

Diagnosis of Alzheimer's disease using structure highlighting key slice stacking and transfer learning

Yanjiao Ban¹ | Xuejun Zhang^{1,2} | Huan Lao³

¹School of Computer, Electronics and Information, Guangxi University, Nanning, Guangxi, P. R. China

²Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning, Guangxi, P. R. China

³School of Artificial Intelligence, Guangxi Minzu University, Nanning, Guangxi, P. R. China

Correspondence

Xuejun Zhang, School of Computer, Electronics and Information, Guangxi University, 100, Daxue Dong Road, Nanning 530004, Guangxi, P. R. China.
Email: xjzhang@gxu.edu.cn

Huan Lao, School of Artificial Intelligence, Guangxi Minzu University, 188, Daxue Dong Road, Nanning 530006, Guangxi, P. R. China.
Email: laohuanlh@st.gxu.edu.cn

Funding information

Guangxi Key Projects of Science and Technology, Grant/Award Number: 2020AA2177007; Innovation Project of Guangxi Graduate Education, Grant/Award Number: YCBZ2020026

Abstract

Background: In recent years, two-dimensional convolutional neural network (2D CNN) have been widely used in the diagnosis of Alzheimer's disease (AD) based on structural magnetic resonance imaging (sMRI). However, due to the lack of targeted processing of the key slices of sMRI images, the classification performance of the CNN model needs to be improved.

Purpose: Therefore, in this paper, we propose a key slice processing technique called the structural highlighting key slice stacking (SHKSS) technique, and we apply it to a 2D transfer learning model for AD classification.

Methods: Specifically, first, 3D MR images were preprocessed. Second, the 2D axial middle-layer image was extracted from the MR image as a key slice. Then, the image was normalized by intensity and mapped to the red, green, and blue (RGB) space, and histogram specification was performed on the obtained RGB image to generate the final three-channel image. The final three-channel image was input into a pretrained CNN model for AD classification. Finally, classification and generalization experiments were conducted to verify the validity of the proposed method.

Results: The experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set show that our SHKSS method can effectively highlight the structural information in MRI slices. Compared with existing key slice processing techniques, our SHKSS method has an average accuracy improvement of at least 26% on the same test data set, and it has better performance and generalization ability.

Conclusions: Our SHKSS method not only converts single-channel images into three-channel images to match the input requirements of the 2D transfer learning model but also highlights the structural information of MRI slices to improve the accuracy of AD diagnosis.

KEYWORDS

Alzheimer's disease (AD), structural magnetic resonance imaging (sMRI), structure highlighting key slice stacking (SHKSS), transfer learning, two-dimensional convolutional neural network (2D CNN)

1 | INTRODUCTION

Alzheimer's disease (AD) is the most common neurodegenerative disease and is clinically manifested by memory impairment and other cognitive decline.¹ As of 2018, there were 50 million AD patients worldwide, and this number is expected to reach 152 million in 2050.² In addition, the cost of AD treatment is very high, and related surveys show that the cost of AD treatment

has increased significantly year by year and will reach between 379 and 500 billion in 2040.³ Mild cognitive impairment (MCI) is a stage between normal control (NC) and AD. Many studies have demonstrated that appropriate interventions for MCI patients can delay the progression of AD.^{4–6} Therefore, early diagnosis of AD is crucial.

Early diagnosis of AD can be achieved by automatic analysis of structural magnetic resonance imaging

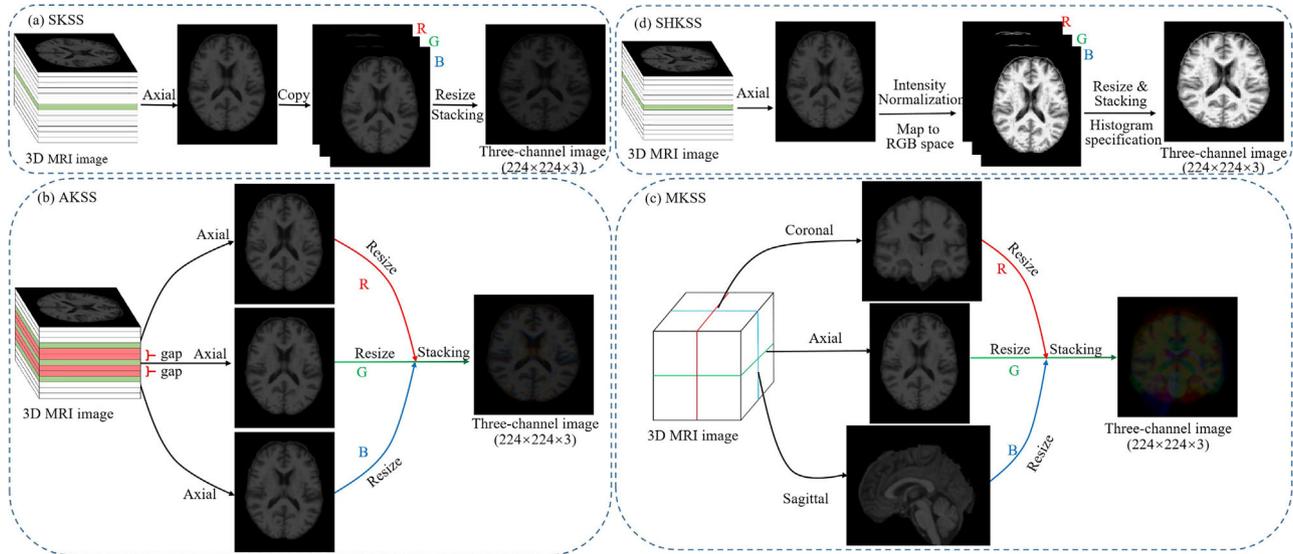


FIGURE 1 Schematic diagram of common key slice pretreatment methods and our proposed key slice pretreatment method: (a) single key slice stacking (SKSS) method, (b) adjacent key slice stacking (AKSS) method, (c) multikey slice stacking (MKSS) method, and (d) structure highlighting key slice stacking (SHKSS) method

(sMRI) through deep learning.^{7–9} As a structural imaging method, sMRI can clearly display structural changes in the brain to physicians, so it is widely used in the clinical evaluation of AD patients. Existing studies have shown that convolutional neural network (CNN) support images as input to network models for higher level image feature extraction, and the capture of subtle lesion sites better than traditional manual feature extraction methods in AD classification studies based on sMRI images.¹⁰ Currently, CNN has been applied in the study of AD classification based on sMRI images. For example, Korolev et al.¹¹ achieved AD versus NC classification by using a CNN for feature extraction of sMRI images, and their classification accuracy was 79%. Khvostikov et al.¹² proposed an adaptive CNN for training on sMRI images for AD classification and achieved accuracies of 96.7%, 80%, and 65.8% in AD versus NC, AD versus MCI, and NC versus MCI classification, respectively. A similar study was performed by Liu et al.¹³ They combined two CNNs to classify AD. However, due to the small number of samples of medical images,¹⁴ the difficulty of training CNN from scratch, and the difficulty of parameter optimization, it is still a difficult task to obtain a highly accurate AD diagnosis model. To overcome these problems, Yosinski et al.¹⁵ proposed a transfer learning method in 2014, which not only effectively avoids overfitting but also captures the internal connections among data well. The essence of transfer learning is to transfer the trained model parameters to the new model to assist in training the new model according to the correlation among objects in nature. Most of the existing publicly available transfer learning models are 2D CNN, and most of the transfer learning models commonly used for AD classification are 2D CNN. Existing

pretrained 2D CNN models require the input image to be a three-channel image, so 2D slice images containing regions of interest (e.g., the hippocampus or ventricles) need to be extracted from 3D sMRI images and converted to three-channel images using key slice extraction techniques.

