

Article

Not peer-reviewed version

Resizer Swin Transformer based Classification using sMRI for Alzheimer's Disease

[Yihang Huang](#) and [Wan Li](#) *

Posted Date: 12 July 2023

doi: 10.20944/preprints202307.0799.v1

Keywords: Swin Transformer; Resizer; CNN; structural MRI; Alzheimer's disease



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Resizer Swin Transformer Based Classification Using sMRI for Alzheimer's Disease

Yihang Huang¹ and Wan Li^{1,*}

¹ School of Computer Science and Engineering, Beijing Technology and Business University, Beijing 100048 China

* Correspondence: wanli@btbu.edu.cn

Abstract: Structural magnetic resonance imaging (sMRI) is widely used in the clinical diagnosis of diseases due to its advantages: high definition and noninvasive. Therefore, computer-aided diagnosis based on sMRI images is broadly applied in classifying Alzheimer's disease (AD). Due to the excellent performance of Transformer in computer vision, Vision Transformer (ViT) has been employed for AD classification in recent years. ViT relies on access to large datasets, while the sample size of brain imaging datasets is relatively insufficient. Moreover, the pre-processing procedures of brain sMRI images are complex and labor-intensive. To overcome the limitations mentioned above, we propose Resizer Swin Transformer (RST), a deep learning model that can extract information from brain sMRI images that are only briefly processed to achieve multi-scale and cross-channel features. In addition, we pre-trained our RST on a natural image dataset and obtained better performance. The experimental results of ADNI and AIBL datasets prove that RST can achieve better classification performance in AD prediction compared with CNN-based and Transformer models.

Keywords: Swin Transformer; Resizer; CNN; structural MRI; Alzheimer's disease

1. Introduction

Classification algorithms applied for medical imaging have been a major field of research over the past few decades. The recent success of deep learning techniques has inspired new research and development efforts to improve classification performance and develop novel models for various complex clinical tasks [1,2].

Alzheimer's disease (AD) is an irreversible chronic neurodegenerative disease that progressively impairs cognitive and behavioral functions. Because of the high definition and noninvasive visualization, structural magnetic resonance imaging (sMRI) is one of the most common imaging techniques for AD identification in both clinical and research [3].

Therefore, deep learning algorithms have been applied increasingly frequently for AD classification using sMRI images since they can replace time and labor-consuming procedures such as feature extraction. In particular, there is no need for feature selection in the deep learning model, but it can also automatically learn sophisticated features by itself [4].

In this study, we propose a deep learning model for AD classification called Resizer Swin Transformer (RST) using sMRI images. The Resizer module creates a learnable image scaling, acquiring characteristics supporting Swin Transformer classification [5]. The cross-window connection realized by the moving window mechanism of the Swin Transformer and the patch merging provides multi-scale learning. Additionally, a convolutional neural network (CNN) model is employed to colorize the sMRI images to enable multi-channel learning [6].

Furthermore, a majority of current AD-related classification methods demand pre-processing procedures like calibration, skull stripping, and alignment to standard templates [7]. Particularly, skull stripping is regarded as a crucial pre-processing step that cannot be skipped since it can lessen the influence of irrelevant data on classification outcomes and simplify the computing process of the model [8]. However, insufficient skull stripping or over-processing caused by the toolkit might result in the loss of edge information, and the data must then be manually checked after processing, which takes time and significantly lowers the already limited data set relevant to AD.

2. Background and related work

2.1. CNN-based classification for AD

The main types of existing AD classification techniques based on the CNN model include the region of interest (ROI), voxel, patch, and attention mechanism. Noticeably, this categorization does not imply that the four approaches mentioned above are entirely distinct.

The ROI-based methods demand the pre-segmentation of brain regions based on prior knowledge, such as brain atlases. For instance, Wang et al. segmented the hippocampus area, one of the most AD-sensitive regions, and employed a dense convolutional neural network (Dense CNN) model to classify NC and AD [9]. On the other hand, one study tried each brain region to train a 3D-CNN ensemble model [10].

The voxel-based technique, which does not require any prior knowledge or laborious preparation, obtains features directly from sMRI images and fully exploits the global characteristics [11]. For instance, Hazarika et al. employed LeNet, AlexNet, VGG, DenseNet, and other models to classify AD while evaluating their effectiveness [12]. In addition, the classification accuracy may be increased by combining the CNN model with transfer learning [13] and data augmentation [14].

