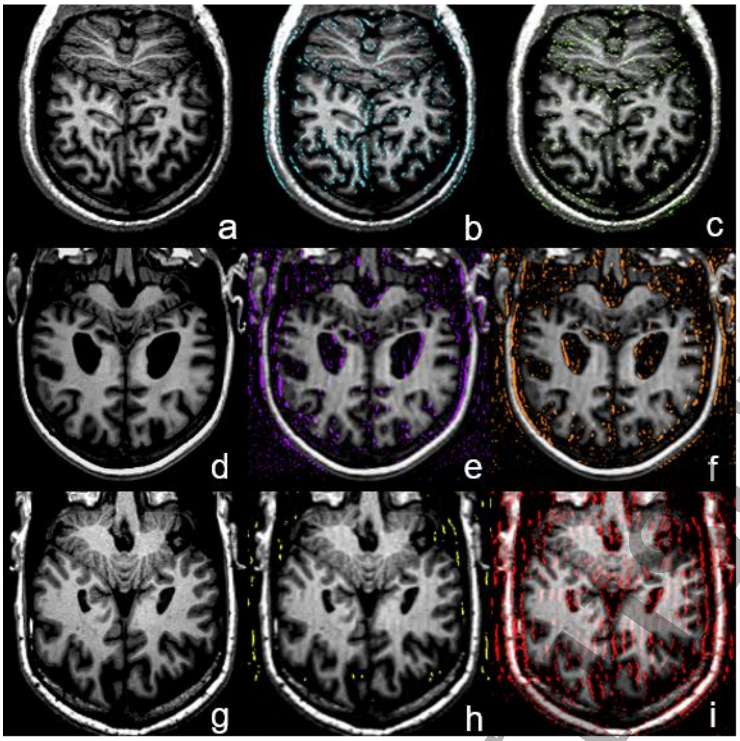

Fig. 3 Example of filter distribution on pixels. (Different colors represent different types of filters. where a, d and g represent target images. b, c, e, f, h and i show Laplace filter, unsharp filter, sobel_left filter, sobel_right filter, box filter and subtraction filter, respectively.)

In Fig 3, the DRL model can explain the individual filter for each pixel of the CS-MRI. Different filters are marked with different colors. In other words, the DRL model can show which processing method is selected for each pixel values. For example, the blue pixels are shown in Fig 3b represents the pixels processed using the Laplace filter. The well-trained DRL model can select the filter according to the anatomical structure of the brain. The filter with contrast enhancement function is mainly used to optimize regions such as the skull and cerebrospinal fluid as shown in Fig. 3b and Fig. 3c. Sobel_ Left and sobel_right filters strengthen the left and right sides of the skull, as shown in Fig. 3e and Fig 3f. In this way, researchers in related fields can intuitively judge whether the model reconstruction strategy is reasonable. Fig 4 shows the distribution of Gaussian filters with different time steps. It can be found that as the quality of the picture improves, the frequency of use of the Gaussian filter is also decreasing. Through reinforcement learning to visualize the optimization strategy of each step of the model, the interpretability of the model is improved to a certain extent.

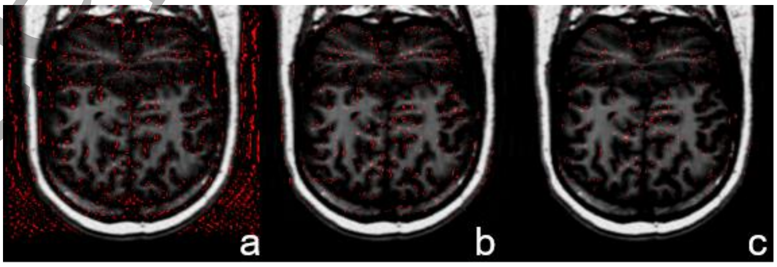

Fig. 4 Distribution of Gaussian filters at different time steps

We also compared the performance of our model with other advanced models on the same dataset, such as DAGAN [30] and Unet [31]. Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are used to evaluate the model. The principles are shown in Equation 8-9.

$$SSIM = \frac{\left(2\mu_x\mu_y + c_1\right)\left(\sigma_{xy} + c_2\right)}{\left(\mu_x{}^2 + \mu_y{}^2 + c_1\right)\left(\sigma_x{}^2+\sigma_y{}^2 + c_2\right)} \tag{8}$$

Here, μ_x and μ_y represent the average of the two images respectively. σ_x and σ_y represent the standard deviation of the two images respectively. σ_{xy} represents the covariance between two images. c_1 and c_2 are constants.

$$PSNR = 20\log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right) \tag{9}$$

MAX means maximum value of pixels. MSE means mean square error. The performances of the models in the three test sets are presented in Table 3.

Table 3 Performance of Different Algorithms in Different Test Sets

	PSNR	SSIM		PSNR	SSIM		PSNR	SSIM
DAGAN			Unet			DRL(our)		
ADNI	36.01	0.98		36.77	0.99		36.78	0.98
AIBL	35.27	0.97		35.98	0.97		36.03	0.98
NACC	35.96	0.97		36.69	0.98		36.76	0.98

We randomly selected samples from three test sets to show the effect of the DRL model on MRI image reconstruction, as shown in Fig 5.

