



Research article

Improved neural network with multi-task learning for Alzheimer's disease classification

Xin Zhang^a, Le Gao^{a,*}, Zhimin Wang^a, Yong Yu^b, Yudong Zhang^c, Jin Hong^d^a School of Electronic and Information Engineering, Wuyi University, Jiangmen, 529000, China^b School of Computer Science, Shaanxi Normal University, Xi'an, 710062, China^c School of Computing and Mathematic Sciences, University of Leicester, Leicester, LE17RH, UK^d School of Information Engineering, Nanchang University, Nanchang, 330031, China

ARTICLE INFO

Keywords:

Alzheimer's disease
VGG16 network
Multi-task learning

ABSTRACT

Alzheimer's disease (AD) poses a significant challenge due to its widespread prevalence and the lack of effective treatments, highlighting the urgent need for early detection. This research introduces an enhanced neural network, named ADnet, which is based on the VGG16 model, to detect Alzheimer's disease using two-dimensional MRI slices. ADNet incorporates several key improvements: it replaces traditional convolution with depthwise separable convolution to reduce model parameters, replaces the ReLU activation function with ELU to address potential issues with exploding gradients, and integrates the SE (Squeeze-and-Excitation) module to enhance feature extraction efficiency. In addition to the primary task of MRI feature extraction, ADnet is simultaneously trained on two auxiliary tasks: clinical dementia score regression and mental state score regression. Experimental results demonstrate that compared to the baseline VGG16, ADNet achieves a 4.18% accuracy improvement for AD vs. CN classification and a 6% improvement for MCI vs. CN classification. These findings highlight the effectiveness of ADnet in classifying Alzheimer's disease, providing crucial support for early diagnosis and intervention by medical professionals. The proposed enhancements represent advancements in neural network architecture and training strategies for improved AD classification.

1. Introduction

Alzheimer disease (AD) [1] is a degenerative disease that hides the progressive nervous system. AD patients often suffer from dementia symptoms such as memory decline, communication disorder, and confusion after illness, which seriously affect daily life and can lead to irreversible brain damage. Research has shown that in the past 20 years, dementia has become the seventh leading cause of death among all diseases, with one-third of elderly people dying from Alzheimer's disease or other forms of dementia [2]. However, people are still unclear about the pathogenic factors of AD. If a correct diagnosis can be made in the early stages of a patient's attack, some modern medical treatment methods may effectively alleviate the condition and delay its development [3–5]. In the past, most AD detection methods relied on clinical observation, which mainly relied on the experience of clinical doctors, and often had to wait until the condition worsened and symptoms became apparent before being diagnosed. With the continuous development and maturity of medical imaging technology, neuroimaging technology can detect subtle changes in the brain structure of AD patients, among which

* Corresponding author.

E-mail addresses: Le.gao@nscg-gz.cn (L. Gao), hongjin@ncu.edu.cn (J. Hong).

magnetic resonance imaging (MRI) technology can display high-resolution brain anatomy, and can effectively highlight images with AD characteristics [6].

Magnetic resonance imaging (MRI) technology offers high-resolution and non-invasive capabilities, enabling the detection of subtle alterations in the brain structure of individuals with Alzheimer's disease (AD). By employing specialized imaging sequences, MRI can unveil distinctive brain characteristics linked to AD, which is crucial for the accurate diagnosis of AD and the differentiation from other forms of dementia. Furthermore, MRI technology can assess brain function and activity, with functional magnetic resonance imaging (fMRI) techniques allowing for the observation of changes in brain activity in AD patients during task performance or stimulation. The Mini-Mental State Examination (MMSE) [7] is a widely utilized cognitive assessment instrument for evaluating cognitive function in individuals with AD and other forms of dementia. The MMSE offers several advantages, including its ease of use, standardization, high reliability, and ability to aid in the early identification of dementia in patients. By comparing MMSE scores to established normative values, healthcare professionals can gauge the extent of cognitive decline in a patient and track the progression of the disease. The Clinical Dementia Rating (CDR) [8] is an assessment tool designed to evaluate the overall functioning and disease severity in individuals with dementia. The comprehensive nature of the CDR allows for a thorough assessment of the patient's overall functioning, aiding in the determination of disease severity, development of treatment plans, and monitoring of disease progression.

Most current research focuses on a single test (such as MRI), but should focus on combining multiple tests to improve the diagnosis of Alzheimer's disease dementia. This article proposes a classification method for Alzheimer's disease based on an improved VGG16 network and multi-task learning, named ADnet, with data sourced from MRI imaging data in ADNI1 and patient related clinical information. The study used an improved VGG16 network for feature extraction of MRI data as the main task, and used the CDR score regression and the MMSE score regression as two auxiliary tasks to jointly train the model. The model parameters were updated based on gradient backpropagation to improve the classification performance of the Alzheimer's disease classification algorithm.

2. Related work

2.1. Background

In recent decades, machine learning and pattern recognition methods have been widely used in neuroimaging analysis of AD. The machine learning method has a good effect in analyzing MRI [9]. Meng et al. [10] used MRI in ADNI to obtain 649 voxel based morphometry (VBM) methods, including early Mild cognitive impairment (MCI), cognitive normal(CN), Late Mild cognitive impairment (LMCI) and Alzheimer's disease (AD) construct new voxel based features, and propose a feature detection method based on random support vector machine (RS-SVM) for feature extraction and classification. Elaheh et al. [11] used regularized Logistic regression to conduct Feature selection on the MRI data of AD samples, and used Random forest classifier to classify the data. Kumari et al. [12] proposed an adaptive hyperparametric tuned Random forest fusion classifier (HPT-RFE). The simulation results show that the performance of the adaptive HPT-RFE classifier is the best in binary classification. The research shows that the machine learning method has achieved initial results in the field of Medical imaging, and then there are some limitations and shortcomings in this method. For example, machine learning requires a complex process for manual feature extraction, and extracting important features from MRI images often requires experienced clinical staff to operate. In addition, the mechanism of AD pathogenesis is not fully understood, and machine learning methods are also difficult to effectively achieve manual feature extraction. The research shows that the machine learning method has achieved initial results in the field of Medical imaging, and then there are some limitations and shortcomings in this method. For example, machine learning requires a complex process for manual feature extraction, and extracting important features from MRI images often requires experienced clinical staff to operate. In addition, the mechanism of AD pathogenesis is not fully understood, and machine learning methods are also difficult to effectively achieve manual feature extraction.

Deep learning is an important branch of Artificial Intelligence, which can automatically extract image information through efficient algorithms and complete complex task learning through abstract features of images [13,14]. Compared with machine learning, deep learning has fewer image preprocessing steps and can automatically extract the features of Medical imaging data to obtain a more objective and less biased result. Spasov et al. [15] used magnetic resonance imaging (MRI), Demographics, Neuropsychology and APOe4 genetic data as input data, and proposed a new deep learning architecture based on dual learning and 3D separable convolution temporary layer. Zhang et al. [16] proposed a 3D generated countermeasure network (BPGAN), which can synthesize brain positron emission tomography (PET) from magnetic resonance imaging (MRI), which can be used as a potential data completion scheme for multimodal medical image research. Kim et al. [17] proposed a spectral GCN method with CCA loss function (GCN-CCA), which can extract important image features from genetic data. Jin et al. [18] proposed using an improved 3D residual network (ResNet) to diagnose AD, achieving good classification results. By introducing attention mechanisms into the network, classification performance was improved, while potential biomarkers were studied. Oh et al. [19] used Unsupervised learning of convolutional automatic encoder (CAE) to complete the classification task of AD and CN, with an accuracy of 86.60%.

