Fine-Grained and Multiple Classification for Alzheimer's Disease With Wavelet Convolution Unit Network

Jinyu Wen, Yang Li, Meie Fang, Lei Zhu, David Dagan Feng, *Life Fellow, IEEE*, and Ping Li, *Member, IEEE*

***Abstract*—In this article, we propose a novel wavelet convolution unit for the image-oriented neural network to integrate wavelet analysis with a vanilla convolution operator to extract deep abstract features more efficiently. On one hand, in order to acquire non-local receptive fields and avoid information loss, we define a new convolution operation by composing a traditional convolution function and approximate and detailed representations after single-scale wavelet decomposition of source images. On the other hand, multi-scale wavelet decomposition is introduced to obtain more comprehensive multi-scale feature information. Then, we fuse all these cross-scale features to improve the problem of inaccurate localization of singular points. Given the novel wavelet convolution unit, we further design a network based on it for fine-grained Alzheimer's disease classifications (i.e., Alzheimer's disease, Normal controls, early mild cognitive impairment, late mild cognitive impairment). Up to now, only a few methods have studied one or several fine-grained classifications, and even fewer methods can achieve both fine-grained and multi-class classifications. We adopt the novel network and diffuse tensor images to achieve fine-grained classifications, which achieved state-of-the-art accuracy for all eight kinds of fine-grained classifications, up to 97.30%, 95.78%, 95.00%, 94.00%, 97.89%, 95.71%, 95.07%, 93.79%. In order to build a reference standard for Alzheimer's disease classifications, we actually implemented all twelve coarse-grained and fine-grained classifications. The results show that the proposed method achieves solidly high accuracy for them. Its classification ability greatly exceeds any kind of existing Alzheimer's disease classification method.**

***Index Terms*—Alzheimer's disease, wavelet analysis, fine-grained, multiple classification.**

I. Introduction

ALZHEIMER'S disease (AD) is a common disease in the elderly. At present, there is no effective treatment that can cure AD or change its progression. Mild Cognitive Impairment (MCI) is an intermediate stage between AD and normal controls (NC). Clinical studies show that mild cognitive impairment (MCI) can be divided into early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI). EMCI stage is reversible, detection and intervention timely can avoid the development of AD, while diagnosis and treatment timely at the LMCI stage can delay the development of AD or cure it. Therefore, Early detection and diagnosis of dementia will become the main goal. Accurate early diagnosis of AD is a meaningful and challenging task.

With the successful applications of convolutional neural networks for natural image-oriented classification tasks, many studies transferred these methods to medical image classification and computer-aided diagnosing diseases. In the recent ten years, many deep learning method based AD classification methods listed in Table I were continuously proposed. However, these existing methods have some limitations.

- They all directly used vanilla convolution operations designed for natural images. The local receptive field is limited. Dilated convolution can enlarge local receptive domain to a certain extent, but it will be accompanied by much information loss. Different from natural images, detailed features may fill the whole medical image. The

Manuscript received 17 November 2022; revised 26 January 2023; accepted 22 February 2023. Date of publication 17 March 2023; date of current version 30 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 62072126 and 61772164, in part by the Fundamental Research Projects Jointly Funded by Guangzhou Council and Municipal Universities under Grant 202102010439, in part by the Postgraduate Innovation Project of Guangzhou University under Grant 2021GDJC-D16, in part by the PolyU Research Institute for Sports Science and Technology under Grant P0044571, and in part by The Hong Kong Polytechnic University under Grants P0030419, P0042740, P0044520, P0043906, and P0035358. *(Corresponding author: Meie Fang.)*

Jinyu Wen is with the School of Computer Science and Cyber Engineering, Guangzhou University, China.

Yang Li is with the School of Mechatronic Engineering and Automation, Shanghai University, China.

Meie Fang is with the School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 511400, China (e-mail: fme@gzhu.edu.cn).

Lei Zhu is with the ROAS Thrust, The Hong Kong University of Science and Technology (Guangzhou), China, and also with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong.

David Dagan Feng is with the Biomedical and Multimedia Information Technology Research Group, School of Computer Science, The University of Sydney, Australia.

Ping Li is with the Department of Computing, the School of Design, and the Research Institute for Sports Science and Technology, The Hong Kong Polytechnic University, Hong Kong.

This article has supplementary downloadable material available at https://doi.org/10.1109/TBME.2023.3256042, provided by the authors.

Digital Object Identifier 10.1109/TBME.2023.3256042

TABLE I
RELATED WORKS UPON COARSE-GRAINED AD CLASSIFICATION AND THEIR CORRESPONDING ACCURACY RATES. THE HIGHEST ACCURACY AND SECOND HIGHEST ACCURACY FOR EACH KIND OF COARSE-GRAINED CLASSIFICATIONS ARE RESPECTIVELY MARKED IN RED AND BLUE (D-1: AD VS. NC, D-2: AD VS. MCI, D-3: MCI VS. NC, T-1: AD VS. NC VS. MCI)

Methods	Modality	Accuracy ("-": not classified)			
		D-1	D-2	D-3	T-1
(Ben Ahmed et al., 2015) [1]	MRI	0.838	0.695	0.621	-
(Payan and Montana, 2015) [2]	MRI	0.954	0.868	0.921	0.895
(Aderghal et al., 2017) [3]	MRI	0.914	0.695	0.656	-
(Yang et al., 2018) [4]	MRI	0.91	0.877	0.855	-
(Xiao et al., 2017) [5]	MRI	0.8571	0.7944	0.8611	0.75
(Madusanka et al., 2019) [6]	MRI	0.8661	0.8205	0.7896	-
(Bi et al., 2020) [7]	MRI	0.8915	0.9701	0.926	0.9125
(Duc et al., 2020) [8]	MRI	-	-	-	0.8715
(Xing et al., 2020) [9]	MRI	0.92	-	-	-
(Pan et al., 2020) [10]	MRI	0.89	0.83	0.68	-
(Ebrahimi and Luo, 2021) [11]	MRI	0.9375	-	-	-
(Liu et al., 2022) [12]	MRI	0.9896	0.9537	0.926	-
(Lei et al., 2016) [13]	MRI+PET	0.969	-	0.866	-
(Vu et al., 2018) [14]	MRI+PET	0.988	0.93	0.95	0.9113
(Ben Ahmed et al., 2017) [15]	MRI+DTI	0.902	0.766	0.794	-
(Ebadi et al., 2017) [16]	DTI	0.8	0.833	0.7	-
(Qu et al., 2021) [17]	DTI	0.8235	-	-	-
(Lella et al., 2021) [18]	DTI	0.885	-	-	-
(Bigham et al., 2022) [19]	DTI	0.958	0.833	0.833	-
Ours	DTI	0.992	0.9668	0.9616	0.9362

implementation and accuracy of fine-grained AD classification have encountered a bottleneck. A novel convolution operation with non-local receptive field is needed to more effectively extract the depth features of medical images.

- Most of the existing methods can only achieve coarse-grained classifications. Only a few methods, for example, [20], [21], studied some kinds of fine-grained classifications. While fine-grained classifications have more important clinical significance. In addition, there are many AD classification methods, but the classification combinations they can achieve are uneven. A method that can realize twelve kinds of full classifications is needed to provide a reference for the research in this field.
- Most existing works adopted Magnetic Resonance Imaging (MRI) for AD classifications. According to clinical medicine theory, Diffusion Tensor Imaging (DTI) reflects the continuity of tissue structure between fiber bundles in the brain, which is closely related to Alzheimer's Disease. Thus, we chose DTI for fine-grained AD classifications.

About choosing DTI data, we give more detailed explanations here. The modalities of neuroimaging include DTI scans, Positron Emission Tomography (PET) scans, MRI scans, etc. are commonly used in the diagnosis of AD. DTI [22] also known as special MRI modality which is based on water molecule motion. The principle of DTI imaging is that the diffusion of water molecules in the gradient field will change the magnetic moment. Therefore, it can display the walking direction of nerve conduction bundles in white matter and realize fine imaging of human central nerve fibers. The continuity of tissue structure between fiber bundles in the brain can be inferred from imaging of diffused fiber bundles in water [23]. In the study of patients with AD, the computed DTI images show an abnormal decrease in FA (Fractional anisotropy) and an abnormal increase in MD (Mean diffusivity) values in gray matter, white matter, etc. Therefore, compared with other image data, DTI data is more conducive to the realization of the fine-grained classification of MCI.

In order to break through the above-mentioned limitations of existing AD classification methods, we apply the wavelet theory into the convolution operations to define a novel convolution, and integrate features from single-scale and multi-scale wavelet decomposition, to acquire non-local respective field and avoid information loss. Based on the novel wavelet convolution unit (WCU), we adopt DTI data to implement twelve kinds of full combination classifications with improved accuracy, especially for fine-grained classifications.