Others noticed that only localized brain areas in early AD patients exhibit minor structural abnormalities, which leads to a possibility that features obtained at the voxel or region level cannot contribute to AD identification completely. Patch, the intermediate level between voxel and region, has gained more attention. Flexibility in size and location, patch-based models improve classification accuracy and avoid laborious pre-processing procedures. However, the choice of patches significantly impacts the categorization outcomes. By using anatomical marker detectors, Liu et al. identified the patches discriminative to AD first, then trained a CNN model to learn from those patches [15]. The landmark-based deep multi-instance learning (LDMIL) system was introduced the following year to learn local patch information as well as global information from all patches [16].

Another frequently employed module recently is the attention mechanism, which is quite useful for pinpointing AD-sensitive areas. For example, Zhang et al. added the attention mechanism to the ResNet framework, which effectively enhances the gray matter features information and increases the accuracy of AD diagnosis [17].

2.2. Transformer-based classification for AD

One of the strongest deep learning models available today is Transformer, and a key component is the attention mechanism [18]. Despite the original purpose of the Transformer being for natural language processing (NLP), abundant studies have now demonstrated that Transformer-based models may reach superior performance in computer vision (vision Transformer, ViT). Because of the multi-headed attention mechanism of ViT, Li et al. integrated CNN to capture the relationships between distant brain areas. In this study, ViT received input from the feature maps extracted using the convolutional layer [19]. Another approach proposed by Jang et al. combined the ViT with a CNN structure that has an inductive bias, and the feature maps were generated using 3D ResNet. The 3D information provided by sMRI images can efficiently assess local aberrant characteristics associated with AD and link markers from multiplanar and multilayer slices to gather distant details [20]. Due to the inherent lack of inductive bias in the ViT-related study, a significant quantity of data is needed to train the model. Natural images share similar fundamental properties with brain sMRI images, including texture, edges, shape, etc. Hence, Lyu et al. applied the ImageNet [21] dataset to pre-train the ViT model using joint transfer learning first to address the limited brain imaging data [22].

2.3. Limitations of current research

1. ImageNet is a natural image dataset that each image contains three colors, although the coronal slices taken from 3D sMRI images are only gray-scale. Thereby, using sMRI images directly as the input of the ViT model implies delivering the same images into all three channels, which is a

total waste of computational resources. In addition, the gray-scale image deviates significantly from the original RGB color image on each channel.

2. The majority of the current studies demand strict procedures for sMRI image preparation. Skull stripping is one of the crucial components. However, there are still some issues with employing SPM12 for skull stripping, such as partial stripping or loss of edge information due to over-processing. Therefore, the data must be visually checked and manually selected afterward. While eliminating the background (black region) of sMRI images has also become a critical step of pre-processing, the capacity of the deep learning models is hampered by the growing complexity of the pre-processing process.

3. The ViT model can only extract features at the same scale because its computational complexity is proportional to the square of the image size. Besides, it can only extract features on one scale at one time. However, the damaged brain areas of AD patients are typically subtle, making it possible to overlook some regionally specific details.

3. Methods

3.1. Slices options

The center slices of the sMRI images were chosen for this study because it contains most of the brain information. The classification effect of the coronal plane was found to be slightly inferior to that of the sagittal and axial planes in the comparison experiments. Still, the classification performance of the sagittal and axial planes was not significantly different. Therefore, 30 consecutive 2D slices of size 127×181 in the axial plane were employed in this study.

3.2. Model structure

3.2.1. Overview of Resize Swin Transformer Network

Li et al. applied joint transfer learning approaches to classify AD by first using 3D gray matter volume images as input, extracting local information via ResNet to generate feature maps, and finally adding location encoding to the maps that were input to ViT [19]. Swin Transformer performs multi-scale learning through cross-window connection and patch merging, as well as decreases the computational complexity of ViT from square level to linear level [23]. Therefore, this study proposes the ReSwin Transformer (RST), a novel network structure that combines CNN and Swin Transformer to realize the AD classification using sMRI images.

The RST model proposed in this study is shown in Figure 1. Axial plane sMRI slices X are the input to the RST. A CNN model is first utilized to map the gray-scale image from one channel to three channels to generate color images. The Resizer module then scales the input image proportionally and accentuates information essential to classification according to the Swin Transformer during the training procedure. The images are separated into non-overlapping patches using the patch-splitting module. Each patch is regarded as a separate *token*. A linear embedding layer then maps the *token* to a size C channel as the Swin Transformer input. Finally, a SoftMax layer summarizes the AD and NC classification predictions from all the slices.

