IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 October 2023, accepted 14 January 2024, date of publication 22 January 2024, date of current version 29 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3356824

RESEARCH ARTICLE

A Novel Artificial-Intelligence-Based Approach for Automatic Assessment of Retinal Disease Images Using Multi-View Deep-Broad Learning Network

AO ZHANG[1], XIN QIAN[2], CHENGCHENG XU[3], AND JIE ZHANG[1]
[1]School of Computing and Internet of Things, Chongqing Institute of Engineering, Chongqing 400056, China
[2]School of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 401331, China
[3]School of Science, Guangxi University of Science and Technology, Liuzhou 545006, China

Corresponding author: Xin Qian (17756276202@163.com)

This work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJ202301992966111-KJQN202301903.

ABSTRACT Retinal disease detection and diagnosis relying solely on artificial retinal diseases will put great pressure on doctors and increase the rate of misdiagnosis. Therefore, the development of computer vision technology has brought the possibility for ophthalmologists to use computer-aided diagnosis. In recent years, most models for retinal disease recognition have been based on deep learning, which has the disadvantage of requiring large amounts of data and training time. It is also partly based on broad learning and its disadvantages are that feature extraction ability is limited and poor scalability. To overcome these limitations, we propose a novel artificial intelligence-based approach for the automatic assessment of retinal disease images called a multi-view deep-broad learning network (MDBL-Net), which absorbs the advantages of deep learning and broad learning. MDBL-Net comprises a Multi-view and Multi-scale Feature Extraction (MMFE) module and a Multi-scale Aggregation (MA) block. The MMFE module extracts features of different scales by learning feature representations from multiple views, while the MA block fully aggregates multi-scale deep-broad features from low-level to high-level representations. Experiments were conducted on two public datasets, UCSD and OCT2017. Experiments were conducted on two public datasets, UCSD and OCT2017, and results demonstrate that MDBL-Net achieves high accuracy even with limited training data (only 1%) and significantly reduces training time compared to traditional deep learning models. Specifically, MDBL-Net achieved an accuracy of 96.93% on the UCSD dataset and 99.90% on the OCT2017 dataset, outperforming state-of-the-art models in both cases. These findings suggest that the proposed MDBL-Net approach holds great promise for the task of retinal disease screening and recognition.

INDEX TERMS Broad learning, image classification, medical image, retinal disease, computer aided diagnosis.

I. INTRODUCTION

According to the World Health Organization, at least 1 billion people worldwide have near or distance vision impairment [1]. There are many causes of vision loss, and retinal disease is one of the main causes of blindness, the main diseases include age-related macular degeneration (AMD) and sugar urine disease macular edema (DME). With the help of optical coherence tomography (OCT), the early detection and timely treatment of fundus diseases can be realized, which is an effective means to prevent blindness. However, due to the large number of patients, only relying on manual screening

The associate editor coordinating the review of this manuscript and approving it for publication was Mingbo Zhao.

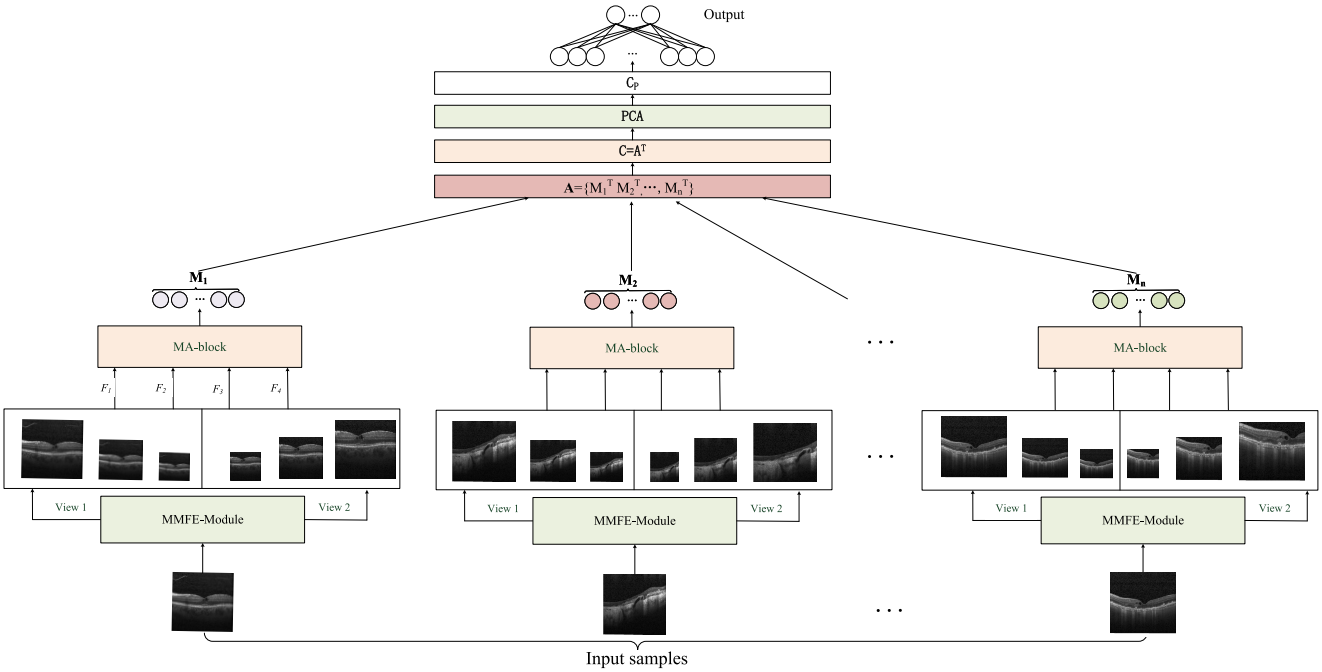


FIGURE 1. Overall architecture of MDBL-Net, it consists of three sub-modules: MMFE, MA and multi-view broad learning.

and diagnosis of eye diseases will put great work pressure on doctors and increase the misdiagnosis rate. To relieve the pressure of reading films, computer-aided diagnosis technology has been paid more and more attention. The development of computer vision technology makes it possible for ophthalmologists to use computer-aided diagnosis.

In recent years, deep learning and broad learning have been widely used in medical image analysis. Deep learning has achieved a series of achievements on OCT medical images [2], [3], [4], [5]. Wang et al. [6]. proposed a weakly supervised deep learning framework with uncertainty estimation to address the macula-related disease classification problem from OCT images with the only volume-level label being available. Wang et al. [7]. developed a computer-aided diagnosis (CADx) approach based on a self-supervised ViT-based model to classify cervical OCT images effectively. Based on the above research, the advantages of the deep learning model are as follows: (1) Processing large-scale data. Learn and model from a large amount of data, and use the distributed features of the data for efficient pattern recognition and feature extraction. (2) Automatic learning features. Compared with traditional machine learning methods, which require manual selection and construction of features, deep learning models can automatically learn feature representations. (3) Dealing with nonlinear relationships. Through the connection of activation function and multi-layer neural network, the model can capture nonlinear changes in data, which improves the flexibility and expressiveness of the model. However, there are also some limitations: (1) A large amount of labeled data is required. (2) Long iterative training time is required. (3) The model parameter tuning is complex.

