



# ADVIAN: Alzheimer's Disease VGG-Inspired Attention Network Based on Convolutional Block Attention Module and Multiple Way Data Augmentation

Shui-Hua Wang<sup>1,2</sup>, Qinghua Zhou<sup>3</sup>, Ming Yang<sup>4\*</sup> and Yu-Dong Zhang<sup>1,3\*</sup>

<sup>1</sup> Key Laboratory of Child Development and Learning Science (Southeast University), Ministry of Education, Nanjing, China,

<sup>2</sup> School of Mathematics and Actuarial Science, University of Leicester, Leicester, United Kingdom, <sup>3</sup> School of Informatics, University of Leicester, Leicester, United Kingdom, <sup>4</sup> Department of Radiology, Children's Hospital of Nanjing Medical University, Nanjing, China

## OPEN ACCESS

### Edited by:

Rong Chen,  
University of Maryland, Baltimore,  
United States

### Reviewed by:

Xuyun Wen,  
Nanjing University of Aeronautics and  
Astronautics, China  
Dimas Lima,  
Federal University of Santa  
Catarina, Brazil

### \*Correspondence:

Ming Yang  
yangming19710217@163.com  
Yu-Dong Zhang  
yudongzhang@ieee.org

Received: 29 March 2021

Accepted: 18 May 2021

Published: 18 June 2021

### Citation:

Wang S-H, Zhou Q, Yang M and  
Zhang Y-D (2021) ADVIAN:  
Alzheimer's Disease VGG-Inspired  
Attention Network Based on  
Convolutional Block Attention Module  
and Multiple Way Data Augmentation.  
*Front. Aging Neurosci.* 13:687456.  
doi: 10.3389/fnagi.2021.687456

**Aim:** Alzheimer's disease is a neurodegenerative disease that causes 60–70% of all cases of dementia. This study is to provide a novel method that can identify AD more accurately.

**Methods:** We first propose a VGG-inspired network (VIN) as the backbone network and investigate the use of attention mechanisms. We proposed an Alzheimer's Disease VGG-Inspired Attention Network (ADVIAN), where we integrate convolutional block attention modules on a VIN backbone. Also, 18-way data augmentation is proposed to avoid overfitting. Ten runs of 10-fold cross-validation are carried out to report the unbiased performance.

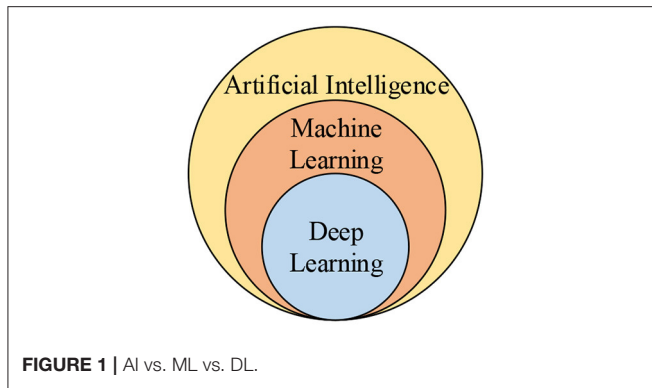
**Results:** The sensitivity and specificity reach  $97.65 \pm 1.36$  and  $97.86 \pm 1.55$ , respectively. Its precision and accuracy are  $97.87 \pm 1.53$  and  $97.76 \pm 1.13$ , respectively. The F1 score, MCC, and FMI are obtained as  $97.75 \pm 1.13$ ,  $95.53 \pm 2.27$ , and  $97.76 \pm 1.13$ , respectively. The AUC is 0.9852.

**Conclusion:** The proposed ADVIAN gives better results than 11 state-of-the-art methods. Besides, experimental results demonstrate the effectiveness of 18-way data augmentation.

**Keywords:** Alzheimer's disease, convolutional block attention module, VGG, transfer learning, deep learning, attention network, data augmentation

## BACKGROUND

Alzheimer's disease (AD) is a neurodegenerative disease, which affects 60%–70% of all cases of dementia (Alhazzani et al., 2020). The main symptom of AD is difficulty in short-term memory. As AD progressively worsens, patients exhibit symptoms such as mood and cognition (Lee et al., 2019), motivation loss, speech and language problems (Petti et al., 2020), spatial disorientation (Puthusserypady et al., 2020), sleep behaviors (Mather et al., 2021), etc. These symptoms lead to a significant decline in quality of life and an increase in care-taker burden (Scheltens et al., 2016; Fulton et al., 2019). AD's etiology is damage to brain cells observable on imaging scans



(Fulton et al., 2019) as the atrophy of anatomical structures like the cerebral cortex. The atrophy is caused by amyloid plaque (Ferreira et al., 2021) formation and neurofibrillary tangles (Kumari and Deshmukh, 2021). Manual differential diagnosis of AD is lab-intensive, onerous, and expensive due to various mental and physical tests, laboratory and neurological tests, and neuroimaging scans (Senova et al., 2021) [computed tomography (CT), positron emission tomography (PET), or magnetic resonance imaging (MRI)] which requires professional experts.

Therefore, scholars tend to use artificial intelligence (AI) approaches to create automatic models to identify AD. AI enables machines to mimic human behaviors. Machine learning (ML) is a subset of AI, which uses statistical methods to enable machines to improve. Deep learning (DL) is a subset of ML. DL makes the computation of deep neural networks feasible. Their relationship is displayed in **Figure 1**.

For instance, Plant et al. (2010) used brain region cluster (BRC) as a feature extractor. The authors tested three classifiers and found Bayesian classifier (BC) achieved the best performance. Their average accuracy of BRC-BC reached 92.00%. Savio and Grana (2013) employed the trace of Jacobian matrix (TJM) approach. Their method's average accuracy reached  $92.83 \pm 0.91\%$  over the Open Access Series of Imaging Studies (OASIS) dataset. Gray et al. (2013) presented a random forest (RF)-based similarity measures for multiple modality classification of AD. The authors included CSF biomarker measures, regional MRI volumes, voxel-based FDG-PET signal intensities, and categorical genetic information. Lahmiri and

**Abbreviations:** AD, Alzheimer's disease; ADNI, Alzheimer's disease neuroimaging initiative; AI, artificial intelligence; AP, average pooling; AUC, area under the curve; CAM, channel attention module; CBAM, convolutional block attention module; CDR, clinical dementia rating; CH, configuration of hyperparameters; CT, computed tomography; CV, cross-validation; DL, deep learning; FCL, fully connected layer; FM, feature map; FMI, Fowlkes–Mallows index; GF, gain field; HS, histogram stretching; HVS, human visual system; MC, motion correction; MCC, matthews correlation coefficient; ML, machine learning; MLP, multilayer perceptron; MMSE, mini-mental state exam; MP, max pooling; MRI, magnetic resonance imaging; MSD, mean and standard deviation; NWL, number of weighted layers; OASIS, open access series of imaging studies; OI, original image; PDNN, pretrained deep neural network; PET, positron emission tomography; ReLU, rectified linear unit; ROC, receiver operating characteristic; SAM, spatial attention module; SES, socioeconomic status; SN, speckle noise; TL, transfer learning; VGG, visual geometry group.

Boukadoum (2014) used fractal multiscale analysis (FMSA) to extract features. However, their dataset is small, with only 33 images. Zhang (2015) mingled displacement field (DF) with three different support vector machines, and they observed that the twin support vector machine yielded the best performance. Gorji and Haddadnia (2015) combined pseudo-Zernike moment (PZM) with a scaled conjugate gradient (SCG) algorithm. The experimental outcomes showcased that PZM with the order of 30 gave the paramount performance. Li (2018) presented a novel method to combine wavelet entropy (WE) with biogeography-based optimization (BBO). The interclass variance criterion was employed to pick out the single slice from the 3D image. Du (2017) reused PZM for feature extraction. They extracted 256 features from each brain image and substituted SCG with a linear regression classifier (LRC). Sui (2018) presented an eight-layer convolutional neural network (CNN). In traditional CNN, rectified linear unit (ReLU) is the default activation function. The authors replaced ReLU with a new activation function—leaky ReLU (LReLU). They tested three different pooling methods and found that max pooling gave the best performance. Jiang and Chang (2020) further improved the CNN structure and included batch normalization and dropout (BND) technique. Their method is abbreviated as CNN-BND in this paper. Dua et al. (2020) suggested a combination of DL models, which chose some primary models as CNN, recurrent neural networks (RNNs), and long short-term memory (LSTM). Its amalgamation achieved an accuracy of 92.22%. Sutoko et al. (2021) utilized a deep neural network with optimized stepwise feature selection and cross-validation method.

From previous studies, we can observe DL methods can have better performance than traditional ML methods. As mentioned before, DL is a subfield of ML (see **Figure 1**), but DL powers itself by using a human-like artificial deep neural network to learn and make decisions by itself from given data (Saood and Hatem, 2021).

To further improve the performance of DL, there are three possible ways: (i) depth, (ii) width, and (iii) cardinality of the deep neural networks. We try to improve the performance from the fourth way—the attention mechanism. In all, we propose a novel DL model termed Alzheimer's Disease VGG-Inspired Attention Network (ADVIAN). The contributions of our paper are listed as following four points:

1. A VGG-inspired network (VIN) is particularly designed as the backbone model to identify AD.
2. Convolutional block attention modules are integrated to introduce attention to the VIN.
3. Multiple-way data augmentation is introduced to make test performance more reliable.
4. The test results prove our ADVIAN model is better than 11 state-of-the-art methods.

## SUBJECTS

The dataset we used is already reported in the work of Sui (2018), where 28 AD patients and 98 healthy control (HC) subjects were selected from the OASIS-1 dataset (Ardekani et al.,

**TABLE 1** | Demographics of dataset in this study.

Trait	OASIS		Local hospitals
	AD (28)	HC (98)	AD (70)
Gender (M/F)	9/19	26/72	24/46
Age	77.75 ± 6.99	75.91 ± 8.98	76.34 ± 7.81
SES	2.87 ± 1.29	2.51 ± 1.09	2.89 ± 1.16
Education	2.57 ± 1.31	3.26 ± 1.31	2.63 ± 1.42
MMSE	21.67 ± 3.75	28.95 ± 1.20	21.12 ± 4.62
CDR	1	0	1

SES, socioeconomic status; CDR, clinical dementia rating; MMSE, mini-mental state exam.

2013). The selection criterion is to remove individuals under 60 and incomplete observations. Meanwhile, 70 AD subjects were enrolled from local hospitals. Hence, we have a balanced dataset, of which the demographics are itemized in **Table 1**, where SES means Socioeconomic Status, MMSE Mini-Mental State Exam, and CDR Clinical Dementia Rating.

There are AD researchers favoring Alzheimer's disease neuroimaging initiative (ADNI) (Abuhmed et al., 2021), and many others use OASIS, which is freely accessible, grants sensible demographics for proof of concept, and generalizes easily for forthcoming longitudinal studies.