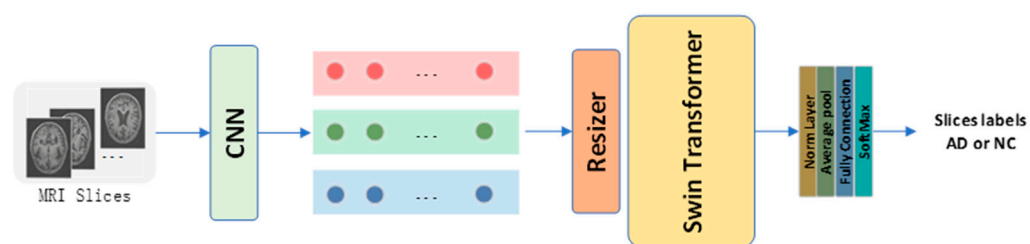


Figure 1. The overall structure of the proposed model. The CNN module is used to convert a single-channel gray-scale image of the axial sMRI slices into a three-channel RGB image. The Resizer module

resizes the inputting image and removes invalid parts. The Swin Transformer identifies features that are useful for classification. Finally, the Softmax layer generates classification scores for each slice.

3.2.2. CNN module

sMRI scans only contain gray-scale information, generally known as a single-channel imaging technique. On the other side, various widely used CNN models, such as AlexNet [24], ResNet [25], and EfficientNet [26], are three-channel models. Hence, the single-channel gray-scale images are often repeated three times as input when the joint transfer learning algorithms are involved. Inspired by the colorization of lung MRI images to restore the color images of lungs observed by human eyes proposed in [27] and to address the issue of wasted computational resources, our study converts brain sMRI images to color images using the CNN networks to achieve cross-channel learning from single-channel to three-channels [6]. Figure 2 illustrates the CNN structure.

According to the given brightness Y on the gray-scale image, two chroma channels a and b were generated based on the CIELab color space. Then, the brightness-chroma color space is transformed to RGB color space using the image and OpenCV library. For a given brightness channel $X \in \mathbb{R}^{H \times W \times 1}$, we converted it to $Y \in \mathbb{R}^{H \times W \times 2}$ via the map $\hat{Y} = \mathcal{F}(X)$, where H and W are the dimensions of the image. In addition, for a given X , its probability distribution was also obtained $\hat{Z} = H(X)$. Where $\hat{Z} \in [0,1]^{H \times W \times Q}$ and Q represents the number of output spaces from the channel a and b . The following equation compares the true value with \hat{Z} through polynomial cross-entropy loss $L_{cl}(\cdot, \cdot)$:

$$L_{cl}(\hat{Z}, Z) = - \sum_{h,w} v(Z_{h,w}) \sum_q Z_{h,w,q} \log(\hat{Z}_{h,w,q})$$

In particular, the real color Y is converted into a vector Z via $Z = \phi_{gt}^{-1}(Y)$, $v(\cdot)$ is the weight that measures the rarity of the color class and thus rebalances the loss. Finally, the probability distribution \hat{Z} is mapped to color values by the function $\hat{Y} = \phi(\hat{Z})$.

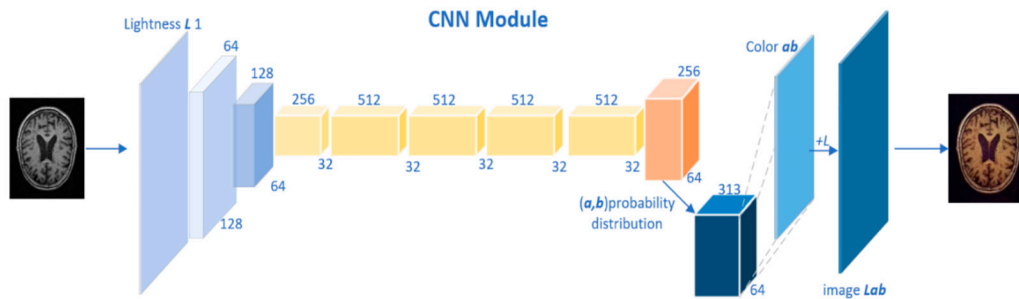


Figure 2. Colorization CNN network structure. Each block includes 2 or 3 conv and ReLU layers, followed by a BatchNorm layer.