In sum, we propose a wavelet convolution unit to build an effective WCU-Net for fine-grained and multiple AD classification. The deep learning method with WCU and the computed DTI data mainly guarantee that our proposed method achieved fine-grained AD classification with high accuracy. Our work makes the following three main contributions as:

- We propose a novel WCU, where a novel kind of convolution is defined by integrating wavelet transformation into the traditional convolution function to acquire non-local receptive fields and avoid information loss. The WCU-embedded network dramatically improves the performance of the convolutional neural network. In the network with WCU, wavelet transformation runs through the process of convolution, batch normalization (BN), and activation to obtain new wavelet coefficients after each level of the convolution operation. Due to the locality, multi-resolution, and multi-scale properties of wavelet transformation, the new convolution operator can fully take advantage of these optimal properties. WCU expands the receptive field to non-locality and captures comprehensive cross-scale features to improve the effect of extracting deep features. To the best of our knowledge, WCU is the first work embedding wavelet analysis into a vanilla convolution neural network.

- We propose a new AD classification framework via WCU-Net and brain DTI data, which achieves all coarse-fined and fine-grained combination classifications for the first time. Most existing AD classification methods are limited to coarse-grained classifications among AD vs. MCI vs. NC. A few methods achieve one or several kinds of fine-grained classifications. The proposed framework achieved all eight kinds of fine-grained classifications, where fine-grained triple classifications emerge for the first time. Our method provided a reference standard for later AD classification research.
- Our work solidly achieves high accuracy for all twelve kinds of classifications, especially improves the accuracy for eight kinds of fine-grained classifications from 92.5%, 92.6%, 93.5%, 90.9%, 81%, none, none, 92.6% to 97.30%, 95.78%, 95.00%, 94.00%, 97.89%, 95.71%, 95.07%, 93.79%. Firstly, emerged fine-grained triple classifications reach an accuracy higher than 95%. The classification accuracy for fine-grained double classification EMCI vs. LMCI is lifted by 16.89%. The accuracy for fine-grained quadruple classification reaches 93.79%, 1.19% higher than state-of-the-art methods. In addition, the accuracy can be further lifted by using initial weights from suitable pre-trained classic network models.

The subsequent content of this paper is organized as follows: Section II briefly introduces frequency domain analysis method including the wavelet analysis method and its application in image processing and introduces the existing work on AD aided diagnosis and classification based on neural network. Section III describes the novel wavelet convolution unit, the neural network framework based on WCU and the training method in detail. Section IV describes the realization of AD fine-grained multiple classification methods based on WCU-Net using brain DTI image data. Section V carries out a series of experiments and results analysis. Finally, Section VI concludes the work and gives future research directions.

II. Related Work

A. Frequency Domain Analysis Method

Frequency domain analysis is a view from frequency, which can find something not in the time domain. Among them, Fourier transform and wavelet transform are two typical frequency domain analysis methods. Wavelet is composed of a family of basis wavelet functions, which can describe the local properties of signal time (spatial) and frequency (scale) domains, while Fourier transform only has the properties of frequency analysis. In addition, the wavelet transform is an order of magnitude faster than the Fast Fourier transform. Assuming that the signal length is L, the computational complexity of Fourier transform and wavelet transform is $Llog_2L$ and L respectively. Wavelet analysis is widely used in image analysis, such as image feature extraction, image compression and image watermarking.

In recent years, there are several papers introduced wavelet transform into CNN [24], [25], [26]. Liu et al. [24] applied wavelet decomposition to obtain high and low frequency information. Then low frequency information is continuously decomposed 4 times. And the high frequency information is continuously concated with the low frequency information in the next layer to be updated. In the end, the last high and low frequency information is concated. All these operations emerged before convolution. Fujieda et al. [25] used wavelet analysis to replace the max pooling operation to form a modified U-Net architecture, which is the generalization of dilated filtering and subsampling. But in this existing work, wavelet transform only acts as a filter with fixed coefficients to participate or substitute a certain procedure, for example downsampling, in the structure of the network. They didn't integrate the wavelet transform with convolution operation as a new unit, just combined them together.

B. AD Classification Based on Conventional Method

In the task of AD classification, most of conventional method focus on coarse classification of AD, NC and MCI [2]. Xiao et al. [5] proposed a new classification framework to accurately identify patients with Alzheimer's disease. The method analyzed the multi-feature combination correlation technology and improved the SVM-RFE algorithm by using the covariance method, which showed that the multi-feature combination method was better than the single feature method. White matter (WM) injury is an important part of the AD pathological cascade, which always be neglected in the diagnosis of AD [27]. The most powerful tool in White matter is DTI, which is a non-invasive in vivo imaging technique. Identification of White matter fiber trends and degree of damage by dispersion characteristics can reveal the structural integrity of AD and delineates the deterioration of white matter [28]. Ben Ahmed et al. [15] proposed to extract local image-derived biomarkers from DTI and sMRI to construct multimodal AD features. Few scholars using conventional method to do fine-graind double classification [29]. Ashburner and Friston [30] proposed one method based on conventional feature representation of voxel-based morphometry. Zhang et al. [31] proposed one landmark-based morphometrical feature method.

C. AD Classification Based on Deep Learning Method

In general, the patient is in the middle and late stages when he is diagnosed with AD. Therefore, early diagnosis of AD using artificial intelligence requires a sensitive and efficient diagnosis method, such as [32], [33], [34]. In recent years, with excellent achievements in various fields, deep learning has gradually been widely applied in the field of medicine. CNN, the most common deep learning method, has received a lot of attention due to its success in the field of image analysis and classification [35], [36]. However, using deep learning methods to diagnose AD is still a great challenge due to the lack of pre-processed medical image acquisition, errors and knowledge. Aderghal et al. [3] proposed a data enhancement strategy adapted to sMRI scan specificity for training and classification of limited continuous sections. Lei et al. [13] proposed a discriminant feature based learning and canonical correlation analysis in MRI and PET modes. Multimodal images help to improve the diagnosis of AD, but these multimodal combined methods need to take a long time

TABLE II
THE ACCURACY ACHIEVED BY EXISTING FINE-GRAINED AD CLASSIFICATION METHODS AND OURS. THE RED VALUE REFERS TO THE HIGHEST ACCURACY, AND THE BLUE VALUE REFERS TO THE SECOND HIGHEST ACCURACY. OBVIOUSLY, OUR METHOD ACHIEVES HIGHEST ACCURACY FOR ALL THESE LISTED CLASSIFICATIONS

Methods	Modality	D-1	D-4	D-5	D-6	D-7	D-8	T-2	T-3	QC
(Ashburner and Friston, 2000) [30]	MRI	0.884	-	-	-	-	-	-	-	0.404
(Zhang et al., 2016) [31]	MRI	0.870	-	-	-	-	-	-	-	0.431
(Liu et al., 2019) [34]	MRI	0.937	-	-	-	-	-	-	-	0.518
(Basaia et al., 2019) [39]	MRI	0.992	0.754	0.859	0.871	0.761	0.751	-	-	-
(Wen et al., 2020) [38]	MRI	0.91	-	-	-	-	0.81	-	-	-
(Prasad et al., 2015) [29]	DTI	0.78	-	-	-	-	0.63	-	-	-
(De and Chowdhury, 2021) [21]	DTI	-	-	-	-	-	-	-	-	0.926
(Fang et al., 2022) [20]	DTI	0.946	0.925	0.926	0.935	0.909	0.808	-	-	-
Ours	DTI	0.9920	0.9730	0.9578	0.9500	0.9400	0.9789	0.9571	0.9507	0.9379

in the image processing stage. To solve these problems, Fang et al. [37] proposed a novel framework that integrates three state-of-the-art deep convolutional neural networks with multimodal images for AD classification to achieve higher accuracy.

In the task of AD classification, MRI data are processed mostly, such as double classification of AD and NC [9], [11], and three pairwise combinations in AD, MCI and NC [1], [4], [6], [10], [12]. [14] described an automated and robust method for detecting and identifying AD MRI and PET images. Bi et al. [7] proposed a new deep-learning technique for the prediction of AD based on MRI images. The accuracy of the method in AD vs. NC vs. MCI classification was 91.25%. DTI has been used intermittently for the classification task of identifying Alzheimer's disease [17], [18], achieving double classification (AD vs. NC) with the accuracy 0.885, 0.8235 respectively. Ebadi et al. [16] reported studies on the potential of applying brain connectivity patterns as an aid in diagnosing AD and MCI by using an integrated classification module to perform classification tasks. Table I lists these methods and the corresponding classification accuracy.