## PREPROCESSING

The same preprocessing procedure (shown in **Figure 2**) applies to all the images in this dataset. First,  $1 \leq n \leq 4$  multiple raw scans of the same structural protocol within a single session of the same person is carried out; we obtain  $n$  volumetric images as  $V_R(n)$ .

Second, motion correction (MC) is performed over all the  $n$  raw images. The motion-corrected images are symbolized as  $V_{MC}(n)$ .

Third, an average image  $V_A$  is obtained by averaging all the  $n$  motion-corrected images, i.e.,

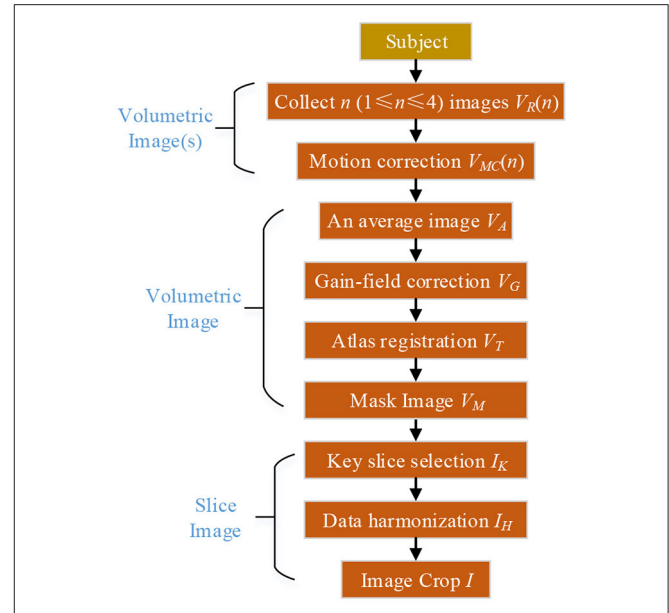
$$V_A = \frac{1}{n} \sum_{i=1}^n V_{MC}(n) \quad (1)$$

Fourth, gain field (GF) correction is performed. The GF is intensity variations unrelated to the subject's anatomical information. GF may relate to movement, nearly static fields, radiofrequency turbulence, or additional nonsubject causes (Hou, 2006). The image is now symbolized as  $V_G$ .

Fifth, atlas registration will spatially normalize the image  $V_G$  to Talairach atlas (Saletin et al., 2019) and obtain the image  $V_T$ .

Sixth, a masked image  $V_M$  is obtained by removing all the nonbrain voxels. We do not do gray matter/white matter/CSF segmentation at this stage.

Seventh, a key slice is selected  $I_K$  from the masked volumetric image  $V_M$ . There are three view angles: axial, sagittal, and coronal view angles, as shown in **Figure 3**. In this study, we chose the 80th

**FIGURE 2** | Pipeline of preprocessing.

axial  $I_K$  out of 176 slices. The key slice is considered the original image (OI).

Eighth, data harmonization is performed *via* histogram stretching (HS) (Luo et al., 2021) to counter intersource variability from the difference between our dataset's two sources. The HS is indispensable to normalize the interscan images by increasing the difference between the maximum intensity value and the minimum one in an image. Mathematically, HS (Luo et al., 2021) altered OI  $x$  to a different image  $y$  as:

$$y(i, j) = \frac{x(i, j) - x_{min}}{x_{max} - x_{min}} \quad (2)$$

where  $x_{min}$  and  $x_{max}$  stand for the minimum and maximum intensity values of OI, respectively.

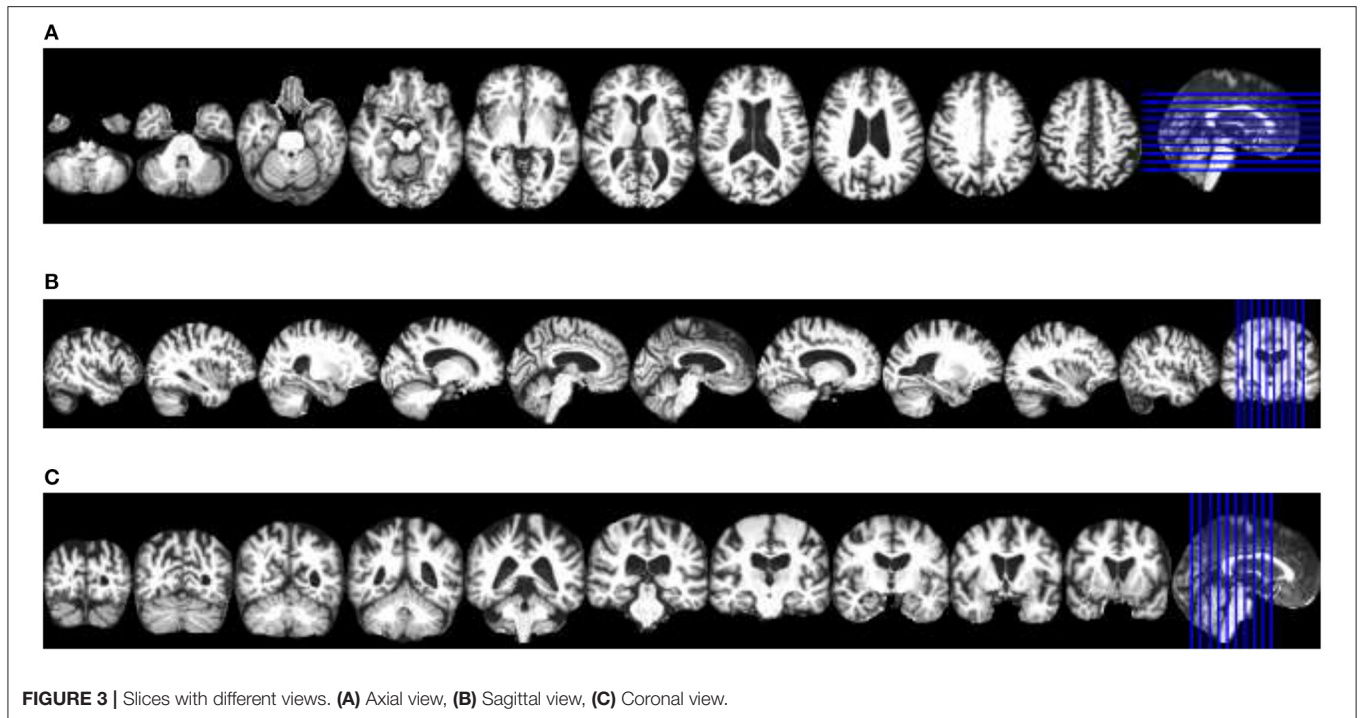
Traditionally, the minimum and maximum correspond to 0 and 100% of the whole grayscale range. In this study, 5 and 95% are employed to replace 0 and 100%, respectively. The motivation is the pixels with the least (0%) and the greatest (100%) values are more susceptible to noises. Using the  $95 - 5 = 90\%$  interval can make HS more dependable than using the 100% interval. After this step, we get harmonized image  $I_H$ .

Finally, the image  $I_H$  is cropped. The cropped image  $I$  has the size of  $[176 \times 176]$ . Two key slices of one AD sample and one HC sample are displayed in **Figure 4**.

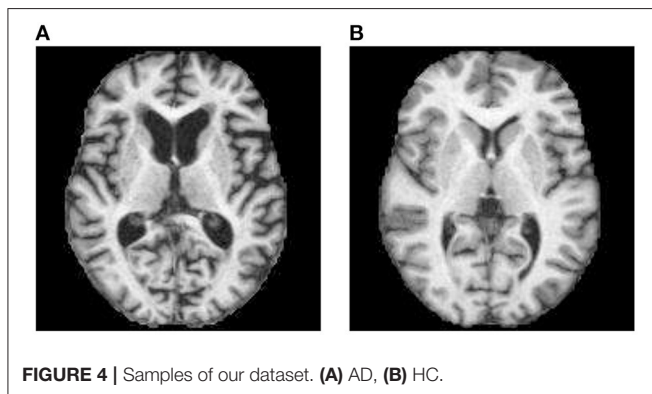
## METHODOLOGY

### Background of VGG-16

Transfer learning (TL) stores knowledge gained while solving one problem and applies it to solve a different but related problem (Santana and Silva, 2021). Most pretrained deep neural networks (PDNNs) are trained on a subset of ImageNet database. Those



**FIGURE 3** | Slices with different views. (A) Axial view, (B) Sagittal view, (C) Coronal view.



**FIGURE 4** | Samples of our dataset. (A) AD, (B) HC.

PDNNs could classify images into 1,000 object categories. Hence, using PDNNs for TL is easier and faster than training networks from scratch.

VGG stands for Visual Geometry Group, an academic group at Oxford University. This team presented two famous networks: VGG-16 (Jahangeer and Rajkumar, 2021) and VGG-19 (Sudha and Ganeshbabu, 2021), which are included as library packages of popular programming languages such as Python and MATLAB. This study chooses VGG-16 because it is easier to implement and has less layers, while VGG-16 has similar performance of VGG-19.

**Figure 5A** displays the structure of VGG-16, which is composed of five conv blocks and three fully connected layers (FCLs). The input of VGG-16 is  $224 \times 224 \times 3$ . After the 1st convolution block (CB), the output is  $112 \times 112 \times 64$ . Components of 1st CB are shown in **Table 2**. The 1st CB can be written as