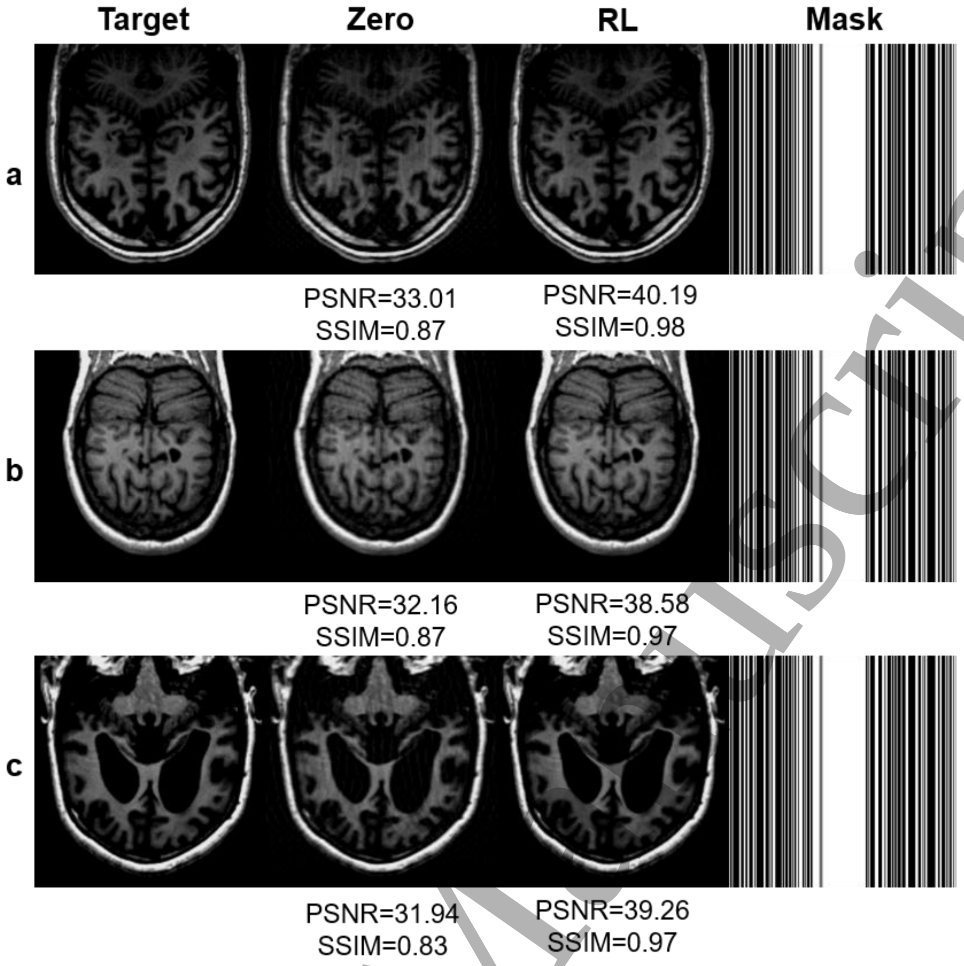

Fig. 5 Results of CS-MRI reconstruction by deep reinforcement learning (Figure a, b and c show the reconstruction of three random CS-MRI samples randomly obtained from ADNI, AIBL and NACC. Zero represents the experimental results of the traditional zero filling method, and DRL represents the experimental results of the reinforcement deep learning framework in this paper. Mask shows the image information of Cartesian mask. Target represents the raw data. It also is ground truth)

According to the experimental results, the quality of HQ-MRI is higher than CS-MRI, so HQ-MRI is more suitable for the diagnosis of AD. The performance of using CS-MRI and HQ-MRI to diagnose AD is shown in Supplemental Material Figure S3 and S4.

3.2 The Generation of 3D DPM of Brain

The AD DPM of the whole brain region can be generated quickly. AD DPM at the pixel level can help neurologists find early pathological features in the patient's brain and provide evidence of early AD diagnosis. CS-MRI was reconstructed to generate HQ-MRI, and HQ-MRI was used to generate 3D DPM, as shown in Fig 6. Fig 6a shows the DPM of AD and NL in each test set, with red for high-risk areas and blue for low-risk areas. In the three test datasets, it is obvious that the high-risk area (red) of the AD sample is very obvious, while the NL sample is dominated by the low-risk area (blue). High-risk area refers to the area where the inferred probability of Alzheimer's disease is greater than 0.5, and low-risk area refers to the area where the inferred probability is less than 0.5. The color depth has a linear relationship with the probability of brain diseases. Fig 6. b-d shows the DPM of different MRI slices of the same AD sample from axial, coronal, and sagittal images. Fig 6. e-g shows the DPM of different slices of the same NL sample. The experimental results show that the model can effectively predict the AD risk at any position of the brain according to MRI images, and the model can be used for pathological interpretation. To a certain extent, the model in this paper can assist neurologists to discover more pathological features of AD. At the same time, the model can provide the basis for the diagnosis of AD, and it can also allow relevant clinicians to judge whether the AI diagnosis result is reasonable.

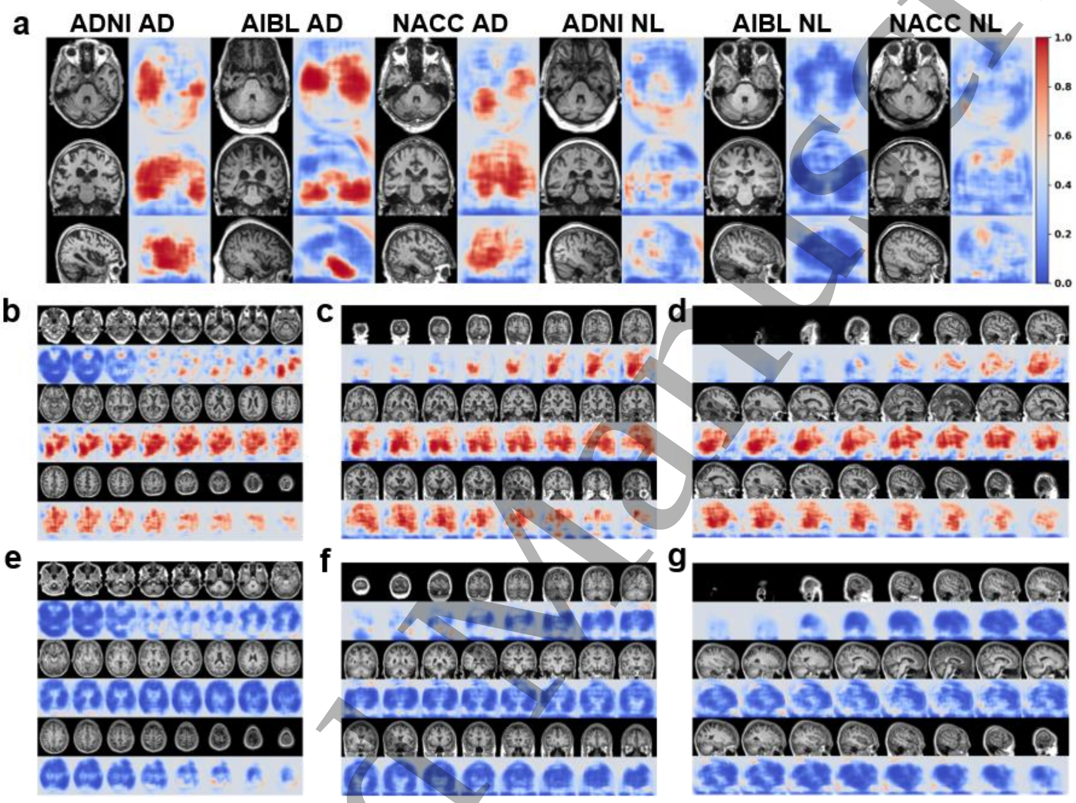

Fig. 6 Disease probability of risk map of different subject. (The DPM of different health samples from different datasets is shown in the chart a of DPM of different health conditions from axial, coral and digital. Blue represents low risk, red for high-risk. Red refers to the area where the inferred probability of Alzheimer's disease is greater than 0.5, and blue refers to the area where the inferred probability is less than 0.5. The DPM highlights the anatomical areas of the brain associated with AD pathology. Figure 5. b-d shows DPM of different MRI sections of the same ad sample from axial, coral and digital. Figure 3. e-g shows the DPM of different MRI sections in NL samples.)