With more and more scholars focusing on the coarse-grained double classifications concerning AD, NC and MCI, the methods are constantly improved and reached satisfied accuracy rate. A few scholars used the deep-learning method to do fine-graind classification of AD [38], which further refine MCI into EMCI and LMCI. Liu et al. [34] proposed a joint classification and regression framework for AD diagnosis via a deep multi-task multi-channel learning framework. Basaia et al. [39] proposed a CNN-based approach to fine-grained classification of MCI using MRI data, which achieved six kinds of combination double classifications among AD, LMCI, EMCI and NC. Fang et al. [20] proposed a fine-grained method via re-transfer learning, which achieved five kinds of accuracy has been improved. De and Chowdhury [21] proposed quadruple classification methods of AD vs. NC vs. EMCI vs. LMCI in DTI data. In this method, VoxCNNs [40] network was used to train on FA, MD and EPI data respectively, and random forest classification was used on average data of FA and MD. For fusing output results of the four network models, the author proposed a hierarchical average fusion decision to achieve the classification task, and the classification accuracy reached 92.6% which better than the existing methods. Table II lists these methods and the corresponding classification accuracy.

III. APPROACH

In the task of fine-grained and multiple classification, we propose a novel method for AD prediction. Next, we will introduce our self-defined WCU, our fine-grained multiple classification network (WCU-Net), and prediction problems on AD by WCU-Net.

A. WCU (Wavelet Convolution Unit)

Wavelet analysis has good localization properties in both the spatial domain and frequency domain, and wavelet transform has the characteristics of multi-resolution, which is conducive to the extraction of different features of each resolution. The low frequency part of a certain scale is decomposed into four parts by two-dimensional wavelet function decomposition: approximate information of higher order scale and detailed information of three directions (i.e., horizontal, vertical and diagonal), as (1):

$$\begin{aligned} f_{wav}(x,y) = &\frac{1}{\sqrt{MN}}\sum_m\sum_n W_\varphi(0,m,n)\varphi_{0,m,n}(x,y) \\ &+\frac{1}{\sqrt{MN}}\sum_{j=0}^{\infty}\left(\sum_m\sum_n W_\psi^H(j,m,n)\psi_{j,m,n}^H(x,y)\right. \\ &+\sum_m\sum_n W_\psi^V(j,m,n)\psi_{j,m,n}^V(x,y) \\ &\left.+\sum_m\sum_n W_\psi^D(j,m,n)\psi_{j,m,n}^D(x,y)\right), \end{aligned} \tag{1}$$

where M and N represent the size of a certain image. $\varphi_{0,m,n}(x,y)$ represents the scale function and the calculation is shown as (2). $\psi_{j,m,n}^{Dir}(x,y)$ represents wavelet primary function and the calculation is shown as (3). j represents the order, which determines the extent and narrowing. Dir shows the direction, which can be horizontal, vertical or diagonal. m and n represent the position of the movement. In (1), $W_\varphi(0,m,n)$ represents the approximate coefficient and its calculation is shown as (4). $W_\psi(j,m,n)$ represents the detail coefficient and its calculation is shown as (5).

$$\varphi_{0,m,n}(x,y) = 2^{\frac{j}{2}}\varphi\left(2^j x - m, 2^j y - n\right), \tag{2}$$

$$\psi_{j,m,n}^{Dir}(x,y) = 2^{\frac{j}{2}}\psi^{Dir}\left(2^j x - m, 2^j y - n\right), Dir = \{V,H,D\}, \tag{3}$$

$$W_{\varphi}(0,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\varphi_{0,m,n}(x,y), \tag{4}$$

$$W_{\psi}(j,m,n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\psi^{Dir}_{j,m,n}(x,y). \tag{5}$$

Wavelet transform can decompose an image into components of different sizes, positions and directions. Based on the wavelet transform, we improved it shown as (6). After wavelet decomposition, we used $\xi(\cdot)$ function in approximation and detail coefficient. $\xi(\cdot)$ represents the combination of convolution, BN and activation as (7), which can make the coefficients of wavelet decomposition more local and more complete than the overall information obtained. In (7), the Fm means feature map.

$$\begin{aligned}\xi_{wav}(x,y)&=\xi(Approx)+\xi(Detail)\\&=\xi\left(\frac{1}{\sqrt{MN}}\sum_{m}\sum_{n}W_{\varphi}(0,m,n)\varphi_{0,m,n}(x,y)\right)\\&\quad+\xi\left(\frac{1}{\sqrt{MN}}\sum_{j=0}^{\infty}\left(\sum_{m}\sum_{n}W^{H}_{\psi}(j,m,n)\psi^{H}_{j,m,n}\right.\right.\\&\quad(x,y)+\sum_{m}\sum_{n}W^{V}_{\psi}(j,m,n)\psi^{V}_{j,m,n}(x,y)\\&\quad\left.\left.+\sum_{m}\sum_{n}W^{D}_{\psi}(j,m,n)\psi^{D}_{j,m,n}(x,y)\right)\right),\end{aligned} \tag{6}$$

$$\xi(Fm)=ReLU(BN(Conv(Fm))). \tag{7}$$

As we all know, vanilla convolution has limited local receptive field, which often results in discarding some non-local structure information in deep abstract features. Dilated convolution is usually adopted to expand local receptive field by sampling, with much information loss. While for medical images, both non-local structure information and local detail information often have important clinical meanings. Thus, we define the above novel convolution, integrating wavelet transform into convolution operation, to simultaneously acquire non-local receptive field and avoid information loss.

Fig. 1 take a piece of original MRI as an example to illustrate the implement process. According to the novel defined convolution, four coefficient components (A: approximate representation of the original image, D: diagonal edge features of the original image, V: singular features in the vertical direction, H: singular features in the horizontal direction) are generated by wavelet transformation at first. Then doing convolution for each component, we totally get four pieces of map images. Assuming the output feature map being 1×1, the stride being 1 and adopting 5×5 convolution kernels, we acquire four different 5×5 local receptive fields, as illustrated in Fig. 2(c). Certainly, there is some overlap for three receptive fields corresponding to diagonal features, horizontal features and vertical features. For comparison, Fig. 2(a) and (b) respectively illustrated the local receptive fields about vanilla convolution and dilated convolution with factor 2. Except for the overlapping part, we can see that the novel receptive field is 3.24 times larger than that of vanilla convolution, as illustrated in Fig. 2. Since the four coefficients of wavelet decomposition represent different image features, the region after stacking them marked with the yellow box in the original MRI of Fig. 1). The local receptive field of vanilla convolution is also marked with the red box for comparison. Further point-wise update in the frequency domain globally affects all input features involved in wavelet transform according to wavelet transform theory. When we carry out inverse wavelet transformation to obtain the final feature map (see Fig. 1(b)), the expanded local receptive field further diffuses to non-local receptive field. And, we can see that the feature map image is very close to the original image, without much information loss. The final feature map image of vanilla convolution is shown in Fig. 1(a) for comparison, which is quite different from the original image, with much information loss.

Algorithm 1: Wavelet Convolution Unit.

WCU (X)
  $W_s \leftarrow \text{SinScaT}\,(X,\ wave)$
  $W_m \leftarrow \text{MulScaT}\,(X,\ wave)$
  $W \leftarrow \text{Concatenate}\,([W_s,\ W_m],\ 1)$
  $X_r \leftarrow \xi_{wav}\,(W)$
  return X_r
Function SinScaT $(X,\ wave)$
  $C : \{c_i\}_{i=0}^{3} \leftarrow \text{dwt}\,(X,\ wave)$
  $\hat{C} : \{\hat{c}_i\}_{i=0}^{3} \leftarrow \xi_{wav}\,(c_0,\ c_1,\ c_2,\ c_3)$
  $\Gamma_s \leftarrow \text{idwt}\,((\hat{c}_0,\ \hat{c}_1,\ \hat{c}_2,\ \hat{c}_3),\ wave)$
  return Γ_s
Function MulScaT $(X,\ wave)$
  $\hat{C} \leftarrow \text{wavedec}\,(X,\ wave,\ 3)$
  $\Gamma_{0,1} \leftarrow \text{idwt}\,(\hat{c}_0,\ \hat{c}_1,\ wave)$
  $\Gamma_{3,2} \leftarrow \text{idwt}\,(\Gamma_{0,1},\ \hat{c}_2,\ wave)$
  $\Gamma_m \leftarrow \text{idwt}\ \,(\Gamma_{3,2},\ \hat{c}_3,\ wave)$
  return Γ_m

Based on self-defined wavelet decomposition, we propose a wavelet convolution unit as shown yellow square in Fig. 3, which consists of four parts: Encoder, WCU, Decoder, Fully connection. WCU (wavelet convolution unit) is self-defined module for extracting information of cross-scale and effective by using wavelet analysis and embedding vanilla convolution. In order to acquire non-local receptive field, we implement convolution operations to coefficients of single-scale wavelet decomposition. Besides, multi-scale wavelet decomposition used to obtain more comprehensive multi-scale feature information. What's more, we fuse all these cross-scale features in the wavelet convolution unit to solve the problem of inaccurate localization of singular points. Our WCU is shown in Algorithm 1.