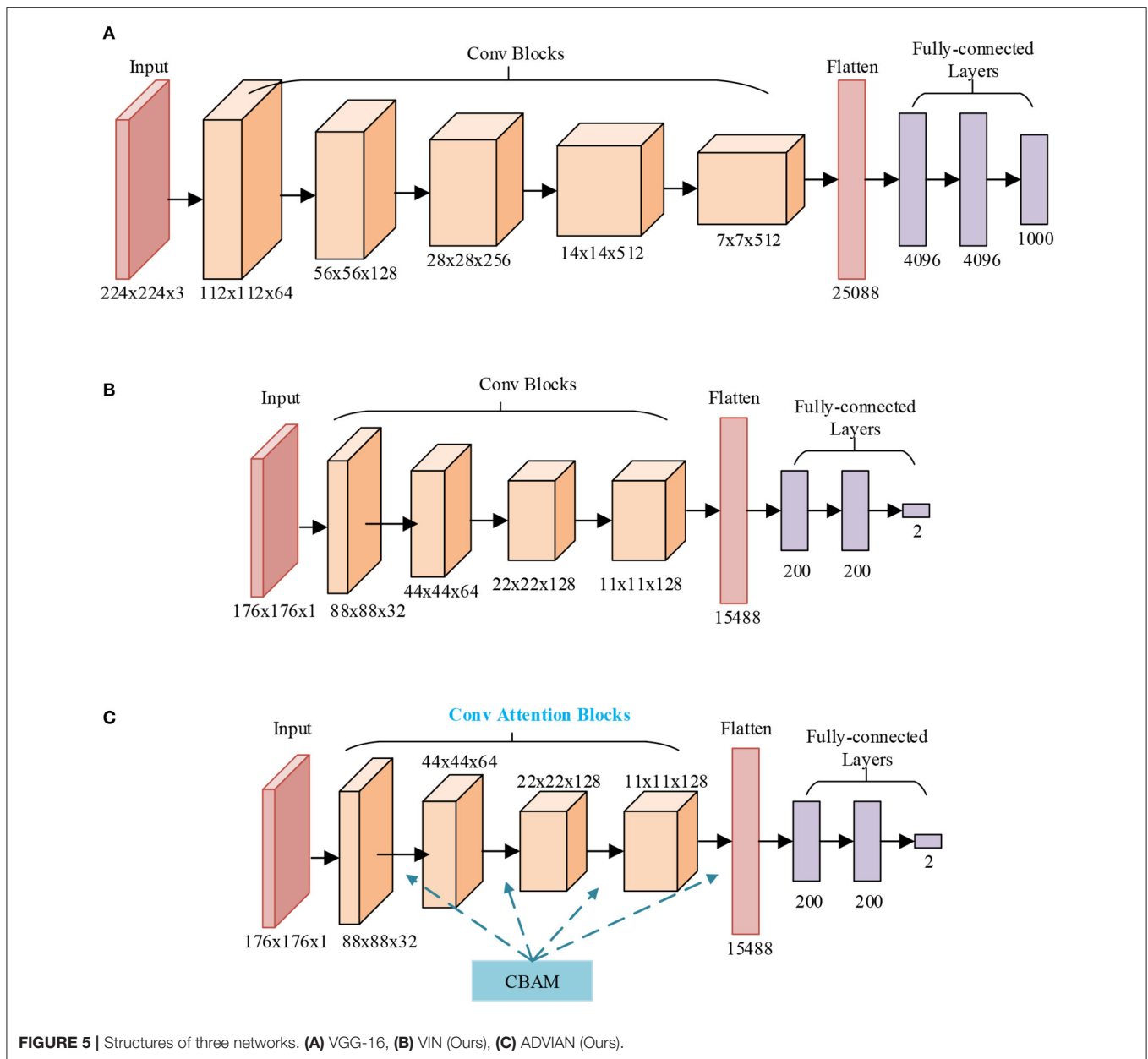
$“2 \times (64 \ 3 \times 3) / 2,”$  which means “2 repetitions of 64 kernels with sizes of  $3 \times 3$  followed by a max pooling with a kernel size of  $2 \times 2$ .” Note that (i) ReLU layers are skipped in the following texts as default. (ii) Stride and padding are not included since they can be calculated easily.

The 2nd CB “ $2 \times (128 \ 3 \times 3) / 2,$ ” 3rd CB “ $3 \times (256 \ 3 \times 3) / 2,$ ” 4th CB “ $3 \times (512 \ 3 \times 3) / 2,$ ” and 5th CB “ $3 \times (512 \ 3 \times 3) / 2,$ ” produce the feature maps (FMs) with sizes of  $56 \times 56 \times 128,$   $28 \times 28 \times 256,$   $14 \times 14 \times 512,$  and  $7 \times 7 \times 512,$  respectively. Afterward, FM is compressed into a column vector of 25,088 neurons and sent into three FCLs with 4,096, 4,096, and 1,000 neurons, respectively.

## VGG-Inspired Network

A VIN is designed, shown in **Figure 5B**, as our task’s backbone network. The VIN is inspired by VGG-16. The VIN contains four CBs and three FCLs. The first CB “ $2 \times [3 \times 3, 32] / 2$ ” contains two repetitions of 32 kernels with sizes of  $3 \times 3$  followed by a max pooling with a kernel size of  $2 \times 2$ . After four CBs, the size of FM becomes  $11 \times 11 \times 128$ . The flattening layer vectorizes the FM into a vector with a size of  $1 \times 1 \times 15,488$ . After three consecutive FCLs, we output a binary code that represents either AD or HC. The structure of the proposed 13-layer VIN is depicted in **Table 3**, where NWI represents the number of weighted layers, and CH configuration of hyperparameters.

The similarities between the proposed VIN and VGG-16 are itemized in **Table 4**. Apart from those six similarity aspects (Fernandes, 2021), there are several differences between the proposed VIN and VGG-16. The input of VGG-16 is  $224 \times 224 \times 3$ , while the input of VIN is  $176 \times 176 \times 1$ . The output of VGG-16 is 1,000 neurons corresponding to 1,000



categories to be classified, while the output of VIN is 2 neurons because our task is a binary-coded problem. Also, some structural differences exist between those two networks, which can be observed from **Figure 5** and **Table 4**.

## Human Visual System and Attention Mechanism

To increase the functioning of the recent deep neural networks, numerous investigations are carried out in terms of either width, or depth, or cardinality. For examples, (i) the network structures reported in recent ResNet (He et al., 2016) and DenseNet (Huang et al., 2017) show that deeper network (over 1,000 weighted layers) will have better performance in general; (ii) GoogleNet demonstrates

that width (Szegedy et al., 2015) is another critical factor to improve the implementation; Zagoruyko and Komodakis (2016) present wide residual networks, in which the authors reduce the depth and enlarge the width of residual networks; (iii) Xie et al. (2017) expose a new dimension “cardinality” defined as the size of the set of transformations and proves increasing cardinality is more effective than going wider or going deeper.

“Attention” is the fourth possible way to improve the network’s performance. There are many papers using attention to improve their networks. Lee et al. (2021) proposed an attention recurrent neural network to estimate severity. Song et al. (2021) presented a coarse-to-fine dual-view attention network for click-through rate prediction. Arora et al. (2021)

**TABLE 2** | Components of 1st CB “2 × (64 3 × 3) /2” in VGG\_16.

Layer	Component
1	1 convolutional layer with 64 kernels with sizes of 3 × 3 and stride [1, 1] and padding [1, 1, 1, 1]
2	1 ReLU layer
3	1 convolutional layer with 64 kernels with sizes of 3 × 3 and stride [1, 1] and padding [1, 1, 1, 1]
4	1 ReLU layer
5	1 max pooling layer with a kernel size of 2 × 2

**TABLE 3** | Arrangement of our 13-layer VIN.

Index	Tag	NWL	CH	Size of FM
1	Input	0	0	176 × 176 × 1
2	CB-1	2	2 × [3 × 3, 32] / 2	88 × 88 × 32
3	CB-2	2	2 × [3 × 3, 64] / 2	44 × 44 × 64
4	CB-3	3	3 × [3 × 3, 128] / 2	22 × 22 × 128
5	CB-4	3	3 × [3 × 3, 128] / 2	11 × 11 × 128
6	Flatten	0	0	15,488
7	FCL-1	1	200 × 15,488, 200 × 1	200
8	FCL-2	1	200 × 200, 200 × 1	200
9	FCL-3	1	2 × 200, 2 × 1	2

NWL, number of weighted layers; CH, configuration of hyperparameters; FM, feature map.

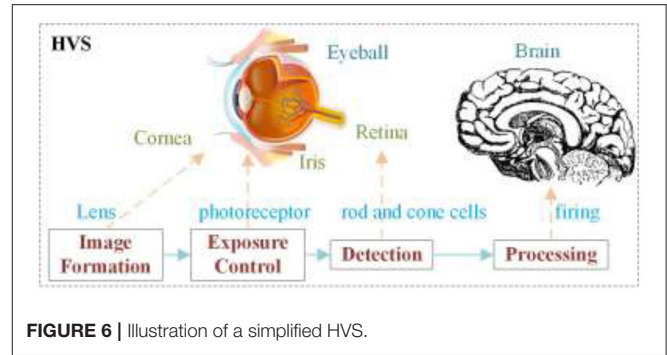
**TABLE 4** | Similarity facets between proposed VIN and VGG-16.

Key	Similarity facet
A	Employing small convolution kernels with size of (3 × 3)
B	Employ small max pooling kernel with size of (2 × 2)
C	Each CB contains a few repetitions of conv layers followed by a max pooling layer
D	Fully connected layers are put at the end of the deep network
E	Channel number increase as it goes from input to the last conv layer, later decreases as to output.
F	Size of FMs shortens as it goes from input to output

offered an attention-based deep network for automated skin lesion segmentation.

In all, attention acts an essential role within the human visual system (HVS) (Choi et al., 2020). **Figure 6** displays a simplified instance of HVS, in which image formation is first seized by the lens of the human eye’s cornea. Thenceforth, the iris makes use of the photoreceptor sensitivity to control the exposure. Afterward, the information stream is passed to cone and rod cells in the retina. At long last, the neural firing is forwarded to the brain for additional handling.

Human eyes do not endeavor to sort out the whole scenarios captured at one time. In contrast, human beings take the full practice of partial glimpses and fix on salient features selectively to grab a sounder pictorial structure. Thus, the recent attention networks (Oh et al., 2021) embedding attention mechanism will have the advantages of (a) focusing on those critical and salient



**FIGURE 6** | Illustration of a simplified HVS.

features, (b) performing more successful than networks without attention mechanism, and (c) become more reliable to noisy inputs than networks without attention mechanism.

### ADVIAN

Woo et al. (2018) presented a new convolutional block attention module (CBAM), which not only informs the neural network model of the regions to focus but also perfects the representation of interests. In their paper, the core idea of CBAM is to improve the 3D FMs by being trained with channel attention and spatial attention, respectively.

CBAM is composed of two consecutive submodules: (i) channel attention module (CAM) and (ii) spatial attention module (SAM). The complete relation between CBAM and its two submodules is exposed in **Figure 7**.

Suppose we have a provisional input FM of  $F \in \mathbb{R}^{C \times H \times W}$ . The CBAM applies 1D CAM  $N_{CAM} \in \mathbb{R}^{C \times 1 \times 1}$  and a 2D SAM  $N_{SAM} \in \mathbb{R}^{1 \times H \times W}$  in sequence to the input  $F$ , as illustrated in **Figure 7**. Thus, the channel-refined FM and the final FM are obtained as:

$$\begin{cases} Q = N_{CAM}(P) \otimes P \\ R = N_{SAM}(Q) \otimes Q \end{cases} \quad (3)$$

where  $\otimes$  means the element-wise multiplication.

If the two operands are not with the same dimension, then the values are transmitted (copied) in such tactics that the spatial attentional values are transmitted by the channel dimension, and the channel attention values are transmitted by the spatial dimension (Fernandes, 2021).