3.2.3. Resizer Module

In the field of deep learning for image processing, the input image size is typically scaled to 224×224 , and both training and inference are carried out at that resolution. Image scaling often uses both bilinear and trilinear interpolation currently. In actuality, this modification does not improve the image of the network, which somewhat even reduces the performance of the model [28]. Therefore, this experiment substitutes the learnable Resizer module for the original linear interpolation after CNN transfers the single-channel to the three-channel color space. Resizer seeks to significantly improve classification performance by learning attributes favorable to Swin categorization by collaborative training with the backbone network, in contrast to other approaches that resize images to enhance human eye perception. The Resizer module is shown in Figure 3.

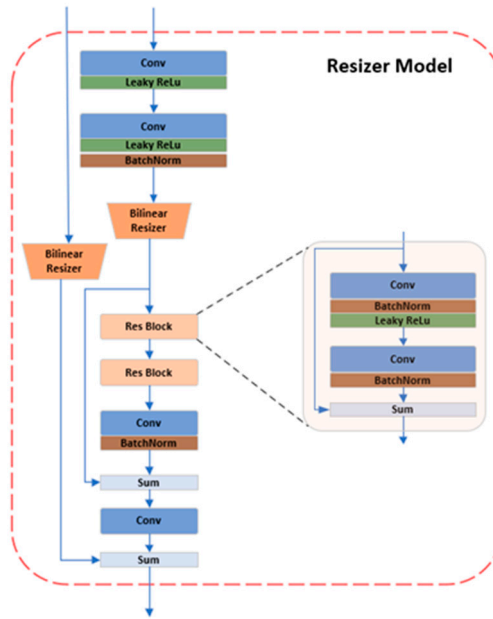


Figure 3. Resizer module in our proposed model for resizing images.

3.2.4. Swin Transformer

Swin Transformer utilizes within-window calculation of self-attention to increase modeling efficiency. Beginning in the top left corner, the window evenly and non-overlappingly divides the image into sections. Assuming that there are $M \times M$ patches in a window, the next module utilizes a different window than the previous layer and moves $(\lfloor \frac{M}{2} \rfloor, \lfloor \frac{M}{2} \rfloor)$ patch from the original window when the window-based self-attention module completes its computation. The calculation process of the Swin Transformer block is as follows:

$$\begin{aligned}\hat{Z}^l &= \text{W-MSA} \left(\text{LN}(Z^{l-1}) \right) + (Z^{l-1}), \\ Z^l &= \text{MLP} \left(\text{LN}(\hat{Z}^l) \right) + \hat{Z}^l, \\ \hat{Z}^{l+1} &= \text{SW-MSA} \left(\text{LN}(Z^l) \right) + Z^l, \\ Z^{l+1} &= \text{MLP} \left(\text{LN}(\hat{Z}^{l+1}) \right) + \hat{Z}^{l+1}\end{aligned}$$

Where \hat{Z}^l and Z^l are multi-headed self-attentive based on the window and shifted window, respectively, and characterized by the output of the multilayer perceptron (MLP) module. Window movement enables the reciprocal learning of patches between several windows, so achieving the goal of global modeling. Figure 4 depicts the model of the Swin Transformer structure as well as the calculating procedure of the block.

Windows multi-head self-attention (W-MSA) adds relative position bias $B \in \mathbb{R}^{M^2 \times M^2}$ for each head in the calculation of multi-headed self-attention, as follows:

$$\text{Attention}(Q, K, V) = \text{SoftMax} \left(Q \frac{K^T}{\sqrt{d}} + B \right) V$$

Where, $Q, K, V \in \mathbb{R}^{M^2 \times d}$ are the query, key, and value matrices, respectively. d is the dimension of the key. The number of patches in the window is M^2 . Given that the relative location is between $[-M+1, M-1]$, the bias matrix $\hat{B} \in \mathbb{R}^{(2M-1) \times (2M-1)}$ is set. \hat{B} from whence the values for B are derived. Q, K and V are calculated from W_Q, W_K , and W_v by applying a linear transformation of Z^{l-1} .

On the other hand, shifted windows multi-head self-attention (SW-MSA) uses the circular window movement rule to compute the multi-headed self-attentiveness.