Firstly, the image is decomposed by a single-scale to obtain the low-frequency approximation coefficient and three (horizontal, vertical and diagonal) high-frequency detail coefficients, and apply ξ function in each coefficient, and then invert the coefficients. Before the inverse transformation, the local coefficients in the wavelet transform domain are changed by ξ function, which can selectively enlarge the classification of local details

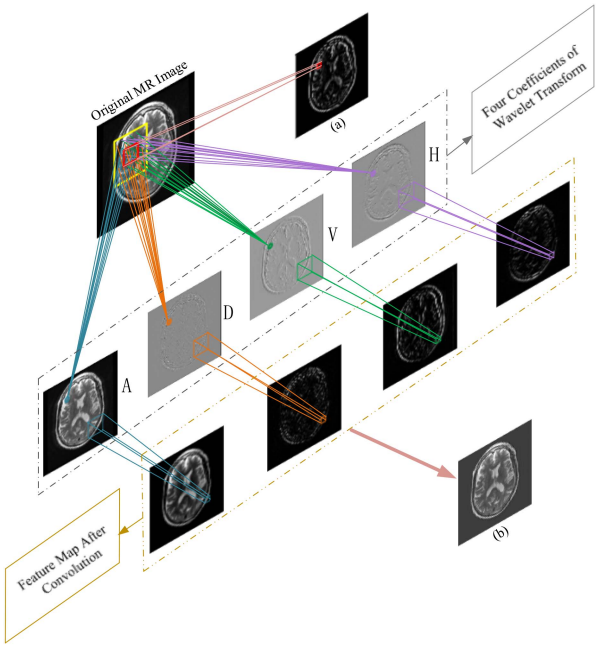


Fig. 1. Take a piece of original MRI as an example to illustrate how to acquire non-local receptive field by our convolution embed wavelet. In the process of our convolution embed wavelet operation, four different coefficients images A, D, V, and H are obtained by wavelet transform, and then four different feature maps are generated by convolution, with four times local receptive field of vanilla convolution. Furthermore, point-wise update in the frequency domain globally affects all input features involved in wavelet transform. Hence, the actual receptive field doesn't limit in the region marked by yellow box as illustrated in the original MRI, but diffuses to non-local receptive field of the whole image. The receptive field of vanilla convolution operation limits in the red box marked in the original MRI. In addition, the feature map images obtained by vanilla convolution and our wavelet convolution are illustrated as image (a) and image (b) for comparison.

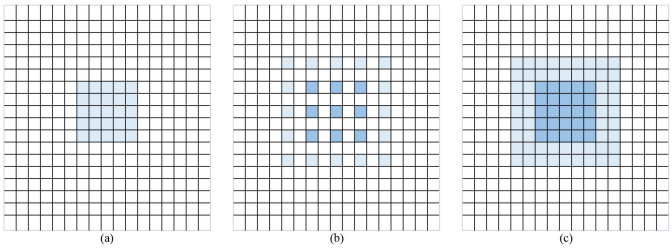


Fig. 2. Illustration of receptive field marked with blue blocks of different convolution operations. (a) The limited local receptive field of vanilla Convolution operation. (b) The expanded local receptive field of dilated convolution operation, a dilated filtering with factor 2. (c) The expanded local receptive field of our proposed convolution operation.

and reduce the useless components. At the same time, the image is decomposed to three scales and reconstructed according to different scale coefficients to obtain the transformed feature map. Finally, the results of the two reconstructions are connected together and the ξ function is applied again to obtain the output result of the wavelet convolution unit. In the experimental test, the wavelet primary function we choose is the Haar wavelet (First-order Daubechies wavelet) which is the only discontinuous wavelet directly suitable for discrete 2D images. Haar wavelet has both orthogonality and symmetry. In addition, in the multi-scale decomposition part, we set 3-scale decomposition. In the convolution of single-scale wavelet decomposition, kernel $size = 1, stride = 1, padding = 0; kernelsize = 3, stride = 1, padding = 1$ were set in the convolution of single-scale and multi-scale combination.

B. Architectural Design (WCU-Net)

CNN as a typical network structure in deep learning, which consists of several layers, including but not limited to the convolutional layer, pooling layer, activation layer and full connection layer. The framework of our method is mainly composed of four parts, including encoding, wavelet convolution unit, decoding and full connection layer, as shown in Fig. 3. To deal with variable length/short sequences efficiently, we use the idea of encoding and decoding. In the stage of encoding and decoding, the general convolution operation is mainly used to constantly adjust the depth of network layers. In these two processes we have used a combination of the convolutional layer, BN, and activation layer to improve the fitting capability of the network model. The convolutional layer of deep CNN is used to extract scale/displacement invariant features of local areas of images. The main benefit of the convolutional layer is the idea of weight sharing in the same feature map, which reduces parameters and leads to the simplicity of the model. The number of input and output channels in different layers of the convolution layer is shown as the values around the blue box in Fig. 3. In all convolution layers, $kernelsize = 4, stride = 2, padding = 1$. The convolutional layer is followed by BN layer, which is essentially a normalized network layer.

The input's distribution of all hidden layer neurons is then mapped to a non-linear function by degrees, finally to the limit saturation area of value interval. It is then designed to be back to the general normal distribution with 0 for the mean value and 1 for the variance value. Then, to avoid gradient disappearing, the inputs for the nonlinear function is designed to be in the area where is input sensitive. This is followed by the introduction of an activation layer for the model to learn complex representations. In order to increase the convergence rate, the activation function used in this paper (except for the final output) is nonlinear activation function ReLU. The performance of CNN mainly depends on the layer structure and filter set, and many studies show that network structure design is an effective way to improve the performance of CNN [36]. In this paper, we design a wavelet convolution unit based on the localization and multi-resolution characteristics of the wavelet, details in Section III-A. The last part is the full connection layer. The feature map processed in the previous parts has turned into a vector, which is no longer spatially located. The nonlinear relationship between local features in the convolution layer can be found by adding the full connection layer. We used 3 layers of full connectivity, and output the predicted categories by softmax. In order to improve the generalization ability of the network model, we adopt the dropout method before and after full connection, and the coefficient is set to 0.5. In the process of training, we used the classical cross entropy loss function to

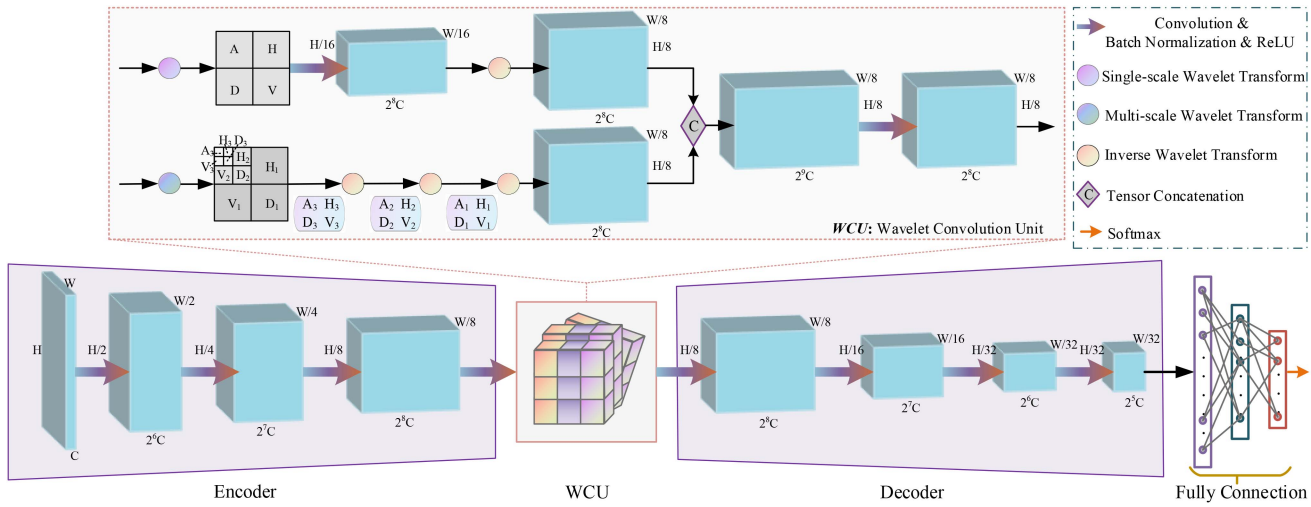


Fig. 3. The framework of our proposed network which consists of four parts: Encoder, WCU, Decoder, Fully connection. WCU (wavelet convolution unit) is self-defined module for extracting information of cross-scale and effective by using wavelet analysis and embedding vanilla convolution. Encoder and Decoder are common convolutions for feature extraction, which mainly reflected in the depth of features. Fully connection is the last step for predicting category labels and softmax is helped to achieve output classification results.