Firstly, CAM is defined. Both max pooling (MP)  $f_{mp}$  and average pooling (AP)  $f_{ap}$  are applied, breeding two features  $S_{ap}$  and  $S_{mp}$ .

$$\begin{cases} S_{ap} = f_{ap}(P) \\ S_{mp} = f_{mp}(P) \end{cases} \quad (4)$$

Both are thenceforth sent on to a shared shallow neural network—multilayer perceptron (MLP) (Tiwari, 2021), to produce the output FMs, that are thenceforth united *via* element-wise summation  $\oplus$ . Normally, MLP consists of three layers of nodes: an input layer, a hidden layer, and an output layer, as

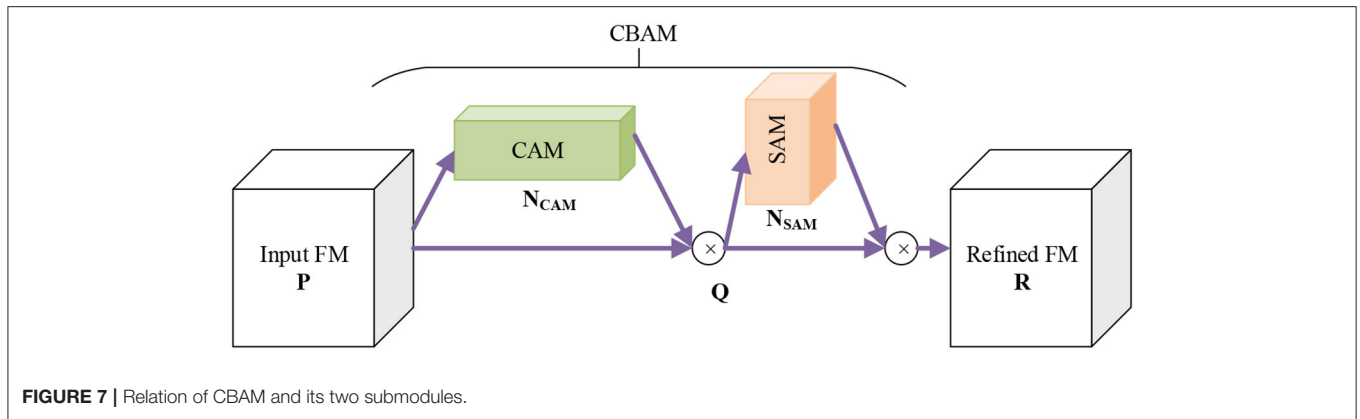


FIGURE 7 | Relation of CBAM and its two submodules.

shown in **Figure 8A**. The united sum is then sent to the sigmoid function  $\beta$ . Precisely,

$$N_{CAM}(P) = \beta \{MLP[S_{ap}] \oplus MLP[S_{mp}]\} \quad (5)$$

To decrease the parameter reserves, the number of hidden neurons of MLP is arranged to  $\mathbb{R}^{C/e_r \times 1 \times 1}$ , where  $e_r$  is identified as the reduction ratio. Let  $W_0 \in \mathbb{R}^{C/e_r \times C}$  and  $W_1 \in \mathbb{R}^{C \times C/e_r}$  mean the MLP weights, respectively, Equation (5) is updated as:

$$N_{CAM}(P) = \beta \{W_1[W_0(S_{ap})] \oplus W_1[W_0(S_{mp})]\} \quad (6)$$

See  $W_0$  and  $W_1$  are shared by both  $S_{ap}$  and  $S_{mp}$ . **Figure 8A** shows the flowchart of CAM.

Second, SAM is defined. The spatial attention module  $N_{SAM}$  is a paired phase to the preceding channel attention module  $N_{CAM}$ . The AP operation  $f_{ap}$  and MP operation  $f_{mp}$  are harnessed to the channel-refined FM  $Q$ , and we gain

$$\begin{cases} T_{ap} = f_{ap}(Q) \\ T_{mp} = f_{mp}(Q) \end{cases} \quad (7)$$

Both  $T_{ap}$  and  $T_{mp}$  are two-dimensional FMs:  $T_{ap} \in \mathbb{R}^{1 \times H \times W} \wedge T_{mp} \in \mathbb{R}^{1 \times H \times W}$ , which are concatenated jointly along the channel dimension as

$$T = f_{con}^{cha}(T_{ap}, T_{mp}) \quad (8)$$

where  $f_{con}^{cha}$  stands for the concatenation along channel dimension.

The concatenated FM  $T$  is thenceforth sent into a typical convolution with a size of  $7 \times 7 f_{conv}$ . The resultant FM is sent to the sigmoid function  $\beta$ . Altogether, we find:

$$N_{SAM}(Q) = \beta \{f_{conv}[T]\} \quad (9)$$

The yielded  $N_{SAM}(Q)$  is subsequently element-wisely multiplied with  $Q$ , as displayed in Equation (3). **Figure 8B** portrays the diagram of SAM.

The previously introduced CBAM is integrated into the proposed VIN network, which renders the proposed ADVIAN shown in **Figure 5C**, which has the same FM structure as VIN in **Figure 5B**. The difference between ADVIAN and VIN is that we

add CBAM after each CB, and thus we called each block as “conv attention block (CAB),” as shown in **Figure 9**.

For any FM  $P$  of each previous CB, the two uninterrupted attention modules (channel and spatial) are attached, coupled with the refined FM  $R$  which is driven to the succeeding block. Now CAB is made up of one CB and succeeding CBAM module. Comparing **Figures 7, 9**, we can observe the relationship among CAB, CBAM, and CB.

As default, the softmax function  $f_s: \mathbb{R}^K \mapsto \mathbb{R}^K$  is appended at the end of our model. Suppose the input to the softmax is  $z = (z_1, \dots, z_i, \dots, z_K) \in \mathbb{R}^K$ , we have

$$f_s(z)_i = \frac{\exp(z_i)}{\sum_{j=1}^K \exp(z_j)} \quad (10)$$

The softmax function can be regarded as the output unit activation function. For classification-oriented deep neural networks, a softmax layer and a classification layer must follow the last FCL. Also, batch normalization (Vrzal et al., 2021) layers are embedded as assisting layers.

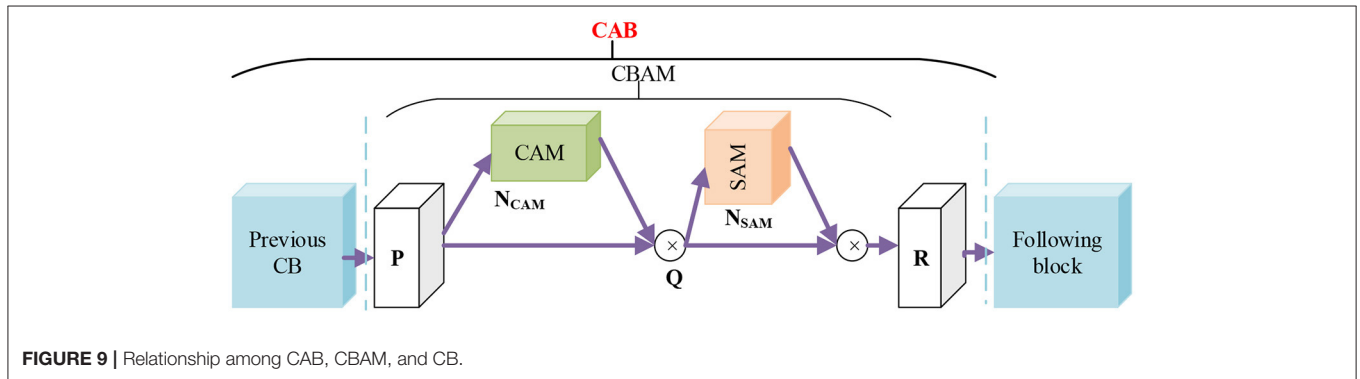
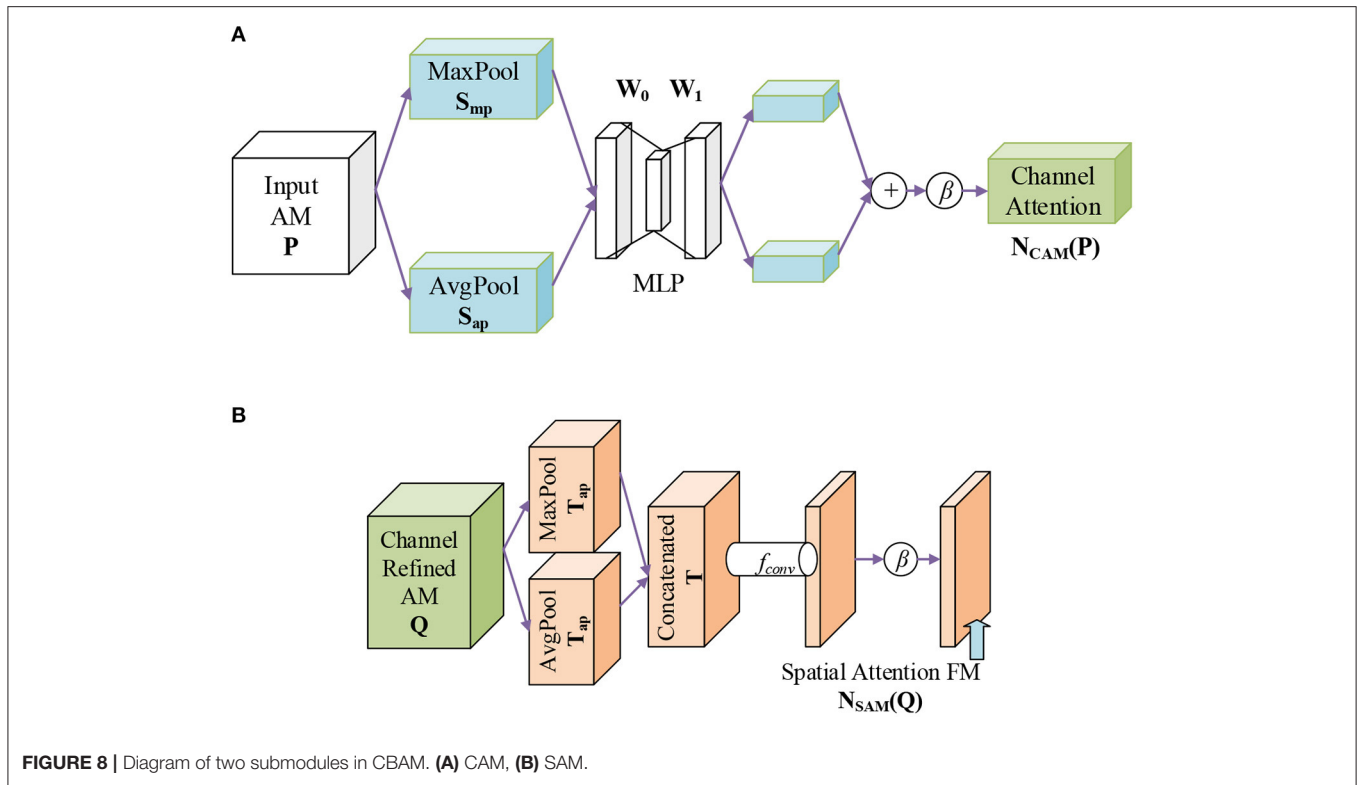
### Cross-Validation

Cross-validation (CV) (Albashish et al., 2021) is a resampling route to evaluate AI models on a limited-size dataset. **Figure 10** shows the diagram of the  $K$ -fold CV. The whole dataset is split into  $K$  folds evenly. Then for  $k$ th ( $k = 1, \dots, K$ ) trial, the  $k$ th fold is used for test, and all the other folds ( $1, \dots, k - 1, k + 1, \dots, K$ ) for training. We repeat  $K$  trials to facilitate each fold used for test only once. The above  $K$ -fold cross-validation will repeat  $R$  times. In this study, we set  $K = R = 10$ .