To evaluate the anatomic consistency of the AD regions concerned by the model, the MCC heat map was established. Fig 7a shows the distribution of the MCC scores in different datasets. The MCC thermogram can show which pathological changes play an important role in the diagnosis of AD, indicating that the model has the ability to analyze the disease from the perspective of anatomy and pathology. Fig 7b shows that the model gives different attention to different morphological positions. The experimental results show that the model has focused attention on the temporal lobe, hippocampus, cingulate cortex, corpus callosum, parietal lobe, and frontal lobe. The region of interest of the model is similar to the diagnosis basis of the clinician. This proves the interpretability of the method proposed in this paper to a certain extent. For morphological information that is not helpful in the diagnosis of AD, the model does not show great interest, such as the skull and cervical spine. This helps neurologists to observe neuropathological changes in AD patients.

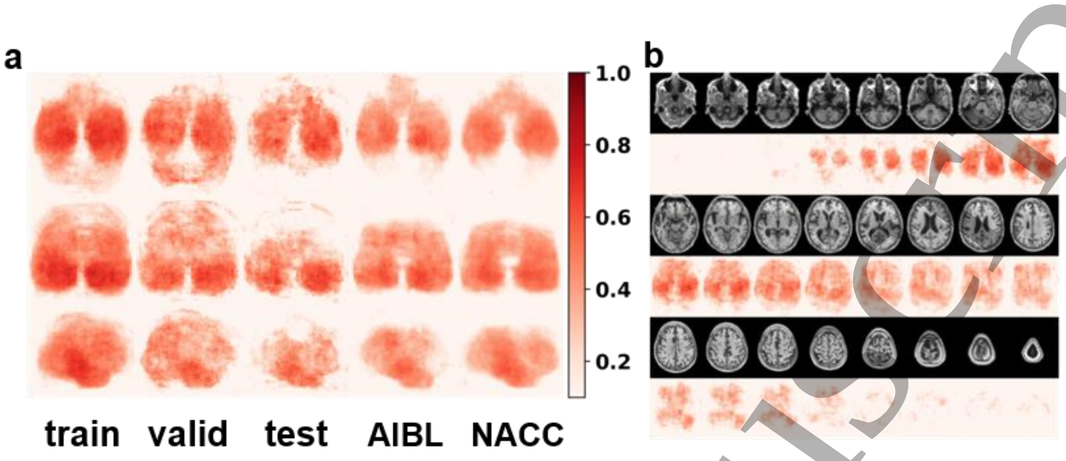

Fig. 7 Distribution of MCC scores of brains in different data sets (Figure 6a shows the Matthew's correlation coefficient (MCC) of all regions of the brain in different dataset samples, which shows the brain area that the AI model focuses on. Figure 7a shows three lines of images from top to bottom, corresponding to axial, national and digital stacks. Figure 7b shows the distribution of MCC scores for different MRI sections in the same sample.)

In order to further evaluate the rationality of the model, we also compared the DPM generated by different methods. The attention mechanism is proven to be an effective method for processing medical images [32-34]. Therefore, in addition to the cube sampling method proposed in this paper, this paper also constructs an attention-driven model to generate DPM. More theoretical analysis can be found in [32-34]. The results as shown in Fig 8.

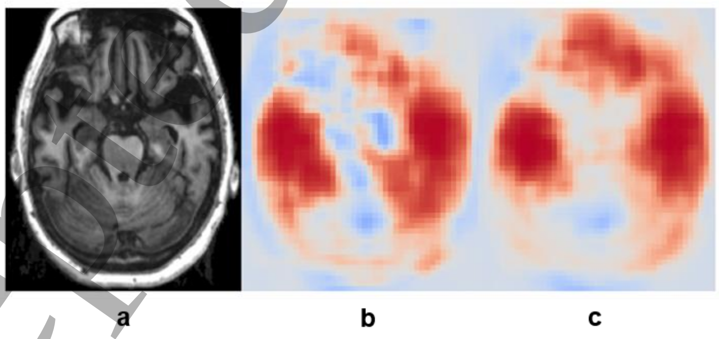

Fig.8 DPM generated by different methods. (a is MRI of random case. b is the DPM generated by model designed in this paper. c is the DPM generated by attention model. The attention mechanism model refers to a three-dimensional image attention mechanism model. The model takes the entire 3D MRI as input and then generates DPM through a 3D attention full convolutional network.)

As shown in Fig. 8, both models believe that the frontal lobes and temporal lobes have a higher risk of AD. This result is consistent with the current clinical diagnostic criteria. This phenomenon shows that the AD analysis method based on cube sampling designed in this paper is reasonable to a certain extent. In addition, we found that compared with the 3D attention method, the method in this paper may be more granular in predicting the risk of diseases in different regions of the brain. It depicts the edges of high-risk areas more clearly. We speculate that this is because the small cube sampling strategy can help the model better learn the disease risk of each pixel. The attention mechanism pays more attention not to pixels, but to areas that are more important to the diagnosis. Although the model in this paper shows potential, the above viewpoints still need to be further verified. Since the etiology of AD is not yet clear, it is difficult for even human experts to accurately draw all the lesion areas in MRI. Autopsy is an effective method to verify neuropathology. Therefore, in future work, we will further collect cases including autopsy reports to further study the reliability of the model.

3.3 Multi-modal AD Diagnosis

In this study, we used both morphological and clinical information to diagnose AD. To eliminate the influence of the initial data distribution on the model performance, this study uses the t-SNE method to study the original data distribution, as shown in Fig 9A and 9B. The pixel matrix of the registered image was employed for the t-SNE analysis. In Fig 9A, the original data distribution of the three datasets is consistent. The original data of all datasets did not show obvious polarization. Fig 9B shows the t-SNE analysis of the DPM for each dataset. In Fig 9B, an obvious polarization between AD samples and NL samples is illustrated, indicating the high efficiency of learning the AD pathological characteristics and the high accuracy of the generated brain AD DPM.