There are three key slice processing techniques commonly used for AD classification tasks. The first is single key slice stacking (SKSS).^{16–18} This method is the most commonly used. First, the key slice is extracted and copied twice, and then the three slices are stacked to generate a three-channel image (Figure 1a). The second technique is adjacent key slice stacking (AKSS).^{19–21} This method is also widely used in key slice processing. First, the key slice and its adjacent slices in the gap above and below (the gap is generally greater than or equal to 0 and less than or equal to 1/2 of the image size) are extracted, and then they are stacked to generate a three-channel image (Figure 1b). Compared with the SKSS method, the images synthesized by this method can contain more spatial information and have a certain theoretical robustness. The last technique is multikey slice stacking (MKSS).^{22,23} The 3D MRI images can be viewed in three visualizations, which are perpendicular to three standard image axes: axial, coronal, and sagittal. This method extracts key slices from the axial, coronal, and sagittal planes. Then, the three key slices are stacked to obtain three-channel images (Figure 1c). Compared with the SKSS and AKSS methods, this method is less frequently used, but its synthesized images contain information from the three planes of MRI images, which plays a role in information supplementation to some extent. However, the above three methods only generate three-channel images

through simple slice stacking, which does not fully use the structure of MRI images. Previous studies^{24–27} have shown that intensity normalization of MRI images can enhance the discriminative information in the images, and histogram specification^{28,29} is effective in enhancing MRI images to highlight image differences and ensure the consistency of image intensity. Based on this, this study proposes a method including intensity normalization and histogram specification to the training process of a CNN for AD diagnosis, which is called structural highlighting key slice stacking (SHKSS). This method first normalizes the intensity of the image, then maps the single-channel image into the RGB space, and finally performs histogram specification on the obtained RGB image to generate the final three-channel image (Figure 1d). To evaluate the performance of these methods, two experiments were conducted in this study on the Alzheimer’s Disease Neuroimaging Initiative (ADNI) data set using pretrained ResNet-18 as a classification model: (1) The effectiveness of these methods was evaluated. The SKSS method, the AKSS method, the MKSS method, and the proposed SHKSS method were applied to the ADNI-1 data set to compare the performance of the four methods in AD classification tasks. (2) The generalization ability of these methods was evaluated. The above four methods were applied in four additional sets of experiments: (a) the models were trained and tested on the ADNI-2 data set, (b) the models trained on the ADNI-1 data set and tested on the ADNI-2 data set, (c) the models trained on the ADNI-2 data set and tested on the ADNI-1 data set, and (d) a mixture of ADNI-1 and ADNI-2 data set were used for training and testing. The experimental results show that our proposed method can effectively identify AD, and its diagnostic performance is better than those of existing methods under the same conditions.

The rest of this paper is organized as follows: In Section 2, we briefly introduce the studied data sets. In Section 3, we describe in detail the overall framework of this study, including key slice processing techniques and transfer learning methods. In Section 4, our proposed method is evaluated and compared with existing methods. In Section 5, we first visualize and analyze the results of the SHKSS method with three commonly used methods and then discuss the effects of transfer learning methods and data leakage on the experimental results. In Section 6, we analyze the limitations of this study. Finally, we conclude in Section 7.

2 | MATERIALS AND MRI PREPROCESSING

2.1 | ADNI data set

The sMRI data used in this study were obtained from the ADNI database (www.adni.loni.usc.edu), which includes

TABLE 1 Characteristics of the subjects used in this study. Values are reported as mean \pm standard deviation

	Diagnosed	Gender (M/F)	Age (Mean \pm SD)	MMSE (Mean \pm SD)
ADNI-1	AD	83/76	75.28 \pm 7.59	23.37 \pm 1.59
	MCI	173/99	76.07 \pm 5.41	29.05 \pm 1.08
	NC	76/76	74.80 \pm 7.15	27.04 \pm 1.76
ADNI-2	AD	83/61	74.56 \pm 8.16	23.06 \pm 2.10
	MCI	117/86	73.50 \pm 6.27	29.04 \pm 1.22
	NC	87/94	71.79 \pm 6.96	27.87 \pm 1.70

the ADNI-1 and ADNI-2 phases. ADNI is a publicly available AD clinical imaging database that was established in 2003, and it is convenient for researchers and clinicians, effectively reducing unnecessary time and cost in the development of new treatments. In this study, MR T1-weighted images of 583 subjects in the ADNI-1 and 528 subjects in the ADNI-2 phase were selected for the experiment. The ADNI-1 data set contained 159 AD subjects, 272 MCI subjects, and 152 NC subjects, and the ADNI-2 data set contained 144 AD subjects, 203 MCI subjects, and 181 NC subjects. These MR T1-weighted images were preprocessed with specific image preprocessing steps by the ADNI study group, including multiplanar reconstruction (MPR), gradient warping, B1 nonuniformity correction, and N3 intensity normalization. The details of all the subjects in this study are shown in Table 1.

2.2 | Image preprocessing

To reduce the anatomical differences between individuals, the “Segment Data” module provided by the computational anatomy toolbox (CAT12)³⁰ can be used to preprocess the MRI before the experiment (CAT12 is found at neuro.uni-jena.de/cat/). CAT12 is a MATLAB toolkit based on the statistical parametric mapping (SPM)³¹ software package (SPM is found at fil.ion.ucl.ac.uk/spm/), which was jointly developed by Dr. Christian Gaser and Dr. Robert Dahnke of the Department of Psychiatry and Neurology of Jena University Hospital, Germany. The preprocessing includes three steps, as shown in Figure 2. First, skull stripping is performed on original 3D MRI images to produce a brain volume with less noise and irrelevant information (called SS-MRI). Second, all SS-MRI images are registered to the MNI space (MNI152 T1 1.5 mm brain) via Dartel registration to achieve spatial standardization. Then, the standardized SS-MRI images are modulated to compensate for the effects of spatial standardization. Finally, a whole-brain image with a size of 121 \times 145 \times 121 (called MNI-MRI) is obtained.

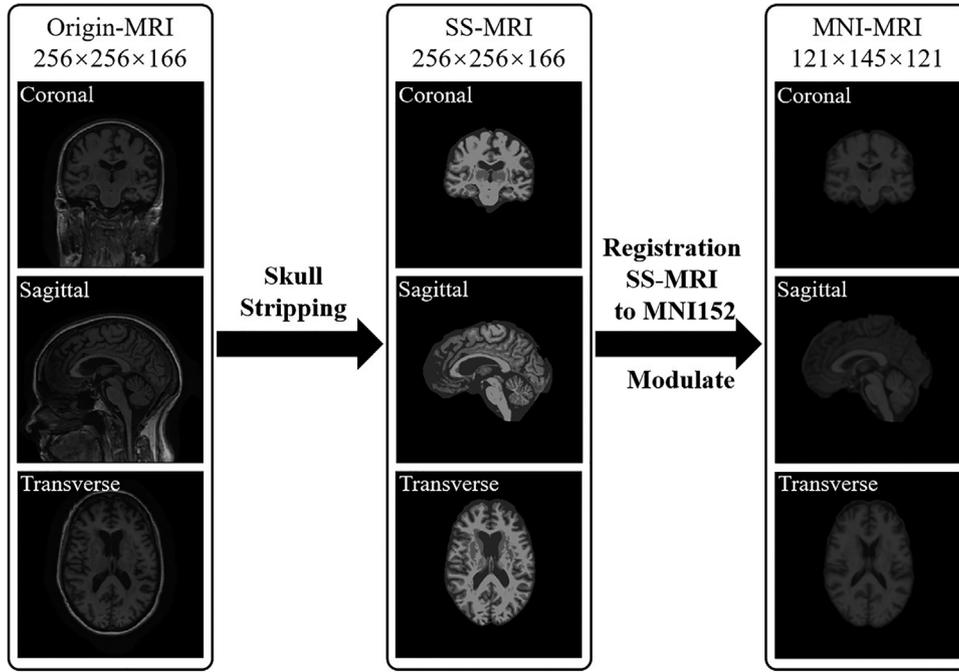


FIGURE 2 Flowchart of sMRI image preprocessing

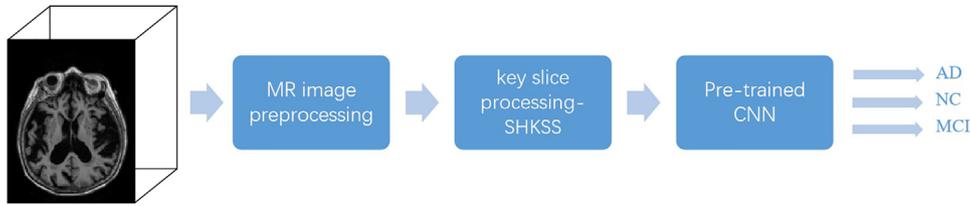


FIGURE 3 Block diagram of our method

3 | METHODOLOGY

The overall flowchart of this study is shown in Figure 3, which consists of three main steps: image preprocessing, key slice processing, and CNN classification. Specifically, first, the 3D MR images were preprocessed (see Section 2.1 for the specific steps). Second, the 2D axial middle layer image was extracted from the 3D brain volume as a key slice, and the image was normalized by intensity and mapped to the RGB space. Finally, the processed key slice was input to a pretrained CNN model for classification. In what follows, we describe the proposed method.