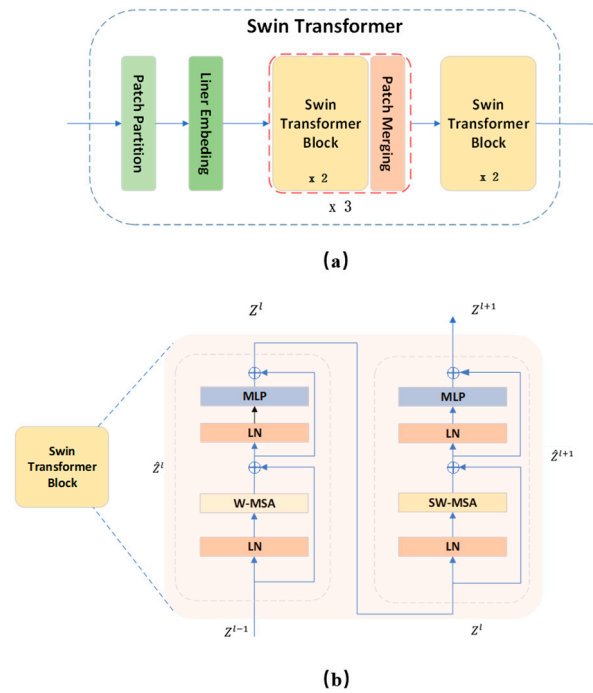


Figure 4. Swin Transformer and its block structure. Figure (a) shows the overall structure of the Swin Transformer. A Swin Transformer block contains two attention calculations. The specific computation process is shown in Figure (b).

4. Evaluation

4.1. Introduction of the datasets

The Australian Imaging, Biomarker & Lifestyle (AIBL, aibl.csiro.au) [29] and the Alzheimer's Disease Neuroimaging Initiative (ADNI, adni.loni.usc.edu) [30] were used in this study. The National Institutes of Health and the National Institute on Aging provided funds to establish the ADNI, the premier data center for AD research. Both datasets gather sMRI, fMRI (functional MRI), and positron emission computed tomography (PET) from AD and NC participants.

The data used in the study are collected from an MRI scanner built by MRI manufacturer SIEMENS. The slice thickness is 1.2 mm; the field strength is 3.0 Tesla. All the sMRI images were downloaded from ADNI-GO, ADNI1, and AIBL. The standard sMRI image pre-processing was conducted using SPM12 (fil.ion.ucl.ac.uk) on Matlab (R2022a), including format conversion, AC-PC correction, non-parametric non-uniform intensity normalization (N3), and alignment to the MNI standard template. The reconstructed images are $181 \times 217 \times 217$, the voxel size is $1 \times 1 \times 1 \text{ mm}^3$, and normalized intensity values are in the range of [0,1]. Table 1 displays the demographic information of the datasets.

The ADNI and AIBL data sets are openly accessible, but access to the information still needs official authorization. Additionally, no data may be shared without consent, and only approved researchers may utilize it for study.

Table 1. Demographics information of ADNI and AIBL datasets

Image dataset	AD	NC	Age	Sex(F/M)
ADNI (N=1188)	388	800	75.76±6.75 [56–96]	388
AIBL (N=847)	196	651	74.56±6.88 [52–96]	196

4.2. Training Setup

This experiment used the PyTorch deep learning framework to construct the proposed network model. Two NVIDIA 3080 TI GPUs were implemented on a server for training the classification task. The model is first transferred to the sMRI dataset after being pre-trained on the ImageNet-1K dataset. Since the Resizer module may shrink the images to eliminate a tiny amount of incorrect information (dark parts of sMRI images), cropping the input images is unnecessary. With a learning rate of 5×10^{-5} , a patch size of 4×4 , a batch size of 16, the number of epochs set to 50, and using cross-entropy loss as a loss function, all networks were trained using the Adam optimizer. The results of this study are based on the test set, with epoch.

4.3. Experimental Results

The classification performance was assessed using the following four common specificity indicators: accuracy (ACC), sensitivity (SEN), specificity (SPE), and precision (PRE). We compared our approach to every type of recent study in this area to demonstrate its superiority, and the findings are displayed in Table 2. Our strategy obviously outperforms earlier research in every index.

Table 2: Comparison of our proposed model with related studies

Models	Types	Classification Results			
		ACC%	SEN%	SPE%	PRE%
DenseCNN [9]	ROI	89.80	98.50	85.20	--
CNN [14]	Whole	93.00	92.00	94.00	--
LDMIL [16]	Patch	92.02±0.93	90.76±2.72	92.40±1.10	--
ResNet+Attention [17]	Attention	90.00	92.80	87.50	--
ResNet+ViT [19]	Transformer	92.26	88.98	94.04	--
CNN+ViT [20]	Transformer	90.58	--	--	--
CNN+ViT [22]	Transformer	96.80	--	--	97.20
Ours	Transformer	99.59	99.58	99.59	99.83

4.4. Ablation Experiments

We carried out several ablation experiments to maximize the experimental outcomes.