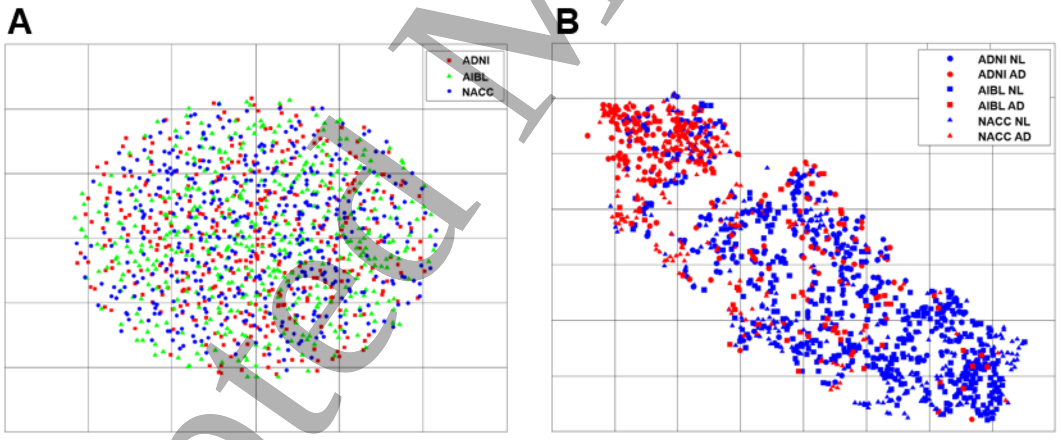

Fig. 9 The t-SNE analysis of image data and ROC of model (Figure 9A shows t-SNE analysis of MRI images of three datasets (ADNI, AIBL, NACC). The pixel matrix of MRI image is used as input, and t-SNE output two-dimensional data distribution map. Figure 9B shows the t-SNE analysis of the disease risk graph for three datasets. The red data represents the AD sample, and the blue data represents the NL sample.)

After analyzing the raw data using t-SNE, the performance of various models is evaluated. The sensitivity-specificity curve (SSC) of our model and other ablation experiments are shown in Fig 10. The area under the curve (AUC) was used to evaluate the performance of the framework and related ablation experiments. To explore the robustness and universality of the model, two additional datasets (AIBL and NACC) were also used to evaluate the performance of the model.

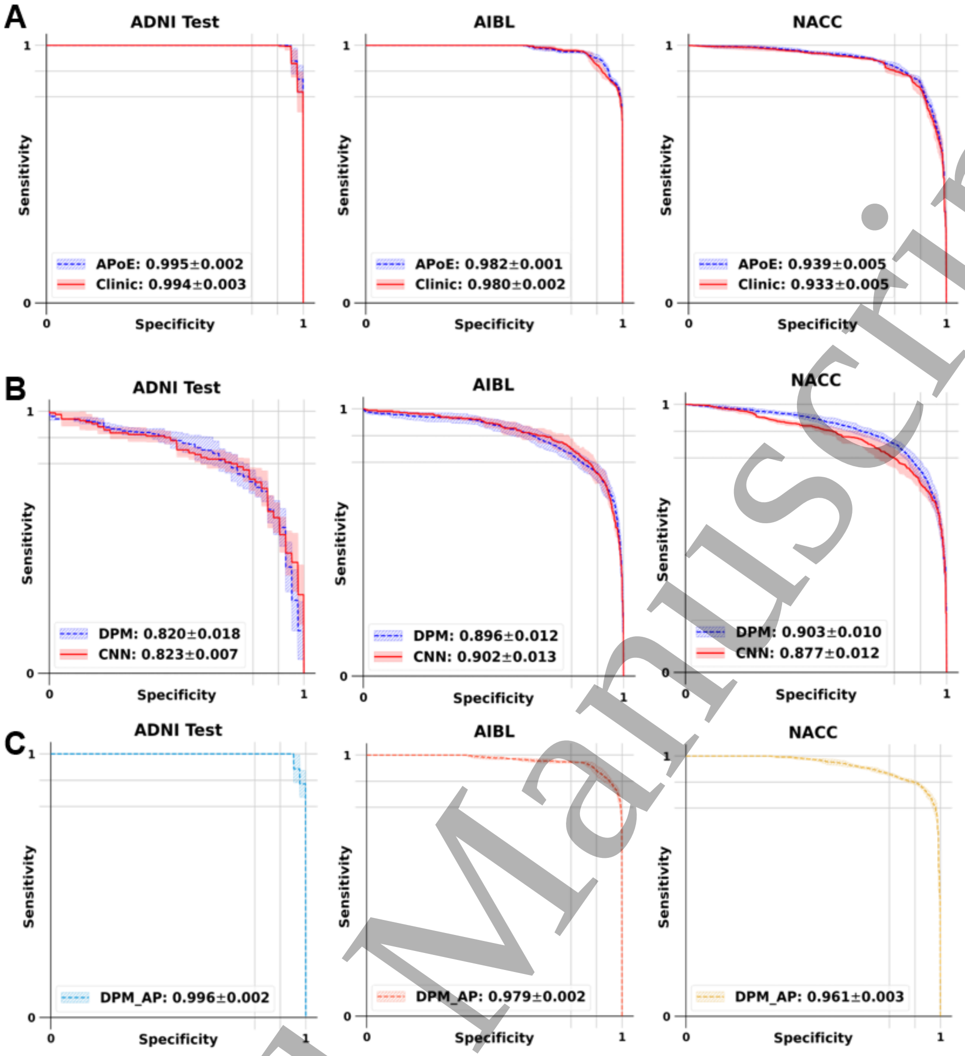

Fig. 10 The sensitivity-specificity curve of our AI model and other models (Figure A shows the performance of the model considering only individual clinical information, APoE represents the performance of the model considering APoE status, and clinic represents the performance of the model not considering APoE status. Figure B shows the performance of the model considering only image information, DPM represents the performance of the model using only disease risk map, and CNN represents the performance of the convolutional neural network using only MRI. Figure C is SS of our model)

The receiver operating characteristic curve (ROC) of each model on ADNI is shown in Figure S3. The precision-recall curve of each model is shown in Fig 11.