3.1 | Key slice processing technology

In this study, we use a pretrained 2D CNN model to perform classification. Therefore, 2D slice images need to be transformed into three-channel images using key slice processing techniques. That is, the key slice

image is mapped to the RGB space to obtain a three-channel image that meets the input requirements of the pretrained CNN. The key slice processing technique proposed in this study, the SHKSS method, can be divided into three steps.

(1) A three-channel color mapping table M is constructed: First, we generate a single-channel color mapping table X based on the grayscale difference of the image. Assuming that there is a grayscale image I (with a size of $m \times n$), I_{max} represents the maximum grayscale value, and I_{min} represents the minimum grayscale value, its grayscale difference value can be expressed as:

$$d = |I_{max} - I_{min}| + 1. \quad (1)$$

The single-channel color mapping table X can be expressed as:

$$X = [0, x_i, \dots, 1]^T, \quad (2)$$

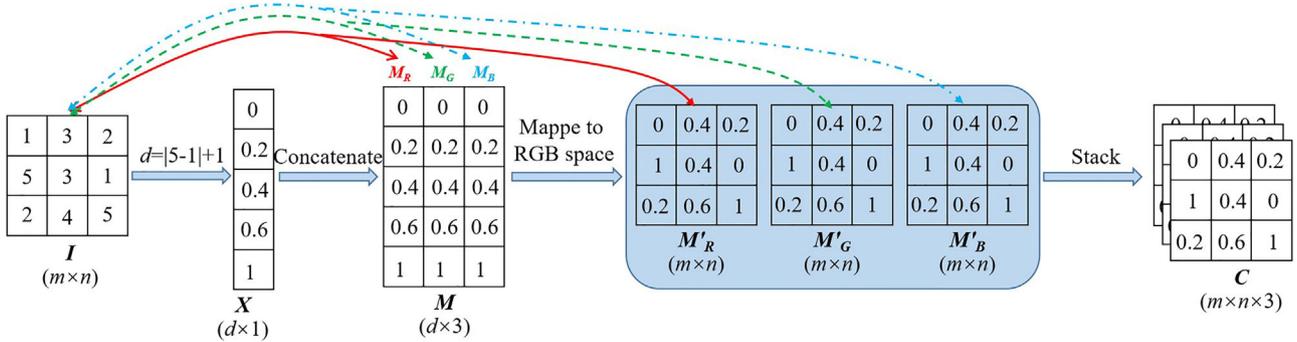


FIGURE 4 Generating a three-channel image

$$x_i = \frac{i-1}{d}, i = 2, \dots, d-1, \quad (3)$$

where X is a matrix of size $d \times 1$ and its values are in the range $[0, 1]$.

Second, X is concatenated to obtain a three-channel color mapping table M . Thus, M is a matrix of size $d \times 3$, and each column of M corresponds to a color channel; the channels are denoted as M_R , M_G , and M_B .

(2) The grayscale image I is mapped to the RGB space to generate a three-channel image C . Specifically, first, the values of each pixel in images M'_R , M'_G , and M'_B are obtained by extracting the corresponding values in M_R , M_G , and M_B using the pixel values in I as the index, and the size is $m \times n$. Thus far, the normalized images M'_R , M'_G , and M'_B have been obtained. Second, the images M'_R , M'_G , and M'_B are superimposed to obtain the desired three-channel image C , and the size is $m \times n \times 3$. The specific process is shown in Figure 4.

(3) Histogram specification^{29,32} is performed on the three-channel image C to enhance the image contrast and highlight the structural features of the image. Specifically, first, the histogram of the three-channel image C is calculated as follows:

$$h(k) = n_k, k = 0, 1, 2, \dots, L_1 - 1, \quad (4)$$

where n_k is the number of pixels at gray level k and L_1 is the total number of gray levels of image C , normalized by:

$$p_s(s_k) = \frac{h(k)}{N}, \quad (5)$$

where s_k is the k -th level gray value of the image and N is the total number of pixels. The cumulative distribution of the histogram of image C is:

$$t_k = \sum_{t=0}^k p_s(s_t). \quad (6)$$

Next, the average histogram of the three-channel images of the K key slices is used as the target his-

toграм z , and its cumulative distribution is obtained. The calculation formula is as follows:

$$z(l) = \frac{1}{K} \sum_{i=1}^K h_i(l), l = 0, 1, 2, \dots, L_2 - 1, \quad (7)$$

$$p_u(u_l) = \frac{z(l)}{N}, \quad (8)$$

$$v_l = \sum_{j=0}^l p_u(u_j), \quad (9)$$

where L_2 is the total number of gray levels of the target histogram z .

Finally, a single mapping law (SML) is used to establish the mapping relationship between t_k and v_l :

$$\left| \sum_{t=0}^k p_s(s_t) - \sum_{j=0}^l p_u(u_j) \right|. \quad (10)$$

When formula (10) takes the minimum value, the u value that matches the s value can be found, that is, the mapping from s to u is obtained. After completing the histogram specification of the three-channel image C , the final three-channel image can be obtained.

Figure 5 shows the three-channel images generated by the SKSS method, the AKSS method, the MKSS method, and the proposed SHKSS method. Compared with those of the other three methods, the three-channel image generated by SHKSS is brighter, and the contrast between tissues is stronger.

3.2 | Transfer learning

The small number of samples in this study is not sufficient to train 2D CNN models with high accuracy from scratch, and overfitting may occur during training. Therefore, we use the transfer learning technique with pretrained CNN models for the AD classification task. Considering that deeper networks may reduce the performance of the model and complicate the optimization process,^{33,34} this study uses a pretrained ResNet-18

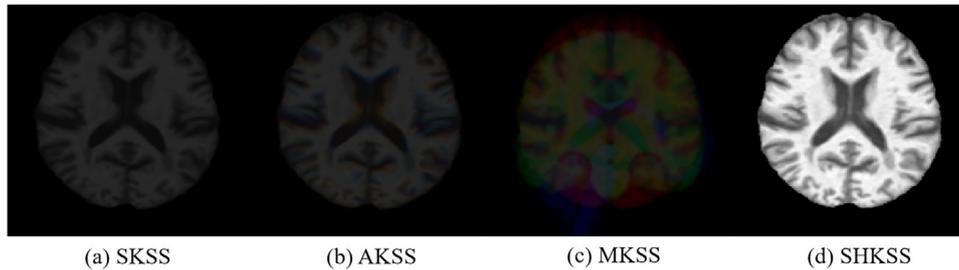


FIGURE 5 Three-channel images processed by (a) the SKSS method, (b) the AKSS method, (c) the MKSS method, and (d) the proposed SHKSS method

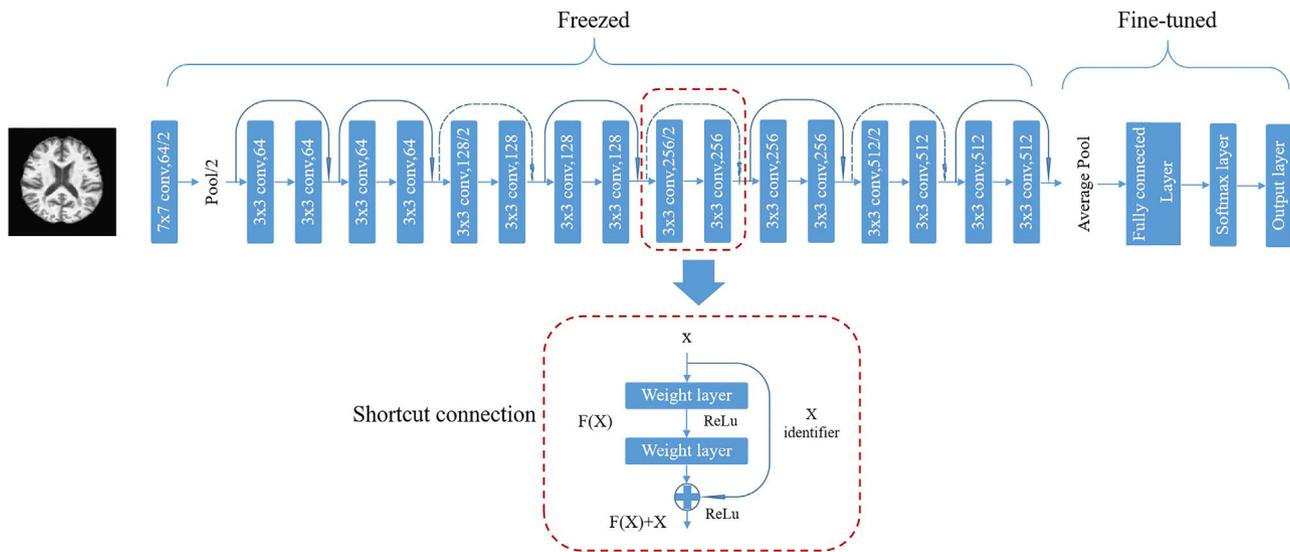


FIGURE 6 Our ResNet-18 model

model as the feature extractor and classifier. The ResNet network was proposed by He et al.³³ in 2015, solving the problem of gradient disappearance and degradation by skipping some layers (the red dashed box at the bottom of Figure 6), called “shortcut connections.” At present, the reliability of ResNet-18 has been verified on the ImageNet database, and it learns rich features representing a wide range of images. Therefore, ResNet-18 can be extended to our AD diagnosis tasks.

