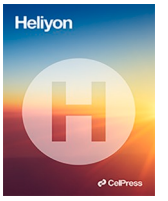
Heliyon 10 (2024) e23008

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



SVFR: A novel slice-to-volume feature representation framework using deep neural networks and a clustering model for the diagnosis of Alzheimer's disease

Rubing Wang [b], Linlin Gao [a,b,*], Xiaoling Zhang [c], Jinming Han [d], the Alzheimer's Disease Neuroimaging Initiative

[a] School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China
[b] Faculty of Electrical Engineering and Computer Science, Zhejiang Engineering Research Center of Advanced Mass Spectrometry and Clinical Application, Ningbo University, Ningbo, China
[c] Ningbo Medical Center Lihuili Hospital, Ningbo, China
[d] Department of Neurology, Xuanwu Hospital, Capital Medical University, Beijing, China

ARTICLE INFO

Keywords:
Slice-to-volume feature representation
Deep neural networks
Clustering model
Informative slice images
Spatial pyramid set pooling module
Diagnosis of Alzheimer's disease

ABSTRACT

Deep neural networks (DNNs) have been effective in classifying structural magnetic resonance imaging (sMRI) images for Alzheimer's disease (AD) diagnosis. In this study, we propose a novel two-phase slice-to-volume feature representation (SVFR) framework for AD diagnosis. Specifically, we design a slice-level feature extractor to automatically select informative slice images and extract their slice-level features, by combining DNN and clustering models. Furthermore, we propose a joint volume-level feature generator and classifier to hierarchically aggregate the slice-level features into volume-level features and to classify images, by devising a spatial pyramid set pooling module and a fusion module. Experimental results demonstrate the superior performance of the proposed SVFR, surpassing the majority of the state-of-the-art methods and achieving comparable results to the best-performing approach. Experimental results also showcase the efficacy of the slice-level feature extractor in the selection of informative slice images, as well as the effectiveness of the volume-level feature generator and classifier in the integration of slice-level features for image classification. The source code for this study is publicly available at https://github.com/gll89/SVFR.

1. Introduction

As one of the most common neurodegenerative diseases found in the elderly, Alzheimer's Disease (AD) accounts for about two-thirds of dementia [1]. The predominant clinical symptoms of AD contain progressive memory loss and cognitive deficits, which can severely affect the daily life of AD patients. An individual converts into AD every 5 s worldwide, and over 33 million people are living with AD globally and the number will be 102 million by 2050 [2], which make AD one of the leading causes of mortality among the elderly. Even though AD is incurable and worsens over time due to its irreversible damage to brain cells, treatments, including medications and management strategies, are helpful to delay the deterioration of the disease [3,4]. Diagnosing AD as early and

* Corresponding author. Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China.
E-mail address: gaolinlin@nbu.edu.cn (L. Gao).

https://doi.org/10.1016/j.heliyon.2023.e23008
Received 9 December 2022; Received in revised form 30 September 2023; Accepted 23 November 2023
Available online 3 December 2023
2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

accurately as possible is thus critical for AD patients to receive timely treatments and slow down the deterioration of the disease.

In clinical practice, two common ways to diagnose AD are measuring the changes of brain tissues and testing the loss of cognition. Studies demonstrate that the changes of brain tissues have begun 20 years or more before cognition loss [5]. Structural magnetic resonance imaging (sMRI) images are widely used for assisting doctors to diagnose AD because of their safe and non-invasive way to visualize brain tissues [6,7]. Numerous studies have devoted to automatic AD diagnosis using sMRI images in the last several decades. Conventionally, the pipelines for these methods mainly include image preprocessing, hand-designed feature extraction, and diagnosis using standard machine learning methods, such as support vector machine [8] or logistical regression [9]. Nevertheless, these traditional methods are limited by the subjectivity of hand-designed features and their complicated extraction process. Recently, a large number of deep neural networks (DNNs) have been developed for medical image analysis due to their automatic abstraction of low-to-high level latent feature representations [10]. Many of these DNNs are used for AD diagnosis. These methods can be divided into two main categories, volume-related 3D DNN-based methods and slice-related 2D DNN-based methods.

The volume-related 3D DNN-based methods take 3D sMRI images as input and design 3D DNNs to extract useful image features for AD diagnosis. These methods are divided into three sub-categories: the regions-of-interest-based, patch-based, and whole-image-based methods. Regions-of-interest-based methods typically commence by segmenting pre-defined 3D regions-of-interest, such as hippocampus, from each 3D sMRI image under the guidance of AD experts. Subsequently, these methods extract features from the segmented regions-of-interest using specifically designed 3D convolutional neural networks (CNNs) [11,12]. Nonetheless, a notable drawback of these approaches lies in the fact that the pre-defined regions-of-interests exhibit variability across different AD experts, consequently leading to subjectivity and partiality in the extracted features. The patch-based methods first select informative image patches in a data-driven manner and then extract and fuse the features of these image patches for AD diagnosis using deep general or multi-instance CNNs [13–19]. This kind of method registers each sMRI image into a brain template, losing the individual-specificity. Whole--image-based methods in AD diagnosis directly extract salient features from entire 3D sMRI images using fine-tuned state-of-the-art 3D CNNs or designing new CNNs [20–27]. These methods do not require additional guidance from AD experts. However, it is noteworthy that such approaches are prone to overfitting due to the limited number of available 3D sMRI images, the substantial volume of each sMRI image, and the relatively smaller volume of lesion regions.

The slice-related 2D DNN-based methods utilize 2D slice images, selected from 3D sMRI images, as input. These methods devise 2D DNNs to extract slice-level features and integrate these slice-level features for AD diagnosis. One type of the slice-related 2D CNN method takes the slice images from the three views, i.e, the axial, sagittal, and coronal views, as input. For instance, Aderghal et al. proposed a 2D+ε approach for AD diagnosis [28]. Specifically, they first segmented 3D hippocampal image patches and then selected the three middle slice images of the three views from the 3D hippocampal image patches, together with their individual two closest slice images, to form three input images. After that, they devised three parallel CNNs to learn the features of three input images, respectively, and an FC layer to fuse these features for final classification. Similarly, Islam and Zhang [29] and Mehmood et al. [30] also took the slice images from the three views as three input images. Specifically, Islam and Zhang [29] first trained three parallel CNNs for the three input images and then utilized the majority voting to fuse the results of the three CNNs for AD diagnosis. Mehmood et al. [30] developed a deep siamese CNN for AD diagnosis. However, the selection of slice images was not well introduced in Refs. [29, 30]. The other kind of slice-related 2D CNN-based method solely considers slice images from one of the three views as input. For example, Valliani and Soni took the median slice image in the axial view of each sMRI image as input and fine-tuned the advanced ResNet [31] for AD diagnosis [31,32]. Gao et al. utilized the middle 50 slice images in the sagittal view of each MRI image as input under the guidance of neurologists [33]. This approach fine-tuned ResNet to extract the features of these slice images first. After that, the bag-of-words strategy [34] was utilized to integrate these slice-level features. Qiu et al. selected a "signature" slice image in the axial view of each sMRI image in a semi-automatic manner, together with its two adjacent slice images, as input. They designed three individual CNNs for each of the three slice images for slice-level feature extraction and slice image classification. After that, they employed the majority voting strategy for sMRI image classification [35].