2.2. Convolutional neural network

Convolutional neural network (CNN) is one of the most concerned research topics in the field of artificial intelligence depth learning [20,21]. It has a broad application prospect in target recognition, image detection, segmentation and other fields. CNN includes various network structures, such as AlexNet [22], VGGNet [23], GoogLeNet [24], ResNet [25], and other frameworks. The basic structure of CNN usually includes input layer, convolutional layer, activation layer, pooling layer, fully connected layer, and output layer. By superimposing these layers, a complete convolutional neural network can be constructed.

2.2.1. Convolutional layer

Convolution layer is the core layer of Convolutional neural network, which mainly performs convolution operations. Convolutional verification iteratively calculates the input or output features of the previous layer, extracts features through filtering, and summarizes the extracted local feature maps to obtain the feature matrix extracted by the convolution of that layer. The eigenvalues calculated by convolution kernel corresponding to the local area of the image are called Receptive field. In the case of two-dimensional convolution, the convolution kernel slides left and right first, then up and down on the feature map, and multiplies and sums the pixel values in the receptive field to obtain the mapping values of each Receptive field. Fig. 1 is a schematic diagram of the convolution calculation process, with input features of 4x4, convolution kernel size of 3x3, and stride size of 1.

2.2.2. Pooling layer

The main function of the pooling layer is to downsample data to reduce its spatial dimension. This can be used to reduce the model's size and improve its computational efficiency. Commonly used pooling functions include: Mean pooling, Max pooling, and Stochastic pooling. These three pooling methods are shown in Fig. 2.

2.2.3. Fully connected layer

In a convolutional neural network, the function of fully connected layer is equivalent to that of a classifier. Each neuron in the fully connected layer is connected to all neurons in its previous layer. The fully connected layer can integrate local information with class distinctiveness from convolutional or pooling layers, and utilize filters for convolutional operations to merge the extracted feature information, and finally achieve classification calculation.

2.3. Depthwise separable convolution

Depthwise separable convolution [26] is a common convolution method in convolutional neural networks. As shown in Fig. 3, depthwise separable convolution decomposes a standard convolution into two steps: depthwise convolution and pointwise convolution. Depthwise convolution involves applying convolutional kernels to each channel of input data, and then overlaying all results to obtain a set of depthwise convolution outputs. The computational complexity of depthwise convolution is relatively small, as it only requires convolution operations on each channel. Pointwise convolution is the convolution of the outputs of depthwise convolution, which uses a convolution kernel to multiply each depthwise convolution output with a set of learnable weight matrices to obtain the final convolution output.

2.4. Channel attention module

Channel attention [27] is used to mine dependencies in convolutional channels, model data channels, and adjust channel weights for features. The utilization of the channel attention module can weight the importance of key data features in the model, while also suppressing other irrelevant features to improve the network's representation ability [28].

The SE (Squeeze Excitation) module [29] mainly captures correlations among feature channels, enabling the network to focus on effective features. This leads to enhanced extraction of target features and an overall improvement in the model's expression ability. As shown in Fig. 4, the SE module mainly comprises two operations: Squeeze and Excitation.

The squeeze operation of the SE module compresses features based on spatial dimensions, using global average pooling to encode the spatial features of different channels into corresponding global features. The calculation formula is shown in Eq. (1).

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (1)$$

Among them, Z_c is the global feature on the compressed channel, u_c is the feature before compression, and W and H are the width and height of the feature map.

The excitation function in the SE module can use full connection, ReLU activation function, sigmoid activation function and other operations on the global feature information Z obtained in the previous step, and finally obtain the relationship between channels, and calculate the normalized weight of (0,1). Its calculation formula is shown in Eq. (2).

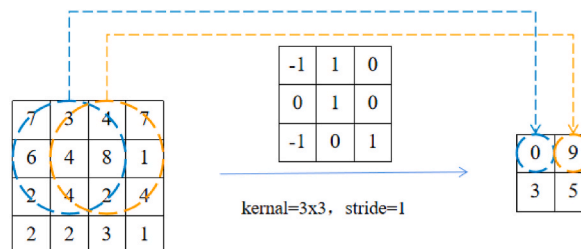


Fig. 1. Schematic diagram of convolution operation process.

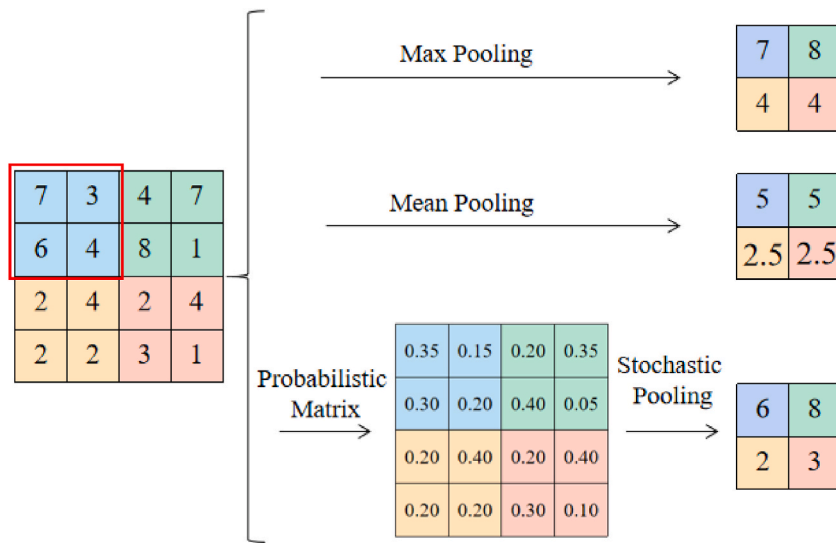


Fig. 2. Three pooling methods.

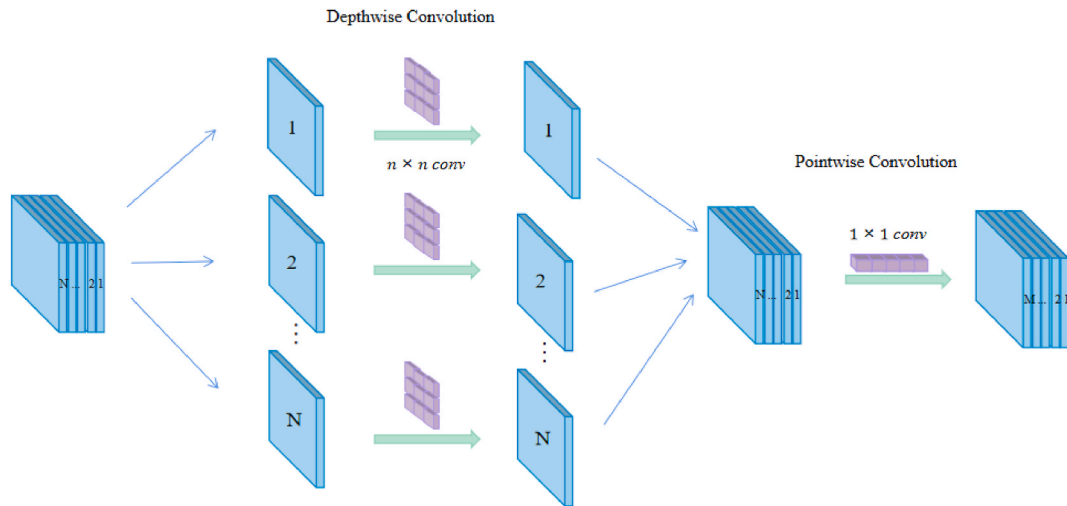


Fig. 3. Depthwise separable convolution.