In this study, the final classification is achieved using a pretrained ResNet-18 model, which contains 18 layers with learnable weights (sums of convolutional and fully connected layers from the input layer to the output layer). To allow the original ResNet-18 model that identified 1000 classes to train the data used in this study, the last layer with learnable weights (fully connected layer) needs to be replaced with a new fully connected layer. That is, the number of neurons in the new fully connected layer should be the number of classes of the samples in the classification experiment. For example, the number of neurons in the binary clas-

sification experiment is 2, and the number of neurons in the multiclassification experiment is 3. In addition, a new softmax layer and a classification output layer are needed to replace the corresponding original layers. The specific structure of the model is shown in Figure 6. Each convolutional layer in Figure 6 is represented with the size of the convolution kernel and the number of channels. The weights of the layers indicated by “Freezed” are not updated during backpropagation, and the weights of the layers indicated by “Fine-tuned” are updated during backpropagation. A solid-line shortcut means that a “shortcut connection” can be used directly when the dimensions of the input and output are the same. A dotted-line shortcut means that a “shortcut connection” can be used directly when the dimension increases. Therefore, during training, all layers before the fully connected layer are frozen to prevent overfitting, and the fully connected layer and its subsequent layers (i.e., the fully connected layer, softmax layer, and output layer) are randomly initialized and retrained to fit the AD classification task.

4 | EXPERIMENT AND ANALYSIS

4.1 | Experimental setup

In our experiments, we used a fine-tuning approach to modify the last three layers (i.e., the fully connected layer, softmax layer, and output layer) of the ResNet-18 model. The weights of the other layers of the model were frozen during fine-tuning to prevent overfitting. The training model parameters were set as follows: Gradient descent was performed by stochastic gradient descent with momentum optimizer (SGDWMO), where the momentum value was set to 0.9000; the minimum batch size was 20; the maximum number of iterations was 40; and the initial learning rate was set to 0.003. The learning rate was kept constant throughout the model training process. To ensure a fair comparison, the same parameter configurations were used for all four key slice techniques.

In this study, we conducted classification experiments and generalizability experiments. In the classification experiments, the models were trained and tested on the ADNI-1 data set; there were three binary classification experiments (AD vs. NC, NC vs. MCI, and AD vs. MCI) and one multiclassification experiment (AD vs. NC vs. MCI). The performance metrics for all classifications included classification accuracy (ACC) (Equation 11, where TP , TN , FP , and FN denote the true positives, true negatives, false positives, and false negatives, respectively), sensitivity (SEN) (Equation 12), specificity (SPE) (Equation 13), and area under the receiver operating characteristic curve (AUC). In general, a method has low classification performance at AUC values of 50%–70%; at 70%–90%, the classification performance can be considered moderate; if the AUC is greater than 90%, the classification performance is considered high. In addition, we applied a fivefold cross-validation strategy in our experiments to split the data into training and test sets. In the generalizability experiments, we performed four sets of experiments: (1) The models were trained and tested on the ADNI-2 data set alone, (2) models trained using the ADNI-1 data set were used to predict the ADNI-2 data, (3) models trained using the ADNI-2 data set were used to predict the ADNI-1 data, and (4) a mixture of ADNI-1 and ADNI-2 data were used for training and testing. The reliability of the proposed method is demonstrated by comparing the results of the classification experiments with those of the generalization experiments. In particular, the trained models and parameters in the generalization experiments were kept the same as in the classification experiments, and testing was performed on AD versus NC, NC versus MCI, AD versus MCI, and AD versus NC versus MCI.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \quad (11)$$

$$SEN = \frac{TP}{TP + FN}, \quad (12)$$

$$SPE = \frac{TN}{TN + FP}. \quad (13)$$

4.2 | Classification results

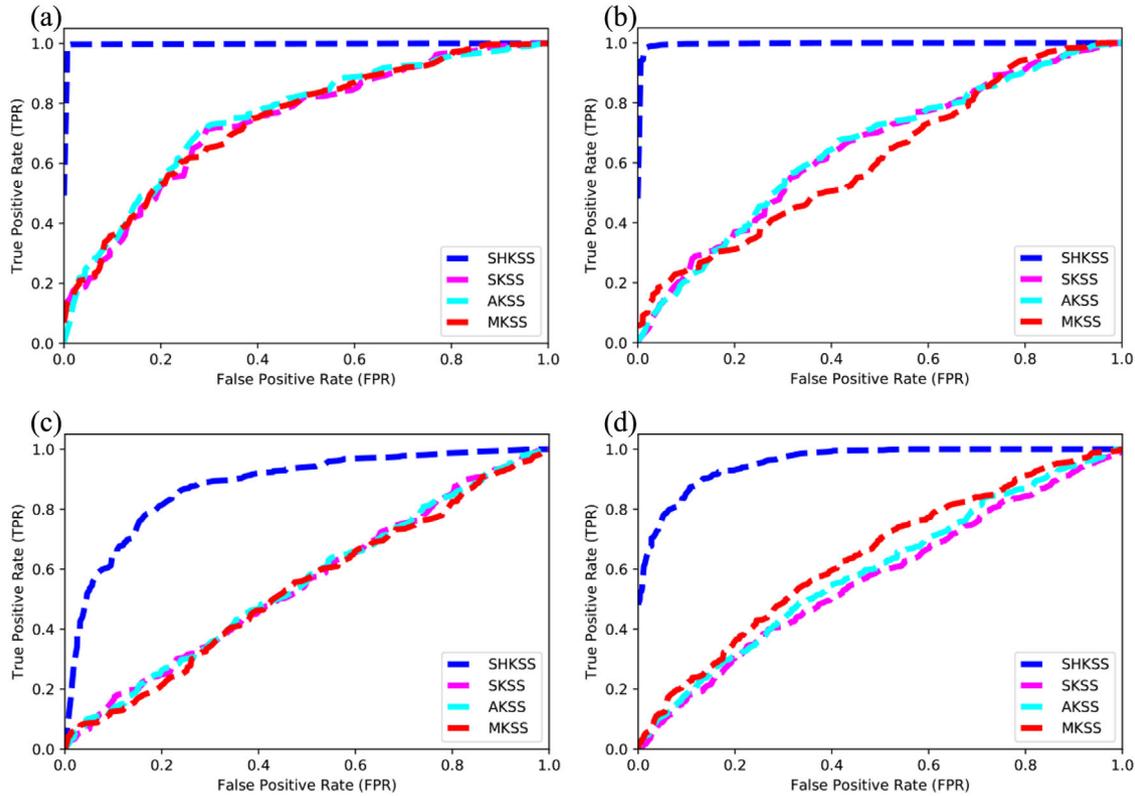
In this experiment, we compared the classification results of three commonly used key slice processing techniques (the SKSS method, AKSS method, and MKSS method) and the proposed SHKSS method. Table 2 shows the classification results of these methods for the AD versus NC, NC versus MCI, AD versus MCI, and AD versus NC versus MCI classification tasks. The ROC curves of these methods are also compared in Figure 7a–d. From Table 2 and Figure 7, it can be observed that (1) the classification performance of the proposed method is better than those of the three commonly used methods. Specifically, for AD and NC classification, the ACC of the SHKSS method was improved by at least 28%. In the NC versus MCI and AD versus MCI classification tasks, the ACC of the SHKSS method was improved by at least 26% and 29%, respectively. For AD versus NC versus MCI classification, the ACC of the SHKSS method was improved by at least 38%. In summary, the SHKSS method improved the accuracy by at least 26% over the commonly used methods, and the AUC values of the SHKSS method were all greater than 88%, indicating that the proposed method has good classification performance. (2) Our method performs well on small sample data. Specifically, without database augmentation, our method achieved ACCs of 98.77% and 98.60% for AD versus NC and AD versus MCI tasks, respectively. This indicates that the proposed method can accurately distinguish AD patients from NC and MCI patients. In the classification of NC and MCI, our method achieved an ACC of 81.63%, SEN of 86.05%, SPE of 77.95%, and AUC of 88%, which were much greater than those of the three common methods, indicating that the proposed method can distinguish NC and MCI more efficiently. (3) The proposed method has great potential for multiclassification task; the accuracy of multiclassification was 82.82%, and the SEN, SPE, and AUC values were greater than 80%.