In clinic, when diagnosing a patient with an sMRI image, neurologists first identify the disease-related slice images from the sMRI image, then analyze these slice images to collect useful information, and finally integrate such information to make a decision for the patient. Motivated by this diagnosing process, we explore a novel slice-related 2D DNN-based framework, namely an automatic slice-to-volume feature representation (SVFR) framework, to distinguish AD from normal controls (NC), as depicted in Fig. 1. SVFR consists of a slice-level feature extractor (SFE) and a joint volume-level feature generator and classifier (VFGC). SFE aims to automatically select informative slice images from sMRI images and extract the features of these informative slice images. It is achieved by combining a clustering model and CNNs. The goal of VFGC is to hierarchically fuse the slice-level features of each sMRI image into a volume-level feature and make a final decision. VFGC is realized through a devised spatial pyramid and set pooling (SPSP) module and a fusion module. Our key contributions are as below.

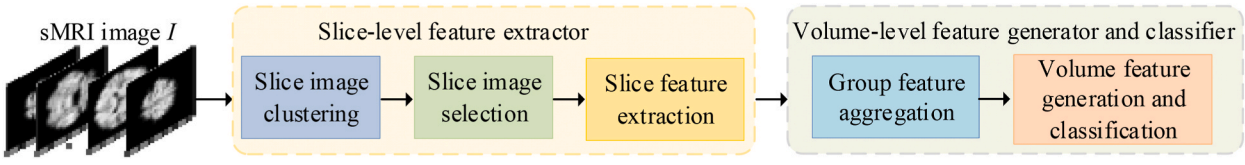


Fig. 1. Framework of SVFR. It consists of a slice-level feature extractor and a joint volume-level feature generator and classifier.

- We propose a novel SVFR framework for AD diagnosis. This framework is motivated by the diagnosing process of neurologists and it makes full use of the merits of the slice images, i.e, the large number of slice images and the small size of slice images.
- In SFE, we combine a clustering model and CNNs for automatic selection and feature extraction of informative slice images.
- In VFGC, we design an SPSP module and a fusion module to hierarchically fuse the slice-level features into volume-level features for AD diagnosis.

The rest of the paper is organized as follows. Section 2 describes the proposed method and the dataset used. Section 3 shows experimental results and discussion. At last, the paper is concluded in Section 4.

2. Methodology

The proposed SVFR fully leverages the benefits of slice images, i.e, the large number and small size of slice images. These benefits enhance the rapid convergence and better performance of SVFR. SFE and VFGC are the two phases of SVFR. The goal of SFE is to select informative slice images and extract their discriminative slice-level features. The target of VFGC is to generate volume-level features and to further make final classification.

2.1. Slice-level feature extractor

The challenge of slice-level feature extraction is the lack of slice-level labels. To deal with this issue, the simplest way is to assign each slice image the label of its corresponding 3D sMRI image. However, this assignment manner brings in numerous noisy labels for the slice images of AD samples, because disease-related regions generally occur in several brain structures yet not the whole brain. However, the above assignment manner gives the slice images with normal brain structures the label of AD. To deal with the noisy labels in AD samples, we devise SFE motivated by the prior knowledge that disease-related regions occur in certain brain structures such as hippocampi, amygdalae, and ventricles.

SFE consists of slice image clustering, slice image selection, and slice feature extraction. Fig. 2 shows its pipeline. To be specific, slice images are first clustered according to brain structures. After that, informative slice image groups are preserved and noisy slice image groups are eliminated based on the classification performance of slice groups. At last, the features of informative slice images are extracted. We depict the details of each step as below.

Slice image clustering. Different levels of slice images in the axial view of sMRI images present diverse brain structures, as shown in Fig. 3. Clustering based on the phenotype of slice images can divide slice images with different brain structures into various groups. Specifically, each slice image is first represented by a slice vector by zooming out the slice image 10 times and sequentially concatenating the row pixels of the zoomed image. After that, all the slice vectors are clustered into K groups by using K-means [36]. The K groups of slice images are denoted as $\{S_1, .., S_k, .., S_K\}$, displaying K groups of distinct brain structures. It is noted that each slice group has the uncertain number of slice images.

Slice image selection. As aforementioned, several brain structures are closely related to AD, while other brain structures are not impacted by the disease. This indicates that different slice groups present various abilities in disease diagnosis. The power of a slice group in disease diagnosis can be determined based on the classification performance of the slice group. In detail, K advanced pre-trained ResNets, i.e, $\{\text{ResNet}_1, .., \text{ResNet}_k, .., \text{ResNet}_K\}$, are employed as the classifiers of the K groups of slice images. Each S_k is divided into the training and validation sets to train ResNet_k. Moreover, cross-entropy is employed as the loss function, which is described in the following Eq. (1),

$$L = \sum_{i=1}^{T} \sum_{j=1}^{J} y_j \log(p_j) \tag{1}$$

where L denotes the loss, T is the number of the training samples, J is the number of categories, y_j is the one-hot format of the true label of a slice image, and p_j represents the probability of the slice image belonging to the jth category.

The classification accuracy of the well-trained ResNet_k on the corresponding validation set of S_k is represented as acc_v_k. An informative slice group SI_m is defined as

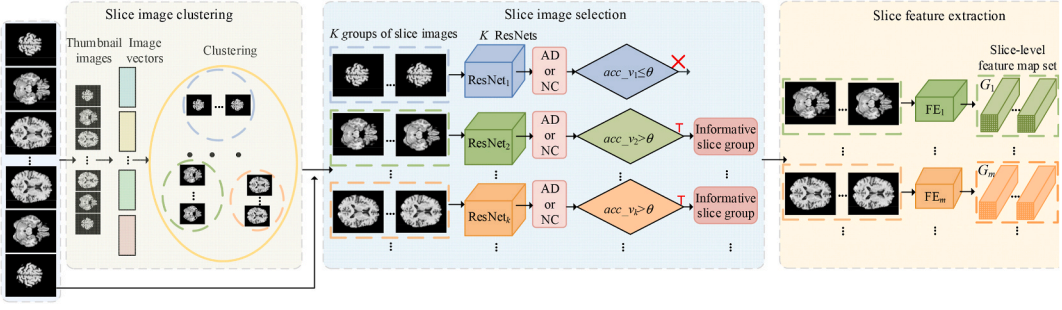


Fig. 2. Pipeline of SFE. It is comprised of slice image clustering, slice image selection, and slice feature extraction.