In the field of medical image analysis, broad learning systems have emerged as a promising approach for improving the accuracy and efficiency of image recognition and classification tasks [8], [9], [10]. Broad Learning System (BLS) is a type of single-layer incremental neural network proposed by Professor Junlong Chen of the University of Macau in 2017 based on Random Vector Functional Link Neural Network (RVFLNN) and Single-Layer Feedforward Neural Network (SLFN) [11]. There have been a lot of studies on this basis. Wu and Duan [12]. proposed a novel collaborative-competitive broad learning system for COVID-19 detection from radiology images. Han et al. [13]. proposed a new variant model of the BLS for accurate diagnosis of AD and MCI is presented for MRI images. The advantages of broad learning model are: (1) Easy to understand and explain. Compared with the deep learning model, the structure of the width learning model is relatively simple and the number of layers is less. (2) The training speed is fast. broad learning models typically take less time to train than deep learning models. (3) Parameter tuning is relatively simple. The parameters of the broad learning model are few, so the complexity of adjusting and optimizing the parameters during training is relatively low. (4) Strong robustness.The broad learning model is relatively shallow and thus has high robustness to noisy and abnormal data. But there are also some limitations: (1) Feature extraction ability is limited. (2) Poor scalability. (3) Need to design features manually.

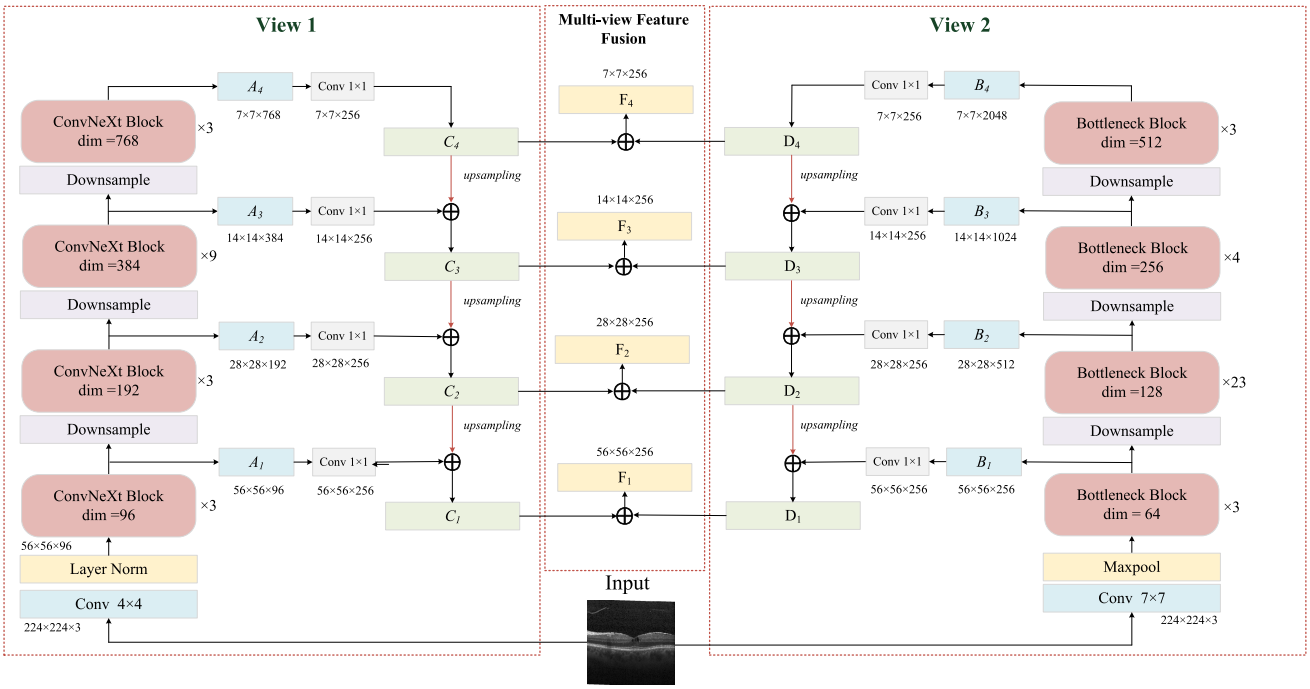


FIGURE 2. Overall architecture of multi-view and multi-scale feature extraction(MMFE) module.

Therefore, in the field of retinal diseases OCT image recognition, there are some problems in both the deep learning model and the broad learning model. From the above, it can be concluded that the advantages and disadvantages of deep learning and broad learning are complementary. If we can combine the two machine learning approaches, we can achieve a good performance in medical image analysis. Inspired by the above, a novel artificial intelligence-based approach for the automatic assessment of retinal disease images using broad learning and deep learning is proposed. The contributions of this paper are summarized as follows:

(1) We propose a novel artificial intelligence-based approach for the automatic assessment of retinal disease images called a Multi-view Deep-Broad Learning Network (MDBL-Net).

(2) A multi-view and multi-scale feature extraction(MMFE) module is proposed, which can extract features of different scales by learning feature representations of multiple views. Mining potential information from different perspectives can improve the model's perception of input data, thereby improving performance and accuracy.

(3) A multi-scale aggregation(MA) block is proposed, which can fully aggregate multi-scale deep-broad features from low level to high level.

(4) MDBL-Net can achieve accuracy comparable to or better than existing methods in classification tasks while showing significant performance improvements for models with very limited training samples (only 1%).

II. RELATED WORK

A. RETINAL DISEASE OCT IMAGES CLASSIFICATION

Optical Coherence Tomography (OCT) is a novel and effective screening tool for ophthalmic examination. OCT images have the characteristics of high resolution, high contrast, and non-intrusion [14]. The challenges of OCT image classification mainly include category similarity, data noise, a large amount of data and high dimension, a small amount of annotated data and benchmark standard controversy, etc. [14]. Advanced algorithms and technologies are needed to overcome these challenges. Wang et al. [15]. proposed a novel fundus-enhanced disease perception distillation model (FDDM) for the classification of retinal diseases from OCT images. Karthik et al. [2]. proposed a residual module for OCT image classification of retinal diseases, which enhances the contrast of feature maps, resulting in clearer retinal layer boundaries. Muhammad Junaid Umer [16] proposed an automatic method to detect and classify retinal ophthalmopathy from OCT images using fusion and selection techniques, and the proposed retinal ophthalmopathy detection method can be reliably used for automatic ophthalmopathy detection in OCT images. Sunija et al. [17] proposed a deep convolutional neural network with six convolutional blocks for the classification of retinal OCT images. The proposed method has an