4.3 | Validation of the generalization ability of the model

In the above experiments, the classification network was trained and tested on the ADNI-1 data set. To evaluate and compare the generalization abilities of the proposed method and the other three methods, we performed four sets of experiments. (1) The classification network was trained and tested on the ADNI-2 data set, and the

TABLE 2 The classification results of four methods for different classification tasks. Values are reported as mean (%)

Methods	AD vs. NC				AD vs. MCI				NC vs. MCI				AD vs. NC vs. MCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SKSS	70.35	43.16	86.77	73.30	67.24	1.05	99.48	64.32	55.54	34.82	72.31	54.67	42.66	42.26	69.91	58.01
AKSS	70.33	60.00	76.61	74.84	66.55	50.52	74.36	64.46	53.28	35.26	67.69	54.81	43.54	44.51	70.91	64.69
MKSS	66.84	66.32	67.28	73.70	68.79	16.84	94.36	60.97	55.26	47.28	61.54	53.35	44.43	50.42	72.72	67.05
SHKSS	98.77	100	97.24	99.00	98.60	98.89	98.06	99.00	81.63	86.05	77.95	88.00	82.82	83.60	90.73	95.71

**FIGURE 7** Comparison of the ROC curves of four methods for classifying: (a) AD vs. NC classification, (b) AD vs. MCI classification, (c) NC vs. MCI classification, and (d) AD vs. NC vs. MCI classification**TABLE 3** Classification results with fivefold cross-validation on the ADNI-2 data set. Values are reported as mean (%)

Methods	AD vs. NC				AD vs. MCI				NC vs. MCI				AD vs. NC vs. MCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SKSS	69.54	59.68	77.34	77.25	68.59	21.53	89.52	72.95	62.57	87.64	17.61	56.40	44.68	36.58	67.70	60.52
AKSS	74.77	66.53	81.21	83.78	71.58	35.17	87.63	75.48	57.43	71.88	31.41	51.81	44.52	38.35	68.05	61.26
MKSS	68.00	68.25	68.11	78.58	67.73	38.47	80.89	71.59	56.83	74.77	25.00	51.19	44.99	35.64	66.99	56.96
SHKSS	98.46	98.02	98.85	99.00	97.11	96.10	99.03	99.00	81.25	86.71	77.01	80.00	82.58	84.39	90.98	98.42

results are shown in Table 3. As seen from Table 3, the classification performance of the proposed method on this data set was also better than those of the three commonly used methods. From Table 2 and Table 3, we can observe that all four methods have comparable AD diagnosis capabilities on ADNI-1 and ADNI-2. The maximum difference between the classification results of the three commonly used methods is approximately

5% (AKSS: AD vs. MCI). However, the maximum difference in the classification results of our SHKSS method is less than 2%. Specifically, the difference in the classification results is 0.31% for AD versus NC, 0.38% for NC versus MCI, 1.49% for AD versus MCI, and 0.24% for AD versus NC versus MCI. This experiment shows that the proposed experimental method does not depend on a specific data set. We also find that the classification

TABLE 4 Results for AD classification on the ADNI-2 data set, using ADNI-1 as the training set. Values are reported as mean (%)

Methods	AD vs. NC				AD vs. MCI				NC vs. MCI				AD vs. NC vs. MCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SKSS	68.18	65.42	70.39	75.64	64.87	32.92	79.07	66.89	57.74	69.01	37.57	55.40	42.68	40.26	69.07	59.86
AKSS	68.12	67.36	68.73	75.97	68.68	23.61	88.70	67.69	60.48	78.21	28.73	57.41	42.87	43.84	70.38	61.97
MKSS	70.65	63.75	76.13	77.15	67.35	34.03	82.16	66.11	61.98	90.68	10.61	54.12	43.27	41.77	69.70	61.10
SHKSS	98.58	100	96.81	100	98.20	99.88	94.44	99.00	71.41	96.57	48.97	95.00	71.89	74.23	85.61	94.23

TABLE 5 Results for AD classification on the ADNI-1 data set, using ADNI-2 as the training set. Values are reported as mean (%)

Methods	AD vs. NC				AD vs. MCI				NC vs. MCI				AD vs. NC vs. MCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SKSS	66.17	56.86	75.92	74.63	62.23	32.96	79.34	63.19	60.90	82.94	21.45	56.86	45.21	35.63	67.60	58.28
AKSS	68.94	62.77	75.39	76.40	59.68	44.40	68.60	60.39	61.84	78.46	32.11	59.90	43.57	39.31	68.83	57.66
MKSS	65.79	47.17	85.26	77.26	60.46	28.68	79.04	59.80	62.17	87.13	17.50	58.55	43.36	37.98	68.45	57.01
SHKSS	96.59	100	94.08	99.00	96.79	100	95.23	99.00	75.24	77.68	87.73	90.00	71.01	76.18	85.88	95.57

results on the ADNI-2 data set are slightly lower than those on the ADNI-1 data set, which may be because the number of ADNI-1 subjects is larger than the number of ADNI-2 subjects.

(2) The model was trained on the ADNI-1 data set and tested on the ADNI-2 data set. As above, the classification results of the SHKSS method are far superior to those of the other three methods, and the results are shown in Table 4. For the SHKSS method, by comparing Tables 2 and 4, we can observe that the models trained on the ADNI-1 data set can better classify the data from the ADNI-2 data set. The difference between their classification results and those of the models trained and tested on ADNI-1 is within an acceptable range of approximately 10%. For our SHKSS method, the difference between the classification results for AD versus NC is 0.19%, that of NC versus MCI is 10.22%, that of AD versus MCI is 0.4%, and that of AD versus NC versus MCI is 10.93%.

(3) The model was trained on the ADNI-2 data set and tested on the ADNI-1 data set; the results are shown in Table 5. From Table 5, we can observe that for the SHKSS method, the results are better than those of the other three methods, and the model trained on the ADNI-2 data set can better classify the data in the ADNI-1 data set. Specifically, by comparing Table 3 and Table 5, the difference between the classification results for AD versus NC is 1.87%, that of NC versus MCI is 6.01%, that of AD versus MCI is 0.32%, and that of AD versus NC versus MCI is 11.57%. By comparing Table 4 and Table 5, we find that the ADNI-1 and ADNI-2 data sets have comparable mutual diagnostic capabilities when they are used as each other's training and test sets. In the SHKSS method, the difference between the classification results of AD versus NC is 1.99%, that of NC versus MCI is 3.83%, that of AD versus MCI is 1.41%, and that of AD versus NC versus MCI is 0.88%.

(4) The classification network was trained and tested using mixed data from ADNI-1 and ADNI-2; the results are shown in Table 6. We can see that the classification performance of the proposed method is also better than those of the three commonly used methods. Moreover, by comparing the results with those of the previous generalization experiments, it can be seen that the classification results using mixed data do not differ greatly from those obtained by training and testing on the ADNI-1 (or ADNI-2) data set (Tables 2 and 3), and both outperform the classification results for training on the ADNI-1 data set and testing on the ADNI-2 data set (Table 4) as well as the classification results for training on the ADNI-2 data set and testing on the ADNI-1 data set (Table 5). In conclusion, our proposed method has good generalization ability in sMRI-based AD diagnosis tasks.