### Multiple-Way Data Augmentation

Overfitting may occur due to the small-size dataset in this study. To avoid this, multiple-way data augmentation (MDA) is employed. MDA is a variant of the traditional data augmentation (DA) method. Cheng (Cheng, 2021) presented a 16-way DA to identify COVID-19 chest CT image. In their method, the number of DA is set to  $J_1 = 8$ , i.e., eightway different DA were applied to original raw image  $r(x)$  and the horizontally mirrored version  $r^h(x)$ .

In this method, we propose an 18-way DA, of which the diagram is displayed in **Figure 11**. The difference of our 18-way



DA against 16-way DA (Cheng, 2021) is that we add the speckle noise (SN) to both  $r(x)$  and  $r^h(x)$ , respectively. the SN altered image is defined as

$$r^{SN}(x) = r(x) + N_R^* r(x), \tag{11}$$

where  $N_R$  is uniformly distributed random noise. In this study, we set the mean and variance of  $N_R$  to 0 and 0.05, respectively.

First,  $J_1$ -different DA methods as displayed in **Figure 11** are applied to raw training image  $r(x)$ . Let  $H_j, j = 1, \dots, J_1$  denotes each DA operation, we have the augmented images of raw image  $r(x)$  as

$$H_j[r(x)], j = 1, \dots, J_1 \tag{12}$$

Suppose  $J_2$  means the size of generated new images for each DA method, then,

$$|H_j[r(x)]| = J_2 \tag{13}$$

where  $||$  represents the number of elements in the set.

Second, horizontally mirrored image  $r^h(x)$  is generated by

$$r^h(x) = f_{HM}[r(x)] \tag{14}$$

where  $f_{HM}$  stands for horizontal mirror function.

Third, all the  $J_1$  different DA methods are performed on the mirror image  $r^h(x)$  and generate  $J_1$  different datasets.

$$\begin{cases} H_j[r^h(x)], j = 1, \dots, J_1 \\ |H_j[r^h(x)]| = J_2, j = 1, \dots, J_1 \end{cases} \tag{15}$$



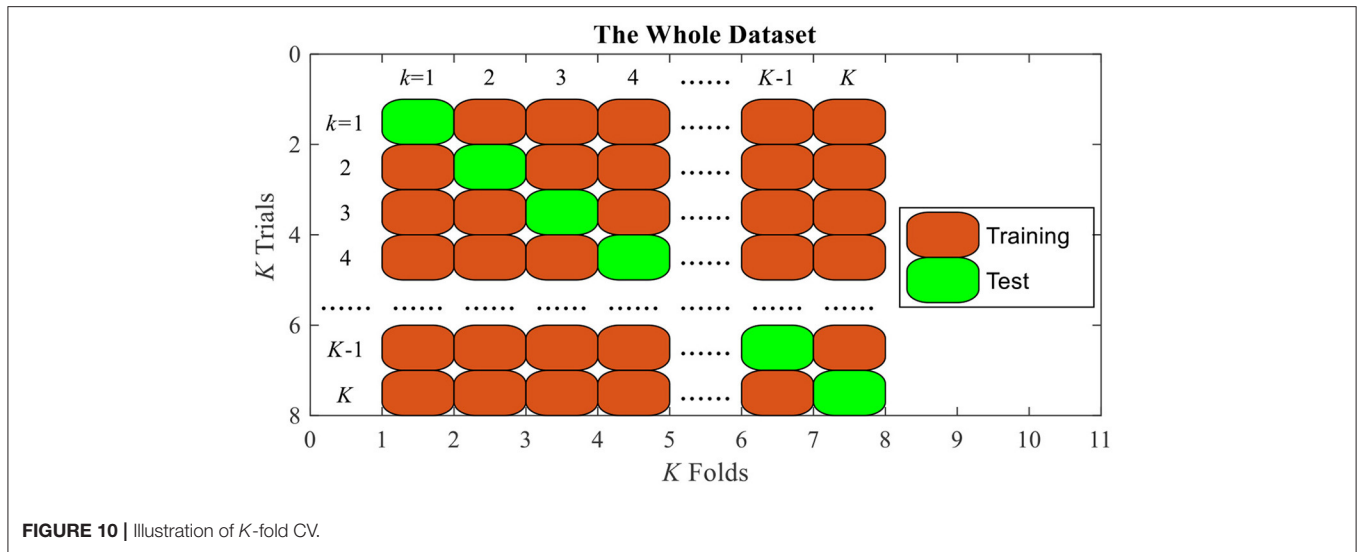


FIGURE 10 | Illustration of K-fold CV.

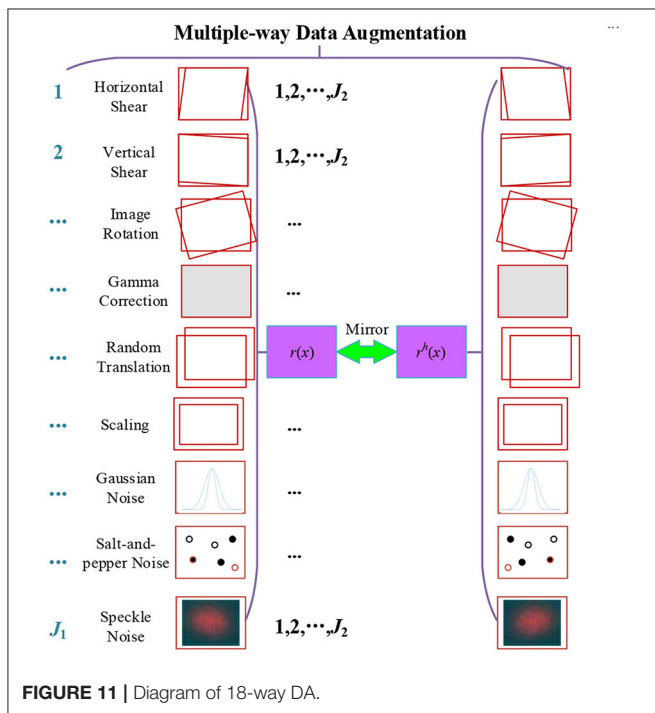


FIGURE 11 | Diagram of 18-way DA.

Fourth, the raw image  $r(x)$ , the horizontally mirrored image  $r^h(x)$ ,  $J_1$ -way datasets of raw image  $\mathbf{H}_j[r(x)]$ , and  $J_1$ -way datasets of horizontally mirrored image  $\mathbf{H}_j[r^h(x)]$  are combined. The final generated dataset from  $r(x)$  is defined as  $\mathbf{R}(x)$ :

$$r(x) \mapsto \mathbf{R}(x) = f_{fuse} \left\{ \begin{array}{cc} \underbrace{\begin{matrix} r(x) \\ \mathbf{H}_1[r(x)] \end{matrix}}_{J_2} & \underbrace{\begin{matrix} r^h(x) \\ \mathbf{H}_1[r^h(x)] \end{matrix}}_{J_2} \\ \dots & \dots \\ \underbrace{\begin{matrix} \mathbf{H}_{J_1}[r(x)] \end{matrix}}_{J_2} & \underbrace{\begin{matrix} \mathbf{H}_{J_1}[r^h(x)] \end{matrix}}_{J_2} \end{array} \right\} \quad (16)$$

where  $f_{fuse}$  is the concatenation function.

Suppose augmentation factor is  $J_3$ , which represents the number of images in  $\mathbf{R}(x)$ , we get

$$J_3 = \frac{|\mathbf{R}(x)|}{|r(x)|} = \frac{(1 + J_1 \times J_2) \times 2}{1} = 2 \times J_1 \times J_2 + 2 \quad (17)$$

Algorithm 1 recaps the pseudocode of the 18-way DA method. We set  $J_1 = 9$ ,  $J_2 = 30$ ; thus,  $J_3 = 542$ .

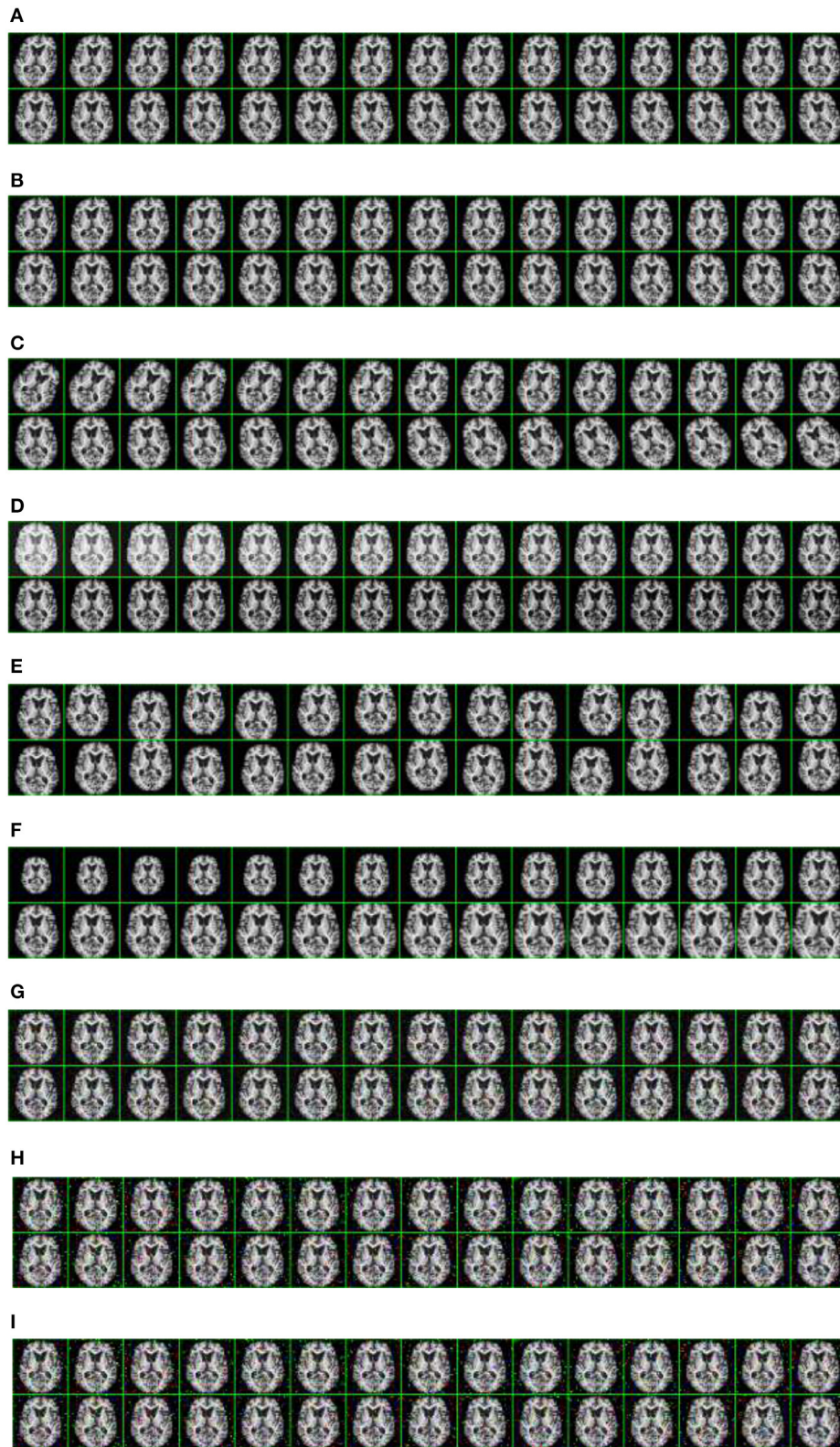
**Algorithm 1 | Pseudocode of 18-way data augmentation.**