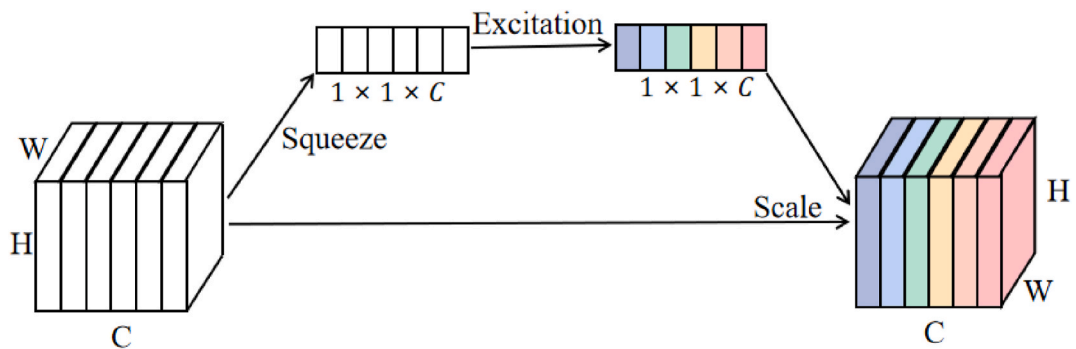


Fig. 4. SE module.

$$S = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z))$$

(2)

Where σ and δ are respectively sigmoid activation function and ReLU activation function, $W_1 \in \mathbb{R}^{c \times c}$, $W_2 \in \mathbb{R}^{c \times c}$, the hyperparameter represents the dimensionality reduction parameter, and r represents the scaling ratio. Its purpose is to simplify the computational workload by reducing the count of channels.

The scale operation of SE module can multiply the weight value with the original features of the image to obtain different contribution degrees of different channels. The calculation formula is shown in Eq. (3):

$$X_c = F_{\text{Scale}}(u_c, S_c) = S_c \bullet u_c \quad (3)$$

Among them, S_c is the weight value of the channel obtained during the excitation process, and u_c is the feature information of the channel obtained during the compression operation.

2.5. Multi-task learning

Multi-task learning was proposed by Caruana [30]. Traditional machine learning methods are usually based on a single task learning pattern, which decomposes into multiple independent single tasks when dealing with complex tasks. This approach only focuses on a single model and often overlooks important information that exists in related tasks [31]. In multi-task learning, the original task can have better generalization ability by sharing parameters between different tasks. As shown in Fig. 5, according to the MTL sharing method, it can be divided into hard sharing and soft sharing [32]. Hard sharing is to place the network layer of a specific task on the sharing layer and combine multiple Loss function to train the model [33,34]. Soft sharing is where each task has an independent model and parameter set, and interaction between tasks is achieved through a sharing mechanism [35,36].

Compared to single tasks, multi-task learning have the following advantages: 1) Parameter sharing between tasks can greatly reduce the memory occupied by the network. 2) The features in the shared layer are not recalculated. One task iteration can display higher inference speed. 3) Sharing complementary information from related tasks can help improve network performance potential.

3. Data and methods

3.1. Data sources

The data for this study is sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [37]. ADNI is an open source medical dataset currently the authoritative data center for studying Alzheimer's disease. This study conducted experiments using magnetic resonance data from test subjects in the ADNI1 standard database. As shown in Table 1, the dataset comprises 208 AD patients, 264 individuals with normal cognition and 222 patients with Mild Cognitive Impairment(MCI). Among the AD patients, there are 108 male and 100 female patients, with an average age of 76.3 years. The average MMSE score is 20.56, and the average CDR score is 1.09; Among normal cognitive individuals (CN), there are 132 males and 132 females, with an average age of 75.84 years. The average MMSE score is 27.94, and the average CDR score is 0.07; There were 110 male patients and 112 female patients with Mild cognitive impairment (MCI), with an average age of 75.58 years. The mean MMSE score was 27.06, and the mean CDR score was 0.56. In clinical practice, doctors will use their MMSE and CDR levels as a basis for evaluation, and the diagnostic criteria are shown in Table 2.

3.2. MRI data preprocessing

Use Matlab 2016b's SPM12 (Statistical Parameter Mapping 12) and CAT12 (Computational Anatomy Toolbox 12) toolkits to preprocess the data of MRI images downloaded from ADNI. The preprocessing steps included: (1) Correction: AC-PC (Anterior

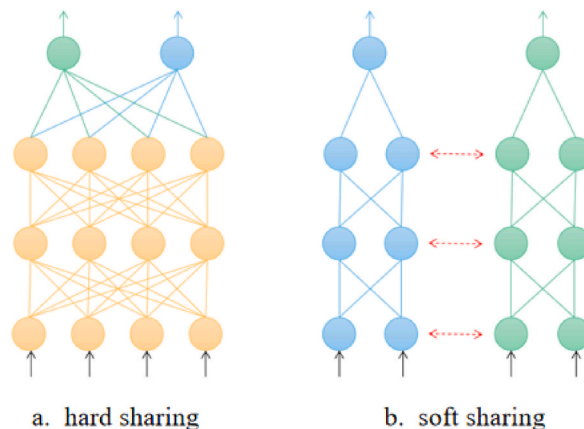


Fig. 5. Multi-task learning sharing method.

Table 1
Experimental data.

	AD	CN	MCI
Gender (male/female)	108/100	132/132	110/112
Age (mean ± standard deviation)	76.3 ± 7.24	75.84 ± 6.39	75.58 ± 8.05
MMSE score (mean ± standard deviation)	20.56 ± 4.03	27.94 ± 2.81	27.06 ± 2.43
CDR score (mean ± standard deviation)	1.09 ± 0.77	0.07 ± 0.32	0.56 ± 0.59

Table 2
Diagnostic criteria.

AD	MCI	CN
CDR 1–3 & MMSE < 24	CDR 0.5–1 & MMSE ≥ 24	CDR 0–0.5 & MMSE ≥ 24

Commissure-Posterior Commissure) correction [38] is performed on each MRI. Because there is no significant difference in the AC-PC lines of different human brains, a three-dimensional coordinate system can be established with the midpoint of AC-PC as the origin, and the image data of different human brains obtained from this three-dimensional space can be compared. (2) Skull removal and cerebellectomy: The original MRI images contain some non brain structures, such as the skull. To avoid increasing computational complexity and avoiding subsequent image preprocessing that may affect experimental results, it is necessary to remove non brain structures such as the skull from the image during the image preprocessing operation. (3) Registration: use Affine transformation and nonlinear registration to normalize all subjects' images and register them on the MNI (Montreal Neurological Institute) standard template. (4) Modulation: After spatial registration, the volume of gray tissue and other tissues changes, so it is necessary to compensate for the volume of gray tissue after nonlinear registration. (5) Segmentation: The brain image is divided into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). In this article, gray matter data is used as experimental data, so removing white matter and cerebrospinal fluid can reduce the number of parameters and avoid excess data affecting experimental accuracy. (6) Slicing: Slice the segmented images, calculate the information entropy of each sliced image, and select the 30 images with the highest information entropy as the experimental dataset.