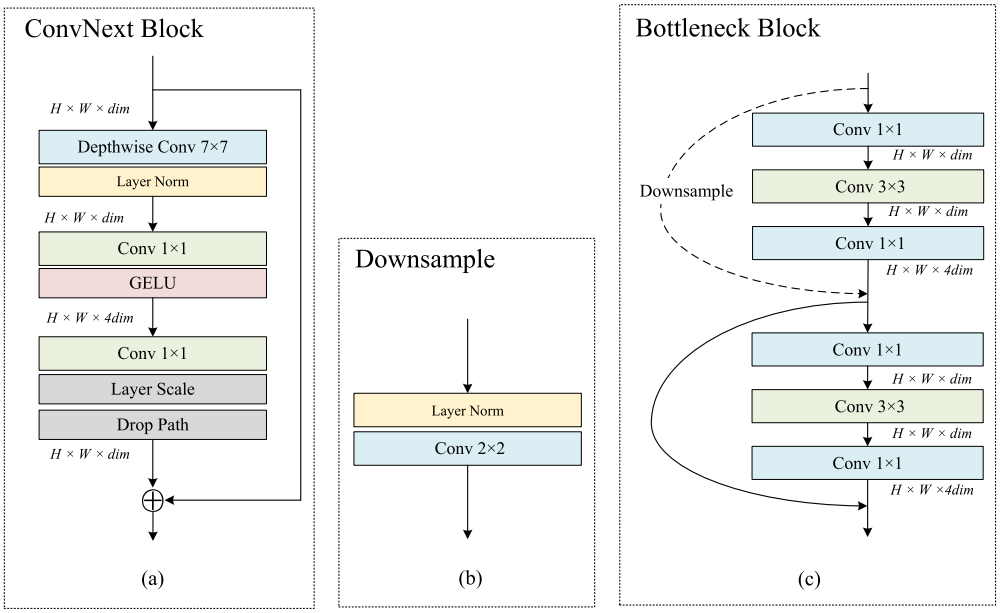


FIGURE 3. The three sub-module structure of MMFF; (a) ConvNeXt block; (b) Downsample;(c) Bottleneck block.

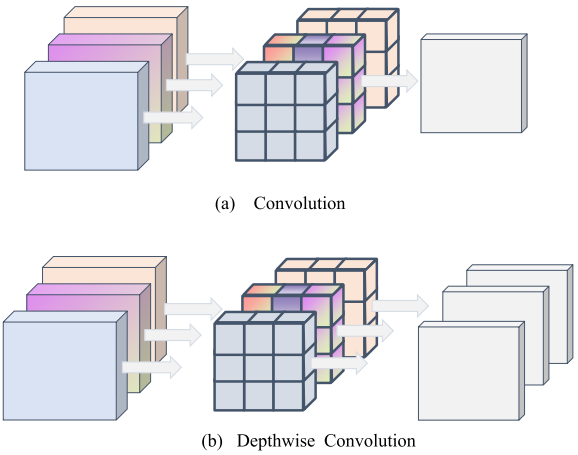


FIGURE 4. Comparison of different convolution blocks; (a)Traditional convolution; (b)Depthwise convolution.

accuracy of 99.69%, a specificity of 99.69%, and a sensitivity of 99.69%, and only 968 misclassifications in three test cases.

In this paper, we use multi-scale structures to fully extract features and use broad learning to improve the performance of the model without incurs excessive computational costs.

B. THE DEVELOPMENT OF DEEP LEARNING MODELS

Deep learning uses artificial neural networks to perform complex computations on large amounts of data, and it is one of the latest trends in machine learning and artificial intelligence research. General deep learning algorithms include CNN [18], LSTM [19], RNN [20], GAN [21], MLP [22] and AutoEncoding [23], etc, which are used in different scenarios. In recent years, deep learning models have made great progress. In 2014, VGG-Net [24] was proposed, which outstanding contribution lies in proving that using a small convolution (3 × 3) and increasing the depth of the network can effectively improve the effect of the model. In 2015, ResNet [25] was proposed to solve the problem of gradient disappearance and gradient explosion during deep neural network training, making deep neural network training no longer difficult. In 2017, transformer was proposed [26], which is a model that uses the attention mechanism to improve the speed of model training. In 2020, Vision Transformer (ViT) [27] was proposed. Different from traditional convolutional neural networks (CNNS), ViT utilizes techniques such as self-attention mechanism and Transformers to convert images into a series of vectors for classification.

The above model has some obvious advantages, such as processing large-scale data, automatic learning features, automatic learning features, etc. At the same time, it also has some disadvantages, such as large amount of labeled data, long training time, and complex model parameter tuning. Through the above inspiration, the proposed MDBL-Net absorbs the advantages of deep learning models and compensates the corresponding disadvantages using broad learning.

C. BROAD LEARNING SYSTEMS IN MEDICAL IMAGES

Deep learning networks have complex structures and involve a large number of hyperparameters, so most of them are plagued by extremely time-consuming training processes [28]. To solve this problem, the broad learning system [11] is

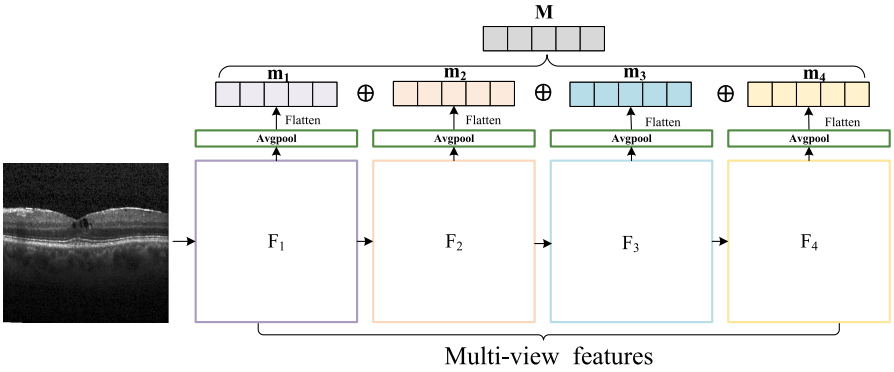


FIGURE 5. The architecture of MA-block.

proposed. The broad learning system is another improvement based on a random vector functional-link neural network (RVFLNN), especially the incremental learning process. Since broad learning systems were proposed, there has been a lot of research on their application to medical image analysis [8], [9], [10], [12], [13]. BLSNet [29] uses an incremental learning algorithm for the classification of non-melanoma and melanoma skin lesions from dermoscopic images, which achieves a performance trade-off between classification accuracy and execution time. DeepBLS [30] combines a comprehensive learning system with deep feature extraction to identify different tissue components in colon cancer histological images. SPRBF-ABLS [31] is a cascaded neural network framework based on a sparse polynomial-based RBF neural network and an attention-based broad learning system.

From above, we found that the broad learning model compared with the model structure of deep learning has the following advantages, such as easy to understand and explain, easy to understand and explain, better performance for a small amount of data, and strong robustness. However, there are also some disadvantages, including limited feature extraction ability, poor scalability, and features that need to be designed manually. It is worth noting that these shortcomings can be well compensated by deep learning models. Drawing upon the aforementioned background, this paper harnesses the strengths of deep learning and broad learning models to enhance the performance of OCT image recognition for retinal diseases.