Input	Import raw preprocessed training image $r(x)$ .
Step A	$J_1$ geometric/photometric/noise-injection DA transforms $\mathbf{H}_j$ are utilized on $r(x)$ . We obtain datasets $\mathbf{H}_j[r(x)]$ , $j = 1, \dots, J_1$ . See Eq. (12). Each enhanced dataset comprises $J_2$ new images. See Eq. (13).
Step B	Horizontally mirrored image is obtained by $r^h(x) = f_{HM}[r(x)]$ . See Eq. (14).
Step C	$J_1$ -way DA transforms are implemented on $r^h(x)$ , we obtain datasets $\mathbf{H}_j[r^h(x)]$ , $j = 1, \dots, J_1$ . See Eq. (15).
Step D	$r(x)$ , $r^h(x)$ , $\mathbf{H}_j[r(x)]$ , $j = 1, \dots, J_1$ , and $\mathbf{H}_j[r^h(x)]$ , $j = 1, \dots, J_1$ are combined. See Eq. (16).
Output	Output a new dataset $\mathbf{R}(x)$ . Its image number is $J_3 = 2 \times J_1 \times J_2 + 2$ . See Eq. (17).

**Evaluation**

The evaluation was reported on the  $R$  runs of  $K$ -fold CV of our 98–98 image dataset. Suppose the image number of each class is  $T_k$  ( $k = 1, 2$ ). The perfect confusion matrix (CM) is

$$O^{ideal} = \left\{ o^{ideal} \right\} = R \times \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix}, \quad (18)$$



**FIGURE 12 |** Results of data augmentation. **(A)** Horizontal shear, **(B)** Vertical shear, **(C)** Image rotation, **(D)** Gamma correction, **(E)** Random translation, **(F)** Scaling, **(G)** Gaussian noise, **(H)** Salt-and-pepper noise, **(I)** Speckle noise.

where the off-diagonal entries of ideal  $O^{ideal}$  are all 0 s, viz.,  $o^{ideal}(i, j) = 0, \forall i \neq j$ . The realistic confusion matrix is

$$O = \{o\} = \begin{bmatrix} o(1,1) & o(1,2) \\ o(2,1) & o(2,2) \end{bmatrix}. \tag{19}$$

Now, we define positive (P) and negative (N) classes. The meaning of TP, TN, FP, and FN are shown in **Table 5**.

Nine measures are used: sensitivity, specificity, precision, accuracy, F1 score, Matthews correlation coefficient (MCC) (Daines et al., 2020), Fowlkes–Mallows index (FMI) (Monteiro et al., 2018), receiver operating characteristic (ROC), and area under the curve (AUC). The first four measures are defined as

$$\begin{cases} \text{Sen} = \frac{o(1,1)}{o(1,1)+o(1,2)} & \text{Spc} = \frac{o(2,2)}{o(2,2)+o(2,1)} \\ \text{Prc} = \frac{o(1,1)}{o(1,1)+o(2,1)} & \text{Acc} = \frac{o(1,1)+o(2,2)}{o(1,1)+o(2,2)+o(1,2)+o(2,1)} \end{cases} \tag{20}$$

and the middle three measures are defined as:

$$F_1 = 2 \times \frac{\text{Sen} \times \text{Prc}}{\text{Sen} + \text{Prc}} = \frac{2 \times o(1,1)}{2 \times o(1,1) + o(1,2) + o(2,1)} \tag{21}$$

$$\text{MCC} = \frac{o(1,1) \times o(2,2) - o(2,1) \times o(1,2)}{\sqrt{[o(1,1) + o(2,1)] \times [o(1,1) + o(1,2)] \times [o(2,2) + o(2,1)] \times [o(2,2) + o(1,2)]}} \tag{22}$$

$$\text{FMI} = \sqrt{\text{Sen} \times \text{Prc}} = \sqrt{\frac{o(1,1)}{o(1,1) + o(1,2)} \times \frac{o(1,1)}{o(1,1) + o(2,1)}} \tag{23}$$

The above measures are calculated in the mean and standard

deviation (MSD) format. Besides, ROC is a curve to measure a binary classifier with varying discrimination thresholds. The ROC curve is created by plotting the sensitivity against 1-specificity. The AUC is calculated based on the ROC curve.

## EXPERIMENTS AND RESULTS

### Multiple-Way Data Augmentation

**Figure 12** displays the part of 18-way DA results (i.e.,  $H_j[r(x)], j = 1, \dots, J_1$ ) if we take **Figure 4A** as the raw image  $r(x)$ . From **Figure 12**, we can observe that this 18-way DA improves the diversity of our training set, which will make our classifier model more robust. In the following experiments, we shall prove this robustness.

### Statistical Analysis

The results of 10 runs of 10-fold cross-validation of our model ADVIAN are itemized in **Table 6**. The sensitivity and specificity

**TABLE 5** | Meanings in measures.

Abbreviation	Full form	Symbol	Meaning
P	Positive	AD	
N	Negative	HC	
TP	True positive	$o(1,1)$	AD images are classified correctly.
FP	False positive	$o(2,1)$	HC images are wrongly classified as AD.
TN	True negative	$o(2,2)$	HC images are classified correctly.
FN	False negative	$o(1,2)$	AD images are wrongly classified as HC.

**TABLE 6** | Results of proposed ADVIAN model.

Run	Sen	Spc	Prc	Acc	F1	MCC	FMI
1	100.00	97.96	98.00	98.98	98.99	97.98	98.99
2	97.96	96.94	96.97	97.45	97.46	94.90	97.46
3	97.96	95.92	96.00	96.94	96.97	93.90	96.97
4	97.96	100.00	100.00	98.98	98.97	97.98	98.97
5	95.92	98.98	98.95	97.45	97.41	94.94	97.42
6	97.96	95.92	96.00	96.94	96.97	93.90	96.97
7	97.96	98.98	98.97	98.47	98.46	96.94	98.46
8	95.92	96.94	96.91	96.43	96.41	92.86	96.41
9	95.92	96.94	96.91	96.43	96.41	92.86	96.41
10	98.98	100.00	100.00	99.49	99.49	98.98	99.49
MSD	97.65 ± 1.36	97.86 ± 1.55	97.87 ± 1.53	97.76 ± 1.13	97.75 ± 1.13	95.53 ± 2.27	97.76 ± 1.13

reach  $97.65 \pm 1.36$  and  $97.86 \pm 1.55$ , respectively. Its precision and accuracy are  $97.87 \pm 1.53$  and  $97.76 \pm 1.13$ , respectively. The F1 score, MCC, and FMI are obtained as  $97.75 \pm 1.13$ ,  $95.53 \pm 2.27$ , and  $97.76 \pm 1.13$ , respectively. We can see that all the seven indicators of our model are above 95%. The ROC curve is displayed in **Figure 14B**, and the AUC is 0.9852.

### Effect of 18-Way DA

To validate the importance of 18-way DA, we carry out an ablation study in which we remove 18-way DA from our model and observe the performance change. After another 10 runs of 10-fold CV, the performances decrease to a sensitivity of  $92.45 \pm 2.21$ , a specificity of  $94.18 \pm 1.99$ , a precision of

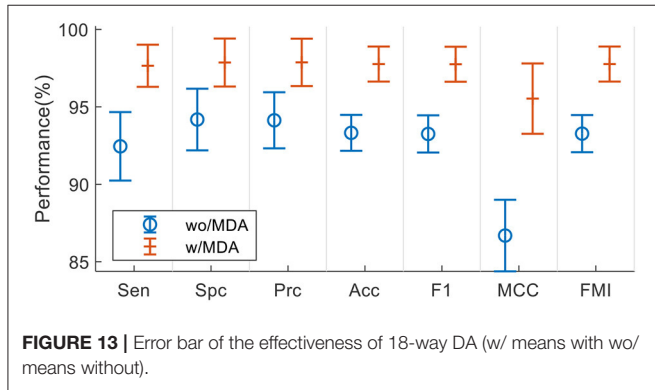
94.13 ± 1.81, an accuracy of 93.32 ± 1.16, and an F1 score of 93.25 ± 1.20. The MCC and FMI decrease to 86.69 ± 2.31 and 93.27 ± 1.20, respectively. The result of comparison with and without 18-way DA is shown in **Figure 13**. The ROC curve comparison is shown in **Figure 14**, where we can observe that

AUC without 18-way DA is only 0.9603 (**Figure 14A**) and AUC with 18-way DA is 0.9852 (**Figure 14B**).