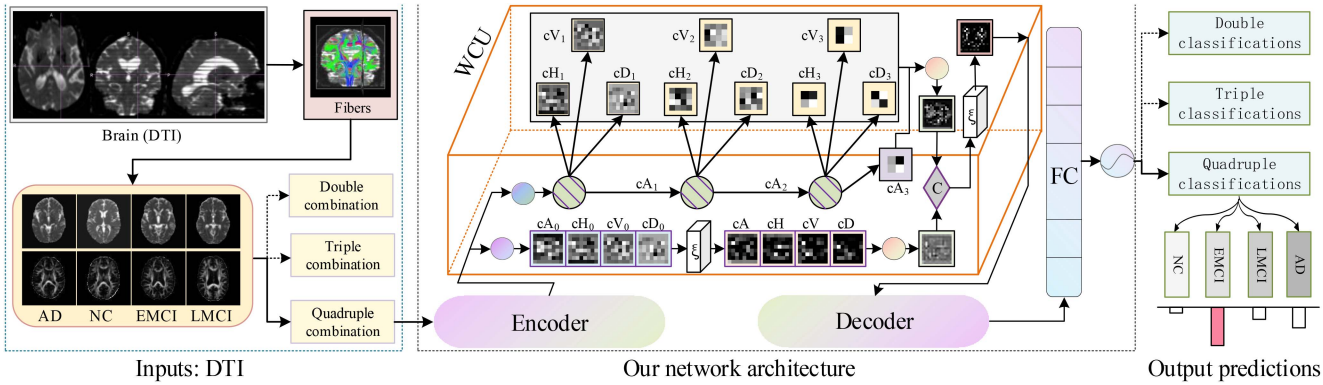


Fig. 4. The outline of fine-grained combination classification on AD, EMCI, LMCI and NC, including DTI data processing, the network architecture with WCU and the classification prediction results. The inputs are MD (the first row in the light yellow block) and FA (the second row in the light yellow block) computed from DTI data of four types of cohorts (i.e., AD, NC, EMCI, LMCI). In the network architecture, WCU is the kernel, whose detailed specific process is shown in the middle part of this figure and Fig. 3.

calculate the loss cost between the real label and the predicted output.

IV. AD Classification Via WCU-Net

A. DTI Data

As we all know, data preprocessing has always been an important step in deep learning models, especially in the medical field. DTI is a special form of MRI, which is the non-invasive method that can observe effectively and track the white matter fiber tracts. MRI is mostly used in the diagnosis of Alzheimer's disease and other symptoms. Generally speaking, DTI can better reflect the characteristics of brain diseases, which is conducive to diagnosis.

The signal data of diffusion anisotropy are quantified by the concept of symmetric matrix in three dimensional view, which is called the D, as shown in (8). Where, D_{xx}, D_{yy}, D_{zz} are the dispersion coefficient applied along the three mutually perpendicular directions of x, y and z axes of the space cartesian coordinate system. The DTI is a 3×3 symmetric, positive definite matrix, which contains three eigenvalues (λ_1, λ_2, λ_3) and associated eigenvectors $V = (v_1, v_2, v_3)^T$. The three eigenvectors reflect the three dispersion directions of water molecules. The size of the eigenvalue indicates the degree of dispersion of water molecules in each direction.

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix} = V^T \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} V. \quad (8)$$

In our method, the FA and MD indices Ind_{FA}, Ind_{MD} of DTI are used, these two indices are images pre-processed from DTI data using FSL software or scripts. As shown in the yellow square of inputs in Fig. 4, the MD image is in the first line, and the

TABLE III
THE ACCURACY OF NINE COMBINATION CLASSIFICATIONS ACHIEVED BY PROPOSED WUC-NET WITH OR WITHOUT ADOPTING PRE-TRAINED CNNS DURING WEIGHT INITIALIZATION PROCESS. EACH PRE-TRAINED MODEL CAN IMPROVE THE ACCURACY FOR SEVERAL KINDS OF COMBINATION CLASSIFICATIONS TO A CERTAIN EXTENT. THE RED VALUE MEANS THE HIGHEST ACCURACY, AND THE GREEN VALUES MEANS THAT ADOPTING THE CORRESPONDING PRE-TRAINED MODELS IMPROVED THE ACCURACY TO A CERTAIN EXTENT (D-1: AD VS. NC, D-4: AD VS. EMCI, D-5: AD VS. LMCI, D-6: NC VS. EMCI, D-7: NC VS. LMCI, D-8: EMCI VS. LMCI, T-2: AD VS. EMCI VS. LMCI, T-3: NC VS. EMCI VS. LMCI, QC: AD VS. NC VS. EMCI VS. LMCI)

Class	ResNet101	DenseNet161	SqueezeNet1_1	Inception_v3	AlexNet	VGG13	GoogLeNet	**w/o pretrained CNNs**
D-1	0.9840	0.9890	0.9790	0.9710	0.9701	0.9850	0.9890	0.9920
D-4	0.9630	0.9550	0.9690	0.9670	0.9500	0.9480	0.9680	0.9730
D-5	0.9533	0.9556	0.9533	0.9511	0.9478	0.9611	0.9522	0.9578
D-6	0.9750	0.9720	0.9650	0.9710	0.9690	0.9870	0.9700	0.9500
D-7	0.9611	0.9533	0.9456	0.9489	0.9567	0.9089	0.9590	0.9400
D-8	0.9800	0.9800	0.9722	0.9833	0.9644	0.9778	0.9900	0.9789
T-2	0.9350	0.9586	0.9490	0.9521	0.9600	0.9586	0.9636	0.9571
T-3	0.9257	0.9243	0.9464	0.9833	0.9193	0.9407	0.9429	0.9507
QC	0.9353	0.9168	0.9284	0.9068	0.9358	0.9337	0.9053	0.9379

second line is the FA image. FA in white matter was positively correlated with the integrity of myelin sheath, fiber density and parallelism. Hence, the fiber structure of white matter in the brain was most clearly observed by the FA image, and the gray matter boundary was obvious. The calculation formula of Ind_{FA} is shown in (9). Where $\bar{\lambda}$ is the mean of three eigenvalues. MD reflects the overall diffusion level and resistance of the molecule. MD only represents the size of diffusion and has irrelevant to the direction of diffusion. The larger Ind_{MD} is, the more free water molecules are contained in the tissue. The calculation formula of Ind_{MD} is shown in (10). $Tr(D)$ refers to the trace of D, which is a matrix invariant.

$$Ind_{FA} = \sqrt{\frac{\sum_{i=1}^{3} 3\left(\lambda_i - \bar{\lambda}\right)^2}{\sum_{i=1}^{3} 2\lambda_i^2}}, \tag{9}$$

$$Ind_{MD} = Tr(D)/3 = \left(D_{xx} + D_{yy} + D_{zz}\right)/3. \tag{10}$$

B. Fine-Grained and Multiple AD Classification Framework

In this paper, we build a new framework for fine-grained and multiple AD classification via WCU-Net. For double, triple and quadruple classifications, the framework is almost the same. In the experimental setup, only their input and output are different. As shown in Fig. 4, the framework includes three parts: the data input, the network architecture of Encoder, WCU, Decoder, FC and Activation function, and the output of class prediction. In the first part, we preprocessed DTI data. FSL (https://fsl.fmrib.ox.ac. UK/fsl/fslwiki/Fsl Installation/Linux) tools were used to extract b0 image, peel brain, eddy correction, tensor calculation and other operations on the original DTI data. FA and MD are selected as input data from the obtained tensor indices. In the second part, we input FA and MD indices of 4 categories (AD, NC, EMCI, LMCI) via Encoder, WCU, Decoder, FC and Activation function (Softmax) successively. In this part, we mainly expand and present the processes in WCU. WCU uses the wavelet convolution unit self-defined by us and combines the characteristics of single-scale wavelet decomposition and multi-scale wavelet decomposition to obtain complete cross-modal characteristics. The last part is the prediction output of the network, which converts the probability distribution into label data. The final prediction class depends on the category with the highest probability.