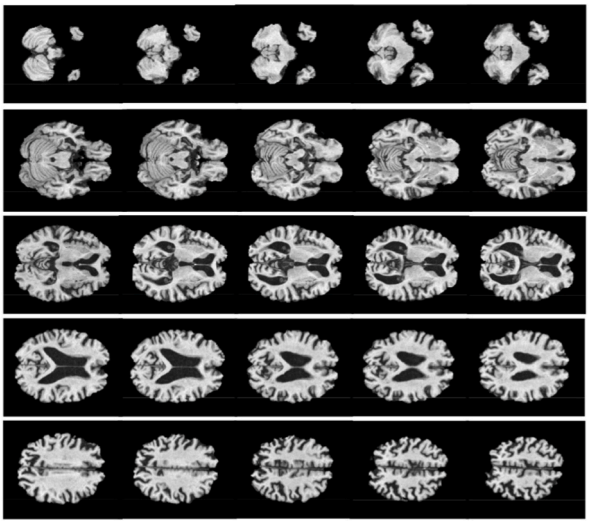


Fig. 3. Different levels of slice images of an sMRI image in the axial view. The slice images from the 1st to 5th rows show various brain structures.

$$SI\_(m) = S\_(k) \ \& \ \left(acc\_v\_(k) > \theta\right) \tag{2}$$

where θ is a threshold and its value will be discussed in the section of Results and Discussion; m is the index of the informative slice group. Inversely, the slice group with the classification accuracy on the validation set not larger than θ is called an uninformative slice group.

The validation sets of the informative slice groups are more accurately classified than the validation sets of the uninformative slice groups, even though all the slice groups are from the same dataset and are trained using the same CNN architectures and settings. We thus obtain that the classification accuracy differences among the validation sets are caused by the only variable factor, i.e, the assigned labels of slice images. That is, the labels of the slice images belonging to the informative slice groups tend to be right, while the uninformative slice groups contain more noisy labels and are thus discarded.

Slice feature extraction. The layers from the 1st to the last convolutional layers of the well-trained ResNet_m are utilized as the feature extractor of the corresponding informative slice groups and denoted as FE_m. The feature map set of SI_m generated through FE_m is denoted as $G_m = \{g_{m1}, \ldots, g_{mn}, \ldots, g_{mN}\}$, where $g_{mn} \in \text{R}^{L*W*D}$ is the feature map of the nth slice image in SI_m; N is the number of slice images in SI_m and it is a variable, indicating various numbers of slice images among informative slice groups; L, W, and D are the length, width, and depth of a feature map, respectively. The slice-level feature map sets of all the informative slice groups, $\{SI_1,..,SI_m,..,SI_M\}$, are denoted as $\{G_1 \ldots, G_m \ldots, G_M\}$, where M is the number of informative slice groups.

2.2. Volume-level feature generator and classifier

Since intra-group slice images have similar brain structures and inter-group slice images have different brain structures, the features extracted from intra-group slice images (i.e, intra-group features) are homogeneous and the features extracted from inter-group slice images (i.e, inter-group features) are heterogeneous. Based on this, VFGC is designed to hierarchically aggregate the homogeneous and heterogeneous slice-level feature maps for sMRI image classification. Specifically, each intra-group feature map set is first fused into a feature vector using the proposed SPSP module. These inter-group feature vectors are then fused for volume-level feature generation and final classification using the devised FC module. The structure of VFGC is shown in Fig. 4.

Intra-group feature map fusion. Each intra-group feature map set, $G_m = \{g_{m1}, \ldots, g_{mn}, \ldots, g_{mN}\}$, is fused by using the proposed SPSP module. The detailed structure of the SPSP module is displayed in Table 1. It includes two convolutional blocks, a spatial pyramid pooling block, set pooling layers, global average pooling (GAP) layers, and a concatenating layer. The two convolutional blocks are to extract high-level semantic feature maps. The spatial pyramid pooling block aims to extract different-scale semantic features, considering the various sizes of disease-related regions. The set pooling layer is to deal with the variable numbers of feature maps in G_m and meanwhile extract prominent features. It is realized by the maximum pooling along the channel dimension, i.e, the first dimension of $G_m \in \text{R}^{N*L*W*D}$, and generates a feature map with the size of $1*L*W*D$. GAP generates a feature vector to emphasize discriminative features for classification. The concatenating layer is used to cascade multiple feature vectors into a feature vector.

A FC layer and an softmax layer are connected after the SPSP module for the training of SPSP. Cross-entropy is employed as the loss function. After training, the well-trained SPSP module is employed for intra-group feature map fusion. Formally,

$$F_m = \text{SPSP}(G_m)$$

where the generated feature vector $F_m \in \text{R}^{1*MS*D}$ and M_S represents the number of scales in the spatial pyramid pooling block. The

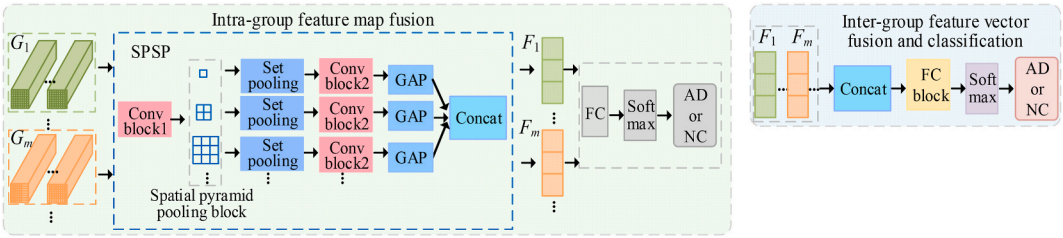


Fig. 4. Structure of VFGC. It consists of two parts, intra-group feature map fusion and inter-group feature vector fusion and classification.