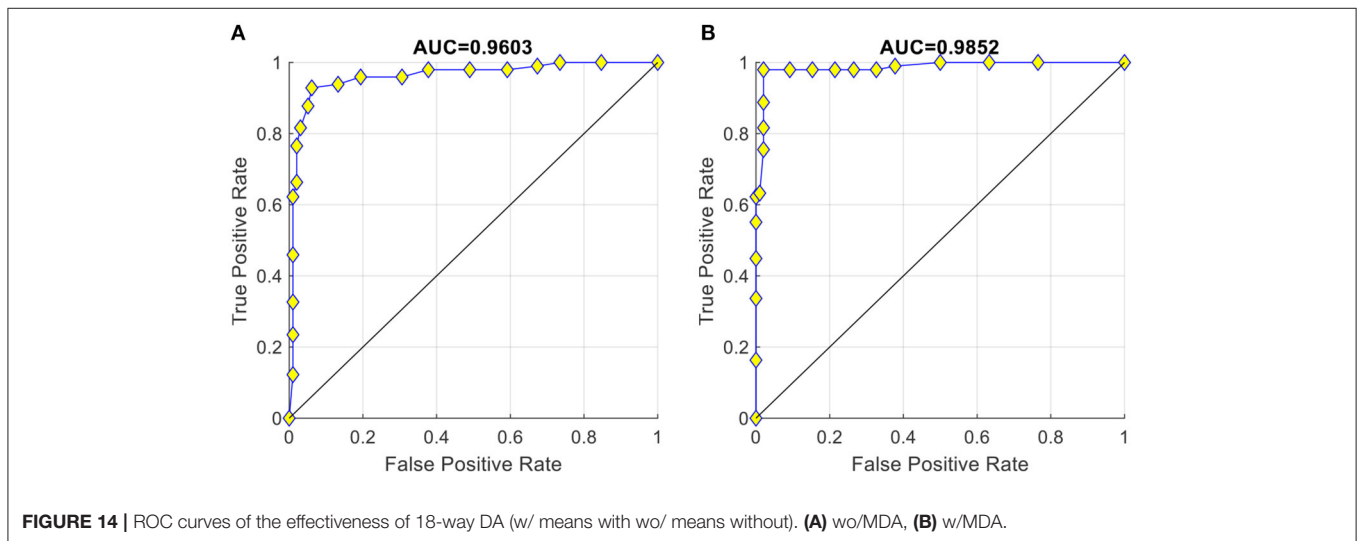
### Method Comparison

To further show the proposed ADVIAN model's effectiveness, we compare it with 11 existing algorithms on the same dataset by 10 runs of 10-fold CV. The comparison methods include BRC-BC (Plant et al., 2010), TJM (Savio and Grana, 2013), RF (Gray et al., 2013), FMSA (Lahmiri and Boukadoum, 2014), DF (Zhang, 2015), PZM-SCG (Gorji and Haddadnia, 2015), BBO (Li, 2018), PZM-LRC (Du, 2017), CNN-LReLU (Sui, 2018), CNN-BND (Jiang and Chang, 2020), and CNN-RNN-LSTM (Dua et al., 2020). The comparison is displayed in **Table 7**, with the bar plot shown in **Figure 15**.

In **Figure 15**, we move the MCC to the leftmost since its value range is smaller than the other six measures. We sort all algorithms in terms of MCC, and the sorted list can be observed at the bottom left corner of **Figure 15**. The 3D bar plot clearly shows that our method achieves better results than all 11 state-of-the-art methods.



**FIGURE 13** | Error bar of the effectiveness of 18-way DA (w/ means with wo/ means without).



**FIGURE 14** | ROC curves of the effectiveness of 18-way DA (w/ means with wo/ means without). **(A)** wo/MDA, **(B)** w/MDA.

**TABLE 7** | Comparison with other methods.

Algorithm	Sen	Spc	Prc	Acc	F1	MCC	FMI
BRC-BC (Plant et al., 2010)	92.96 ± 1.63	88.78 ± 1.86	89.25 ± 1.59	90.87 ± 1.11	91.05 ± 1.09	81.83 ± 2.22	91.08 ± 1.09
TJM (Savio and Grana, 2013)	88.27 ± 3.27	92.45 ± 2.37	92.20 ± 2.03	90.36 ± 1.31	90.13 ± 1.44	80.88 ± 2.53	90.18 ± 1.41
RF (Gray et al., 2013)	87.86 ± 2.18	88.67 ± 1.70	88.60 ± 1.55	88.27 ± 1.36	88.21 ± 1.41	76.56 ± 2.72	88.22 ± 1.41
FMSA (Lahmiri and Boukadoum, 2014)	90.31 ± 2.32	87.86 ± 2.47	88.21 ± 1.93	89.08 ± 1.08	89.21 ± 1.08	78.25 ± 2.13	89.23 ± 1.07
DF (Zhang, 2015)	90.61 ± 1.65	93.16 ± 1.18	93.00 ± 1.08	91.89 ± 0.70	91.78 ± 0.75	83.83 ± 1.39	91.79 ± 0.74
PZM-SCG (Gorji and Haddadnia, 2015)	92.96 ± 1.63	92.65 ± 1.79	92.72 ± 1.57	92.81 ± 0.70	92.82 ± 0.69	85.65 ± 1.40	92.83 ± 0.69
BBO (Li, 2018)	91.73 ± 1.83	91.43 ± 2.21	91.52 ± 1.89	91.58 ± 0.60	91.60 ± 0.56	83.22 ± 1.21	91.61 ± 0.57
PZM-LRC (Du, 2017)	93.37 ± 1.82	92.76 ± 2.01	92.83 ± 1.81	93.06 ± 1.30	93.08 ± 1.29	86.15 ± 2.59	93.09 ± 1.29
CNN-LReLU (Sui, 2018)	97.35 ± 1.88	96.94 ± 1.08	96.97 ± 1.01	97.14 ± 0.87	97.14 ± 0.90	94.31 ± 1.72	97.15 ± 0.89
CNN-BND (Jiang and Chang, 2020)	97.04 ± 1.55	97.35 ± 1.29	97.36 ± 1.24	97.19 ± 0.88	97.19 ± 0.89	94.41 ± 1.73	97.19 ± 0.88
CNN-RNN-LSTM (Dua et al., 2020)	92.65 ± 1.65	92.35 ± 1.30	92.38 ± 1.19	92.50 ± 1.02	92.51 ± 1.04	85.02 ± 2.04	92.51 ± 1.04
Ours	97.65 ± 1.36	97.86 ± 1.55	97.87 ± 1.53	97.76 ± 1.13	97.75 ± 1.13	95.53 ± 2.27	97.76 ± 1.13

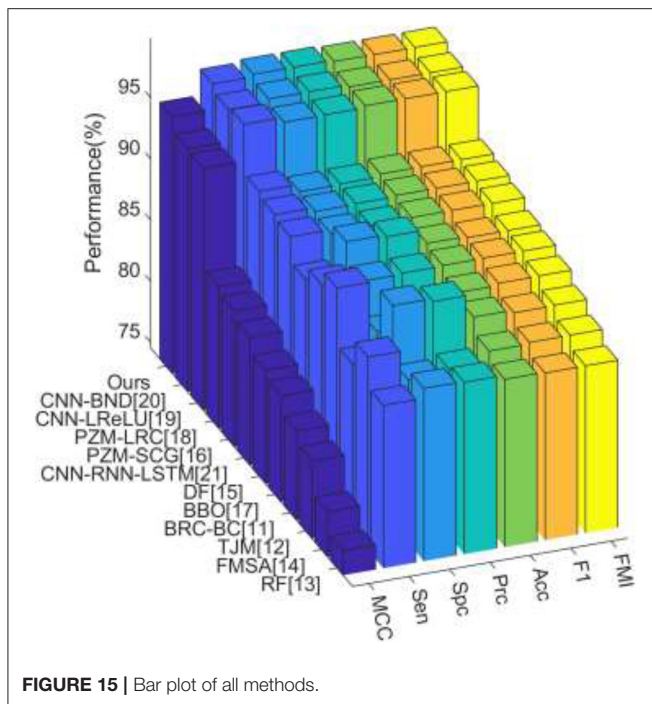


FIGURE 15 | Bar plot of all methods.

This paper is mainly focusing on methodological improvements. We shall try to combine DL with individual anatomical brain regions [such as medial temporal lobe (Chen et al., 2016a), etc.] and brain network connectively patterns (Chen et al., 2016b) in AD patients.

## CONCLUSIONS

This paper proposes a novel VGG-inspired network as the mainstay and combines the attention mechanism with VIN to produce a new ADVIAN deep-learning model to detect AD. The 18-way DA is harnessed to prevent overfitting in the training set. The experiments revealed the usefulness and superiority of this proposed ADVIAN method.

Nevertheless, there are several shortcomings. First, this model did not go through strict clinical environment tests. Second, the dataset is relatively small. Third, the AI output is hard to understand for human experts.

## REFERENCES

- Abuhmed, T., El-Sappagh, S., and Alonso, J. M. (2021). Robust hybrid deep learning models for Alzheimer's progression detection. *Knowl. Based Syst.* 213:106688. doi: 10.1016/j.knsys.2020.106688
- Albashish, D., Hammouri, A. I., Braik, M., Atwan, J., and Sahran, S. (2021). Binary biogeography-based optimization based SVM-RFE for feature selection. *Appl. Soft Comput.* 101:107026. doi: 10.1016/j.asoc.2020.107026
- Alhazzani, A. A., Alqahtani, A. M., Alqahtani, M. S., Alahmari, T. M., and Zarbah, A. A. (2020). Public awareness, knowledge, and attitude toward Alzheimer's

Correspondingly, we may carry out the following researches in the future. We shall deploy our ADVIAN to hospitals to receive feedback directly from clinical doctors. Meanwhile, we will try to collect more AD data. Finally, explainable AI will be included in our future studies.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

S-HW: conceptualization, methodology, software, data curation, writing (original draft), and funding acquisition. QZ: writing (original draft), writing (review and editing), and visualization. MY: resources, writing (review and editing), supervision, project administration, and funding acquisition. Y-DZ: methodology, software, formal analysis, validation, resources, writing (original draft), writing (review and editing), supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by a Royal Society International Exchanges Cost Share Award, UK (RP202G0230); Medical Research Council Confidence in Concept Award, UK (MC\_PC\_17171); Hope Foundation for Cancer Research, UK (RM60G0680); British Heart Foundation Accelerator Award, UK; Sino-UK Industrial Fund, UK; Global Challenges Research Fund, UK (P202PF11); Fundamental Research Funds for the Central Universities, CN (2242021k30014, 2242021k30059); and Key Laboratory of Child Development and Learning Science (Southeast University), Ministry of Education, CN (CDLS-2020-03).