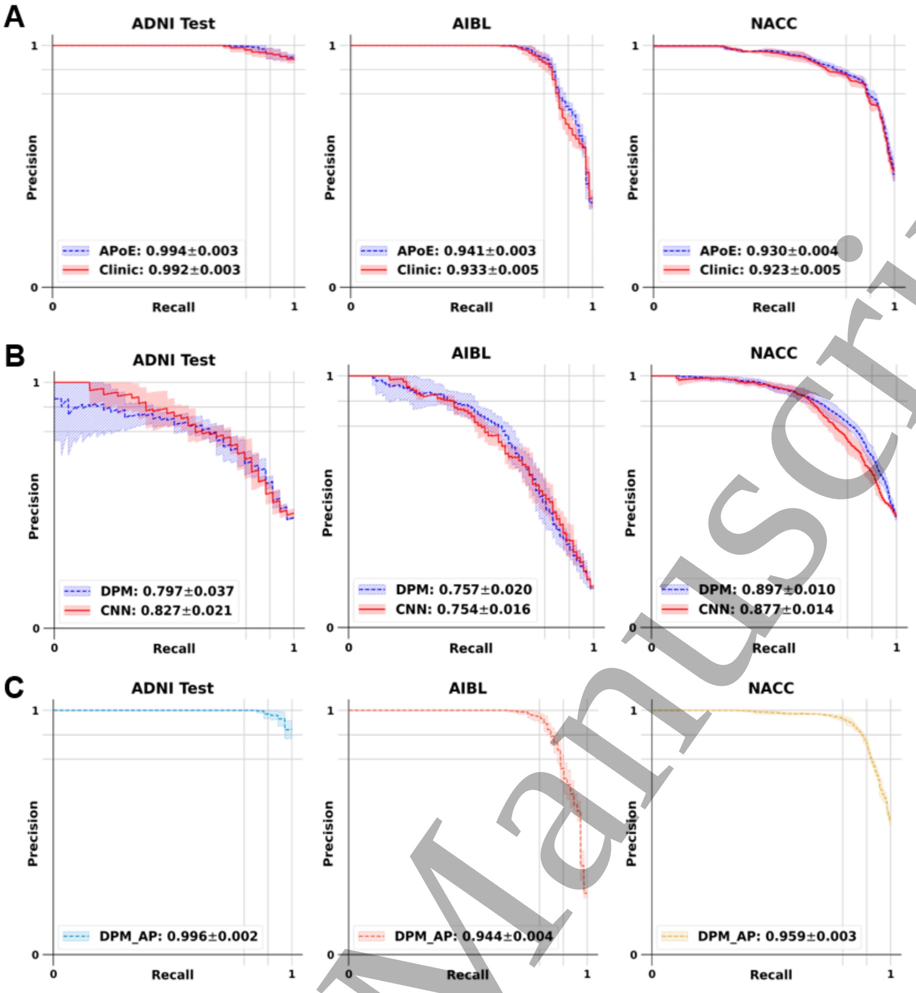

Fig. 11 Precision-Recall curve of each model (Figure A shows the performance of a model that only considers individual clinical information. Here, APoE represents the performance of a model that considers the APoE state, and Clinic represents the performance of a model that does not consider the APoE state. Figure B shows the performance of the model considering only image information. Here, DPM represents the performance of the model using only the disease risk map, and CNN represents the performance of the convolutional neural network using only the MRI image. Figure C shows the performance of the multi-modal diagnostic model designed in this paper.)

Finally, the multi-modal model designed in this study achieved an AUC of 99.6% in the ADNI test set, 97.9% AUC for AIBL, and 96.1% AUC for NACC. The comprehensive performance of the model was superior to that of other comparative experiments. To evaluate the performance of the models more comprehensively, in addition to AUC, a variety of indicators are used to evaluate the performance of each model, as shown in Table 4. The calculation principles of sensitivity, specificity, and F1 score are shown in Equation 10-11.

$$\mathrm{SEN}=\frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad \mathrm{SPE}=\frac{\mathrm{TN}}{\mathrm{TN+FP}} \tag{10}$$

Where, SEN means sensitivity, SPE means specificity. TN means true negative, TP means true positive, FN means false negative and FP means false positive..

$$F1 = \frac{2\times \mathrm{TP}}{(2\times \mathrm{TP}+\mathrm{FP}+\mathrm{FN})} \tag{11}$$

In Equation 11, F1 means F1 score. The meaning of other parameters is the same as Equation 10.

Table 4 Performance Statistics of Different Models (Here, Clinic represents the performance of a model that does not consider the APoE state, and APoE represents the performance of a model that considers the APoE state. CNN represents the performance of the convolutional neural network using only the MRI image, and DPM represents the performance of the model using only the disease risk map. DP+AP shows the performance of the multi-modal diagnostic model designed in this paper.)

	ACC	*AUC*	*Sen*	*Spe*	*F1*	*MCC*
clinic						
Test	0.947±0.020	0.994±0.003	0.969±0.054	0.930±0.036	0.942±0.023	0.898±0.039
AIBL	0.916±0.029	0.980±0.002	0.893±0.050	0.920±0.045	0.772±0.052	0.736±0.052
NACC	0.863±0.011	0.933±0.005	0.814±0.046	0.899±0.028	0.840±0.017	0.721±0.023
APoE4						
Test	0.949±0.015	0.995±0.002	0.986±0.024	0.938±0.032	0.957±0.015	0.922±0.027
AIBL	0.919±0.029	0.981±0.002	0.914±0.045	0.920±0.041	0.782±0.051	0.750±0.051
NACC	0.872±0.007	0.940±0.003	0.841±0.030	0.895±0.022	0.848±0.010	0.738±0.014
CNN						
Test	0.766±0.019	0.823±0.007	0.724±0.076	0.800±0.029	0.732±0.035	0.527±0.044
AIBL	0.843±0.055	0.902±0.013	0.783±0.097	0.854±0.079	0.617±0.060	0.556±0.067
NACC	0.796±0.044	0.877±0.012	0.738±0.090	0.839±0.131	0.757±0.027	0.597±0.066
DPM						
Test	0.772±0.028	0.820±0.018	0.690±0.034	0.838±0.041	0.730±0.030	0.538±0.057
AIBL	0.913±0.009	0.896±0.012	0.581±0.065	0.973±0.018	0.671±0.029	0.639±0.030
NACC	0.830±0.021	0.903±0.012	0.689±0.060	0.934±0.021	0.774±0.037	0.656±0.039
DP+AP						
Test	0.956±0.025	0.996±0.002	0.961±0.026	0.968±0.031	0.961±0.018	0.930±0.031
AIBL	0.958±0.007	0.979±0.002	0.864±0.022	0.975±0.011	0.863±0.018	0.839±0.022
NACC	0.907±0.010	0.961±0.003	0.873±0.024	0.932±0.027	0.889±0.001	0.810±0.020

When only using clinical information data without APoE4, the model reached 94.7%±0.2% accuracy (ACC) and 99.4%±0.3%AUC in the ADNI test set. When the APoE4 state is considered by the model, the model achieves 94.9%±1.5%ACC and 99.5%±0.2%AUC in the same dataset. When the APoE4 state is added, the performance of the model was improved. When only using DPM to diagnose AD, the diagnosis level of the model is similar to that of a traditional convolutional neural network (CNN). The high-risk areas of AD predicted by our model overlapped with the MRI images, as shown in Fig 12A. According to the experimental results, the model's main focus area is the hippocampus and the medial temporal lobe. Cortical atrophy also occurred in the high-risk area given by the model. This is very consistent with the existing clinical diagnosis basis. This phenomenon proves the rationality of the method in this paper to a certain extent.