Table 1
Detailed structure of the SPSP module.

	layers
Conv block1	1*1*512 convolution layer with stride = 1, ReLU, 1*1*1024 convolution layer with stride = 1, ReLU;
Spatial pyramid pooling block	3*3 max pooling layer with stride = 3, 1*1*1024 convolution layer with stride = 1, ReLU; 2*2 max pooling layer with stride = 2, 1*1*1024 convolution layer with stride = 1, ReLU; 1*1 max pooling layer with stride = 1; 1*1*1024 convolution layer with stride = 1, ReLU;
Set pooling	1D adaptive max pooling layer
Conv block2	1*1*1024 convolution layer with stride = 1, ReLU;
GAP	Global average pooling;
Concat	Concatenating layer.

feature vector set $\{F_1, \ldots, F_m, \ldots, F_M\}$ is produced for $\{G_1 \ldots, G_m \ldots, G_M\}$ by using the well-trained SPSP.

Inter-group feature vector fusion and classification. We devise the FC module to merge the heterogeneous inter-group feature vectors $\{F_1, \ldots, F_m, \ldots, F_M\}$ for volume-level feature generation and sMRI classification. The detailed structure of the FC module is displayed in Table 2. The concatenating layer is to connect M feature vectors into a long feature vector. The FC block is used to fully fuse the inter-group feature vectors. A volume-level feature vector is produced after the FC block. The softmax layer is used to normalize the output for classification. Moreover, cross-entropy is further utilized as the loss function.

2.3. Dataset and preprocessing

We utilize the public Alzheimer's Disease Neuroimaging Initiative (ADNI) [1] as our experimental data. ADNI is launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease. In this study, we select total 788 sMRI images from ADNI, which consist of 367 AD samples and 421 NC ones. These sMRI images are from different subjects at the baseline time and are acquired using 1.5T scanners with the protocol of sagittal T1-weighted MPRAGE. Each sMRI image has a label. However, there is no slice-level label. The demographic and clinical information are summarized in Table 3.

We perform several preprocessing steps on the sMRI images. First, we utilize the brain extraction tool of FMRIB Software Library 5.0 [2] to remove the skull and dura. The exclusive option of "B" is used during this process to achieve accurate neck removal. Next, we utilize FMRIB Software Library 5.0 to resample all sMRI images into a spatial resolution of 1*1*1 mm^3. This is accomplished by linearly aligning the sMRI images to the template of MNI152, resulting in sMRI images with a size of 182*218*182. However, these resampled sMRI images include a large background portion. To remove the unnecessary background and maintain the brain's morphology, we perform cropping operations on each sMRI image. Specifically, we eliminate the background along the minimum vertical external matrix encompassing the brain portion. Consequently, the sizes of sMRI images vary. To ensure a consistent size and preserve the brain's morphology, we apply a scaling operation to each sMRI image, maintaining the image's original ratio, until the size of its maximum side reach 128. Tri-linear interpolation is utilized in this process, Additionally, we pad the other two sides of each sMRI image to achieve a final size of 128*128*128. Through these operations, all the sMRI images are standardized to a size of 128*128*128 without deformations. We can thereby obtain 128 slice images from each sMRI image in its axial view and each slice image has the same label of its corresponding sMRI image. The pixel values of each slice image were normalized into [0, 1].

[1] http://adni.loni.usc.edu/.
[2] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET.

Table 2
Structure of the FC module.

	layers
Concat	Concatenating layer;
FC block	$M \times M_S \times D$ FC layer,
	$M \times M_S \times D$ FC layer,
	$M \times M_S \times D$ FC layer,
	2 FC layer;
Softmax	Softmax function.

Table 3
Demographic and clinic information of the 788 subjects (Age, Edu, and MMSE are defined as mean ± standard deviation).

	Number	Sex(F/M)	Age	Edu	MMSE
AD	367	176/191	75.0 ± 7.9	15.1 ± 3.0	23.2 ± 2.1
NC	421	218/203	74.6 ± 5.8	16.3 ± 2.7	29.1 ± 1.1

3. Results and discussion

3.1. Evaluation metrics and Implementation details

The proposed SVFR is evaluated on the task of classifying AD and NC. In the task, we regard AD subjects as positive cases and NC subjects as negative cases. We utilize accuracy (ACC), specificity (SPE), sensitivity (SEN, also called Recall), precision (PRE), and F1-score (F1) to evaluate our method. The computation of the five metrics is described in Eqs. (3)–(7).

$$\mathrm{ACC} = (\mathrm{TP} + \mathrm{TN}) \ / \ (\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}) \tag{3}$$

$$\mathrm{SPE} = \mathrm{TN} \ / \ (\mathrm{TN} + \mathrm{FP}) \tag{4}$$

$$\mathrm{SEN} = \mathrm{TP} \ / \ (\mathrm{TP} + \mathrm{FN}) \tag{5}$$

$$\mathrm{PRE} = \mathrm{TP} \ / \ (\mathrm{TP} + \mathrm{FP}) \tag{6}$$

$$\mathrm{F1} = (2 \times \mathrm{PRE} \times \mathrm{Recall}) \ / \ (\mathrm{PRE} + \mathrm{Recall}) \tag{7}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

We divide the sMRI images into the training, validation, and test sets with a ratio of 7:2:1. The numbers of the sMRI images with the label of AD in the training, validation, and test sets have the ratio of 7:2:1 and the sMRI images with the label of NC in the three sets also have the same ratio, which avoids bringing into extra data imbalance. Importantly, it should be noted that the splitting of the data is performed on the participant-level, rather than the slice-level. The training and validation sets are used to train SVFR. Specifically, for SFE, we first cluster the slice images in both the training and validation sets into K groups, and then adjust each group of slice images into training and validation sets according to the above divided recording. After that, we separately train K ResNets using the K groups of training and validation sets, with 60 epochs and a batch size of 100. Furthermore, we determine M informative slice groups based on Eq. (2), and generate M slice-level feature map sets for the informative slice groups using their corresponding feature extractors. For VFGC, we first train M SPSP modules together with their FC and softmax layers with 60 epochs and a batch size of 1, by using their corresponding slice-level feature map sets. We then employ M well-trained SPSP modules to fuse their corresponding slice-level feature maps into M feature vectors. After that, we train the FC module using the M feature vectors with 40 epochs and a batch size of 240 for volume-level feature generation and sMRI classification.

We train all the methods, including SVFR and the following comparison methods, on an NVIDIA RTX 2080Ti GPU. During the training process, these methods utilize the optimizer of stochastic gradient descent with the momentum [37] of 0.9 and with the weight decay of 0.0001. We initialize the learning rates of these methods to 0.001 and then lower them by a tenth every 15 epochs.