III. METHOD

This section introduces the specific structure of the proposed MDBL-Net, as shown in Figure 1. Under the background that deep learning (DL) has the powerful feature extraction ability, and broad learning (BL) is good at combining multiple groups of features for fast inference, the MDBL-Net is proposed. MDBL-Net consists of three sub-modules: MMFE, MA and multi-view broad learning. The following specific introduced the network structure.

A. MULTI-VIEW AND MULTI-SCALE FEATURE EXTRACTION (MMFE) MODULE

The MMFE module in MDBL-Net plays an important role. By learning multi-view and multi-scale feature representations, the potential information of the data can be fully mined, and the perception ability and generalization performance of the model can be improved. As shown in Figure 2, the multi-view and multi-scale feature extraction (MMFE) module is used to extract the feature information of retinal disease Images. The MMFE module contains the two substructures of different views, which are based on the ConvNeXt block and bottleNeck block, respectively. The following is an introduction to the two view structures.

1) VIEW 1

As shown in the Figure 2, the structure of view1 is based on ConvNeXt [32]. The ConvNeXt block is shown in Figure 3.(a), and mainly has the following structures: two traditional convolution(Figure 4.(a)), a depthwise convolution [33](Figure 4.(b)), GELU [34], Layer Normalization [35], layer scale, drop path and residual connection. The downsampling(Figure 3.(c)) operation includes Layer Normalization and traditional convolution. The downsample operation includes Layer Normalization and traditional convolution.

First of all, the four feature maps $\{A_1, A_2, A_3, A_4\}$ are obtained by ConvNeXt, then adjust their number of channels by 1×1 convolution. Finally, $\{C_1, C_2, C_3, C_4\}$ is obtained by the following formula.

$$C_i = \begin{cases} Conv_{1\times1}(A_i) + upsampling(Conv_{1\times1}(A_{i+1})) & i = 1, 2, 3 \\ Conv_{1\times1}(A_i) & i = 4 \end{cases} \tag{1}$$

where *upsampling* represents the double upsampling operation to adjust size.

At this point, $\{C_1, C_2, C_3, C_4\}$ is obtained.

2) VIEW 2

As shown in the Figure 2, the structure of view2 is based on ResNet [25]. The Bottleneck block is shown in Figure 3.(c), and mainly has the following structures: traditional convolution, downsample and residual connection. First of all, the four feature maps $\{F_1, F_2, F_3, F_4\}$ are obtained by ResNet, then adjust their number of channels by 1×1 convolution. Finally, $\{D_1, D_2, D_3, D_4\}$ is obtained by the following formula.

$$D_i = \begin{cases} Conv_{1\times1}(B_i) + upsampling(Conv_{1\times1}(B_{i+1})) & i = 1, 2, 3 \\ Conv_{1\times1}(B_i) & i = 4 \end{cases} \tag{2}$$

where *upsampling* represents the double upsampling operation to adjust size.

At this point, $\{D_1, D_2, D_3, D_4\}$ is obtained.

3) MULTI-VIEW FEATURE FUSION

The two views feature matrix information $\{C_1, C_2, C_3, C_4\}$ and $\{D_1, D_2, D_3, D_4\}$ are fused by element-wise addition, as follows:

$$F_i = C_i \oplus D_i \qquad i = 1, 2, 3, 4 \tag{3}$$

where $\oplus$ represents the element-wise addition.

At this point, $\{F_1, F_2, F_3, F_4\}$ is obtained.

B. MULTI-SCALE AGGREGATION (MA) BLOCKS

The proposed MA-block is to fuse the extracted features $\{F_1, F_2, F_3, F_4\}$, and reduce the number of parameters to prevent overfitting, as shown in Figure 5. The deep features $\{F_1, F_2, F_3, F_4\}$ are squeezed into a channel-wise feature vector by adaptive global average pooling in Eq.4.

$$\mathbf{m}_k = \frac{1}{H_k \times W_k} \sum_{i=1}^{H_k} \sum_{j=1}^{W_k} F_k(i, j) \tag{4}$$

where H_k and W_k represent the height and width of the feature maps of F_k,$\{k = 1, 2, 3, 4\}$.

The multi-view deep learning features M_i of the F_k can be obtained by element-wise addition all feature vector $m_i (i = 1, 2, 3, 4)$, as follows:

$$\mathbf{M}_n = \mathbf{m}_1 \oplus \mathbf{m}_2 \oplus \mathbf{m}_3 \oplus \mathbf{m}_4 \tag{5}$$

where$\oplus$ represents the element-wise addition operation, and n denotes the number of training or testing examples.

C. MULTI-VIEW BROAD LEARNING INFERENCE

Inspired by paper [36], multi-view broad learning is used to solve problems with long training times. As shown in Figure 1. First of all, the feature vector M_n is combined using the following two formulas.

$$\mathbf{A} = \delta\left(\left[\mathbf{M}_1^\top, \mathbf{M}_2^\top, \ldots, \mathbf{M}_n^\top\right]\right) \tag{6}$$

$$\mathbf{C} = \mathbf{A}^\top \tag{7}$$

where δ denotes the matrix normalization transformation. At this point, the broad learning feature matrix C is obtained. In order to transform the original high-dimensional data in C into a low-dimensional representation while retaining the most important information in the original data, a principal component analysis was performed on C.

$$\mathbf{C_p} = \mathbf{PCA(C)} \tag{8}$$

where **PCA** denotes the principal component analysis.

Finally, the fully connected layer was used to obtain the final classification probabilities P, can be calculated by:

$$\mathbf{P} = \mathbf{C_p W}_{MDBL} \tag{9}$$

where $\mathbf{W}_{MDBL}$ is the weights of MDBL-Net, which can be calculated by the pseudo-inverse algorithm. Assuming the ground truth label is $\mathbf{Y}$, $\mathbf{W}_{MDBL}$ can be obtained by the following:

$$\mathbf{W}_{\mathrm{MDBL}} = \mathbf{P}^+ Y \tag{10}$$

$$\mathbf{P}^+ = \lim_{\lambda \to 0} \left(\lambda E + \mathbf{P}\mathbf{P}^T\right)^{-1} \mathbf{P}^T \tag{11}$$

where λ and E represent a constant and unit matrix, respec- tively. Due to the weight $\mathbf{W}_{MDBL}$ is calculated using pseudo-inverse and the weight is updated only once, the algorithm running time is shortened.

In conclusion, multi-view broad learning inference enables all training samples to be input at one time, which greatly reduces the training time.The details of MDBL-Net are shown in Algorithm 1.