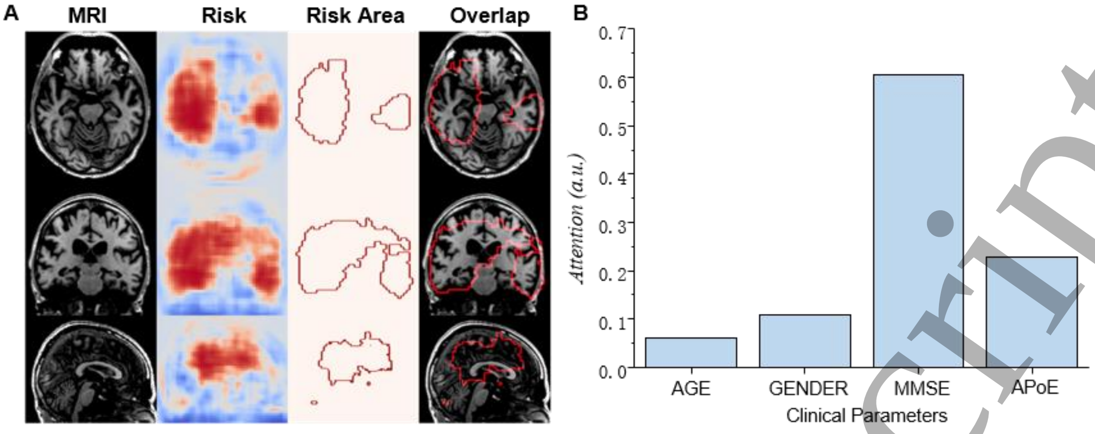

Fig. 12 The pathological analysis of the model (Figure A shows the overlap between the high-risk areas predicted by the model and the MRI images. The risk area in Figure A refers to the area where the risk probability is greater than 0.8. Figure B shows the model's attention to different individual physiological parameters, which shows whether the model speculates that a certain physiological parameter will affect the formation and development of AD)

When using multimodal information (DMP and clinical information containing APoE4) at the same time, the model reached 95.6%±2.5%ACC and 99.6%±0.2%AUC in the ADNI test set. To make the model multimodal and interpretable, the attention mechanism is integrated into the analysis of personal clinical information. The attention map expresses the degree of attention of the model to different physiological parameters, that is, the basis of the diagnosis of AD. The average attention of the model to different physiological parameters on the ADNI is shown in Fig 12B.

MMSE and APoE4 are the two physiological parameters with the highest degree of concern, which are consistent with the clinical diagnostic criteria [35-38]. In addition, the model also shows a certain degree of attention to age and gender. The model suggests that age and gender are also related to AD, and gender is more important than age in the diagnosis of AD. In recent studies, age and sex have also been found to be related to the formation of AD [39-42]. This phenomenon shows that the AI model designed in this study can not only analyze the biomarkers directly related to diseases, but also have the ability to discover potential biomarkers of diseases.

4. Discussion

This paper presents a multimodal AI framework with the entire process interpretability that can diagnose AD accurately. CS-MRI is first used to reduce the scanning time and improve the clinical experience of patients. Second, CS-MRI is converted to HQ-MRI using an MRI reconstruction module based on deep reinforcement learning. Third, HQ-MRI is employed by the DPM generation module to generate AD DPM in the whole brain region. Finally, the DPM of important areas of the brain and individual clinical information are input into the multimodal diagnosis module based on the attention mechanism to obtain the diagnosis results.

In this study, we designed a deep reinforcement learning framework for converting CS-MRI to HQ-

MRI. Compressed sensing technology is applied to acquire CS-MRI to solve the time-consuming problem of continuous sampling in k-space. However, compressed sensing technology will reduce the image quality. Although the end-to-end black box model based on traditional AI can reconstruct CS-MRI, the reconstruction process cannot be explained. The deep reinforcement learning framework designed in this study regards each pixel as an agent, and then each agent will choose different individual filters according to pathological features to change its own pixel value, so as to achieve the purpose of optimizing image quality. This method not only enhances the interpretability of the model but also enhances its reliability. The experimental results show that the selection of the filter by the model is generally consistent with the selection of human experts. An unsharp filter and a Laplace filter will be selected in the skull and cerebrospinal fluid, which need to enhance contrast. For the gap between the skull and brain tissue, the model infers that a box filter is used to smooth the pixels. For aliasing in the image, the model uses a subtraction filter. The final model achieved acceptable results in the ADNI test set and two external test sets, as shown in Fig 5. At the same time, we also compare the performance of our model with other advanced models on the same dataset, as shown in Table 3.

HQ-MRI was used to generate the AD DPM. The core of the DPM generation module is the FCN obtained by reforming the trained classification CNN network. The module can map complex anatomical information into a simple and intuitive DPM. Disease risk in any position in the brain can be accurately predicted, and the prediction granularity reaches the pixel level, as shown in Fig 6. Compared with the traditional method of pathological research using regular-shaped ROIs, this method can determine the high-risk area of AD more accurately. Thus far, the etiology of AD is still unclear, and the above strategies can lay a foundation for further pathological research on AD.

Before training the diagnosis model, we used t-SNE to study the distribution of the original data to avoid the huge difference in the distribution of the original data. T-SNE can project high-dimensional data into two-dimensional space to visualize the data distribution. We used t-SNE to analyze the pixel matrix of the registered MRI image, and found that there was no significant difference in the distribution of the original data, indicating that the AD and NL samples of all datasets do not show obvious pixel value differences, and will not affect the effectiveness of the model, as shown in Fig 9A. After analyzing the DPM of AD disease by t-SNE, polarization was found in AD and NL samples, as shown in Fig 9B. The AD and NL samples without distribution differences showed obvious differences after the model processing. The results show that the model can effectively learn the pathological features of AD and generate a DPM of AD with different degrees. Pixel-level brain AD DPM can help neurologists find the early pathological features of patients' brains and provide evidence of early AD diagnosis.

This study uses multimodal data to diagnose AD. The AD DPM of important brain regions and clinical information with individual characteristics (age, gender, MMSE, and APoE4) were used to diagnose AD. The model reached 95.6%±2.5%ACC and 99.6%±0.2%AUC in the ADNI test set. For multi-center study, the reproducibility of this method is also very important [43]. Therefore, in addition to the ADNI data set, we also verified the performance of the model in AIBL and NACC. The model reached 95.8%±0.7%ACC and 97.9%±0.2%AUC in the AIBL. And it reached 90.7%±1.0%ACC and 96.1%±0.3%AUC in the NACC. In order to verify the influence of different modal information on the diagnosis of AD, ablation experiments were implemented. When using only DPM, the diagnosis level of the model is similar to that of the traditional CNN. This phenomenon means that the model proposed can effectively extract the effective features from the raw data. The original MRI was abstracted as DPM without the loss of a large amount of effective information. The predicted high-risk AD areas overlapped with the MRI images, as shown in Fig 9A. The model showed great attention to the hippocampus and medial temporal lobe. Combining Fig 7 and Fig 12, we found the phenomenon of cerebral cortex atrophy in the high-risk areas predicted by the model, which is the same phenomenon that neurologists are concerned about. The experimental results prove the rationality of this method to a certain extent. In addition, the model also pays attention to other regions, such as parietal lobe, frontal lobe, corpus callosum and so on. We speculate that this is related to amyloid-β and Tau protein deposition in important brain regions. The model in this paper has the ability to assist clinicians in analyzing pathological characteristics to a certain extent. However, since the etiology of AD is still unclear, the result still needs to be further verified by anatomical results. For example, the autopsy report overlaps with the prediction results of the model in this paper. Therefore, in our future work, we will collect samples containing anatomical reports to further verify the reliability of this method from a clinical perspective.