Experiment 1: We tested various batch sizes and learning rates to determine the settings that would produce the best experimental results. The experimental outcomes are displayed in Figure 5. Most curves are generally stable at epoch=50, yet some still vary remarkably. It reaches the most stable when the batch size is 16, and the learning rate is 5×10^{-5} . Therefore, the batch size is set to 16, and the learning rate is set at 5×10^{-5} , considering parameters like processing speed and classification performance. set at 5×10^{-5} , considering parameters like processing speed and classification performance.

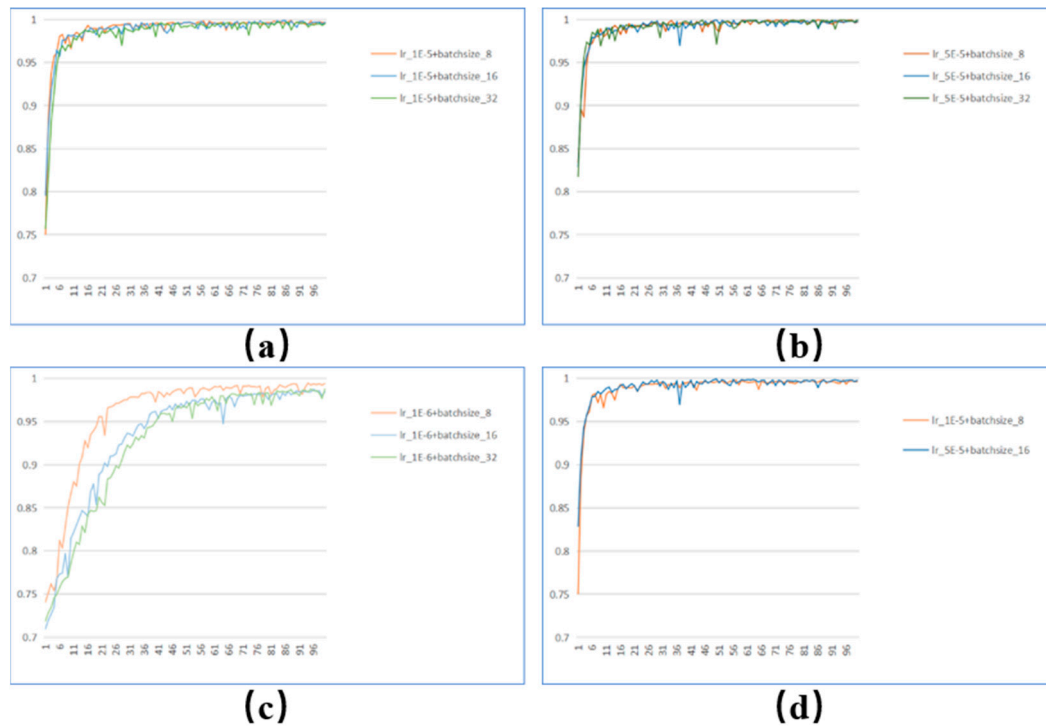


Figure 5. Figures (a), (b), and (c) show the variation of ACC with epoch for different batch size with learning rates of 1×10^{-5} , 5×10^{-5} , and 1×10^{-6} , respectively. Fig. (d) compares the two curves with the optimal variation of ACC

Experiment 2: We used sMRI slices in three different orientations. In Table 3, the experimental findings are displayed. The worst result was using the coronal plane slice without the skull stripping pre-processing procedure, which differed greatly from the findings of the other two orientations. Since most experiments were centered on the axial plane and the differences between the experimental outcomes in the axial and sagittal planes were not particularly apparent, axial plane slices were employed in this experiment.

Table 3: Comparisons of three directional slices using RST model

Types	Classification Results			
	ACC%	SEN%	SPE%	PRE%
Sagittal	99.69	99.74	99.67	99.54
Coronal	99.07	99.46	98.79	98.50
Axial	99.59	99.58	99.59	99.83

Experiment 3: Tests were run to show how well the various parts of our proposed RST model worked together. The test results are listed in Table 4. The RST model, as implemented by CNN, performs noticeably better than the other experiments, as shown in the table. In contrast, the step of skull stripping only slightly improves the performance of the RST model and is not proportional to the additional expense it incurs.