3.2. Comparison results with state-of-the-art methods

The proposed SVFR are compared with seven automatic CNN-based methods. They are the patch-based method in Ref. [18], the residual-based method in Ref. [20], the attention-based method in Ref. [21], the self-attention-based method in Ref. [26], the contrastive-learning-based method in Ref. [27], the ranking-based method in Ref. [38] and the slice-related method in Ref. [28]. As sMRI images the methods utilized are different, we implement all these methods and evaluate them on our downloaded sMRI images. To be specific, the patch-based method [18] devises a participant-specific lesion probability map for patch selection and then design a multilayer perceptron for AD diagnosis. The residual-based method [20] utilizes the 3D ResNet for AD diagnosis. The attention-based

method [21] designs an attention-based 3D ResNet to identify AD and NC. The attention block is realized by a convolutional layer and a rectified linear unit layer. The self-attention-based method [26] introduces a residual self-attention block to capture discriminative local, global and spatial features for AD diagnosis. The contrastive-learning-based method [27] devises a contrastive loss using sMRI images with group categories comparative information. The ranking-based method [38] proposes a triple-based ranking network architecture and loss to learn the ordinal relations among samples. The strategy of triplet loss in this method can be employed for sMRI image-based AD diagnosis, even though this method is for colorectal cancer grading. The slice-related method [28] designs a 2D+ε approach for AD diagnosis, as illustrated in Introduction. The comparison results are summarized in Table 4.

It can be seen that the proposed SVFR surpass the majority of the state-of-the-art methods and achieves comparable results compared with the best-performing approach. Specifically, our SVFR is superior to the patch-based and self-attention-based, and residual-based methods referring to all the metrics, especially outperforming the patch-based and self-attention-based methods a lot. Moreover, our SVFR is better than the attention-based and slice-based methods in terms of ACC, SEN, PRE, and F1, and a little lower than the two comparison methods about SPE. Moreover, our SVFR outperforms the contrastive-learning-based method regarding to ACC, SEN, and F1 by 5.81 %, 11.91 %, and 4.66 %, and is lower than the comparison method regarding to SPE and PRE by 2.27 % and 2.8 %. Additionally, our SVFR achieves better SEN, almost the same ACC and F1, and lower SPE and PRE values compared with the ranking-based method. These findings strongly indicate that the proposed SVFR holds substantial potential for AD diagnosis.

3.3. Selection of the hyper-parameters

There are three hyper-parameters in our proposed framework, which are K in K-means and the threshold θ of the first phase of SFE, and M_S in the SPSP block of the second phase. Among the three hyper-parameters, K and M_S determinate the structures of the framework, and are first evaluated. After that, for ease of description, we determine the value of the threshold θ in the section of 3.4.

We investigate K and M_S based on the classification accuracy on the validation set. Specifically, we vary K in {5, 7, 9, 11, 13} and M_S in {1, 2, 3}, and summarize the results in Fig. 5. We can see that the accuracy achieves the highest when K is set as 9 and M_S is set as 2, respectively. As such, all the experiments were evaluated with the two values.

3.4. Performance of the two phases of S2Veer and related decisions for the final framework

In this section, we display the performance of the two phases of SVFR, i.e, SFE and VFGC, and meanwhile illustrate how we select the value of the hyper-parameter θ by two steps. Moreover, we also illustrate that how we made decisions to obtain the final framework based on the performance of the two phases.

Evaluation for SFE. During this phase, we display the ACC values of the 9 clusters in the validation set first and then illustrate the candidate informative slice groups. Further, we show the ACC values of the candidate informative slice groups in the test set.

Table 5 summarizes the ACC values of the 9 slice groups in the validation set, i.e, S_1, S_2, …, and S_9. We can see that the ACC values among these slice groups vary a lot. We select informative slice groups from the 9 slice groups based on Eq. (2). Specifically, in order to preserve more information, we set the preliminary value of θ to a low value of 70 %. Therefore, S_1, S_4, S_5, and S_7 are regarded as the candidate informative slice groups. Inversely, S_2, S_3, S_6, S_8, and S_9 are the uninformative slice groups.

For the test set, we first cluster the slice images into 9 groups based on the cluster centers, obtained based on the training and validation sets. After that, we select and evaluate the four candidate informative slice groups, i.e, S_1, S_4, S_5, and S_7, using the four corresponding well-trained ResNets. Their ACC values are shown in Table 6. We find that S_1, S_4, and S_7 in the test set hold similar ACC to those in the validation set, demonstrating superior performance. However, the ACC value of S_5 in the test set is lower than that in the validation set by 8.27 %.

From Tables 5 and 6, we can see that, compared with the candidate informative slice group S_5, the three candidate informative slice groups, i.e, S_1, S_4, and S_7**,** hold more stable and better performance, which lays good foundations for the follow-up feature fusion and classification.

Evaluation for VFGC. We first display the ACC values of the four candidate informative groups in the validation set for the slice group classification, by using the SPSP module and the corresponding classifier, and meanwhile determine the value of θ and the informative slice groups based on the classification results. After that, we display the final performance of the FC module for volume-level feature generation and final classification between AD and NC.

Table 7 summarizes the ACC values of the four candidate informative slice groups in the validation set. We find that, in the

Table 4
Comparison results with state-of-the-art methods on the test set (%).

Methods	Publication	ACC	SPE	SEN	PRE	F1
Patch [18]	Brain 2020	78.95	78.57	79.41	75.00	77.14
Residual [20]	ISBI 2017	85.5	80.95	91.18	79.49	84.93
Attention [21]	ISBI 2019	84.21	85.71	82.35	82.35	82.35
Self-attention [26]	JBHI 2020	76.47	70.59	82.35	73.68	77.78
Contrastive [27]	CMPB 2021	82.35	84.62	80.95	89.47	85.00
Ranking [38]	MICCAI 2021	*88.24*	*92.31*	85.71	*94.74*	*90.00*
Slice [29]	ICMR 2017	80.26	83.33	76.47	78.78	77.61
Ours	_	**88.16**	**82.35**	**92.86**	**86.67**	**89.66**

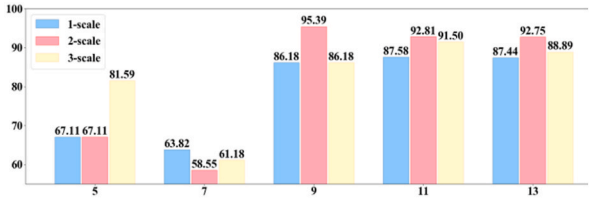


Fig. 5. Classification accuracy of the validation set. The *x*-axis denotes the *K* values, the *y*-axis represents the accuracy values, and the three colors denote different M_S values.

Table 5
ACC values of the 9 slice groups in the validation set (%).