3.3. Method

The ADNet model diagram of this article is shown in Fig. 6. This study takes the VGG16 network as the basic network and improves the model in four aspects: (1) using depthwise separable convolution instead of ordinary convolution to reduce the number of model parameters; (2) Use ELU activation function instead of ReLU activation function to avoid gradient explosion during backpropagation; (3) Embed the channel attention SE module into the network; (4) Add two auxiliary tasks to improve the accuracy of the model.

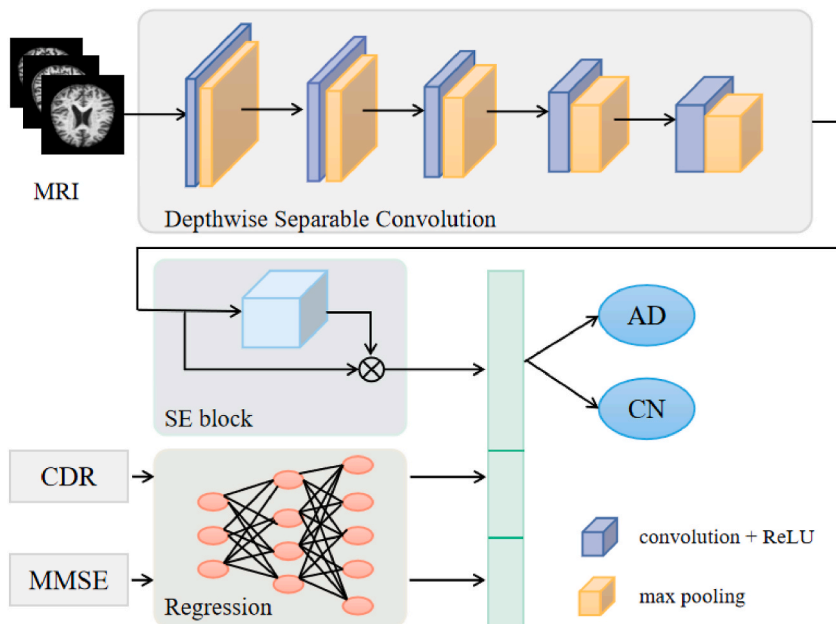


Fig. 6. ADNet model diagram.

The VGGNet architecture was proposed by a research team at the University of Oxford, and the VGG16 network model is shown in Fig. 7. Examining Fig. 7 reveals that the VGG16 network can be divided into five convolutional blocks, where each convolutional block contains two or three convolutional layers, with 3×3 convolutional kernels being employed within each convolutional layer. The initial two convolutional blocks are composed of two convolutional layers, while the subsequent three convolutional blocks are composed of three convolutional layers each. Subsequently, there are three fully connected layers with 4096, 4096 and 1000 neurons. The ReLU function is used for all activation units in the hidden layer, and the softmax activation function is used for classification in the final output layer.

3.3.1. Depthwise separable convolution

In the VGG16 model, the convolutional layer uses standard convolution, which has a relatively large number of parameters and also requires high computational complexity. However, replacing standard convolution with depthwise separable convolution can lead to a significant reduction in both the model's parameters and computational complexity. This enhances the model's training and inference speed. Eqs. (4) and (5) represent the parameter quantities of standard convolution and depthwise separable convolution, respectively. It can be seen that the parameter quantities of depthwise separable convolution are much smaller than standard convolution.

$$P_1 = C_{in} \times K \times K \times C_{out} \tag{4}$$

$$P_2 = C_{in} \times (K \times K + C_{out}) \tag{5}$$

3.3.2. ELU activation function

The activation function realizes the arbitrary mapping between input and output by adding nonlinear factors, thus improving the expression ability of the model and playing an important role in the deep neural network. The activation function in the initial VGG16 network uses the modified linear unit ReLU. This paper introduces an improved ReLU exponential linear unit ELU activation function, whose function curve is shown in Fig. 8, and the expressions of ReLU activation function and ELU activation function are shown in Eqs. (6) and (7). In contrast to the Rectified Linear Unit (ReLU), the Exponential Linear Unit (ELU) offers notable benefits. Primarily, its function curve maintains continuous differentiability across all intervals, preventing the occurrence of vanishing or exploding gradients during the training process. Additionally, ELU allocates values within the negative range, ensuring that the gradient remains non-zero for all negative inputs. This effectively mitigates the problem of "neuron necrosis," wherein specific neurons become inactive during training, resulting in diminished model performance.

$$\text{ReLU}(x) = \begin{cases} x, & x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

$$\text{ELU}(x, \alpha) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & \text{otherwise} \end{cases} \tag{7}$$

3.3.3. Channel attention module

The attention mechanism increases the nonlinearity of feature mapping in different dimensions to enhance the network's feature extraction ability. Therefore, this article improves the network's feature extraction ability by adding channel attention mechanism. The network structure with SE module is shown in Fig. 9. The SE module is composed of global average pooling, FC, ReLU, FC, and sigmoid. The SE module performs transformations on the channel dimension to weaken unimportant features and enhance their

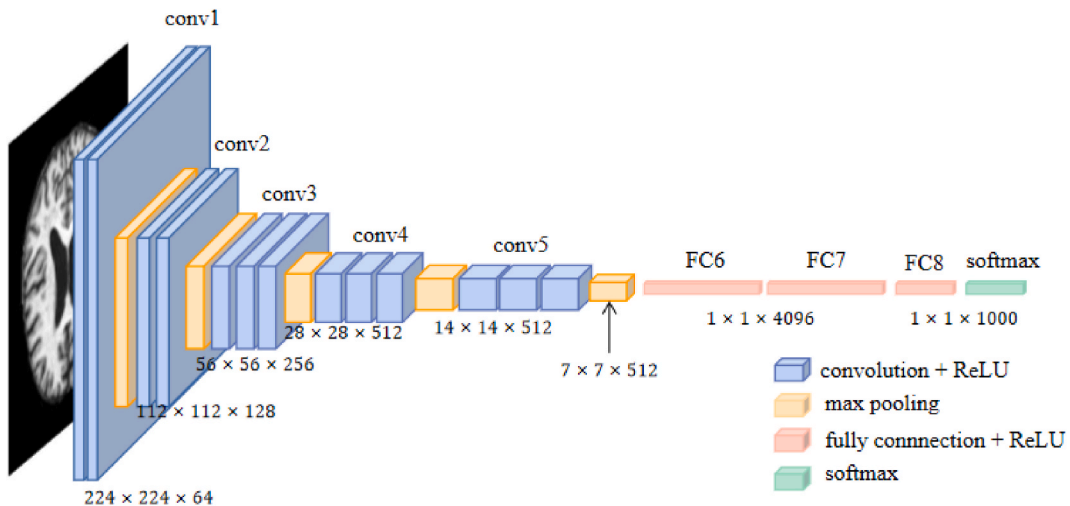


Fig. 7. The structure diagram of VGG16 network.