C. Pretrained Models Help Improve Accuracy

CNN is the most commonly used to detect AD in deep intelligent systems, some studies tend to design the structure of CNN [33]. Classical network structures, such as ResNet, DenseNet, SqueezeNet, Inception, AlexNet, VGGNet and GoogLeNet, have been successfully applied into the task of image classification [41], which is helpful to weight initialization process. In this paper, we adopted these 7 classical network models as different manifestations of weight initialization in our network structure. Table III lists the classification accuracy by WCU-Net with pre-trained CNNs. The classification accuracy without pre-trained models is also listed in the last column for comparison. It can be seen from Table III that each pre-trained model can improve the accuracy for 2 to 4 kinds of combination classifications to a certain extent, marked by red or blue values in column 2 to 8. Especially for fine-grained double classifications NC vs. EMCI and NC vs. LMCI, the effect of pre-trained is obvious. In summary, the highest accuracy values of nine combination classifications reach 0.9920, 0.9730, 0.9611, 0.9870, 0.9611, 0.9900, 0.9636, and 0.9379 respectively, where the best classification accuracy for AD vs. NC, AD vs. EMCI and AD vs. LMCI vs. EMCI vs. NC is persisted by WCU-Net without pre-trained. Therefore, one can decide whether pre-trained is needed and which pre-trained model is chosen to reach the best classification accuracy in applications. While in this paper, we still use the results obtained by WCU-Net without pre-trained to make comparisons with existing AD classification methods.

V. EXPERIMENTAL RESULTS

A. Datasets and Parameters

DTI is a special form of MRI, but its acquisition is more difficult than MRI. DTI can capture white matter tracts and gray matter tracts in the brain and works by measuring the diffusion of water molecules within living tissue. Since the diffusion tensor is a symmetric 3×3 matrix, it can be described by its eigenvalues (λ_i) eigenvectors (V_i). Eigenvalues and eigenvectors are then

TABLE IV
DATASET DETAILS AND AFTER AUGMENTATION DATA SET FOR TRAIN, VALID AND TEST

Diagnosis	AD	EMCI	LMCI	NC
Age	75.23	73.95	73.45	74.52
Gender (M/F)	99 / 54	225 / 138	108 / 59	110 / 109
Subjects	153	363	167	219
MMSE	18-27	24-30	24-30	25-30
Train data	9273	9748	10154	10113
Valid data	2100	2088	2100	2200
Test data	500	400	500	500

used to deal with scalar indices. In the dataset we used, FA and MD are available together. The two main diffusion indices, FA and MD, are based on eigenvalues and represent the magnitude of the diffusion process. Age is a major risk factor for AD, so we selected a population between 73-76 years of age and with significant differences in cognitive function scores (AD: 18-27, EMCI/LMCI: 24-30, NC: 25-30) to research and analysis. The dataset we downloaded comes from 902 samples divided into four categories: 153 AD patients, 167 LMCI, 363 EMCI, and 219 NC. The DTI data set is selected as the training, verification and test set of the model, the number as shown in Table IV. Data used in this paper were from the Alzheimer's Disease Neuroimaging (ADNI: www.adni-info.org) database. Since the amount of data for each class is not equal, and the total number of data is not enough to train the deep learning model, data amplification is used to increase the number of samples per minority class. In addition, the medical image is a gray image after visualization, in order not to change the image greatly, the data enhancement methods mainly include amplification by adding random noise and brightness. In order to increase the size of our data, we will use data enhancement to expand the data we have obtained. The method of data enhancement is consistent with the literature [20], and the situation after data enhancement is shown in Table IV.

In the process of network training, the size of the image to 128 $\times$ 128. In the training process, the initial learning rate is 0.001, 40 epochs, the batch size is 32, we employ batch normalization. The optimization function is SGD. The value of momentum is 0.85. In addition, we adopted an adaptive loss adjustment strategy to update the learning rate to 0.1 times of its original value every 8 epochs. We used Python 3.8.12 and the Pytorch version of 1.10.2 with CUDA version of 10.2 GNU/Linux x86 in Tesla T4 16 G RAM device-64 system. The source code of this article has been shared on GitHub: https://github.com/22385wjy/medicalImageClassification

B. Combination Classifications in AD, LMCI, EMCI, and NC

Clinical studies have shown that the fine-grained AD classification is of great significance. There have been various studies on AD classification. However, most existing methods were limited in coarse-grained double classifications, i.e., AD vs. NC, AD vs. MCI and MCI vs. NC. Some methods implemented coarse-grained triple classification, i.e., AD vs. NC vs. MCI. In recent years, some scholars have further classified MCI with fine granularity. In the early years, few scholars using DTI data to do fine-grained double classification [29], [38], and the higher accuracy is 0.81. Several scholars using MRI data for fine-grained quadruple classification of AD. Ashburner and Friston [30], Zhang et al. [31] and Liu et al. [34] are three methods that all applied in fine-grained quadruple classification of AD and using MRI data, but the accuracy is lower. There are also some methods using MRI data in triple and quadruple classification for AD classification, which can obtain better performance. For example, the triple classification accuracy can reach 68.8%, and the quadruple classification accuracy can reach 59.1% in [42]. However, the triple classification is AD vs NC vs (EMCI+LMEC), which is similar to the traditional coarse-grained triple classification. The accuracy of the quadruple classifications can reach 61.9% in [43]. Basaia et al. [39] proposed a CNN-based approach to fine-grained classification of MCI using MRI data, which achieved six kinds of combination double classifications among AD, LMCI, EMCI and NC. Fang et al. [20] proposed a fine-grained method via re-transfer learning, which improved accuracy of 5 kinds of fine-grained double classifications. At present, De and Chowdhury [21] achieved fine-grained quadruple classification with the accuracy 0.926.

In this paper, in order to facilitate the accurate diagnosis of AD classification as far as possible, we developed twelve combination classifications, including 4 coarse-grained classifications and 8 fine-grained classifications, as listed in the last column of Tables I and II. Two kinds of fine-grained triple classifications (AD vs. LMCI vs. EMCI and NC vs. LMCI vs. EMCI) emerged for the first time.

The first line in Fig. 5 shows the prediction accuracy curve distribution of the proposed method without the pre-trained network model as the initialization weight parameter, and there is a certain deviation in the distribution between the training stage and the verification stage. In the training process, the accuracy curve does not fluctuate, and the accuracy in different classification combinations is closer to 1 with the increase of training epochs. In the process of verification, a small range of oscillation occurred in the epoch 1 to 8, but the overall upward trend is correct. During the validation process, the parameters are constantly updated. With the increase of the number of rounds, the accuracy of AD vs. NC and EMCI vs. LMCI classification were close to 1, and the accuracy of quadruple classification was slightly lower, but also between 0.9 and 1. In the case of other data combination classification, the accuracy also stabilized at the 11th epoch. The second line in Fig. 5 shows the loss curves of the training and verification stages. The loss of double classification is small, while the loss of quadruple classification is relatively large. However, the loss in both the training and verification process decreases obviously and tends to be stable. Therefore, it shows that the network model proposed in this paper has good generalization performance and strong fitting ability.

The accuracy of our method for the classification AD vs. NC and eight kinds of fine-grained classifications is listed in Table II. For comparison, the accuracy of existing methods concerning fine-grained classification is also given in Table II. We can see that the accuracy of our method is much better than that of [20],