	S_1	S_2	S_3	S_4	S_5
ACC	**81.35**	56.92	61.50	**76.31**	**74.42**
	S_6	S_7	S_8	S_9	
ACC	67.91	**79.32**	65.67	62.00	

Table 6
ACC values of the candidates of informative slice groups in the test set (%).

	S_1	S_4	S_5	S_7
ACC	**80.07**	**75.96**	**66.15**	**74.44**

Table 7
ACC values of the candidate informative slice groups in the validation set for the slice group classification (%).

	S_1	S_4	S_5	S_7
ACC	**88.16**	**86.84**	57.24	**87.42**

validation set, the three candidate informative slice groups of S_1, S_4, and S_7 achieve higher ACC values compared with the ACC values of S_1, S_4, and S_7 in Table 5. Specifically, the ACC values of S_1, S_4, and S_7 in Table 7 are 6.81 %, 10.53 %, and 8.10 % larger than those in Table 5. However, the ACC value of S_5 in Table 7 is 57.24 %, much lower than that in Table 5 and slightly better than random guessing. Based on the aforementioned results, we set the final value of θ is set to 75 %. Therefore, S_1, S_4, and S_7 is selected as informative slice groups and S_5 is an uninformative slice group.

Table 8 summarizes the ACC values of the three informative slice groups in the test set. Similarly, we find that the three informative slice groups achieve higher ACC values when compared with the ACC values of the three informative slice groups in the test set in Table 6. Specifically, the ACC values of S_1, S_4, and S_7 in Table 5 are 2.82 %, 8.25 %, and 9.77 % bigger than those in Table 6. These results demonstrate the effectiveness of the SPSP module for intra-group feature map fusion. Furthermore, we find that the ACC values of S_1, S_4, and S_7 in the test set for slice group classification are close to those in the validation set in Table 7. This result indicates the better generalization ability of the SPSP module for intra-group feature map fusion.

Fig. 6 displays the performance of both the validation and test sets using the FC module for volume-level feature generation and final classification, where the blue histograms denote the five metrics of the validation set and the red histograms represent the performance of the test set. We find that the ACC values of the validation set and the test set in Fig. 6 separately increase by 7.23 % and 5.27 % when compared with the ACC values of the validation and test sets in Table 7. This result indicates that the FC block can effectively fuse the inter-group feature vectors and further improve the accuracy of the classification between AD and NC.

It is note that the value of θ is determined by two steps. Initially, we set the preliminary value of θ to a low value of 70 % based on the results of Table 5 (i.e., ACC values of the 9 slice groups in the validation set), leading to the selection of four candidate informative groups, namely, S_1, S_4, S_5, and S_7. In the subsequent phase, we count the ACC values of the four candidate informative groups in the validation set for the slice group classification, as shown in Table 7. The results manifests that three of the candidates– S_1, S_4, and S_7– demonstrate elevated ACC values relative to their preceding values detailed in Table 5. Contrarily, the ACC value for the S_5 group, as indexed in Table 7, stands at 57.24 %, a marked decrement from its initial representation in Table 5. This particular deviation, which

Table 8
ACC values of the informative slice groups in the test set for the slice group classification (%).

	S_1	S_4	S_7
ACC	**82.89**	**84.21**	**84.21**

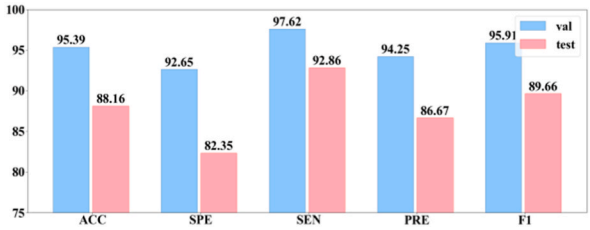


Fig. 6. Performance of the validation and test sets using the FC module (%).

only marginally surpasses random conjecture, signifies the non-informative nature of the S_5 group. Endeavoring to retain high-quality informative slice groups, a synthesis of insights derived from Tables 5 and 7 is executed, culminating in the ultimate value of θ at 75 %. Consequently, S_1, S_4, and S_7 emerge as the elected informative slice groups.

3.5. Ablation study

We conduct an ablation study to validate the effectiveness of the informative slice groups, the intra-group feature map fusion, and the inter-group feature fusion and classification. The results are summarized in Table 9. Specifically, the first line is the results of using the selected informative slices for subject-level classification. Since the numbers of different subjects are variable, we first predicted slice-level labels and utilize the majority voting for subject-level classification. The second line is the results of utilizing the SPSP module to cope with the variable numbers of slices, obtaining intra-group-level labels, and finally employing the majority voting for subject-level classification. The last line is the results of the proposed method.

We can see that the performance of the second line is better than that of the first line. This indicates the effectiveness of the intra-group feature map fusion. Additionally, the performance of the last line is the best. This indicates that the combining the three steps is most effective.

3.6. Visualization of the clustering results and the selected slice groups

We display the clustered 9 slice groups and the selected informative slice groups, which are highlighted in red rectangles, in Fig. 7. We can see that, the slices in each group have similar morphology. Specifically, Cluster 1 mainly contains cortices, hippocampi, cerebellum, and pons. Cluster 2, 3, 8, and 9 are the cortices with different sizes. Cluster 4 mainly includes lateral ventricles, hippocampi, and cortices. Cluster 5 mainly contains lateral ventricles and cortices. Cluster 6 contain the regions of cortices, cerebellums, and pons. Cluster 7 mainly shows cerebellums, where the cerebellums in Cluster 6 is smaller than these in Cluster 7. Moreover, we find that the informative slice groups of Cluster 1 and 4 contain an important AD biomarker, i.e., the obviously atrophic hippocampi. Additionally, Cluster 4 contains the biomarker of obviously atrophic cortices. We infer that the reason of Cluster 6 is selected as the informative slice group is because it contains the large size of cerellums, which are reported to be a potential biomark of AD [39].

4. Conclusions and future work

In this paper, we investigate the slice-related methods for AD diagnosis. Specifically, we propose the novel SVFR framework to classify sMRI image for AD diagnosis. SVFR consists of two phases: SFE and VFGC. SFE aims to select informative slice groups and extract their discriminative slice-level features, by taking advantage of CNNs and the clustering model. The purpose of VFGC is to fuse the slice-level features into volume-level features and further to distinguish AD from NC, by the proposed SPSP module and FC module. Experimental results on the public ADNI dataset demonstrate the effectiveness of SFE in selecting informative slice groups and that of VFGC in aggregating slice-level features for image classification. Moreover, the combination of SFE and VLFC boosts the performance of SVFR for AD diagnosis by leveraging the benefits of slice images, i.e, the large number and the small size.