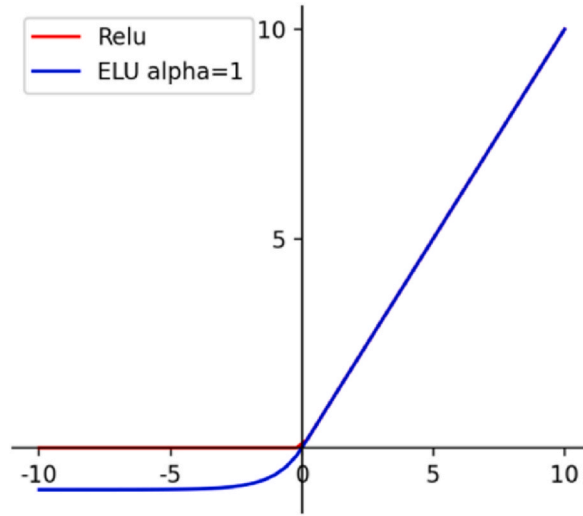


Fig. 8. ReLU and ELU function curves.

directionality.

3.3.4. Multi-task learning

MRI images can evaluate the brain from the perspective of structural changes in the brain, but cannot be evaluated from the patient’s cognitive perspective. CDR is a comprehensive score obtained by doctors based on communication with patients and their families to assess the degree of dementia in patients. Regression using CDR values has a certain degree of correlation with Alzheimer’s disease classification tasks. MMSE is a screening scale for dementia. MMSE is generally used as a reference for doctors in clinical AD diagnosis, so there is a certain correlation between MMSE value regression task and Alzheimer’s disease classification task.

Therefore, this article uses MRI data for AD classification as the main task, CDR score regression task, and MMSE score regression task as auxiliary tasks to jointly train the model. The model parameters are updated based on gradient back propagation to improve the classification performance of the AD classification algorithm.

The AD classification task uses the Cross entropy Loss function. The calculation formula is shown in Eq. (8).

$$L_{ce} = \frac{1}{n} \sum_{i=0}^n -[y_i \log p_i + (1 - y_i) \log(1 - p_i)] \tag{8}$$

Where n represents the total number of samples, y_i represents the true value of the i-th sample, and p_i represents the probability that the i-th sample is predicted to be a positive class.

The Mean squared error (MSE) Loss function was used for CDR score regression task and MMSE score regression task. The calculation formula is shown in Eqs. (9) and (10),

$$L_{cdr} = \frac{1}{n} \sum_{i=0}^n (y_i - y_i^p)^2 \tag{9}$$

$$L_{mmse} = \frac{1}{n} \sum_{i=0}^n (y_i - y_i^p)^2 \tag{10}$$

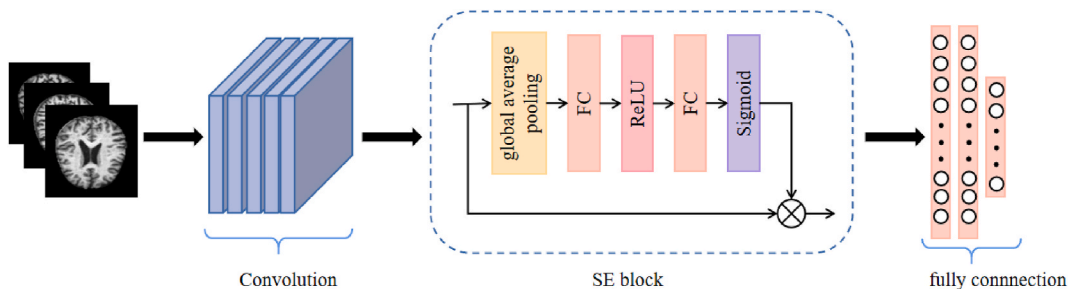


Fig. 9. The structure diagram of VGG16 network integrating SE module.

Where n represents the total number of samples, y_i represents the true value of the i -th sample, and y_i^p represents the predicted value of the model.

The loss function of multi task learning is shown in Eq. (11).

$$L = L_{ce} + L_{cdr} + L_{mmse} \quad (11)$$

3.4. Evaluation indicators

The evaluation indicators are accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV). In AD and CN classification tasks, accuracy refers to the proportion of correctly classified samples to the total number of samples, that is, the proportion of accurately classified AD and CN samples to the total number of samples. Sensitivity is the proportion of predicted positive classes and actually positive classes to actual positive classes, measuring the classifier's ability to recognize positive samples, that is, the proportion of accurately identified CN samples to all real CN samples. Specificity is the predicted negative class, which is actually the proportion of negative classes to all actual negative classes. It measures the classifier's ability to recognize negative samples, that is, the proportion of accurately identifying AD samples to all actual AD samples.

The calculation formula for accuracy is shown in Eq. (12):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

The calculation formula for sensitivity is shown in Eq. (13):

$$SEN = \frac{TP}{TP + FN} \quad (13)$$

The calculation formula for specificity is shown in Eq. (14):

$$SPE = \frac{TN}{TN + FP} \quad (14)$$

The calculation formula for positive predictive value is shown in Eq. (15):

$$PPV = \frac{TP}{TP + FP} \quad (15)$$

The calculation formula for negative predictive value in Eq. (16):

$$NPV = \frac{TN}{TN + FN} \quad (16)$$

TP is true positive, which refers to the number of AD (or MCI) samples predicted by the model as AD (or MCI), while FN is false negative, which refers to the number of AD (or MCI) samples predicted by the model as CN. FP is false positive, the number of CN samples predicted by the model as AD (or MCI) in this article, and TN is true negative, which is the actual number of CN samples predicted by the diagnostic model as CN.

4. Results and discussion

4.1. Training and testing sets

This experiment established three datasets, namely AD, CN, and MCI. In order to improve the classification performance of the neural network model and enhance its generalization ability, during the model training process, all data is divided into 6:2:2 parts, including the training set, validation set, and test set.

4.2. Experimental environment and parameter settings

The hardware environment of this experiment is: the processor is Intel Core i7-9700 CPU; RAM is 16 GB; The GPU is NVIDIA GeForce RTX 3090. The SPM12 and CAT12 toolkits in Matlab R2016b used for image preprocessing. The network model is built using the open-source deep learning framework Python, with an input image size of $224 \times 224 \times 3$. The hyperparameter settings are epochs = 100, Batch_size = 32, Initial Learning rate Learning_Rate = 0.01, Weight_Decay = 0.0001, Dropout = 0.5.

4.3. Experimental results and discussion

4.3.1. The impact of different activation functions

In this section, we compare the change of Loss function when ReLU and ELU activation function are used for training in VGG16 model. The experimental results are shown in Fig. 10. When using the ReLU activation function, the loss value curve has fluctuated several times in a large range during the training process, which may indicate the occurrence of gradient explosion, resulting in

excessive weight update of the model, thus causing instability of the model. In contrast, the loss value curve of the ELU activation function changes more smoothly. The negative part of the ELU activation function has a soft saturation characteristic, which makes the model more robust to input changes and noise. Therefore, using the ELU activation function can effectively suppress the occurrence of gradient explosion and improve the stability of training. In addition, models using the ELU activation function perform better in terms of accuracy than models using the ReLU activation function. This indicates that the properties of the ELU activation function enable the model to better handle input changes and noise, thereby improving classification performance.

4.3.2. The impact of SE module location on classification results

In this section, we investigate the impact of adding SE modules to different locations in the VGG16 network on the classification results of Alzheimer's disease. The VGG16 network consists of 5 convolutional blocks. We evaluate the performance changes by adding SE modules after the last convolutional layer of each convolutional block and conducting ablation experiments. The experimental results are shown in Fig. 11. It can be seen that the SE module added after the third convolutional block has the highest accuracy and specificity. The SE module after the third convolutional block can better capture key features in MRI images and provide more accurate classification and discrimination capabilities.