When only clinical information was used, the model performance with APoE4 status was higher than that without APoE4 status. This suggests that APoE4 may be a target for AD treatment. When using multimodal information, the model performance is further improved. This shows that not only the anatomical pathology, but also the clinical information with personal characteristics plays a crucial role in the diagnosis of AD, implying the better potential development of multimodal models. The attention mechanism is embedded in the model to improve the interpretability of the multimodal model. The attention map expresses the attention of the model to different physiological parameters as an AD diagnostic basis. In Fig 12B, the model shows the strongest attention to MMSE and APoE4, agreeing with the results of clinical practice and the above ablation experiment. From a biomedical point of view, typical histopathological changes in AD include amyloid deposition and neuronal fiber entanglement [44, 45]. Some studies have shown that APoE4 can reduce the stability of the nerve cell membrane, leading to neurofibrillary tangle and cell death, which are important factors for AD [46, 47]. In addition, the model also shows a certain degree of concern regarding age and gender and pays more attention to gender. Recent studies have shown that gender can regulate the effect of APoE4 on Tau protein precipitation in the brain. Women with APoE4 mutations are more likely to have Rau accumulation than men [48, 49]. The actual biomarker research results were the same as the model analysis results in this study. Studies have also shown that as age increases, the probability of suffering from AD increases year by year. This phenomenon shows that the AI model designed in this study can not only analyze the biomarkers directly related to diseases, but also has great potential to discover potential biomarkers of diseases.

Although the model in this study shows potential, there are still limitations. First, this study only included AD and NL health conditions, which makes it difficult to meet the needs of neurologists in clinical practice. In the future, we will collect more data on the types of diseases and study how to further expand the function of the model. Second, the personal physiological information used in this study only included age, gender, MMSE, and APoE4. With the deepening of research, more biological information related to AD has been found. In the future, we plan to collect more types of physiological information, hoping that the AI model can guide the targeted treatment of AD. At the same time, we will also collect more center data to further verify the reliability of the model. The influence of acquisition sequence on diagnosis will be further studied. Third, the risk area predicted by the model needs further analysis and comparison with the actual anatomic report. So far, the pathogeny of AD is not clear, so it is difficult to get accurate lesion markers. Therefore, the purpose of this paper is to propose a multimodal AI auxiliary diagnosis model that can be explained to a certain extent, rather than a model for accurately segmenting lesions. However, the model's ability to predict disease risk still needs to be further evaluated. We will collect autopsy data with more relevant institutions in future work to specifically evaluate the model's ability to segment the lesion. For example, to evaluate the uncertainty of the model and verify the relationship between β-amyloid deposition in autopsy report and model prediction results. We will collect autopsy data with more relevant agencies in our future work. The importance of each physiological parameter in the development of AD requires further quantitative analysis. Fourth, although the method in this paper achieves a better performance, it still does not meet the clinical requirements. In future work, we will continue to study the optimization methods of the model, such as fusing attention-driven to further increase the performance and interpretability of the model. In addition, non-AI algorithms also have advantages that AI methods do not have in MRI reconstruction. Therefore, how to integrate AI and non-AI is also one of our future works.

In conclusion, a multi-mode deep reinforcement learning with the whole process interpretability is designed, which can not only diagnose AD accurately but also analyze potential biomarkers. The model can speed up the process of patients' medical treatment, improve the experience of patients' medical treatment, and provide a point of view for the combination of AI and medical diagnosis technology.

Acknowledgements
This work was financially supported by the NSFC (National Natural Science Foundation of China) project, authorization number: 61771261.The authors thank the ADNI, AIBL and NACC investigators for providing access to the data. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). 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 serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org. 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. The Australian Imaging Biomakers and Lifestyle (AIBL) study was supported by funding from the Science and Industry Endowment Fund, the Dementia Collaborative Research Centres, the McCusker Alzheimer's Research Foundation, the National Health and Medical Research Council (AUS), and the Yulgilbar Foundation, plus numerous commercial interactions supporting data collection. Details of the AIBL consortium can be found at www.AIBL.csiro.au and a list of the researchers of AIBL is provided at http://aibl.csiro.au/. The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI Robert Vassar, PhD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG005131 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

Funding

This work was financially supported by the National Natural Science Foundation of China (NSFC) project (authorization number: 61771261).

Conflict of Interest

The authors declare no conflict of interest.

Data Access Statement

The data that support the findings of this study are available in http://adni.loni.usc.edu/, https://aibl.csiro.au/ and https://naccdata.org/.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent to clinical testing and neuroimaging prior to participation of the ADNI, AIBL, NACC cohorts are obtained, approved by the institutional review boards (IRB) of all participating institutions. The Institutional Review Board of Nankai University approved this study and informed consents were waived for a retrospective cohort of AD disease patients.

Reference

1. Scheltens P, Blennow K, Breteler MM, et al. Alzheimer's disease. Lancet. 2016; 388: 505–517.
2. Brian Fulton-Howard, Alison M Goate, Robert P Adelson, et al. Greater effect of polygenic risk score for Alzheimer's disease among younger cases who are apolipoprotein E-ε4 carriers. Neurobiology of Aging. 2021, 99: 101.e1-101.e9.
3. Wenhong Chen, Songtao Li, Yangyang Ma, et al. A simple nomogram prediction model to identify relatively young patients with mild cognitive impairment who may progress to Alzheimer's disease. Journal of Clinical Neuroscience. 2021; 91: 62-68.
4. Jack CR Jr, Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 2013; 12: 207–216.
5. Harper L, Barkhof F, Scheltens P, et al. An algorithmic approach to structural imaging in dementia. J Neurol Neurosurg Psychiatry. 2014; 85: 692–698.
6. Mattsson N, Insel PS, Donohue M, et al. Predicting diagnosis and cognition with (18)F-AV-1451 tau PET and structural MRI in Alzheimer's disease. Alzheimers Dement. 2019; 15: 570–580.
7. Ossenkoppele R, Smith R, Ohlsson T, et al. Associations between tau, Abeta, and cortical thickness with cognition in Alzheimer disease. Neurology. 2019; 92:e601–e612.
8. Whitwell JL, Dickson DW, Murray ME, et al. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: a case-control study. Lancet Neurol. 2012; 11: 868–877.
9. Frisoni GB, Fox NC, Jack CR Jr, et al. The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol. 2010; 6:67–77.
10. Raji CA, Lopez OL, Kuller LH, et al. Age, Alzheimer disease, and brain structure. Neurology 2009; 73:1899–1905.
11. Barkhof F, Polvikoski TM, van Straaten EC, et al. The significance of medial temporal lobe atrophy:a postmortem MRI study in the very old. Neurology. 2007; 69:1521–1527.
12. Beach TG, Monsell SE, Phillips LE, et al. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. J Neuropathol Exp Neurol. 2012; 71: 266–273.
13. Yuliang Liu, Guohua Liu, Quan Zhang. Deep learning and medical diagnosis. The Lancet. 2019; 394:1709-1710.
14. Zhang Q, Chen Z, Liu G, er al. Artificial Intelligence Clinicians Can Use Chest Computed Tomography Technology to Automatically Diagnose Coronavirus Disease 2019 (COVID-19) Pneumonia and Enhance Low-Quality Images. Infect Drug Resist. 2021;14:671-687
15. J Xu, Z-A Liu, Y Hou, et al. Pixel-level Non-local Image Smoothing with Objective Evaluation. IEEE Transactions on Multimedia. DOI: 10.1109/TMM.2020.3037535
16. Alexander Craik, Yongtian He and Jose L Contreras-Vidal. Deep learning for electroencephalogram (EEG) classification tasks: a review. J Neural Eng. 2019; 16:031001.
17. Fatemeh Fahimi, Zhuo Zhang, Wooi Boon Goh, et al. Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. J Neural Eng. 2019, 16:026007.
18. Yannick Roy, Hubert Banville, Isabela Albuquerque, et al. Deep learning-based