4.4 | Comparison with state-of-the-art methods

In Table 7, the results of our method are compared with those of several other methods. To facilitate the comparison, all the methods used data from the ADNI data set. However, these methods were performed for different numbers of subjects and different partitions of the training and test samples. Although the image data selected for all methods were not identical, quality control and preprocessing of the image data were performed for the ADNI study group. Therefore, although the results in Table 7 may not be fully comparable, we can roughly compare our method with these state-of-the-art methods to verify the efficacy of our proposed method. From Table 7, we can observe that the proposed method outperforms most existing methods for AD versus NC and AD versus MCI. Specifically, our method obtains a

TABLE 6 Classification results with fivefold cross-validation on the ADNI-1+ADNI-2 data set. Values are reported as mean (%)

Methods	AD vs. NC				AD vs. MCI				NC vs. MCI				AD vs. NC vs. MCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SKSS	71.55	55.78	85.87	78.75	68.52	30.68	87.74	68.75	63.40	92.13	12.06	59.42	46.19	42.78	70.70	61.25
AKSS	75.32	71.25	78.97	81.76	68.41	36.93	84.40	68.76	62.21	82.07	26.79	58.85	45.21	40.72	69.66	60.86
MKSS	71.23	56.53	84.71	79.38	68.97	39.93	83.73	69.25	62.54	83.7	24.88	59.87	44.48	42.18	70.00	61.20
SHKSS	99.69	99.37	100	100	99.23	99.67	98.89	99.00	82.54	84.35	79.93	89.00	82.70	86.05	91.66	99.03

TABLE 7 Comparison with state-of-the-art methods

Method	Data (AD/NC/MCI)	Classification results (%)			
		AD vs. NC	NC vs. MCI	AD vs. MCI	AD vs. NC vs. MCI
Korolev et al. ¹¹	ADNI-1:50/61/120	79.00	-	-	-
Lian et al. ³⁷	ADNI-1:199/229/393 ADNI-2:159/200/277	90.30	-	-	-
Khvostikov et al. ¹²	ADNI-1:48/58/108	85.40	65.80	76.00	-
Liu et al. ¹³	ADNI-1:97/119/233	88.90	76.20	-	-
Liu et al. ³⁶	ADNI-1:93/100/204	84.97	-	-	-
Shi et al. ³⁵	ADNI-1:51/52/99	95.44	83.29	-	-
Proposed method	ADNI-1:159/272/152 ADNI-2:144/203/181	98.77	81.63	98.60	82.82

The boldface denotes the best performance in each classification task.

classification accuracy of 98.77% for AD versus NC, which is an improvement of 3.33% (vs. Shi et al.³⁵), 13.8% (vs. Liu et al.³⁶), 9.87% (vs. Liu et al.¹³), 13.37% (vs. Khvostikov et al.¹²), 8.47% (vs. Lian et al.³⁷), and 19.77% (vs. Korolev et al.¹¹). In AD versus MCI, our method achieves a classification accuracy of 98.60%, an improvement of 22.6% (vs. Khvostikov et al.¹²). In NC and MCI classification, although the results of our method are lower than those of existing methods, they differ by only 1.66% from the best results. We conducted an additional multiclassification experiment and obtained a classification accuracy of 82.82%. Therefore, our method has good performance on the AD diagnosis task.

5 | DISCUSSION

5.1 | Visualization using Grad-CAM

To better compare the effects of the three commonly used key slice processing techniques (the SKSS method, AKSS method, and MKSS method) and the proposed SHKSS method on the classification results, we show the information extracted from MRI slices by ResNet-18 through gradient-weighted class activation mapping (Grad-CAM).³⁸ Grad-CAM is a tool that provides interpretability for deep learning models, which can visualize the features learned by the model. Specifically, Grad-CAM uses the gradients of any target concept (e.g., the logits of a class in a

classification category or even the output from a caption task) that flows into the final convolution layer to generate a rough localization map to highlight important regions in the image used for prediction. More details about Grad-CAM generation can be found in literature.³⁸

Figure 8a–d shows the Grad-CAM results of the four key slicing processing techniques for the AD versus NC, NC versus MCI, AD versus MCI, and AD versus NC versus MCI classification tasks, respectively. In the Grad-CAM plots, a redder color indicates a higher contribution to the classification task, while a bluer color indicates a lower contribution to the classification task. Figure 8a shows that for the AD and NC classifications, the SKSS method focuses more on regions such as the thalamus and the anterior horn of the lateral ventricles, the AKSS method focuses more on the corpus callosum pressure region, the MKSS method focuses more on regions such as the precentral gyrus, and the SHKSS method focuses on the middle part of the brain, which contains highly sensitive regions for AD, such as the hippocampus and the ventricles. Thus, compared to the other three methods, the SHKSS method focuses on regions that are more relevant to the neuropathology of AD diagnosis.^{39,40} In Figure 8b, for NC and MCI classification, we find that the SKSS method focuses more on regions such as the thalamus and the anterior horn of the lateral ventricles, the AKSS method focuses on regions such as the ventricles and the thalamus, the MKSS method focuses on regions such as the precentral gyrus, and the SHKSS method focuses on regions

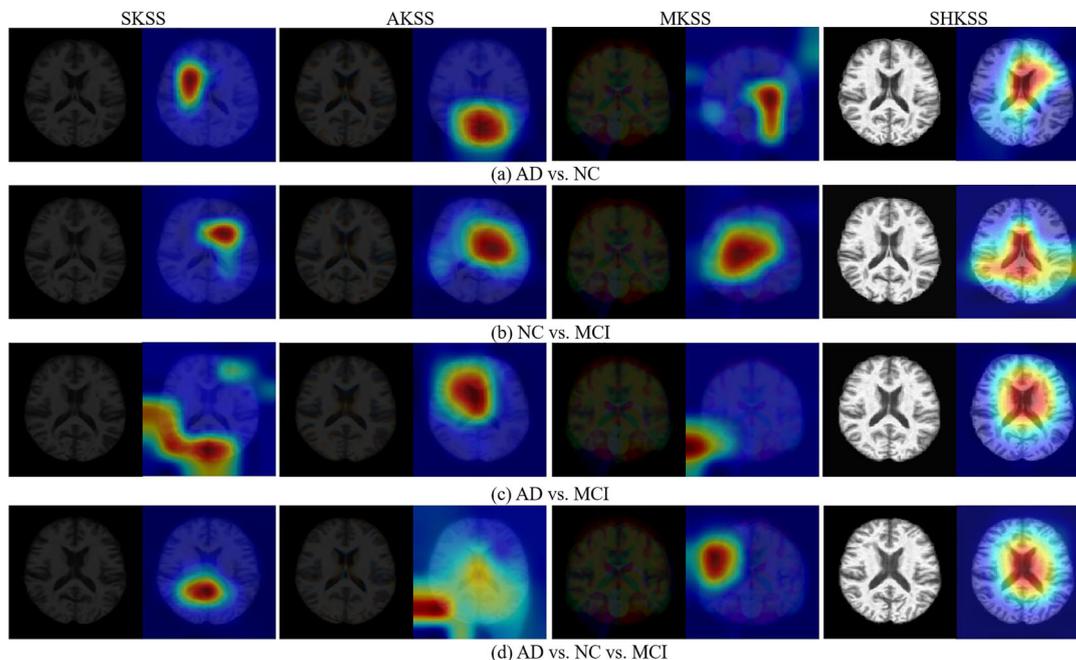


FIGURE 8 Grad-weighted class activation mapping (Grad-CAM) for AD vs. NC classification, NC vs. MCI classification, AD vs. MCI classification, and AD vs. NC vs. MCI classification

such as the hippocampus, the ventricles, and the corpus callosum pressure. Although the SHKSS method also focuses on a small number of background regions, the red areas are mainly concentrated in the middle part of the brain and contain more brain regions. Therefore, compared with the other three methods, SHKSS can provide a more adequate basis for judgment. It can be observed from Figure 8c that for AD and MCI classification, the red areas of the SKSS method contain regions such as the corpus callosum pressure and a small part of the background areas, the AKSS method focuses on changes in the ventricles, the MKSS method focuses more on background areas unrelated to AD pathology, and the SHKSS method still makes decisions based on intermediate areas of the brain that show subtle structural changes in the pathogenesis of AD^{41–44} and are associated with areas of the brain that are closely related to AD. As seen in Figure 8d, for the classification of AD, NC, and MCI, the red areas of the SKSS method contain regions such as the corpus callosum pressure and lateral ventricles, the red areas of the AKSS method are concentrated in background regions unrelated to AD, the red areas of the MKSS method are concentrated in the upper left part of the brain, and the SHKSS method still made judgments based on the middle regions of the brain.