4.3.3. Classification results and discussion of different improvement methods

In order to verify the effectiveness of the improvement strategy based on the VGG16 model in this article, multiple sets of comparative and ablation experiments were conducted. The method used for comparison is as follows: only replace the ordinary convolutional layer in traditional VGG16 with Depthwise Separable Convolution, denoted as VGG16 (1); Only replace ReLU Activation function in VGG16 with ELU Activation function, which is recorded as VGG16 (2); Only embed the attention module SE module into VGG16, denoted as VGG16 (3); Add only two auxiliary tasks, denoted as VGG16 (4); Integrating the above strategies constitutes the method of this article. We compare these methods with the comprehensive improvement methods proposed in this article to evaluate their performance impact on Alzheimer's disease classification.

Table 3 and Table 4 respectively show the classification results of AD and CN, MCI and CN using different methods. It can be seen that the accuracy of the first optimized method VGG16 (1) is only 0.31% lower than VGG16 in AD and CN classification tasks, and 0.46% lower than VGG16 in AD and CN classification tasks, but its running time and parameters are less than VGG16. It is not difficult to see from the statistical data that VGG16 (1) has more advantages in the diagnosis of AD/MCI patients. Its sensitivity reaches 0.8572 and 0.7913, which are higher than 0.8481 and 0.7858 of VGG16, but its specificity is less than VGG16, That is, the misdiagnosis rate of VGG16 (1) is higher than that of VGG. In AD and CN classification tasks, the accuracy and specificity of the other three optimization strategies have been improved, but their specificity will be slightly lower than that of VGG16 network. In this paper, ADNet achieves the highest accuracy, sensitivity and specificity. In MCI and CN classification tasks, the accuracy, sensitivity and specificity of the other three optimization strategies were improved. These results show that ADNet proposed in this paper has achieved significant performance improvement in the classification of Alzheimer's disease. By combining depthwise separable convolution, ELU activation function, SE module and multi-task learning, we can better capture the key features in MRI images and provide more accurate classification and discrimination ability. The combination of these improved strategies plays a key role in improving the generalization ability and robustness of Alzheimer's disease classification model.

4.4. Comparative experiments

4.4.1. Comparison of experimental results and discussion of different models

Under the same experimental conditions, we further compared our method with common classification networks (AlexNet, GoogLeNet, ResNet, MobileNet) and current state-of-the-art models in AD classification. The experimental results are shown in Table 4.

First, the accuracy of VGG16 is second only to MobileNetV3, and it occupies a high position in all common classification networks. It is worth noting that MobileNetV3 includes the attention module. However, according to the results in Table 3, we can see that the

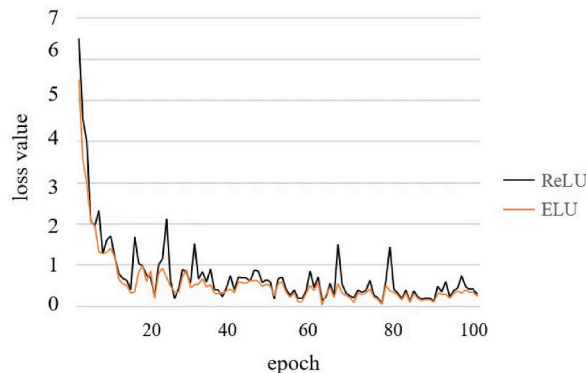


Fig. 10. The impact of different activation functions on the loss value of the model.

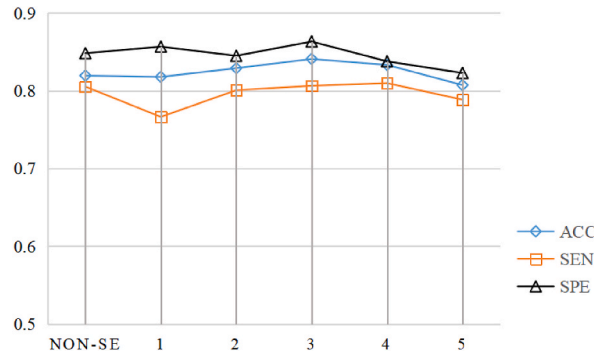


Fig. 11. Model classification effect of SE module in different positions(AD vs. CN).

Table 3

AD and CN classification results of different improved methods.

Methods	AD vs. CN					MCI vs. CN				
	ACC	SEN	SPE	PPV	NPV	ACC	SEN	SPE	PPV	NPV
VGG16	0.8294	0.8152	0.8481	0.8377	0.8553	0.7362	0.6994	0.7458	0.7471	0.7652
VGG16(1)	0.8263	0.7782	0.8572	0.8036	0.8936	0.7316	0.6773	0.7813	0.7243	0.8244
VGG16(2)	0.8451	0.7894	0.8860	0.8536	0.9133	0.7496	0.6977	0.8001	0.7655	0.8192
VGG16(3)	0.8408	0.8063	0.8632	0.8534	0.8956	0.7451	0.7002	0.7863	0.7564	0.8184
VGG16(4)	0.8489	0.7963	0.8912	0.8654	0.9256	0.7553	0.7223	0.7769	0.7847	0.8264
ADNet	0.8712	0.8154	0.9308	0.9336	0.9308	0.7962	0.7776	0.8633	0.8291	0.8655

Table 4

Classification results of different network models.

Methods	AD vs. CN					MCI vs. CN				
	ACC	SEN	SPE	PPV	NPV	ACC	SEN	SPE	PPV	NPV
VGG16	0.8294	0.8152	0.8481	0.8377	0.8553	0.7362	0.6994	0.7458	0.7471	0.7752
AlexNet	0.8006	0.7690	0.8261	0.7772	0.8375	0.7110	0.6431	0.7615	0.6854	0.7788
GoogLeNet	0.8109	0.7456	0.8334	0.7633	0.8464	0.7273	0.6564	0.7733	0.7034	0.7984
ResNet50	0.8252	0.8061	0.8435	0.8335	0.8512	0.7338	0.6690	0.7834	0.7245	0.8048
MobileNetv3	0.8316	0.8004	0.8623	0.8657	0.9127	0.7403	0.6832	0.7955	0.7512	0.8359
Spasov et al. [15]	0.8318	0.8002	0.8701	0.8673	0.9137	0.7432	0.6823	0.7884	0.7478	0.7918
Zhang et al. [16]	0.8099	0.8063	0.8142	0.7894	0.8323	0.7292	0.6685	0.7587	0.7118	0.7653
Kim et al. [17]	0.8300	0.8148	0.8482	0.8571	0.8926	0.7363	0.6538	0.7724	0.7335	0.7814
Jin et al. [18]	0.8454	0.8170	0.8796	0.8833	0.9216	0.7521	0.6556	0.7916	0.7643	0.8035
Oh et al. [19]	0.8165	0.8028	0.8330	0.7892	0.8531	0.7328	0.6294	0.7649	0.7219	0.7727
ADNet	0.8712	0.8154	0.9308	0.9336	0.9308	0.7962	0.7776	0.8633	0.8291	0.8655

accuracy of the attention module embedded in VGG16 is higher than that of MobileNetv3. Therefore, based on the experimental results in this paper, it can be concluded that VGG16 is more suitable for the purpose of this study.

Secondly, compared with other classification network models, ADNet shows better results in accuracy, sensitivity and specificity. This shows that the improved method proposed in this paper has achieved better results in the early diagnosis of AD. By combining depthwise separable convolution, ELU activation function, SE module and auxiliary tasks, we improve the classification performance of the model, so that it can more accurately identify AD patients.