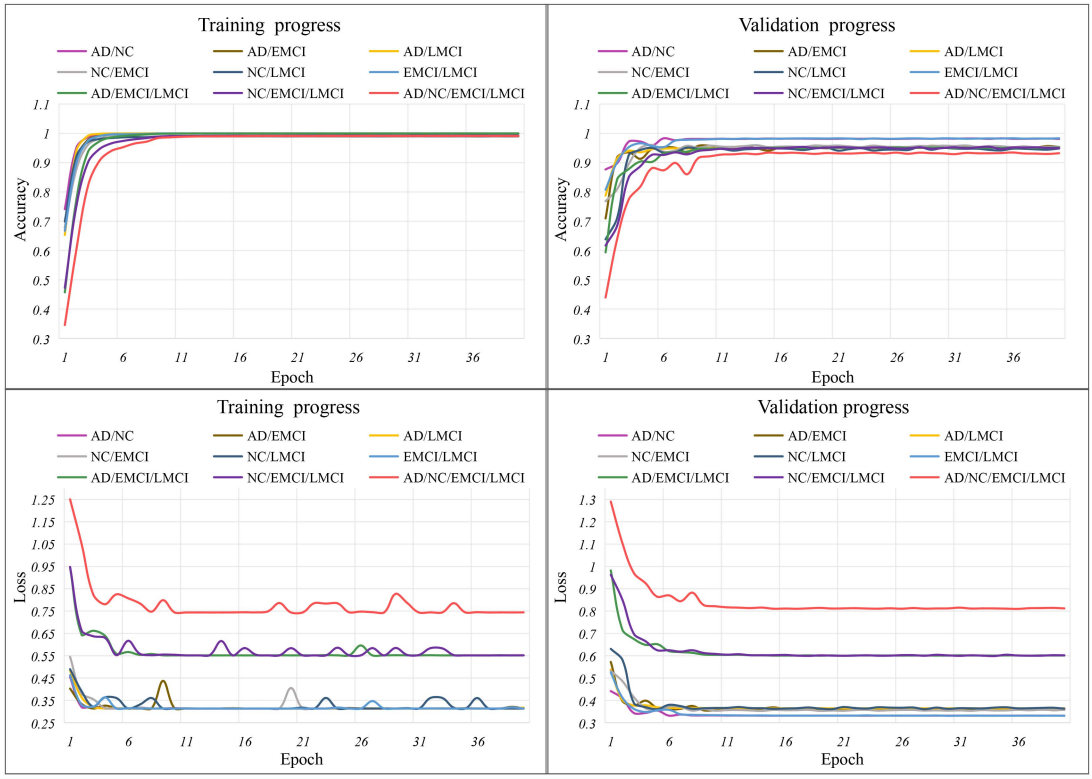


Fig. 5. (Top) The accuracy curves. (Bottom) The loss curves of WCU-Net in the training stage and validating stage for the tasks of AD vs. NC classification and all eight kinds of fine-grained classifications, without using initialization weight parameters from pre-trained network model. In the training process, all accuracy curves rapidly and stably approximate to 1. In the verifying process, all accuracy curves also approximate to 1 or a value higher than 0.9 after 11th epoch. While for loss curves, the more classes, the greater the loss value. But all loss curves still rapidly decrease and converge to a small value. In sum, these curves show that the proposed network model has good generalization performance and strong fitting ability.

[39] for 6 kinds of double classifications achieved in that paper. For fine-grained quadruple classification, the accuracy of our method is 0.9379, also 1.19% higher than that of [21]. Our method firstly achieved fine-grained triple classifications. The accuracy for AD vs. EMCI vs. LMCI and NC vs. EMCI vs. LMCI resectively reach 0.9571 and 0.9507.

C. Result Analysis

Fig. 6 shows the heat map of accuracy compare of all coarse-grained and fine-grained AD classifications concerned in this paper. Row corresponds to class combinations. Column corresponds to classification methods. The darker the color of the matrix cell, the higher the accuracy of its corresponding method, as shown by the colorbar on the right. Each light blue cell without any value means that there is no corresponding combination classification. The cells with values display the accuracy (two decimal places retained) of the corresponding method under different combination classifications. From this heat map, we can make the following conclusions.

- Most of methods can only achieve coarse-grained classifications. A few of methods achieve one or several kinds of fine-grained classifications. Our method achieves all twelve kinds of classifications. Fine-grained triple classifications can only achieved by our method.
- Most of methods achieved unstable accuracy for different class combinations. There even exist many cells with the accuracy less than 0.70. Our method achieves solidly high accuracy higher than 0.935 for all twelve kinds of class combinations.

It seems that there are lots of research works on AD classification. But they usually focused on different data modalities and class combinations. Comparability among all kinds of methods is not very strong. Our method achieves all twelve kinds of coarse-grained and fine-grained classifications with solidly high accuracy, which can build a suitable reference standard for later AD classification researches based on DTI data. However, there is one point that needs special explanation. In this paper, we mainly focused on the accuracy of DL method, but missed Neuroscience part related to the potential target pairs of ROIs linked to the continuum of AD. Actually, this is very important for disease research based on neuroscience and the interpretability of computer-aided diagnosis based on deep learning methods. It will be a meaningful research direction to combine neuroscience and AI methods more deeply in future research.

D. Ablation Experiments

As described in Section III, the optimal performance of our method mainly depends on single-scale wavelet decomposition

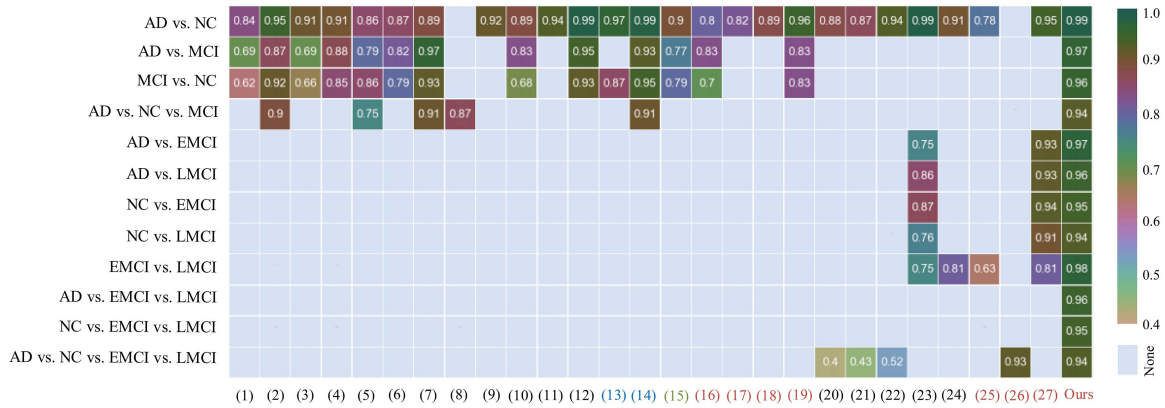
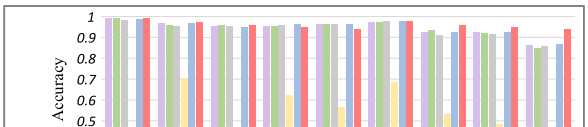


Fig. 6. The heat map of accuracy comparison among all concerned coarse-grained and fine-grained AD classification methods in this paper. The darker the color of the matrix cell, the higher the accuracy of its corresponding method. Light blue cell means no corresponding classification work. A large area of light blue cells shows that most of the existing methods only achieve a few combination classifications. But our method realizes all twelve kinds of classifications, with SOTA accuracy for all fine-grained classifications. Such a method also provides a complete reference for later research work in this field. The fonts with the same color on the horizontal axis in this figure indicate that the data modalities used by these methods are common. Red, blue, green and black respectively correspond to DTI, MRI+PET, MRI+DTI and MRI. (1)–(27) respectively refer to different methods proposed in the following references: Ben Ahmed et al. [1], Payan and Montana [2], Aderghal et al. [3], Yang et al. [4], Xiao et al. [5], Madusanka et al. [6], Bi et al. [7], Duc et al. [8], Xing et al. [9], Pan et al. [10], Ebrahimi and Luo [11], Liu et al. [12], Lei et al. [13], Vu et al. [14], Ben Ahmed et al. [15], Ebadi et al. [16], Qu et al. [17], Lella et al. [18], Bigham et al. [19], Ashburner and Friston [30], Zhang et al. [31], Liu et al. [34], Basaia et al. [39], Wen et al. [38], Prasad et al. [29], De and Chowdhury [21], Fang et al. [20].



	D-1	D-4	D-5	D-6	D-7	D-8	T-2	T-3	QC
w/o uS_iS	0.991	0.966	0.9556	0.953	**0.9656**	0.9733	0.9243	0.9236	0.8621
w/o uM_uS	0.991	0.958	**0.9611**	0.953	0.9633	0.9744	0.9336	0.92	0.85
w/o uC_o	0.985	0.956	0.9533	0.96	0.9633	0.9756	0.9093	0.9136	0.8558
w/o uA_mD	0.2661	0.7035	0.2617	0.6237	0.5658	0.6836	0.5307	0.484	0.402
w/o uC_oMS	0.988	0.969	0.9478	**0.962**	**0.9656**	0.9756	0.9229	0.9271	0.8663
Ours	**0.992**	**0.973**	0.9578	0.95	0.94	**0.9789**	**0.9571**	**0.9507**	**0.9379**

Fig. 7. The accuracy histogram for five groups of ablated methods (without uS_iS, without uM_uS, without uC_o, without uA_mD, without uC_oMS) and our complete method in this paper. Obviously, our complete method has a comparative advantage, especially for fine-grained triple classifications and quadruple classifications.