Therefore, it can be seen from these results that the brain regions obtained by the SHKSS method are consistent with previous studies. For example, Serrano-Pozo et al.⁴⁰ showed that brain regions such as hippocampus and thalamus changed during the conversion from NC

to AD. Dolek et al.⁴¹ found that the severity of dementia in patients increased with the decrease of hippocampal volume. Cavedo et al.⁴² found that the volume of the amygdala in AD patients was smaller than that in NC subjects. Chetelat et al.⁴³ found significant gray matter loss in the hippocampus and corpus callosum pressure in AD patients. In summary, ResNet-18 with the SHKSS method makes a final judgment based on sensitive regions associated with AD, indicating that the SHKSS method can indeed highlight the structural information of MRI slices and improve the accuracy of AD diagnosis. To further analyze the effects of different brain regions on the classification tasks, different brain regions (hippocampus, thalamus, and combined with hippocampus and thalamus) were used for the AD classification tasks. Table 8 shows the classification results of the SHKSS method based on data from different brain regions. As can be seen from Table 8, the highest classification accuracy was achieved using data from combined with hippocampus and thalamus regions for all classification tasks. Meanwhile, by comparing Table 8 with Table 2, it can be seen that the classification results using multiple brain regions outperformed those using a single brain region in most of the classification tasks. This suggests that not only do important brain areas play a role in the AD classification task, but that each brain area has some influence, and that individual brain areas differ from each other in the size (i.e., weight) of their contribution to the classification tasks, and that the simultaneous use of multiple brain areas has a complementary effect on the information.

TABLE 8 The classification results of the SHKSS method based on data from different brain regions. Values are reported as mean (%)

Methods	Diagnosis	Hippocampus				Thalamus				Hippocampus + Thalamus			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SHKSS	AD vs. NC	82.96	82.40	83.63	91.78	63.67	69.25	57.78	71.56	83.93	87.50	80.49	93.02
	AD vs. MCI	61.50	42.72	75.51	63.21	58.95	30.79	75.49	56.62	65.22	41.39	79.13	69.49
	NC vs. MCI	68.41	78.22	50.54	74.36	53.54	50.00	59.85	58.10	69.58	74.53	60.32	72.26
	AD vs. NC vs. MCI	53.35	53.53	75.12	72.79	40.49	37.69	67.89	57.22	54.04	51.28	74.53	73.75

TABLE 9 Classification results for different classification tasks without transfer learning. Values are reported as mean (%)

Methods	AD vs. NC				AD vs. MCI				NC vs. MCI				AD vs. NC vs. MCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SKSS	64.30	39.01	90.75	75.31	60.34	38.31	73.28	60.85	54.74	54.11	55.91	61.55	46.15	40.15	69.77	63.89
AKSS	61.09	45.79	77.18	76.37	60.11	23.81	81.34	59.46	52.61	48.55	59.85	61.02	45.46	38.80	69.02	63.21
MKSS	63.66	40.95	87.66	78.65	59.64	40.93	70.67	58.20	52.13	45.68	63.91	57.40	46.12	40.79	69.68	63.42
SHKSS	95.69	90.34	99.00	99.00	95.29	86.87	98.19	99.00	80.90	72.31	96.09	97.35	80.59	85.50	91.04	97.34

5.2 | Classification tasks without transfer learning

In this experiment, we study the impact of transfer learning methods on each classification task. In general, when the number of data sets is small, a common approach to avoid overfitting is to use a pretrained CNN model and perform classification by simply freezing and fine-tuning some layers. Table 9 shows the classification results for ADNI-1 data without the transfer learning method. By comparing Table 2 and Table 9, it can be found that the results with transfer learning are slightly better than those without transfer learning. Specifically, the least/greatest classification accuracy was improved by 0.8% (NC vs. MCI)/6.9% (AD vs. MCI), 0.67% (NC vs. MCI)/9.24% (AD vs. NC), 1.54% (AD vs. NC vs. MCI)/9.15% (AD vs. MCI), and 0.73% (NC vs. MCI)/3.31% (AD vs. MCI). Although the improvement in classification accuracy using transfer learning was small, the experiments showed that the time required to train the CNN model using the transfer learning method was only approximately 1/2 of that without using the transfer learning method for the same number of epochs (40). Therefore, we can speculate that with the increase in learning samples and training batches, the transfer learning method will save much training time for the model, save computer resources, and help researchers more easily modify the model parameters to optimize the model.

5.3 | Classification results for data augmentation

From the above classification results, it can be seen that the classification performances of the SKSS method, AKSS method, and MKSS method are far lower than

that of the SHKSS method without data augmentation. The highest classification accuracies of the top three methods were 75.32% (AD vs. NC), 71.58% (AD vs. MCI), 63.4% (NC vs. MCI), and 46.19% (AD vs. NC vs. MCI). However, in most existing AD classification studies based on SKSS methods,^{16–18} AKSS methods,²¹ and MKSS methods,^{22,23} these methods achieved high classification accuracy, as shown in Table 10. Specifically, for the SKSS method, the AD versus NC classification task achieved a high accuracy of 98.84% (Sarraf et al.¹⁷), a high accuracy of 94.88% for AD versus MCI (Naz et al.¹⁸), a high accuracy of 91.67% for NC versus MCI (Billones et al.¹⁶), and the highest accuracy of 91.85% for AD versus NC versus MCI (Billones et al.¹⁶). However, these excellent results are due to data leakage, and studies have shown that the incorrect division of the training set and test set is one of the major causes of data leakage.⁴⁵ In literatures,^{16–18} the authors extracted multiple 2D slice images from each 3D MRI image for classification experiments (among them, Billones et al.¹⁶ extracted twenty 2D slice images from each 3D MRI image, Sarraf et al.¹⁷ extracted all 2D slice images with nonzero average pixels from each 3D MRI image, and Naz et al.¹⁸ selected at least three 2D slice images from each 3D MRI image), and these slices were randomly divided into a training set and test set according to a certain ratio. These incorrect data division methods divide different slices of the same subject into a training set and a test set, resulting in data leakage. For the AKSS method, Wu et al.²¹ first extracted 130 slice images from each 3D MRI image, then stacked the slice images into RGB images at four intervals, and finally used a fivefold cross-validation strategy to divide all the RGB images into a training set and test set for experiments. Therefore, this study also suffered from data leakage problems. For the MKSS method, Lin et al.^{22,23} also had the potential for data leakage. Specifically,

TABLE 10 Classification results for different classification tasks with data augmentation. Values are reported as mean (%)

	Methods	Data set (AD/MCI/NC)	Models	AD vs. NC/AD vs. MCI/NC vs. MCIAD vs. NC vs. MCI
Billones et al. ¹⁶	SKSS	300/300/300	VGG16	98.33/93.89/91.67/91.85
Sarraf et al. ¹⁷		211/-/91	GoogleNet	98.84/-/-
Naz et al. ¹⁸		95/146/95	ResNet-18	92.3/94.88/80.61/-
Wu et al. ²¹	AKSS	-/307/150	GoogleNet	-/-/97.58/-
Lin et al. ²²	MKSS	188/-/229	CNN	88.02/-/-
Lin et al. ²³				88.79/-/-
Ours	SKSS	159/152/272	ResNet-18	92.28/86.00/86.17/79.93
	AKSS			91.53/85.22/88.29/80.73
	MKSS			90.57/85.61/89.39/79.36
	SHKSS			99.57/96.68/96.23/91.71

in literature,²² the authors extracted a total of 62,967 ((188+229)×151) 2D slice images from 3D MRI images and used the 10-fold cross-validation strategy to divide them into a training set and test set. In literature,²³ the author randomly divided these 62,967 images into 417 small batches. To further illustrate the impact of data leakage, we extracted three slice images from each 3D MRI image in the ADNI-1 data set to obtain a total of 1,127 2D slice images and randomly divided these slice images into a training set and test set for AD classification experiments using a fivefold cross-validation strategy. The classification results are shown in the last four rows of Table 10. As seen from the table, the accuracy of all classification tasks is significantly improved after incorrect data division (similar to the classification results of existing studies).

6 | LIMITATIONS

In this study, we classified AD, MCI, and NC based on 2D slice images of 3D MRI images. Our method has the following limitations. First, in this study, we used only 2D slice images for the classification experiments, while existing studies have shown that 2D slice images lose more useful information for AD classification than 3D images. Second, demographic data, neuropsychological data, imaging data, genetic data, cerebrospinal fluid data, and blood data of AD subjects, MCI subjects, and NC subjects were collected according to uniform standards in the ADNI data set. In this study, we used only the features of MRI images to classify AD. Finally, the ADNI data set was scanned longitudinally for all recruited subjects; that is, MRI scans and PET scans were performed on subjects at different time points (including baseline, sixth month, 12th month, 18th month, 24th month, 36th month, and 48th month). However, in this study, we used only sMRI images acquired at a single time point (baseline time) to classify AD. In summary, we can

improve the accuracy of the classification task in three respects. Extending the SHKSS method to 3D images further improves the performance of AD classification. In addition, we can combine medical images with other modalities of data (e.g., genetic data, blood data, and cerebrospinal fluid data) to conduct studies on more modalities and improve the classification performance of AD. We can also use images from longitudinal scans to classify AD.