Algorithm 1 Multi-View Deep-Broad Learning Network

Input: All image data:$\{X_0, X_1 \cdots X_n\}$(n: the number of samples)

Output: The final classification probabilities $\mathbf{P}$

for all $i = 1$ to n **do**
  $\{C_1, C_2, C_3, C_4\} \leftarrow$ View 1(X_i)
  $\{D_1, D_2, D_3, D_4\} \leftarrow$ View 2(X_i)
  for all $j = 1$ to 4 **do**
    $F_j \leftarrow C_j \oplus D_j$
    $m_j \leftarrow Flatten(\frac{1}{H_k \times W_k} \sum_{i=1}^{H_k} \sum_{j=1}^{W_k} F_k)$
  end for
  $M_i \leftarrow m_1 \oplus m_2 \oplus m_3 \oplus m_4$
end for
$\mathbf{A} \leftarrow \delta\left(\left[M_1^\top, M_2^\top, \ldots, M_n^\top\right]\right)$
$\mathbf{C} \leftarrow A^\top$
$\mathbf{C_p} \leftarrow \mathbf{PCA(C)}$
$\mathbf{P}^+ \leftarrow \lim_{\lambda \to 0} \left(\lambda E + \mathbf{P}\mathbf{P}^T\right)^{-1} \mathbf{P}^T$
$\mathbf{W}_{\mathrm{MDBL}} \leftarrow \mathbf{P}^+ Y$
$\mathbf{P} \leftarrow \mathbf{C_p W}_{MDBL}$
return $\mathbf{P}$

IV. EXPERIMENTS AND RESULTS

A. DATASETS

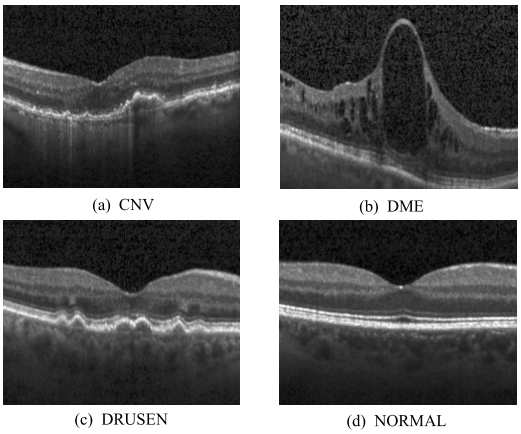
There are three main types of retinal diseases: choroidal neovascularization (CNV), diabetic macular edema (DME)

TABLE 1. The MDBL-Net compares with the other classification models on two datasets.

Dataset	Study	Year	# of images	Method	Performence(%)			Train Time(hours)
					Recall	F1	Accuracy	
UCSD dataset	Liu *et.al* [37]	2021	109309	Swin_transformer	91.02	90.65	92.50	15.20
	Mehta *et.al* [38]	2022	109309	MobileViT	93.03	93.03	94.62	14.10
	Woo *et.al* [39]	2022	109309	ConvNeXt	93.18	92.16	93.68	10.50
	He *et.al* [25]	2023	109309	ResNet101	89.27	91.64	94.39	9.20
	Ours	2023	109309	**MDBL-Net**	**95.45**	**94.84**	**96.93**	0.40
OCT2017	Liu *et.al* [37]	2021	84484	Swin_transformer	99.17	99.18	99.17	15.30
	Mehta *et.al* [38]	2022	84484	MobileViT	99.59	99.59	99.59	9.50
	Woo *et.al* [39]	2022	84484	ConvNeXt	98.55	98.55	98.55	10.50
	He *et.al* [25]	2021	84484	ResNet101	99.49	99.39	99.49	8.50
	Ours	2023	84484	**MDBL-Net**	**99.90**	**99.90**	**99.90**	0.36

TABLE 2. Performance of MDBL-Net under different training sample proportions.

Dataset	Method	1% (Samples)		10% (Samples)		25% (Samples)		100% (Samples)	
		F1(%)	Accuracy(%)	F1(%)	Accuracye(%)	F1(%)	Accuracy(%)	F1(%)	Accuracy(%)
UCSD dataset	Swin_transformer	53.24	74.43	87.06	75.15	87.23	92.05	90.65	92.50
	MobileViT	63.28	78.89	87.46	91.75	84.68	92.50	93.03	97.62
	ConvNeXt	32.45	59.60	55.74	80.31	77.03	87.64	92.16	93.68
	ResNet101	53.70	69.25	94.23	94.21	87.65	92.99	91.64	94.39
	MDBL-Net	**88.99**	**94.78**	**94.69**	**96.82**	**94.68**	**96.84**	**94.84**	**96.93**
OCT2017	Swin_transformer	9.17	49.27	87.84	88.12	92.53	92.51	99.18	99.17
	MobileViT	56.59	57.33	91.77	91.34	**98.77**	**99.27**	99.59	99.59
	ConvNeXt	40.69	45.97	68.21	76.86	90.69	92.04	98.55	98.55
	ResNet101	62.45	67.05	90.88	90.17	98.35	99.07	99.39	99.49
	MDBL-Net	**88.98**	**93.12**	**94.78**	**96.88**	94.68	96.82	**99.90**	**99.90**



(a) CNV (b) DME (c) DRUSEN (d) NORMAL

FIGURE 6. The example of four retinal diseases OCT images.

TABLE 3. Performance of MDBL-Net ablation experiment.

Method	Performence(%)		
	Recall	F1	Accuracy
View1	98.55	98.55	98.55
MMFE(View1+View2)	93.03	93.03	94.62
MDBL-Net(MMFE+BL)	**99.90**	**99.90**	**99.90**

and DRUSEN. OCT images of the three diseases and normal retinas are shown in Figure 6. In this paper, experiments are carried out on two public datasets. and the details of all datasets are introduced below.

1) UCSD DATASET [40]