(uS_iS), multi-scale wavelet decomposition (uM_uS), ξ function operation (uC_o) and data amplification uA_mD. To further prove the effectiveness of these four components, we design five groups of methods to do ablation experiments here, i.e., w/o uS_iS, w/o uM_uS, w/o uC_o, w/o uA_mD and w/o uC_oMS (without uM_uS and uC_o). All experimental results are illustrated in Fig. 7 for comparison with our complete method. We can see that four components have their functionality. Data amplification is essential. Other three components are more important for fine-grained classifications. Especially, the accuracy for fine-grained quadruple classification is obviously improved when they work together.

VI. Conclusion

In this paper, a novel WCU-Net is proposed, which combines single-scale and multi-scale transformation in a wavelet convolution unit. This strategy not only obtain non-local receptive fields but also achieve cross-scale information fusion. Then the WCU-Net has been successfully applied to the fine-grained and multiple AD classifications with SOTA accuracy. We know fine-grained classification is of great significance for accurate diagnosis of cognitive disorders and correct treatment. The accuracy of fine-grained quadruple classification still need to be further improved. In fact, the WCU neural network proposed in this paper can be applied to the classification and staging of various diseases based on medical images. We will conduct more studies in this direction and try to apply research results into computer-aided clinical diagnosis in the future.

References

[1] O. Ben Ahmed et al., "Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex," *Computerized Med. Imag. Graph.*, vol. 44, pp. 13–25, 2015.
[2] A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks," *CoRR*, vol. abs/1502.02506, pp. 1–9, 2015.
[3] K. Aderghal et al., "Classification of sMRI for AD diagnosis with convolutional neuronal networks: A pilot 2-D ϵ study on ADNI," in *Proc. Int. Conf. Multimedia Model.*, 2017, pp. 690–701.
[4] J. Yang, S. Li, and W. Xu, "Active learning for visual image classification method based on transfer learning," *IEEE Access*, vol. 6, pp. 187–198, 2018.
[5] Z. Xiao et al., "Brain MR image classification for Alzheimer's disease diagnosis based on multifeature fusion," *Comput. Math. Methods Med.*, vol. 2017, 2017, Art. no. 1952373.
[6] N. Madusanka et al., "Alzheimer's disease classification based on multifeature fusion," *Curr. Med. Imag.*, vol. 15, no. 2, pp. 161–169, 2019.
[7] X. Bi et al., "Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology," *Neurocomputing*, vol. 392, pp. 296–304, 2020.
[8] N. T. Duc et al., "3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI," *Neuroinformatics*, vol. 18, no. 1, pp. 71–86, 2020.
[9] X. Xing et al., "Dynamic image for 3D MRI image Alzheimer's disease classification," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 355–364.

[10] D. Pan et al., "Early detection of Alzheimer's disease using magnetic resonance imaging: A novel approach combining convolutional neural networks and ensemble learning," *Front. Neurosci.*, vol. 14, 2020, Art. no. 259.
[11] A. Ebrahimi and S. Luo, "Convolutional neural networks for Alzheimer's disease detection on MRI images," *J. Med. Imag.*, vol. 8, no. 2, 2021, Art. no. 024503.
[12] Z. Liu et al., "Diagnosis of Alzheimer's disease via an attention-based multi-scale convolutional neural network," *Knowl.-Based Syst.*, vol. 238, 2022, Art. no. 107942.
[13] B. Lei, et al., "Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion," *Front. Aging Neurosci.*, vol. 8, pp. 77:1–77:17, 2016.
[14] T.-D. Vu et al., "Non-white matter tissue extraction and deep convolutional neural network for Alzheimer's disease detection," *Soft Comput.*, vol. 22, no. 20, pp. 6825–6833, 2018.
[15] O. Ben Ahmed et al., "Recognition of Alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning," *Neurocomputing*, vol. 220, pp. 98–110, 2017.
[16] A. Ebadi et al., "Ensemble classification of Alzheimer's disease and mild cognitive impairment based on complex graph measures from diffusion tensor images," *Front. Neurosci.*, vol. 11, 2017, Art. no. 56.
[17] Y. Qu et al., "AI4AD: Artificial intelligence analysis for Alzheimer's disease classification based on a multisite DTI database," *Brain Disord.*, vol. 1, 2021, Art. no. 100005.
[18] E. Lella et al., "An ensemble learning approach based on diffusion tensor imaging measures for Alzheimer's disease classification," *Electronics*, vol. 10, no. 3, 2021, Art. no. 249.
[19] B. Bigham, S. A. Zamanpour, and H. Zare, "Features of the superficial white matter as biomarkers for the detection of Alzheimer's disease and mild cognitive impairment: A diffusion tensor imaging study," *Heliyon*, vol. 8, no. 1, 2022, Art. no. e08725.
[20] M. Fang et al., "Re-transfer learning and multi-modal learning assisted early diagnosis of Alzheimer's disease," *Multimedia Tools Appl.*, vol. 81, no. 20, pp. 29159–29175, 2022.
[21] A. De and A. S. Chowdhury, "DTI based Alzheimer's disease classification with rank modulated fusion of CNNs and random forest," *Expert Syst. Appl.*, vol. 169, 2021, Art. no. 114338.
[22] K. M. Huynh et al., "Multi-site harmonization of diffusion MRI data via method of moments," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1599–1609, Jul. 2019.
[23] G. B. Frisoni et al., "The clinical use of structural MRI in Alzheimer disease," *Nature Rev. Neurol.*, vol. 6, no. 2, pp. 67–77, 2010.
[24] P. Liu et al., "Multi-level wavelet convolutional neural networks," *IEEE Access*, vol. 7, pp. 74973–74985, 2019.
[25] S. Fujieda, K. Takayama, and T. Hachisuka, "Wavelet convolutional neural networks," *CoRR*, vol. abs/1805.08620, pp. 1–10, 2018.
[26] J. B. Salyers, Y. Dong, and Y. Gai, "Continuous wavelet transform for decoding finger movements from single-channel EEG," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1588–1597, Jun. 2019.
[27] S. Rathore et al., "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, 2017.
[28] E. Horgusluoglu-Moloch et al., "Systems modeling of white matter microstructural abnormalities in Alzheimer's disease," *NeuroImage: Clin.*, vol. 26, 2020, Art. no. 102203.
[29] G. Prasad et al., "Brain connectivity and novel network measures for Alzheimer's disease classification," *Neurobiol. Aging*, vol. 36, pp. S121–S131, 2015.
[30] J. Ashburner and K. J. Friston, "Voxel-based morphometry–the methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000.
[31] J. Zhang et al., "Detecting anatomical landmarks for fast Alzheimer's disease diagnosis," *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2524–2533, Dec. 2016.
[32] M. Shehata et al., "Computer-aided diagnostic system for early detection of acute renal transplant rejection using diffusion-weighted MRI," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 2, pp. 539–552, Feb. 2019.
[33] S. Bringas et al., "Alzheimer's disease stage identification using deep learning models," *J. Biomed. Informat.*, vol. 109, 2020, Art. no. 103514.
[34] M. Liu et al., "Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1195–1206, May 2019.
[35] A.-K. Golla et al., "Convolutional neural network ensemble segmentation with ratio-based sampling for the arteries and veins in abdominal CT scans," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1518–1526, May 2021.
[36] Z. Liang et al., "CameraNet: A two-stage framework for effective camera ISP learning," *IEEE Trans. Image Process.*, vol. 30, pp. 2248–2262, 2021.
[37] X. Fang, Z. Liu, and M. Xu, "Ensemble of deep convolutional neural networks based multi-modality images for Alzheimer's disease diagnosis," *IET Image Process.*, vol. 14, no. 2, pp. 318–326, 2020.
[38] J. Wen et al., "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Med. Image Anal.*, vol. 63, 2020, Art. no. 101694.
[39] S. Basaia et al., "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clin.*, vol. 21, 2019, Art. no. 101645.
[40] S. Korolev et al., "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2017, pp. 835–838.
[41] A. Ananda et al., "Classification and visualisation of normal and abnormal radiographs; A comparison between eleven convolutional neural network architectures," *Sensors*, vol. 21, no. 16, 2021, Art. no. 5381.
[42] L. Sørensen and M. Nielsen, "Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination," *J. Neurosci. Methods*, vol. 302, pp. 66–74, 2018.
[43] S. I. Dimitriadis, D. Liparas, and M. N. Tsolaki, "Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate among healhy elderly, MCI, cMCI and Alzheimer's disease patients: From the Alzheimer's disease neuroimaging initiative (ADNI) database," *J. Neurosci. Methods*, vol. 302, pp. 14–23, 2018.