In the future, we will focus on the following directions. At first, we will explore deep clustering methods to cluster slice images based on their high-level features, such as morphological features. This approach is necessary as the current *K*-means strategy is relatively straightforward and can be influenced by shallow features like grayscale. Secondly, we will optimize the framework of SVFR, since the current framework is a little separate and the FC module is a little over-fitting. Thirdly, we will take the heterogeneity caused by the sub-types of the disease into consideration. Finally, we will conduct more tasks, such as identifying mild cognitive impairment, stable mild cognitive impairment, and progressive mild cognitive impairment, to validate the robustness and generalization of the new framework.

Data availability statement

We utilize the public dataset ADNI, as described in Section 2.3. We also provide the subject ids we utilize in supplementary material.

Table 9
Ablation study results on the test set (%).

Informative slice classfication	Intra-group classification	Inter-group classification	ACC	SPE	SEN	PRE	F1
✓			78.95	73.53	83.33	79.55	81.40
✓	✓		86.67	82.35	90.24	86.05	88.10
✓	✓	✓	**88.16**	**82.35**	**92.86**	**86.67**	**89.66**

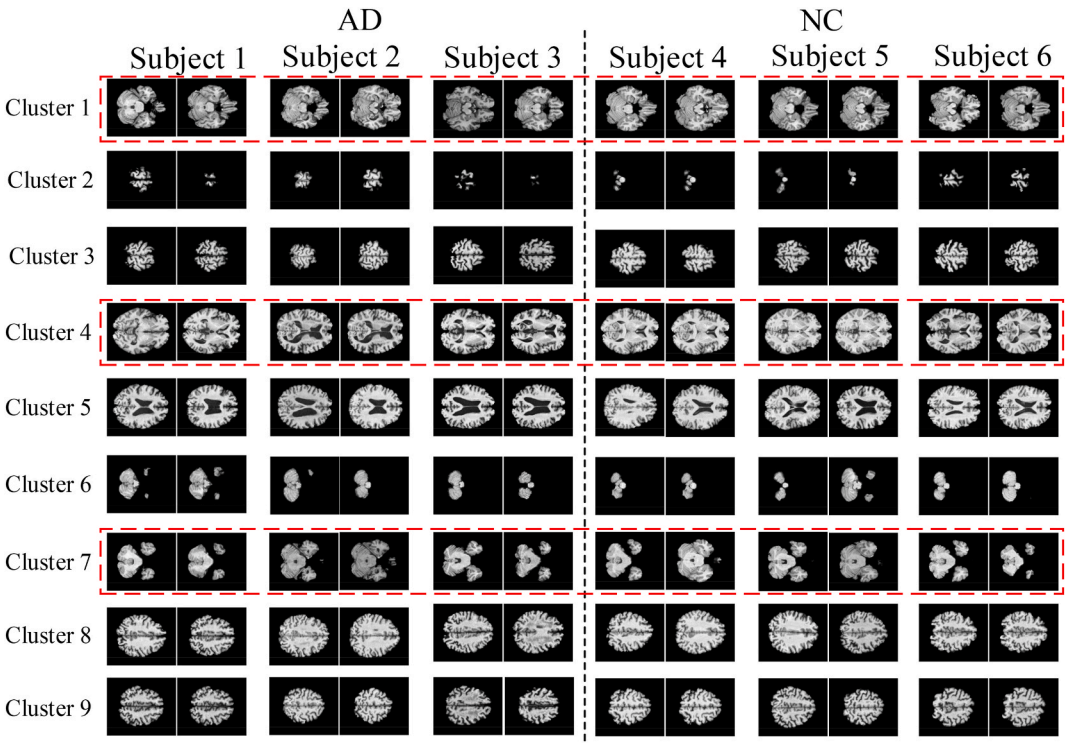


Fig. 7. Visualization of the 9 slice groups and the selected informative slice groups that are marked by red rectangles

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Rubing Wang: Conceptualization, Methodology, Software, Writing – original draft. **Linlin Gao:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Xiaoling Zhang:** Data curation, Formal analysis, Writing – review & editing. **Jinming Han:** Data curation, Formal analysis, Writing – review & editing.

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.

Acknowledgments

We would like to thank the anonymous reviewers for their time and their valuable comments. This work is supported in part by the Natural Science Foundation of Zhejiang Province [grant number LQ20F020013], Medical and Health Technology Project of Zhejiang Province [grant number 2023KY1149], and Science and Technology on Public Welfare Project of Ningbo [grant number:2022S061].