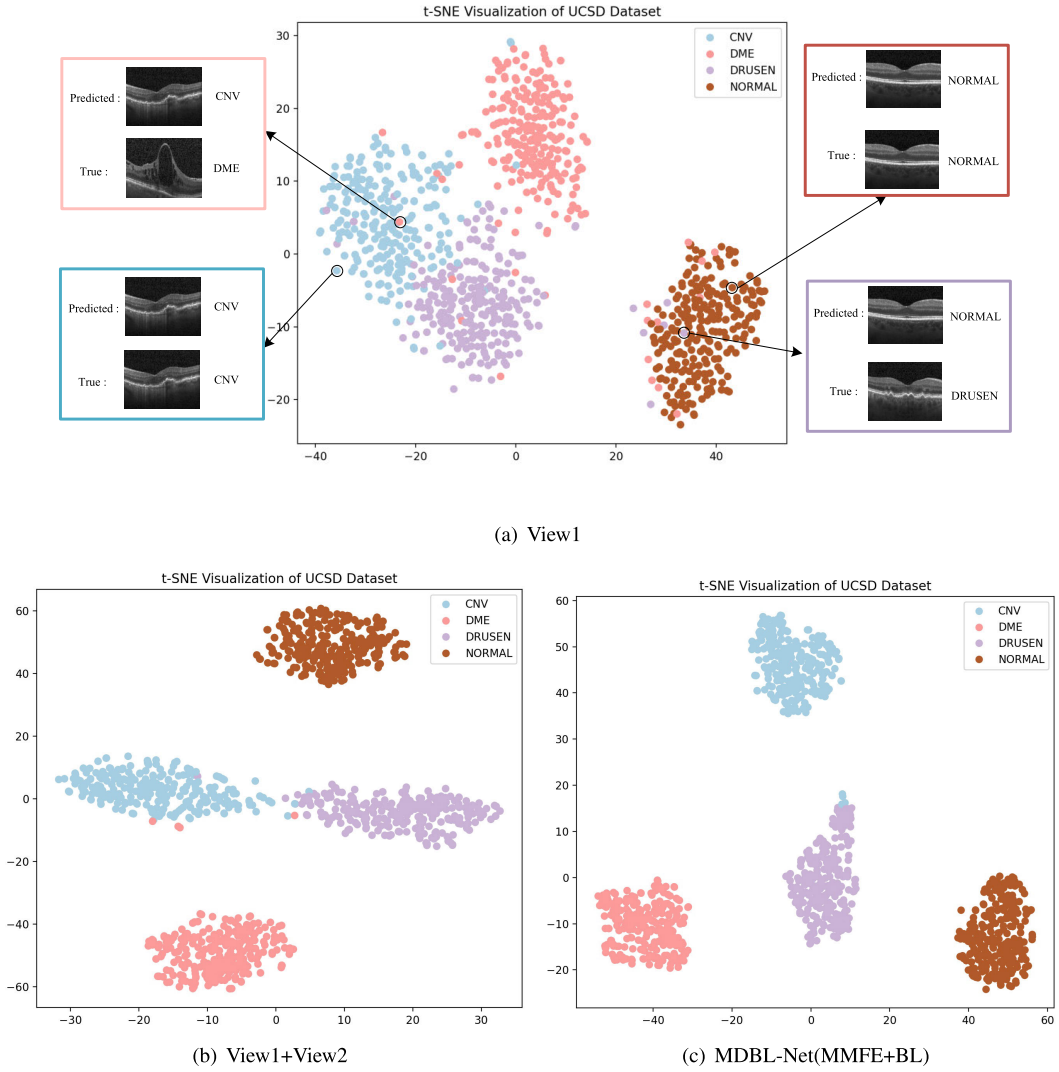
UCSD dataset[1] has been publicly available to the research community on the Mendeley Data website since January 2018. All images were acquired by Heidelberg Spectralis OCT scanner (Heidelberg Engineering, Germany) from multiple centers. The dataset contained (109,309) OCT images from 5319 patients, including CNV, DME, DRUSEN, and NORMAL. All images in the data set have a (jpeg) format that varies in size between (496) and (1536) pixels.We randomly split the data into two groups, where 80% of the data is used for training and the rest (20%) is used for training as our test set.

2) OCT2017 [41]

Retinal optical coherence tomography (OCT2017)[2] is an imaging technique used to capture high-resolution cross sections of the retinas of living patients. OCT2017 contains 84,484 images from four classes, which are choroidal neovascularization (CNV), diabetic macular edema (DME), DRUSEN and NORMAL. The training set contained a total of

[1]https://data.mendeley.com/datasets/rscbjbr9sj/3

[2]https://data.mendeley.com/datasets/rscbjbr9sj/2



(a) View1

(b) View1+View2

(c) MDBL-Net(MMFE+BL)

FIGURE 7. T-SNE visualization results of ablation experiments.

83,484 OCT images, and the test set contained 1,000 images from 633 patients (250 images per category).

B. IMPLEMENTATION DETAILS

All of the coding was done using the Pytorch deep learning framework, using an NVIDIA RTX 3060 TI GPU. The SGD optimizer is used to optimize the algorithm, and the learning rate, batch size and weight decay of the network are set to 0.001, 10 and 5×10^{-4}, respectively. In addition, all the input images are resized to 224 × 224 pixels. The performance parameters used to evaluate the prediction are as follows: recall (Eq. 12), F1-score (Eq. 13), accuracy (Eq. 14), and the area under the curve (AUC).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$F1 - \text{score} == \frac{2TP}{2TP + FP + FN} \tag{13}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

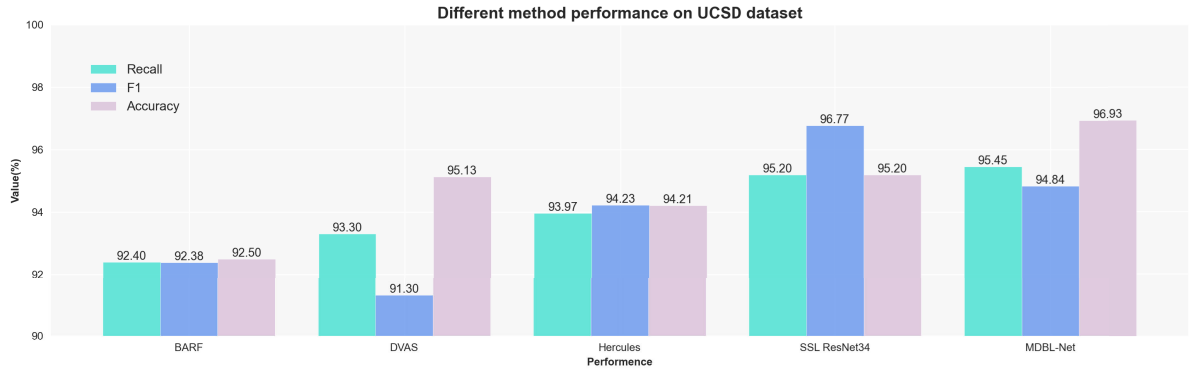
C. EXPERIMENTAL RESULTS

1) COMPARISON WITH OTHER CLASSIFICATION NETWORKS

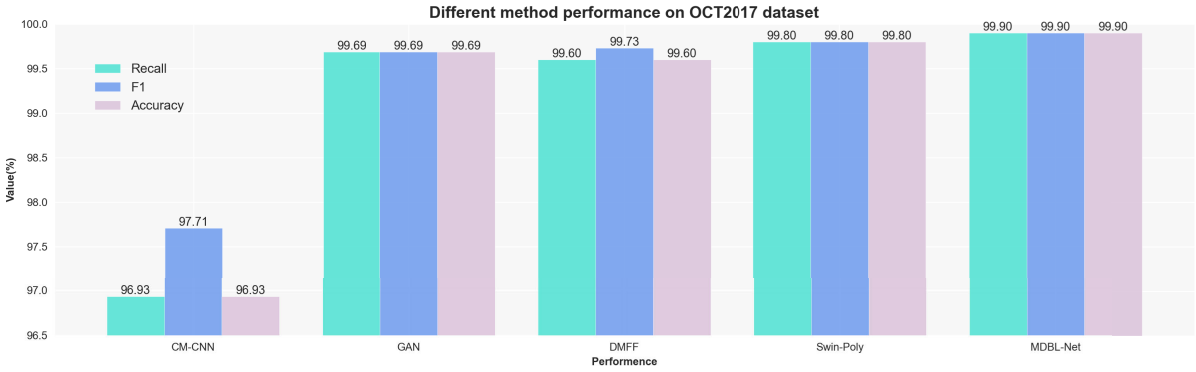
We compare the performance of MDBL-Net with some existing networks, including the traditional ResNet101, Swin_transformer, MobileViT, and ConvNeXt.The experimental results are shown in Table 1, and MDBL-Net achieves the highest performance on all three metrics on both datasets. In particular, MDBL-Net achieved 96.88% accuracy in the UCSD dataset, 2.26% higher than the second-ranked MobileViT. At the same time, 100% accuracy was achieved on the OCT2017 test set. In addition, we can see from the Table 1 that MDBL-Net only takes about 0.5 hour to complete the training, while other models take about 10 hours.