To sum up, compared with common classification networks, the improved VGG16 proposed in this paper has significant advantages in accuracy, sensitivity and specificity. This result further proves the effectiveness of this method in the early diagnosis of AD. By improving the model structure and feature extraction strategy, we can improve the classification performance of the model, so as to provide more reliable support for the early diagnosis of AD.

4.4.2. Experimental results on different datasets

In order to evaluate the robustness and generalization capability of our proposed method, we conducted experiments using the MIRAD dataset and the AIBL dataset. The detailed information of these datasets is shown in Table 5. The MIRAD dataset consists of data from 46 Alzheimer's disease patients and 23 cognitively normal individuals, while the AIBL dataset comprises data from 116 Alzheimer's disease patients and 122 cognitively normal individuals.

The experimental results are shown in Fig. 12, where it can be seen that our model exhibits significant improvements in

performance metrics such as accuracy (ACC), sensitivity (SEN), and specificity (SPE) compared to other base models. These results further validate the effectiveness of our improved approach, especially in identifying patients with Alzheimer's disease. Our model is able to classify and detect potential Alzheimer's disease cases more accurately, providing strong support for early diagnosis and treatment.

4.5. Limitations and future work

Although ADNet approaches demonstrate strong performance on the ADNI dataset, they are subject to certain constraints. Primarily, the method heavily relies on MRI images, which may not be suitable for all patients, particularly those with implanted metal objects or pacemakers. Additionally, there exist alternative data sources for predicting AD, such as resting-state functional magnetic resonance imaging (rs-fMRI) and positron emission tomography (PET), which offer distinct information that could contribute to a more comprehensive understanding of a patient's condition.

To address these limitations, forthcoming research will concentrate on the following areas:

- (1) Integration of multiple data sources: The amalgamation of data from diverse sources, including MRI, PET, and cognitive assessment, can yield a more holistic comprehension of Alzheimer's disease. The amalgamation of this data may facilitate the development of more precise diagnostic models that consider multiple facets of the disease concurrently.
- (2) Enhancement of the interpretability of AI models: Presently, AI models like ADNet can furnish accurate predictions but lack transparency in their decision-making processes. It is imperative to devise methods that can elucidate model predictions, thereby fostering trust in these models and facilitating their integration into clinical practice.

5. Conclusion

This study introduces a novel approach for classifying Alzheimer's disease, referred to as ADNet. Firstly, in this paper, the depthwise separable convolution is applied in VGG16 network to replace the traditional ordinary convolution, which reduces the parameters of the model and improves the computational efficiency. At the same time, using the ELU activation function instead of the ReLU activation function helps to avoid the problems of gradient disappearance and gradient explosion, and enhances the stability of the model. In addition, this paper introduces the SE attention module, which improves the attention of the model to important features and reduces the interference of irrelevant features by adaptively adjusting the channel weight, so as to effectively improve the classification accuracy. The introduction of this attention mechanism enables the model to better capture the key features in MRI images and provide more accurate classification and discrimination ability. In terms of multi task learning, in addition to the main task of Alzheimer's disease classification, this paper also constructed two auxiliary tasks, namely CDR value regression and MMSE value regression. By sharing feature information, multiple tasks promote each other, which improves the generalization ability and classification accuracy of the model. This design of multi task learning makes the model more comprehensive in understanding and representing the clinical information of patients, which is of great significance for the early diagnosis of Alzheimer's disease. Through a large number of experiments on the public data set ADNI, this method shows excellent diagnostic effect. The experimental results show that the classification accuracy of Alzheimer's disease based on ADNet is significantly better than other common classification networks, which provides strong support for the early diagnosis and treatment of Alzheimer's disease.

Funding statement

This work is supported by the National Key R&D Program of China (2022YFC3303200), and Guangdong province teaching reform project (GDJX2020009).

Availability of data and materials

This study utilized multiple datasets, including the following datasets: (1) ADNI Dataset (Alzheimer's Disease Neuroimaging Initiative): The ADNI dataset comprises a wealth of MRI images and clinical information from Alzheimer's disease patients and cognitively normal individuals. For detailed information about the ADNI dataset, as well as guidance on data acquisition and access, please visit the (www.adni-info.org). (2) MIRIAD Dataset (Alzheimer's Disease Imaging Initial Data): The MIRIAD dataset covers a range of MRI images. To request access to the MIRIAD dataset, please visit the (www.nitrc.org/projects/miriad/). (3) AIBL Dataset (Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing): The AIBL dataset includes MRI images and clinical information

Table 5
MIRIAD and AIBL datasets

Datasets	class	Gender(male/female)	Age	MMSE score
MIRIAD	CN	12/11	70.36 ± 7.28	29.39 ± 0.84
	AD	19/27	69.95 ± 7.08	19.20 ± 4.01
AIBL	CN	60/62	72.74 ± 6.14	28.63 ± 1.62
	AD	58/58	71.27 ± 7.08	20.58 ± 5.21

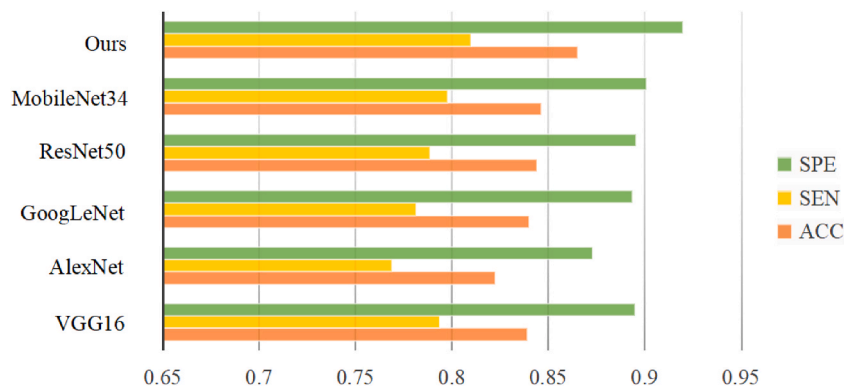


Fig. 12. Classification results of different datasets.

from AD patients and cognitively normal individuals in Australia. For access to the AIBL dataset, please visit the (www.aibl.csiro.au/adni).

CRedit authorship contribution statement

Xin Zhang: Writing – original draft, Software, Methodology. **Le Gao:** Writing – review & editing, Methodology, Conceptualization. **Zhimin Wang:** Validation, Data curation. **Yong Yu:** Visualization, Resources, Conceptualization. **Yudong Zhang:** Validation, Supervision, Conceptualization. **Jin Hong:** Validation, Supervision, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We would like to express our sincere gratitude to all those who have contributed to this work.