7 | CONCLUSION

In this study, to make full use of the structural information of sMRI images, we propose applying the SHKSS key slice processing technique to a 2D transfer learning model for AD classification. Specifically, first, SHKSS key slice processing technology is used to transform a single-channel image into a three-channel image. Then, the three-channel image is used as the input of a pretrained ResNet-18 to perform AD classification. The experimental results on ADNI-1 and ADNI-2 show that the proposed method has good classification performance and generalization ability; in particular, the classification accuracy of AD, NC, and MCI is 82.82%.

ACKNOWLEDGMENTS

The data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). As such, the investigators at the ADNI contributed to the design and implementation of the ADNI and/or provided data but did not participate in the analysis or writing of this report.

This work was supported by the Guangxi Key Projects of Science and Technology (Grant No. 2020AA21077007) and the Innovation Project of Guangxi Graduate Education (Grant No. YCBZ2020026).

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Alzheimer's Disease Neuroimaging Initiative (ADNI) at www.adni.loni.usc.edu.

REFERENCES

- Lombardi A, Amoroso N, Diacono D, Monaco A, Tangaro S. Association between structural connectivity and generalized cognitive spectrum in Alzheimer's disease. *Brain Sci.* 2020;10(11):879-896.
- Alzheimer's Association. 2018 Alzheimer's disease facts and figures. *Alzheimer's Dementia.* 2018;14(3):367-429.
- Hurd MD. Monetary costs of dementia in the United States. *Med. Benefits.* 2013;30(9):5-6.
- Reitz C, Tang MX, Schupf N, Manly JJ, Luchsinger JA. A summary risk score for the prediction of Alzheimer disease in elderly persons. *Arch Neurol.* 2010;67(7):835.
- Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol.* 2011;10(9):819-828.
- Tong T, Wolz R, Gao Q, Guerrero R, Hajnal JV, Rueckert D. Multiple instance learning for classification of dementia in brain MRI. *Med Image Anal.* 2014;18(5):808-818.
- Gupta A, Maida AS, Ayhan M. Natural image bases to represent neuroimaging data. Proceedings of the 30th International Conference on Machine Learning (ICML'13). Atlanta, GA, USA, 16 June 2013;III-987-III-994.
- Suk HI, Lee SW, Shen D. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Struct Funct.* 2015;220(2):841-859.
- Asl EH, Keynton R, El-Baz A. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. IEEE International Conference on Image Processing - ICIP 2016 IEEE, 2016.
- Ji S, Xu W, Yang M, Yu K. 3D convolutional neural networks for human action recognition. *IEEE Trans Pattern Anal Mach Intell.* 2012;35(1):221-231.
- Korolev S, Safiullin A, Belyaev M, Dodonova Y. Residual and plain convolutional neural networks for 3D brain MRI classification. IEEE International Symposium on Biomedical Imaging 2017. Melbourne, Australia, 18-21 April 2017;835-838.
- Khvostikov A, Aderghal K, Benois-Pineau J, Krylov A, Catheline G. 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies. arXiv e-prints, January 2018; arXiv:1801.05968.
- Liu M, Li F, Yan H, Wang K, Xu M. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. *Neuroimage.* 2019;208:116459.
- Tajbakhsh N, Shin JY, Gurudu SR, et al. Liang, Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging.* 2016;35(5):1299-1312.
- Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? arXiv e-prints, November 2014; arXiv:1411.1792.
- Billones CD, Demetria O, Hostallero D, Naval PC. DemNet: a convolutional neural network for the detection of Alzheimer's disease and mild cognitive impairment. TENCON 2016 : IEEE Region 10 Conference. Singapore, 22-25 November 2016;3724-3727.
- Sarraf S, Desouza DD, Anderson JAE, Tofighi G. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *BioRxiv.* 2016;070441.
- Naz S, Ashraf A, Zaib A. Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset. *Multimedia Syst.* 2021;1-10.
- Ebrahimi-Ghahnavieh A, Luo S, Chiong R. Transfer learning for Alzheimer's disease detection on MRI images. 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT). NSW, Australia, 1-3 July 2019;133-138.
- Ebrahimi A, Luo S. Convolutional neural networks for Alzheimer's disease detection on MRI images. *J Med Imaging.* 2021;8(2):024503.
- Wu C, Guo S, Hong Y, Xiao B, Wu Y, Zhang Q. Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks. *Quant Imaging Med Surg.* 2018;8(10):992-1003.
- Lin W, Gao Q, Du M. Convolutional neural network based method for diagnosis of Alzheimer's disease. *J Comput Appl.* 2017;37(12):3504-3508.
- Lin W, Tong T, Gao Q, et al. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Front Neurosci.* 2018;12:777.
- Carré A, Klausner G, Edjlali M, Lerousseau M, Robert C. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Sci Rep.* 2020;10(1):12340.
- Schindler S, Schreiber J, Bazin PL, Trampel R, Schnknecht P. Intensity standardisation of 7T MR images for intensity-based segmentation of the human hypothalamus. *Plos One.* 2017;12(3):e0173344.
- Ma Rcin K, Micha S, Rafa O. Does image normalization and intensity resolution impact texture classification?. *Computerized Med Imaging Graphics.* 2020;81:101716.
- Loizou CP, Pantziaris M, Seimenis I, Pattichis CS. Brain MR image normalization in texture analysis of multiple sclerosis. 2009 9th International Conference on Information Technology and Applications in Biomedicine. IEEE, Larnaca, Cyprus, 5-7 November 2009;1-5.
- Nyul LG, Udupa JK, Xuan Z. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging.* 2000;19(2):143.
- Gonzalez RC, Woods RE, Masters BR. *Digital Image Processing.* Society of Photo-Optical Instrumentation Engineers; 2009.
- Gaser C, Dahnke R. CAT-a computational anatomy toolbox for the analysis of structural MRI data. *Hbm.* 2016;2016:336-348.
- Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. *Statistical Parametric Mapping: The Analysis of Functional Brain Images.* Elsevier; 2011.
- Coltuc D, Bolon P, Chassery JM. Exact histogram specification. *IEEE Trans Image Process.* 2006;15(5):1143-1152.
- He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 7-13 December 2015;1026-1034.
- Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 7-12 June 2015;1-9.
- Shi J, Zheng X, Li Y, Zhang Q, Ying S. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE J Biomed Health Inf.* 2017;22(1):173-183.
- Liu M, Cheng D, Wang K, Wang Y. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. *Neuroinformatics.* 2018;16(3):295-308.
- Lian C, Liu M, Zhang J, Shen D. Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. *IEEE Trans Pattern Anal Mach Intell.* 2018;42(4):880-893.
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, October 22-29 2017;618-626.

39. De Jager P, Chibnik L, Aggarwal N, et al. Relation of the Alzheimer's disease susceptibility allele at the CR1 locus to cognitive decline and Alzheimer's disease neuropathology in older persons. *Neurology*. 2010;74(9):A97-A97.
40. Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT. Neuropathological alterations in Alzheimer disease. *Cold Spring Harbor Perspect. Med*. 2011;1(1):a006189.
41. Dolek N, Saylisoy S, Ozbabalik D, Adapinar B. Comparison of hippocampal volume measured using magnetic resonance imaging in Alzheimer's disease, vascular dementia, mild cognitive impairment and pseudodementia. *J Int Med Res*. 2012;40(2):717-725.
42. Cavado E, Boccardi M, Ganzola R, et al. Local amygdala structural differences with 3T MRI in patients with Alzheimer disease. *Neurology*. 2011;76(8):727-733.
43. Chetelat G, Desgranges B, Sayette VDL, Viader F, Baron JC. Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. *Neuroreport*. 2002;13(15):1939-1943.
44. Kazemifar S, Drozd JJ, Rajakumar N, Borrie MJ, Bartha R. Automated algorithm to measure changes in medial temporal lobe volume in Alzheimer disease. *J Neurosci Methods*. 2014;227:35-46.
45. Yagis E, Atnafu SW, de Herrera AGS, et al. Effect of data leakage in brain MRI classification using 2D convolutional neural networks. *Sci Rep*. 2021;11(1):1-13.

How to cite this article: Ban Y, Zhang X, Lao H. Diagnosis of Alzheimer's disease using structure highlighting key slice stacking and transfer learning. *Med Phys*. 2022;1-15.
<https://doi.org/10.1002/mp.15888>