electroencephalography analysis: a systematic review. J Neural Eng. 2019, 16:051001.

19. Castelvecchi D. Can we open the black box of AI? Nature. 2016; 538:20–23.

20. ELIZABETH A. HOLM. In defense of the black box. SCIENCE. 2019;364(6435):26-27.

21. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology. 2010; 74: 201–209.

22. Ellis KA, Rowe CC, Villemagne VL, et al. Addressing population aging and Alzheimer's disease through the Australian imaging biomarkers and lifestyle study: collaboration with the Alzheimer's Disease Neuroimaging Initiative. Alzheimers Dement. 2010; 6: 291–296.

23. Beekly DL, Ramos EM, van Belle G, et al. The National Alzheimer's Coordinating Center (NACC) Database: an Alzheimer disease database. Alzheimer Dis Assoc Disord. 2004; 18: 270–277.

24. Toghi Eshghi Shadi, Au-Yeung Amelia, Takahashi Chikara, et al. Quantitative Comparison of Conventional and t-SNE-guided Gating Analyses. Frontiers in Immunology. 2019;10:1194.

25. Wentian Li, Jane E Cerise, Yaning Yang. Application of t-SNE to human genetic data. Journal of Bioinformatics and Computational Biology. 2017;15(4):1750017.

26. Li W, Feng X, An H, et al. MRI Reconstruction with Interpretable Pixel-Wise Operations Using Reinforcement Learning. Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34(1):792-799.

27. Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning. 2016; In ICLR.

28. Ye QH, Xia J and Yang G. Explainable AI for COVID-19 CT Classifiers: An Initial Comparison Study. 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Aveiro, Portugal, 2021;521-526. doi: 10.1109/CBMS52027.2021.00103

29. Yang G, Ye QH and Xia J. Unbox the black-box for the medical explainable ai via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. Information Fusion. 2021; 77:29-52.

30. Yang G, Yu S, Dong H, et al. Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE transactions on medical imaging. 2017;37(6):1310–1321

31. Jure Zbontar, Florian Knoll, Anuroop Sriram, et al. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. 2018. arXiv:1811.08839.

32.Liu YK, Yang G, Hosseiny M, et al. Exploring Uncertainty Measures in Bayesian Deep Attentive Neural Networks for Prostate Zonal Segmentation. IEEE Access. 2020; 8:151817-151828.

33.Yang G, Chen J, Gao ZF, et al. Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. Future Generation Computer Systems. 2020;107:215-228.

34.Liu YK, Yang G, Mirak SA, et al. Automatic prostate zonal segmentation using fully convolutional network with feature pyramid attention. IEEE Access. 2019;7:163626-163632.

35. Trzepacz PT, Hochstetler H, Wang S, et al. Relationship between the Montreal Cognitive Assessment and Mini-mental State Examination for assessment of mild cognitive impairment in older adults. BMC Geriatr. 2015;15:107.

36. Creavin ST, Noel-Storr AH, Smailagic N, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's dementia and other dementias in asymptomatic and previously clinically unevaluated people aged over 65 years in community and primary care populations. Cochrane

Database of Systematic Reviews. 2014; 6:CD011145.

37. Philip B Verghese, Joseph M Castellano and David M Holtzman. Apolipoprotein E in Alzheimer's disease and other neurological disorders. Lancet Neurology. 2011, 10(3):241-252.

38. Takahisa Kanekiyo, Huaxi Xu and Guojun Bu. ApoE and Aβ in Alzheimer's Disease: Accidental Encounters or Partners? Neuron. 2014;81(4):740-754.

39. Habib N, McCabe C, Medina S, et al. Disease-associated astrocytes in Alzheimer's disease and aging. Nat Neurosci. 2020;23:701–706.

40. Nebel RA, Aggarwal NT, Barnes LL, et al. Understanding the impact of sex and gender in Alzheimer's disease: A call to action. Alzheimer's & Dementia. 2018;14:1171-1183.

41. Brandalyn C Riedel, Paul M Thompson and Roberta Diaz Brinton. Age, APOE and sex: Triad of risk of Alzheimer's disease. The Journal of Steroid Biochemistry and Molecular Biology. 2016;160:134-147.

42. Mielke M, Vemuri P, Rocca W. Clinical epidemiology of Alzheimer's disease: assessing sex and gender differences. Clin Epidemiol. 2014;6:37-48.

43. Raschke F, Barruck TR, Jones TL et al. Tissue-type mapping of gliomas. NeuroImage: Clinical 2019;21:101648

44. Victor L Villemagne, Samantha Burnham, Pierrick Bourgeat, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurology. 2013;12(4): 357-367.

45. Ju YE, Lucey B, and Holtzman D. Sleep and Alzheimer disease pathology—a bidirectional relationship. Nat Rev Neurol. 2014; 10:115–119.

46. Uddin MS, Kabir MT, Al Mamun A, et al. APOE and Alzheimer's Disease: Evidence Mounts that Targeting APOE4 may Combat Alzheimer's Pathogenesis. Mol Neurobiol. 2019; 56:2450–2465.

47. Safieh M, Korczyn AD and Michaelson DM. ApoE4: an emerging therapeutic target for Alzheimer's disease. BMC Med. 2019;17:64.

48. Altmann A, Tian L, Henderson VW, et al. Sex modifies the APOE-related risk of developing Alzheimer disease. Ann Neurol. 2014;75:563-573.

49. Buckley RF, Mormino EC, Rabin JS, et al. Sex Differences in the Association of Global Amyloid and Regional Tau Deposition Measured by Positron Emission Tomography in Clinically Normal Older Adults. JAMA Neurol. 2019;76(5):542–551.