References

- [1] B. Winblad, P. Amouyel, S. Andrieu, C. Ballard, C. Brayne, et al., Defeating Alzheimer's disease and other dementias: a priority for European science and society, *Lancet Neurol.* 15 (5) (2016) 455–532.
- [2] J. Gaugler, B. James, T. Johnson, J. Reimer, M. Solis, J. Weuve, et al., "2022 Alzheimer's disease facts and figures," *Alzheimer's Dementia* 18 (4) (2022) 700–789.
- [3] B.C. Dickerson, D.A. Wolk, MRI cortical thickness biomarker predicts AD-like CSF and cognitive decline in normal adults, *Neurology* 72 (2) (2012) 84–90.
- [4] T. Muhammad, T. Ali, M. Ikram, A. Khan, S.I. Alam, et al., Melatonin rescue oxidative stress-mediated neuroinflammation/neurodegeneration and memory impairment in scopolamine-induced amnesia mice model, *J. Neuroimmune Pharmacol.* 1 (14) (2019) 278–294.
- [5] W.J. Lee, Y.W. Shin, H. Chang, H.R. Shin, W.W. Kim, et al., Safety and efficacy of dietary supplement (gintonin-enriched fraction from ginseng) in subjective memory impairment: a randomized placebo-controlled trial, *Integrative Medicine Research* 11 (1) (2022) 1–7.
- [6] X.Y. Feng, F.A. Provenzano, S.A. Small, A deep learning MRI approach outperforms other biomarkers of prodromal Alzheimer's disease, *Alzheimer's Res. Ther.* 14 (1) (2022) 1–11.
- [7] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, et al., Handwritten digit recognition with a back-propagation network, in: *Proc. NIPS*, 1990, pp. 396–404. San Francisco, CA, USA.
- [8] J.C. Morris, The Clinical Dementia Rating (CDR): current version and scoring rules, *Neurology* 1 (43) (1993) 2412–2414.
- [9] S. Ghosh, A.P. Hazarika, A. Chandra, R.K. Mudil, Adaptive neighbor constrained deviation sparse variant fuzzy c-means clustering for brain MRI of AD subject, *Visual Informatics* 5 (4) (2021) 67–80.
- [10] X. Meng, Y. Wu, W. Liu, Y. Wang, Z. Xu, et al., Research on voxel-based features detection and analysis of alzheimer's disease using random survey support vector machine, *Front. Neuroinf.* 16 (5) (2022) 56–68.
- [11] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects, *Neuroimage* 104 (21) (2015) 398–412.
- [12] R. Kumari, A. Nigam, S. Pushkar, An efficient combination of quadruple biomarkers in binary classification using ensemble machine learning technique for early onset of Alzheimer disease, *Neural Comput. Appl.* 34 (14) (2022) 11865–11884.
- [13] S.A. Shah, A. Tahir, J. Ahmad, A. Zahid, H. Pervaiz, S.Y. Shah, et al., Sensor fusion for identification of freezing of gait episodes using Wi-Fi and radar imaging, *IEEE Sensor. J.* 20 (23) (2020) 14410–14422.
- [14] S.F. Abbasi, A. Abbas, I. Ahmad, M.S. Alshehri, S. Almakdi, Y.Y. Ghadi, et al., Automatic neonatal sleep stage classification: a comparative study, *Heliyon* (2023).
- [15] S. Spasov, L. Passamonti, A. Duggento, P. Lio, N. Toschi, A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease, *Neuroimage* 189 (14) (2019) 276–287.
- [16] J. Zhang, X. He, L. Qing, F. Gao, B. Wang, Identifying imaging genetics biomarkers in Alzheimer's disease via integrating graph convolutional neural network and canonical correlation analysis, *Comput. Methods Progr. Biomed.* 217 (2022) 106676.
- [17] M. Kim, X. Yao, A.J. Saykin, J.H. Moore, Q. Long, et al., Identifying imaging genetics biomarkers in Alzheimer's disease via integrating graph convolutional neural network and canonical correlation analysis, *Alzheimer's Dementia* 17 (4) (2021) 36–50.

- [18] D. Jin, J. Xu, K. Zhao, F. Hu, Z. Yang, et al., Identifying imaging genetics biomarkers in Alzheimer's disease via integrating graph convolutional neural network and canonical correlation analysis, in: Proc. ISBI, 2019, pp. 1047–1051. Venice, Italy.
- [19] K. Oh, Y.C. Chung, K.W. Kim, W.S. Kim, I.S. Oh, Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning, *Sci. Rep.* 9 (1) (2019) 1–16.
- [20] S.F. Abbasi, J. Ahmad, A. Tahir, M. Awais, C. Chen, M. Irfan, et al., EEG-based neonatal sleep-wake classification using multilayer perceptron neural network, *IEEE Access* 8 (2020) 183025–183034.
- [21] S.F. Abbasi, Q.H. Abbasi, F. Saeed, N.S. Alghamdi, A convolutional neural network-based decision support system for neonatal quiet sleep detection, *Math. Biosci. Eng.* 20 (9) (2023) 17018–17036.
- [22] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Commun. ACM* 60 (6) (2017) 84–90.
- [23] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proc. ICLR, 2015, pp. 147–163. San Diego, CA, USA.
- [24] C. Szegedy, W. Liu, Y.Q. Jia, P. Sermanet, S. Reed, et al., Going deeper with convolutions, in: Proc. CVPR, 2015, pp. 1–9. Boston, USA.
- [25] K.M. He, X.Y. Zhang, S.Q. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. CVPR, 2016, pp. 770–778. Las Vegas, NV, USA.
- [26] L. Sifre, S. Mallat, Rigid-motion Scattering for Texture Classification [3]Ph.D. dissertation, Ecole Polytechnique, France, 2014.
- [27] D.W. Zhang, Z.L. Zheng, M.L. Li, R.X. Liu, CSART: channel and spatial attention-guided residual learning for real-time object tracking, *Neurocomputing* 436 (14) (2021) 260–272.
- [28] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proc. CVPR, 2018, pp. 7132–7141. Salt Lake City, UT, USA.
- [29] D.W. Zhang, Z.L. Zheng, Joint representation learning with deep quadruplet network for real-time visual tracking, in: Proc. IJCNN, 2020, pp. 1–8. Glasgow, UK.
- [30] T.X. Sun, Y.F. Shao, X.N. Li, P.F. Liu, H. Yan, et al., Learning sparse sharing architectures for multiple tasks, in: Proc. AAAI, 2020, pp. 8936–8943. New York, NY, USA.
- [31] L. Duong, T. Cohn, S. Bird, P. Cook, Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser, in: Proc. ACL-IJCNLP, 2015, pp. 845–850. Beijing, China.
- [32] Y.X. Yang, T.M. Hospedales, Trace norm regularised deep multi-task learning, in: Proc. ICLR, 2017, pp. 845–850. Toulon, France.
- [33] A.S. Correia, E.L. Colombini, Attention, please! A survey of neural attention models in deep learning, *Artif. Intell. Rev.* 55 (8) (2022) 6037–6124.
- [34] R. Caruana, Algorithms and applications for multitask learning, in: Proc. ICML, 1996, pp. 87–95. Bari, Italy.
- [35] A. Kendall, Y. Gal, R. Cipolla, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: Proc. CVPR, 2018, pp. 7482–7491. Salt Lake City, UT, USA.
- [36] Z. Chen, V. Badrinarayanan, C.Y. Lee, A. Rabinovich, radnorm: gradient normalization for adaptive loss balancing in deep multitask networks, in: Proc. ICML, 2018, pp. 794–803. Stockholm, Sweden.
- [37] N.N. Chu, H. Gebre-Amlak, Navigating neuroimaging datasets ADNI for Alzheimer's disease, *IEEE Consumer Electronics Magazine* 10 (5) (2021) 61–63.
- [38] Y. Inokuchi, An alternative clue to set axial angle parallel to the AC-PC on brain perfusion SPECT imaging: usefulness of frontal lobe bottom and cerebellum tuber vermis line, *J. Nucl. Med.* 59 (1) (2018), 1849–1849.