Table 4: Comparisons among different data types and network structures

Models	Data	ACC
RST	Not skull-stripping	98.74%

RST	2.5D skull-stripping	96.36%
RST	skull-stripping	98.99%
CNN+RST	skull-stripping	99.98%
CNN+ RST	Not skull-stripping	99.62%

Experiment 4: Investigations were carried out on several datasets to confirm the robustness of our proposed model. The outcomes of our trials are shown in Table 5. The table illustrates how our model performs better for various datasets. The analysis may be caused by the unevenness of the datasets and the image discrepancies between the datasets because the findings of AIBL, on the other hand, are worse than those of ADNI.

Table 5: Experimental Results of RST in Different Data Sets

Training data	Test Data	Classification Results			
		ACC%	SEN%	SPE%	PRE%
ADNI	ADNI	99.59	99.58	99.59	99.83
ADNI+AIBL	ADNI+AIBL	94.05	95.52	95.52	90.95
ADNI+AIBL	ADNI	99.75	99.45	99.45	99.54
ADNI+AIBL	AIBL	87.88	91.40	91.40	82.70
AIBL	AIBL	94.01	95.48	95.48	91.03

5. Conclusion and future work

This paper presents a Resizer Swin Transformer architecture that combines cross-channel learning with CNN and extracts features from two-dimensional axial plane slices of brain MRI. The analysis and visualization of the experimental data demonstrated the accuracy with which RST architecture can accomplish the categorization of AD as well as its strong adaptability to various data sets. This experiment does still have some flaws, though. The RST model has comparatively high model parameters, and cross-channel learning in conjunction with CNN enhances the classification performance using the model. Meanwhile, we discovered that the accuracy of the experimental findings is lower than 85% when the model trained on ADNI is directly assessed by AIBL. The model may undergo further development to overcome these problems and provide a lighter and more scalable model.