References

[1] P. Christina, World Alzheimer's Report 2018. Alzheimer's Disease Internations: World Alzheimer Report, 2018, pp. 1–48.
[2] Alzheimer's Disease International, World Alzheimer Report 2019: Attitudes to Dementia, 2019.
[3] L. An, E. Adeli, M. Liu, J. Zhang, S.W. Lee, D. Shen, A hierarchical feature and sample selection framework and its application for Alzheimer's disease diagnosis, Sci. Rep. 7 (2017), 45269.
[4] Nordberg, Agneta, Dementia in 2014: towards early diagnosis in Alzheimer disease, Nat. Rev. Neurol. 11 (2) (2015) 69.
[5] Alzheimer's Association, 2019, Alzheimer's disease facts and figures 15 (3) (2019) 321–387.
[6] G.B. Frisoni, N.C. Fox, C.R. Jack, P. Scheltens, P.M. Thompson, The clinical use of structural MRI in Alzheimer disease, Nat. Rev. Neurol. 6 (2) (2010) 67–77.
[7] J. Zhang, M. Liu, A. Lee, Y. Gao, D. Shen, Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images, IEEE J. Biomed. Health Inf. 21 (2017) 1607–1616.
[8] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, IEEE Intell. Syst. Their Appl. 13 (4) (1998) 18–28.
[9] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, 1989.
[10] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, O. Colliot, Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation, Med. Image Anal. 63 (2020), 101694.
[11] R. Cui, M. Liu, Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis, IEEE J Biomed Health Informat 23 (2018) 2099–2107.
[12] Z. Xia, G. Yue, Y. Xu, C. Feng, M. Yang, T. Wang, B. Lei, A Novel End-To-End Hybrid Network for Alzheimer's Disease Detection Using 3D CNN and 3D CLSTM. ISBI 2020: 2020 IEEE 17th International Symposium on Biomedical Imaging; 2020 Apr 3-7, IEEE, Iowa City, IA, USA. New York, 2020, pp. 1–4.
[13] M. Liu, J. Zhang, D. Nie, P. Yap, D. Shen, Anatomical landmark based deep feature representation for MR images in brain disease diagnosis, IEEE J Biomed Health Informat 22 (2018) 1476–1485.
[14] M. Liu, J. Zhang, E. Adeli, D. Shen, Landmark-based deep multi-instance learning for brain disease diagnosis, Med. Image Anal. 43 (2018) 157–168.
[15] M. Liu, J. Zhang, E. Adeli, D. Shen, Deep multi-task multi-channel learning for joint classification and regression of brain status, in: MICCAI 2017: Proceedings of the 20th International Conference on Medical Image Computing and Computer-Assisted Intervention; 2017 Sep 11-13, Springer, Quebec City, QC, Canada. Berlin, 2017, pp. 3–11.
[16] M. Liu, J. Zhang, E. Adeli, D. Shen, Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 66 (2018) 1195–1206.
[17] C. Lian, M. Liu, J. Zhang, D. Shen, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE Trans Pattern Anal Mach Intel 42 (2018) 880–893.
[18] S. Qiu, P.S. Joshi, M.I. Miller, C. Xue, X. Zhou, C. Karjadi, G.H. Chang, A.S. Joshi, B. Dwyer, S. Zhu, M. Kaku, Y. Zhou, Y.J. Alderazi, A. Swaminathan, S. Kedar, M. Saint-Hilaire, S.H. Auerbach, J. Yuan, E.A. Sartor, R. Au, V.B. Kolachalama, Development and validation of an interpretable deep learning framework for Alzheimer's disease classification, Brain 143 (2020) 1920–1933.
[19] B. Long, C.P. Yu, T. Konkle, Mid-level visual features underlie the high-level categorical organization of the ventral stream, Proceedings of the National Academy of Sciences of the United States of America 115 (38) (2018) 201719616–201719625.
[20] S. Korolev, A. Safiullin, M. Belyaev, Y. Dodonava, Residual and plain convolutional neural networks for 3D brain MRI classification, in: ISBI 2017: 2017 IEEE 14th International Symposium on Biomedical Imaging; 2017 Ap 18-21, IEEE, Melbourne, Australia. New York, 2017, pp. 835–838.
[21] D. Jin, J. Xu, K. Zhao, F. Hu, Z. Yang, B. Liu, T. Jiang, Y. Liu, Attention-based 3D convolutional network for Alzheimer's disease diagnosis and biomarkers exploration, in: ISBI 2019: 2019 IEEE 16th International Symposium on Biomedical Imaging; Apr 8-11, IEEE, Venice, Italy. New York, 2019, pp. 1047–1051.
[22] Q. Li, X. Xing, Y. Sun, B. Xiao, H. Wei, Q. Huo, M. Zhang, X.S. Zhou, Y. Zhan, Z. Xue, F. Shi, Novel iterative attention focusing strategy for joint pathology localization and prediction of MCI progression, in: MICCAI 2019: Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention; 2019 Oct 13-17, Berlin: Springer, Shenzhen, China, 2019, pp. 307–315.
[23] C. Lian, M. Liu, L. Wang, D. Shen, End-to-end dementia status prediction from brain MRI using multi-task weakly-supervised attention network, in: MICCAI 2019: Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention; 2019 Oct 13-17, Berlin: Springer, Shenzhen, China, 2019, pp. 158–167.
[24] L. Zhang, L. Wang, D. Zhu, Jointly Analyzing Alzheimer's Disease Related Structure-Function Using Deep Cross-Model Attention Network. ISBI 2020: 2020 IEEE 17th International Symposium on Biomedical Imaging; 2020 Apr 3-7, IEEE, Iowa City, IA, USA. New York, 2020, pp. 563–567.
[25] Z. Zhang, L. Gao, G. Jin, L. Guo, Y. Yao, L. Dong, Han J. and the Alzheimer's Disease NeuroImaging Initiative. THAN: task-driven hierarchical attention network for the diagnosis of mild cognitive impairment and Alzheimer's disease, Quant. Imag. Med. Surg. 11 (7) (2021) 3338.
[26] X. Zhang, L. Han, W. Zhu, L. Sun, D. Zhang, An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE Journal of Biomedical and Health Informatics 26 (2020) 5289–5297.
[27] H. Qiao, L. Chen, Z. Ye, F. Zhu, Early Alzheimer's disease diagnosis with the contrastive loss using paired structural MRIs, Comput. Methods Progr. Biomed. 208 (2021), 106282.
[28] K. Aderghal, J. Benois-Pineau, K. Afdel, Classification of sMRI for Alzheimer's disease diagnosis with CNN: single siamese networks with 2D+? Approach and fusion on ADNI, in: ICMR 2017: Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval; June 6-9, ACM, Bucharest, Romania. New York, 2017, pp. 494–498.
[29] J. Islam, Y. Zhang, A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data, in: International Conference on Brain Informatics, Lecture Notes in Computer Science, Springer, Cham, 2017, pp. 213–222.
[30] A. Mehmood, M. Maqsood, M. Bashir, Y. Shuyuan, A deep siamese convolution neural network for multi-class classification of alzheimer disease, Brain Sci. 10 (2) (2020) 84–98.
[31] Kaiming He, et al., Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[32] A. Valliani, A. Soni, Deep residual nets for improved Alzheimer's diagnosis, in: 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics, ACM, 2017, p. 615.
[33] L. Gao, H. Pan, F. Liu, X. Xie, Z. Zhang, J. Han, Brain Disease Diagnosis Using Deep Learning Features from Longitudinal MR Images, APWeb/WAIM, 2018.
[34] J. Sivic, A. Zisserm, Video Google: a text retrieval approach to object matching in videos, Proceedings Ninth IEEE International Conference on Computer Vision (2) (2003) 1470–1477.
[35] S. Qiu, G.H. Chang, M. Panagia, D.M. Gopal, R. Au, V.B. Kolachalama, Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment, Alzheimer's Dementia: Diagnosis, Assessment & Disease Monitoring 10 (2018) 737–749.
[36] J.A. Hartigan, M.A. Wong, A k-means clustering algorithm, Applied Statistics 28 (1) (1979).
[37] N. Qian, On the momentum term in gradient descent learning algorithms, Neural Network. 12 (1999) 145–151.
[38] T.T. Vuong, K. Kim, B. Song, J.T. Kwak, Ranking loss: a ranking-based deep neural network for colorectal cancer grading in pathology images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2021.
[39] V. Gupta, S. Booth, J.H. Ko, Hypermetabolic cerebellar connectome in Alzheimer's disease, Brain Connect. (2020).