TABLE 4. The MDBL-Net compares with state-of-the-art models on two datasets.

Dataset	Study	Year	# of images	Method	Performence(%)			Pure DL Model
					Recall	F1	Accuracy	
UCSD dataset	Moloud *et.al* [42]	2021	109309	BARF	92.40	92.38	92.50	✓
	Wang *et.al* [43]	2021	109309	DVAS^2	93.30	91.30	95.13	✓
	Abdar *et.al* [44]	2022	109309	Hercules	93.97	94.23	94.21	✓
	Shurrab *et.al* [45]	2023	109309	SSL ResNet34	95.20	**96.77**	95.20	✓
	Ours	2023	109309	**MDBL-Net**	**95.45**	94.84	**96.93**	×
OCT2017	Diao *et.al* [5]	2023	84484	CM-CNN	96.93	97.71	96.93	✓
	Sun *et.al* [46]	2023	84484	GAN	99.69	99.69	99.69	✓
	He *et.al* [47]	2023	84484	DMFF	99.60	99.73	99.60	✓
	Mandal *et.al* [48]	2023	84484	Swin-Poly	99.80	92.80	99.80	✓
	Ours	2023	84484	**MDBL-Net**	**99.90**	**99.90**	**99.90**	×



(a)



(b)

FIGURE 8. Visualization of performance compared to other methods on both datasets.

The above statement indicates that MDBL-Net demonstrates superior performance in retinal disease image recognition compared to the majority of commonly used networks. Furthermore, it exhibits a remarkable lead in terms of training speed.

2) EFFECTIVENESS OF MDBL-NET WITH DIFFERENT PROPORTIONS OF TRAINING SAMPLES

In this experiment, we evaluate the performance of MDBL-Net for small sample data. For UCSD datasets, 80% and 20% of the samples are used as training sets and test sets. Then, four subsets were randomly selected from the training set with the proportions of 1%, 10%, 25%, and 50%, respectively. As can be seen from Table 2, when the training sample is limited (less than 50%), MDBL-Net still maintains a good performance. For other models, limited samples do not keep their performance in a normal range. When the training sample was only 1%, MDBL-Net was still able to maintain 80% accuracy and 78% F1-score, while the other models had accuracy and F1 scores of less than 60%. It is worth noting that the proposed MDBL-Net can get the final result by only one batch of training because of the broad learning design,

which greatly reduces the training time. The experimental results are shown in Table 2. It can be seen from the results that each part of MDBL-Net will have a significant impact on the results.

In summary, MDBL-Net has better performance and less training time for small samples.

3) ABLATION EXPERIMENT

In this section, ablation experiments are performed on the OCT2017 dataset to verify the effectiveness of each module in the proposed network. First we used the structure of VIEW1 for performance testing, then we used the MMFE structure for experiments, and finally tested the performance of MDBL-Net (MMFE+broad learning (BL)). The experimental results are shown in Table 3.

In addition, in order to observe the results more clearly,, we performed t-Distributed Stochastic Neighbor Embedding (t-SNE) on each ablation experiment. As shown in Figure 7 (a), the result of the classification experiment is not ideal because the distance between the datas of different colors is relatively close, while in Figure 7 (b) and (c), the distance between the datas of different colors is relatively far, which means that the model can distinguish different classes well.

4) COMPARISON WITH STATE-OF-THE-ART MODELS

MDBL-Net is compared with the latest state of the art methods, and the results are shown in Table 4, the visualized results are shown in Figure 8. On the UCSD dataset, the proposed method has the highest improvement in accuracy, reaching 3.21%. At the same time, F1 and recall also have an improvement of about 1%. The performance on OCT2017 is also state-of-the-art.

It is worth noting that the above comparison models are pure deep learning (DL) models, which require multiple iterations of optimization and more time to train. Our MDBL-Net is a combination of deep learning and broad learning model, which absorbs the advantages of each other and makes up for the disadvantages, and only needs one iteration to complete the training.

The experimental results confirm that the proposed MDBL-Net can effectively handle the task of retinal disease images classification.

V. CONCLUSION

In this paper, to solve a series of problems in the retinal disease image classification model, a novel artificial-intelligence-based approach for automatic assessment of retinal disease images using multi-view deep-broad learning (MDBL-Net) is proposed. MDBL-Net combines the advantage of deep learning in extracting high-level features, and broad learning with shorter training time, thereby complementing each other. Experiments were performed on two datasets, UCSD and OCT2017. In the experiment, we demonstrated that the training time of the proposed model is significantly shorter than that of traditional models. Furthermore, MDBL-Net maintained good performance even when the training data was limited. Achieving an accuracy of 96.93% on UCSD and 99.90% on OCT2017, MDBL-Net outperformed state-of-the-art models in terms of a specific performance metric in both datasets. However, the proposed model also has some disadvantages, such as the need to input all the training data at one time, so the computing performance of the hardware equipment is relatively high. In future studies, we plan to address this question.

In conclusion, the proposed method can maintain high performance with very few training samples. It also has a much faster training time than most existing methods. Finally, the proposed model maintains the leading classification performance and can be well qualified for the task of retinal disease image recognition.

REFERENCES

[1] R. Bourne, J. D. Steinmetz, S. Flaxman, P. S. Briant, H. R. Taylor, S. Resnikoff, R. J. Casson, A. Abdoli, E. Abu-Gharbieh, and A. Afshin, "Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the global burden of disease study," *Lancet Global Health*, vol. 9, no. 2, pp. 130–143, 2021.

[2] K. Karthik and M. Mahadevappa, "Convolution neural networks for optical coherence tomography (OCT) image classification," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104176. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809422006309

[3] X. Wang, F. Tang, H. Chen, C. Y. Cheung, and P.-A. Heng, "Deep semi-supervised multiple instance learning with self-correction for DME classification from OCT images," *Med. Image Anal.*, vol. 83, Jan. 2023, Art. no. 102673. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841522003012

[4] S. Sotoudeh-Paima, A. Jodeiri, F. Hajizadeh, and H. Soltanian-Zadeh, "Multi-scale convolutional neural network for automated AMD classification using retinal OCT images," *Comput. Biol. Med.*, vol. 144, May 2022, Art. no. 105368. [Online]. Available: https://www.proquest.com/scholarly-journals/multi-scale-convolutional-neural-network/docview/2646710632/se-2

[5] S. Diao, J. Su, C. Yang, W. Zhu, D. Xiang, X. Chen, Q. Peng, and F. Shi, "Classification and segmentation of OCT images for age-related macular degeneration based on dual guidance networks," *Biomed. Signal Process. Control*, vol. 84, Jul. 2023, Art. no. 104810. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809423002434

[6] X. Wang, F. Tang, H. Chen, L. Luo, Z. Tang, A.-R. Ran, C. Y. Cheung, and P.-A. Heng, "UD-MIL: Uncertainty-driven deep multiple instance learning for OCT image classification," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3431–3442, Dec. 2020.