## ACKNOWLEDGMENTS

Data were provided in part by OASIS: cross-sectional: principal investigators: D. Marcus, R. Buckner, J. Csernansky, J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, and U24 RR021382.

disease in Aseer region, Saudi Arabia. *Egypt. J. Neurol. Psychiatry Neurosurg.* 56:81. doi: 10.1186/s41983-020-00213-z

- Ardekani, B. A., Figarsky, K., and Sidtis, J. J. (2013). Sexual dimorphism in the human corpus callosum: an MRI study using the OASIS brain database. *Cerebral Cortex* 23, 2514–2520. doi: 10.1093/cercor/bhs253
- Arora, R., Raman, B., Nayyar, K., and Awasthi, R. (2021). Automated skin lesion segmentation using attention-based deep convolutional neural network. *Biomed. Signal Process. Control* 65:102358. doi: 10.1016/j.bspc.2020.102358
- Chen, J., Duan, X. J., Shu, H., Wang, Z., Long, Z. L., Liu, D., et al. (2016a). Differential contributions of subregions of medial temporal lobe to memory

- system in amnesic mild cognitive impairment: insights from fMRI study. *Sci. Rep.* 6:26148. doi: 10.1038/srep26148
- Chen, J., Shu, H., Wang, Z., Zhan, Y. F., Liu, D., Liao, W. X., et al. (2016b). Convergent and divergent intranetwork and internetwork connectivity patterns in patients with remitted late-life depression and amnesic mild cognitive impairment. *Cortex* 83, 194–211. doi: 10.1016/j.cortex.2016.08.001
- Cheng, X. (2021). PSSPNN: PatchShuffle stochastic pooling neural network for an explainable diagnosis of COVID-19 with multiple-way data augmentation. *Comput. Math. Methods Med.* 2021:6633755. doi: 10.1155/2021/6633755
- Choi, C., Leem, J., Kim, M. S., Taqieddin, A., Cho, C., Cho, K. W., et al. (2020). Curved neuromorphic image sensor array using a MoS<sub>2</sub>-organic heterostructure inspired by the human visual recognition system. *Nat. Commun.* 11:5934. doi: 10.1038/s41467-020-19806-6
- Daines, K. J. F., Baddour, N., Burger, H., Bavec, A., and Lemaire, E. D. (2020). “Fall-risk classification in amputees using smartphone sensor based features in turns,” in *42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society: Enabling Innovative Technologies for Global Healthcare Embc’20* (Montreal, QC: IEEE), 4175–4178. doi: 10.1109/EMBC44109.2020.9176624
- Du, S. (2017). Alzheimer’s disease detection by pseudo zernike moment and linear regression classification. *CNS Neurol. Disord.* 16, 11–15. doi: 10.2174/187152731566616111123024
- Dua, M., Makhija, D., Manasa, P. Y. L., and Mishra, P. (2020). A CNN-RNN-LSTM based amalgamation for Alzheimer’s disease detection. *J. Med. Biol. Eng.* 40, 688–706. doi: 10.1007/s40846-020-00556-1
- Fernandes, S. (2021). AVNC: attention-based VGG-style network for COVID-19 diagnosis by CBAM. *IEEE Sens. J.* doi: 10.1109/JSEN.2021.3062442
- Ferreira, S., Raimundo, A. F., Menezes, R., and Martins, I. C. (2021). Islet amyloid polypeptide and amyloid beta peptide roles in Alzheimer’s disease: two triggers, one disease. *Neural Regen. Res.* 16, 1127–1130. doi: 10.4103/1673-5374.300323
- Fulton, L. V., Dolezel, D., Harrop, J., Yan, Y., and Fulton, C. P. (2019). Classification of Alzheimer’s disease with and without imagery using gradient boosted machines and ResNet-50. *Brain Sci.* 9:212. doi: 10.3390/brainsci9090212
- Gorji, H. T., and Haddadnia, J. (2015). A novel method for early diagnosis of Alzheimer’s disease based on pseudo Zernike moment from structural MRI. *Neuroscience* 305, 361–371. doi: 10.1016/j.neuroscience.2015.08.013
- Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., Rueckert, D., and Alzheimer’s Disease Neuroimaging Initiative. (2013). Random forest-based similarity measures for multi-modal classification of Alzheimer’s disease. *Neuroimage* 65, 167–175. doi: 10.1016/j.neuroimage.2012.09.065
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV), 770–778. doi: 10.1109/CVPR.2016.90
- Hou, Z. (2006). A review on MR image intensity inhomogeneity correction. *Int. J. Biomed. Imaging.* 2006:49515. doi: 10.1155/IJBI/2006/49515
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). “Densely connected convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Honolulu, HI)*, 4700–4708. doi: 10.1109/CVPR.2017.243
- Jahangeer, G. S. B., and Rajkumar, T. D. (2021). Early detection of breast cancer using hybrid of series network and VGG-16. *Multimedia Tools Appl.* 80, 7853–7886. doi: 10.1007/s11042-020-09914-2
- Jiang, X., and Chang, L. (2020). Classification of Alzheimer’s disease via eight-layer convolutional neural network with batch normalization and dropout techniques. *J. Med. Imaging Health Inf.* 10, 1040–1048. doi: 10.1166/jmih.2020.3001
- Kumari, S., and Deshmukh, R. (2021). Beta-lactam antibiotics to tame down molecular pathways of Alzheimer’s disease. *Eur. J. Pharmacol.* 895:173877. doi: 10.1016/j.ejphar.2021.173877
- Lahmiri, S., and Boukadoum, M. (2014). New approach for automatic classification of Alzheimer’s disease, mild cognitive impairment and healthy brain magnetic resonance images. *Healthcare Technol. Lett.* 1, 32–36. doi: 10.1049/hlt.2013.0022
- Lee, H., Jeong, H., Koo, G., Ban, J., and Kim, S. W. (2021). Attention recurrent neural network-based severity estimation method for interturn short-circuit fault in permanent magnet synchronous machines. *IEEE Trans. Ind. Electron.* 68, 3445–3453. doi: 10.1109/TIE.2020.2978690
- Lee, J. H., Kim, S. J., Lee, S. H., Suh, I. B., Jang, J. W., and Jhoo, J. H. (2019). Effects of timed light on mood and cognition in Alzheimer’s disease. *Sleep* 42:1. doi: 10.1093/sleep/zsz067.940
- Li, Y.-J. (2018). Single slice based detection for Alzheimer’s disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. *Multimedia Tools Appl.* 77, 10393–10417. doi: 10.1007/s11042-016-4222-4
- Luo, W. L., Duan, S. Q., and Zheng, J. W. (2021). Underwater image restoration and enhancement based on a fusion algorithm with color balance, contrast optimization, and histogram stretching. *IEEE Access.* 9, 31792–31804. doi: 10.1109/ACCESS.2021.3060947
- Mather, M. A., Laws, H. B., Dixon, J. S., Ready, R. E., and Akerstedt, A. M. (2021). Sleep behaviors in persons with Alzheimer’s disease: associations with caregiver sleep and affect. *J. Appl. Gerontol.* 11. doi: 10.1177/0733464820979244
- Monteiro, C., Mendes, V., Comarela, G., and Silveira, S. A. (2018). “Using supervised learning successful descriptors to perform protein structural classification through unsupervised learning,” in *Proceedings 2018 IEEE International Conference on Bioinformatics and Biomedicine* (Madrid: IEEE), 75–78. doi: 10.1109/BIBM.2018.8621332
- Oh, D., Kim, B., Lee, J., and Shin, Y. G. (2021). Unsupervised deep learning network with self-attention mechanism for non-rigid registration of 3D brain MR images. *J. Med. Imaging Health Inf.* 11, 736–751. doi: 10.1166/jmih.2021.3345
- Petti, U., Baker, S., and Korhonen, A. (2020). A systematic literature review of automatic Alzheimer’s disease detection from speech and language. *J. Am. Med. Inform. Assoc.* 27, 1784–1797. doi: 10.1093/jamia/ocaa174
- Plant, C., Teipel, S. J., Oswald, A., C., Böhm, Meindl, T., Mourao-Miranda, J., et al. (2010). Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer’s disease. *NeuroImage* 50, 162–174. doi: 10.1016/j.neuroimage.2009.11.046
- Puthusserypady, V., Emrich-Mills, L., Lowry, E., Patel, M., and Hornberger, M. (2020). Spatial disorientation in Alzheimer’s disease: the missing path from virtual reality to real world. *Front. Aging Neurosci.* 12:550514. doi: 10.3389/fnagi.2020.550514
- Saletin, J. M., Jackvony, S., Rodriguez, K. A., and Dickstein, D. P. (2019). A coordinate-based meta-analysis comparing brain activation between attention deficit hyperactivity disorder and total sleep deprivation. *Sleep.* 42:zsy251. doi: 10.1093/sleep/zsy251
- Santana, M. V. S., and Silva, F. P. (2021). *De novo* design and bioactivity prediction of SARS-CoV-2 main protease inhibitors using recurrent neural network-based transfer learning. *BMC Chem.* 15:8. doi: 10.1186/s13065-021-00737-2
- Saood, A., and Hatem, I. (2021). COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. *BMC Med. Imaging.* 21:19. doi: 10.1186/s12880-020-00529-5
- Savio, A., and Grana, M. (2013). Deformation based feature selection for computer aided diagnosis of Alzheimer’s disease. *Expert Syst. Appl.* 40, 1619–1628. doi: 10.1016/j.eswa.2012.09.009
- Scheltens, P., Blennow, K., M., Breteler, M. B., de Strooper, B., Frisoni, G. B., Salloway, S., and Van der Flier, W. M. (2016). Alzheimer’s disease. *The Lancet.* 388, 505–517. doi: 10.1016/S0140-6736(15)01124-1
- Senova, S., Lefaucheur, J. P., Brugieres, P., Ayache, S. S., Tazi, S., Bapst, B., et al. (2021). Case report: multimodal functional and structural evaluation combining pre-operative nTMS mapping and neuroimaging with intraoperative CT-Scan and brain shift correction for brain tumor surgical resection. *Front. Hum. Neurosci.* 15:646268. doi: 10.3389/fnhum.2021.646268
- Song, K. T., Huang, Q. K., Zhang, F. E., and Lu, J. F. (2021). Coarse-to-fine: A dual-view attention network for click-through rate prediction. *Knowl. Based Syst.* 216:106767. doi: 10.1016/j.knsys.2021.106767
- Sudha, V., and Ganeshbabu, T. R. (2021). A convolutional neural network classifier VGG-19 architecture for lesion detection and grading in diabetic retinopathy based on deep learning. *Comp. Mater. Continua.* 66, 827–842. doi: 10.32604/cmcc.2020.012008
- Sui, Y. X. (2018). Classification of Alzheimer’s disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *J. Med. Syst.* 42:85. doi: 10.1007/s10916-018-0932-7
- Sutoko, S., Masuda, A., Kandori, A., Sasaguri, H., Saito, T., Saido, T. C., et al. (2021). Early identification of Alzheimer’s disease in mouse models: application

- of deep neural network algorithm to cognitive behavioral parameters. *Iscience* 24:102198. doi: 10.1016/j.isci.2021.102198
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, MA: IEEE), 1–9. doi: 10.1109/CVPR.2015.7298594
- Tiwari, S. (2021). Dermatoscopy using multi-layer perceptron, convolution neural network, and capsule network to differentiate malignant melanoma from benign nevus. *Int. J. Healthcare Inf. Syst. Inf.* 16, 58–73. doi: 10.4018/IJHISI.20210701.oa4
- Vrzal, T., Maleckova, M., and Olsovska, J. (2021). DeepRel: deep learning-based gas chromatographic retention index predictor. *Anal. Chim. Acta.* 1147, 64–71. doi: 10.1016/j.aca.2020.12.043
- Woo, S., Park, J., Lee, J.-Y., and So Kweon, I. (2018). "CBAM: convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Munich, Germany: Springer), 3–19. doi: 10.1007/978-3-030-01234-2\_1
- Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 1492–1500. doi: 10.1109/CVPR.2017.634
- Zagoruyko, S., and Komodakis, N. (2016). Wide residual networks. *arXiv preprint. arXiv:1605.07146*. doi: 10.5244/C.30.87
- Zhang, Y. (2015). Detection of Alzheimer's disease by displacement field and machine learning. *PeerJ.* 3:e1251. doi: 10.7717/peerj.1251

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang, Zhou, Yang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.