Author Contributions: Conceptualization, Y.H. and W.L.; methodology, Y.H. and W.L.; software, Y.H.; validation, Y.H. and W.L.; formal analysis, Y.H.; investigation, Y.H.; resources, W.L.; data curation, Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, W.L.; visualization, Y.H.; supervision, W.L.; project administration, W.L.; funding acquisition, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Research Foundation for Youth Scholars of Beijing Technology and Business University (PXM2020_014213_000017) under Grant QNJ2020-29.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: These data used in this paper were derived from the open-source datasets. Data sharing is not applicable to this article as no new data were created in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Burgos, N.; Colliot, O. Machine learning for classification and prediction of brain diseases: recent advances and upcoming challenges. *Current Opinion in Neurology*, 2020, 33(4), 439-450.
2. Segato, A.; Marzullo, A.; Calimeri, F.; De Momi, E. Artificial intelligence for brain diseases: A systematic review. *APL bioengineering*, 2020, 4(4), 041503.
3. Vemuri, P.; Jack, C. R. Role of structural MRI in Alzheimer's disease. *Alzheimer's research therapy*, 2010, 2(4), 1-10.
4. Dharwada, S.; Tembhurne, J.; Diwan, T. (2022, January). Multi-channel Deep Model for Classification of Alzheimer's Disease Using Transfer Learning. In *Distributed Computing and Intelligent Technology: 18th International Conference, ICDCIT 2022, Bhubaneswar, India, January 19–23, 2022, Proceedings* (pp. 245-259). Cham: Springer International Publishing.
5. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp: 10012-10022.
6. Zhang, R.; Isola, P.; Efros, A. Colorful image colorization. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14* (pp. 649-666). Springer International Publishing.
7. Yamanakkanavar, N.; Choi, J. Y.; Lee, B. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: a survey. *Sensors*, 2020, 20(11), 3243.
8. Druzhinina, P.; Kondrateva, E. The effect of skull-stripping on transfer learning for 3D MRI models: ADNI data. In *Medical Imaging with Deep Learning*. 2022.
9. Wang, Q.; Li, Y.; Zheng, C.; Xu, R. DenseCNN: A Densely Connected CNN Model for Alzheimer's Disease Classification Based on Hippocampus MRI Data. In *AMIA Annual Symposium Proceedings* (Vol. 2020, pp. 1277). American Medical Informatics Association.
10. Pan, D.; Zou, C.; Rong, H.; Zeng, A. Early diagnosis of Alzheimer's disease based on three-dimensional convolutional neural networks ensemble model combined with genetic algorithm. *Journal of Biomedical Engineering*, 2021, 38(1), 47-55.
11. Huang, H.; Zheng, S.; Yang, Z.; Wu, Y.; Li, Y.; Qiu, J.; Wu, R. (2023). Voxel-based morphometry and a deep learning model for the diagnosis of early Alzheimer's disease based on cerebral gray matter changes. *Cerebral Cortex*, 2023, 33(3), 754-763.
12. Hazarika, R. A.; Kandar, D.; Maji, A. K. An experimental analysis of different deep learning-based models for Alzheimer's disease classification using brain magnetic resonance images. *Journal of King Saud University-Computer and Information Sciences*, 2022, 34(10), 8576-8598.
13. Morid, M. A.; Borjali, A.; Del Fiol, G. A scoping review of transfer learning research on medical image analysis using ImageNet. *Computers in biology and medicine*, 2021, 128, 104115.
14. Zhang, F.; Pan, B.; Shao, P.; Liu, P.; Shen, S.; Yao, P.; Xu, R. X. An explainable two-dimensional single model deep learning approach for Alzheimer's disease diagnosis and brain atrophy localization. *arXiv preprint arXiv:2107.13200*.
15. Liu, M.; Zhang, J.; Nie, D.; Yap, P. T.; Shen, D. Anatomical landmark based deep feature representation for MR images in brain disease diagnosis. *IEEE journal of biomedical and health informatics*, 2018, 22(5), 1476-1485.
16. Liu, M.; Zhang, J.; Adeli, E.; Shen, D. Landmark-based deep multi-instance learning for brain disease diagnosis. *Medical image analysis*, 2018, 43, 157-168.
17. Zhang, Y.; Teng, Q.; Liu, Y.; Liu, Y.; He, X. Diagnosis of Alzheimer's disease based on regional attention with sMRI gray matter slices. *Journal of neuroscience methods*, 2022, 365, 109376.
18. Han, K.; Xiao, A.; Wu, E.; Guo, J.; Xu, C.; Wang, Y. Transformer in transformer. *Advances in Neural Information Processing Systems*, 2021, 34, 15908-15919.
19. Li, C.; Cui, Y.; Luo, N.; Liu, Y.; Bourgeat, P.; Frapp, J.; Jiang, T. Trans-ResNet: Integrating Transformers and CNNs for Alzheimer's disease classification. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5).
20. Jang, J.; Hwang, D. M3T: three-dimensional Medical image classifier using Multi-plane and Multi-slice Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp: 20718-20729.

21. Deng, J.; Dong, W.; Socher, R.; Li, L. J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255).
22. Lyu, Y.; Yu, X.; Zhu, D.; Zhang, L. Classification of Alzheimer's Disease via Vision Transformer: Classification of Alzheimer's Disease via Vision Transformer. In Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments, 2022, pp: 463-468.
23. Zhang, Z.; Gong, Z.; Hong, Q.; Jiang, L. Swin-transformer based classification for rice diseases recognition. In 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI) (pp. 153-156).
24. Nawaz, W.; Ahmed, S.; Tahir, A.; Khan, H. A. Classification of breast cancer histology images using alexnet. In Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15 (pp. 869-876). Springer International Publishing.
25. Sarwinda, D.; Paradisa, R. H.; Bustamam, A.; Anggia, P. Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. *Procedia Computer Science*, 2021, 179, 423-431.
26. Marques, G.; Agarwal, D.; de la Torre Díez, I. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Applied soft computing*, 2020, 96, 106691.
27. Wang, Y.; Yan, W. Q. Colorizing Gray-scale CT images of human lungs using deep learning methods. *Multimedia Tools and Applications*, 2022, 81(26), 37805-37819.
28. Talebi, H.; Milanfar, P. Learning to resize images for computer vision tasks. In Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp: 497-506.
29. Petersen, R. C.; Aisen, P. S.; Beckett, L. A.; Donohue, M. C.; Gamst, A. C.; Harvey, D. J.; Weiner, M. W. Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology*, 2010, 74(3), 201-209.
30. Ellis, K. A.; Bush, A. I.; Darby, D.; De Fazio, D.; Foster, J.; Hudson, P.; AIBL Research Group. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International psychogeriatric*, 2009, 21(4), 672-687

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.