[7] Q. Wang, K. Chen, W. Dou, and Y. Ma, "Cross-attention based multi-resolution feature fusion model for self-supervised cervical oct image classification," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 4, pp. 2541–2554, Jul. 2023.

[8] Q. Lai, C. M. Vong, P. K. Wong, S. T. Wang, T. Yan, I. C. Choi, and H. H. Yu, "Multi-scale multi-instance multi-feature joint learning broad network (M3JLBN) for gastric intestinal metaplasia subtype classification," *Knowl.-Based Syst.*, vol. 249, Aug. 2022, Art. no. 108960. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705122004658

[9] S. G. Ali, Y. Chen, B. Sheng, H. Li, Q. Wu, P. Yang, K. Muhammad, and G. Yang, "Cost-effective broad learning-based ultrasound biomicroscopy with 3D reconstruction for ocular anterior segmentation," *Multimedia Tools Appl.*, vol. 80, nos. 28–29, pp. 35105–35122, Nov. 2021. [Online]. Available: https://www.proquest.com/scholarly-journals/cost-effective-broad-learning-based-ultrasound/docview/2604657549/se-2

[10] X. Wang, L. Cheng, D. Zhang, Z. Liu, and L. Jiang, "Broad learning solution for rapid diagnosis of COVID-19," *Biomed. Signal Process. Control*, vol. 83, May 2023, Art. no. 104724. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S174680942300157X

[11] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.

[12] G. Wu and J. Duan, "BLCov: A novel collaborative-competitive broad learning system for COVID-19 detection from radiology images," *Eng. Appl. Artif. Intell.*, vol. 115, Oct. 2022, Art. no. 105323. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197622003591

[13] R. Han, Z. Liu, and C. L. P. Chen, "Multi-scale 3D convolution feature-based broad learning system for Alzheimer's disease diagnosis via MRI images," *Appl. Soft Comput.*, vol. 120, May 2022, Art. no. 108660.

[14] K. C. Pavithra, P. Kumar, M. Geetha, and S. V. Bhandary, "Computer aided diagnosis of diabetic macular edema in retinal fundus and OCT images: A review," *Biocybernetics Biomed. Eng.*, vol. 43, no. 1, pp. 157–188, Jan. 2023.

[15] L. Wang, W. Dai, M. Jin, C. Ou, and X. Li, "Fundus-enhanced disease-aware distillation model for retinal disease classification from OCT images," 2023, *arXiv:2308.00291*.

[16] M. J. Umer, M. Sharif, M. Raza, and S. Kadry, "A deep feature fusion and selection-based retinal eye disease detection from OCT images," *Expert Syst.*, vol. 40, no. 6, Jul. 2023, Art. no. e13232. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.13232

[17] A. P. Sunija, S. Kar, V. P. Gopi, and P. Palanisamy, "OctNET: A lightweight CNN for retinal disease classification from optical coherence tomography images," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105877. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260720317107

[18] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[20] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014, *arXiv:1409.2329*.

[21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, Jun. 2014, pp. 139–144.

[22] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "MLP-Mixer: An all-MLP architecture for vision," 2021, *arXiv:2105.01601*.

[23] U. Michelucci, "An introduction to autoencoders," 2022, *arXiv:2201.03898*.

[24] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2015, pp. 730–734.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Long Beach, CA, USA: Red Hook, NY, USA: Curran, 2017, pp. 6000–6010.

[27] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[28] T. T. Khoei, H. O. Slimane, and N. Kaabouch, "Deep learning: Systematic review, models, challenges, and research directions," *Neural Comput. Appl.*, vol. 35, no. 31, pp. 23103–23124, Nov. 2023.

[29] V. R. Gottumukkala, N. Kumaran, and V. C. Sekhar, "BLSNet: Skin lesion detection and classification using broad learning system with incremental learning algorithm," *Exp. Syst.*, vol. 39, no. 9, Nov. 2022, Art. no. e13938. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12938

[30] A. Baidar Bakht, S. Javed, S. Q. Gilani, H. Karki, M. Muneeb, and N. Werghi, "DeepBLS: Deep feature-based broad learning system for tissue phenotyping in colorectal cancer WSIs," *J. Digit. Imag.*, vol. 36, no. 4, pp. 1653–1662, Apr. 2023.

[31] J. Wang, S. Lyu, C. L. P. Chen, H. Zhao, Z. Lin, and P. Quan, "SPRBF-ABLS: A novel attention-based broad learning systems with sparse polynomial-based radial basis function neural networks," *J. Intell. Manuf.*, vol. 34, no. 4, pp. 1779–1794, Apr. 2023.

[32] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976.

[33] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

[34] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.

[35] J. Lei Ba, J. Ryan Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.

[36] J. Lin, G. Han, X. Pan, Z. Liu, H. Chen, D. Li, X. Jia, Z. Shi, Z. Wang, Y. Cui, H. Li, C. Liang, L. Liang, Y. Wang, and C. Han, "PDBL: Improving histopathological tissue classification with plug-and-play pyramidal deep-broad learning," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2252–2262, Sep. 2022.

[37] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[38] S. Mehta and M. Rastegari, "MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer," 2021, *arXiv:2110.02178*.

[39] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "ConvNeXt v2: Co-designing and scaling ConvNets with masked autoencoders," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16133–16142.

[40] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, and J. Dong, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, Feb. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0092867418301545

[41] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (OCT) and chest X-ray images for classification," *Mendeley Data*, vol. 2, no. 2, p. 651, 2018.

[42] M. Abdar, M. A. Fahami, S. Chakrabarti, A. Khosravi, P. Plawiak, U. R. Acharya, R. Tadeusiewicz, and S. Nahavandi, "BARF: A new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification," *Inf. Sci.*, vol. 577, pp. 353–378, Oct. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025521007143

[43] X. Wang, H. Chen, H. Xiang, H. Lin, X. Lin, and P.-A. Heng, "Deep virtual adversarial self-training with consistency regularization for semi-supervised medical image classification," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841521000566

[44] M. Abdar, M. A. Fahami, L. Rundo, P. Radeva, A. F. Frangi, U. R. Acharya, A. Khosravi, H.-K. Lam, A. Jung, and S. Nahavandi, "Hercules: Deep hierarchical attentive multilevel fusion model with uncertainty quantification for medical image classification," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 274–285, Jan. 2023.

[45] S. Shurrab, Y. Shannak, and R. Duwairi, "Retina disorders classification via OCT scan: A comparative study between self-supervised learning and transfer learning," *Int. Arab J. Inf. Technol.*, vol. 20, no. 3, 2023.

[46] L.-C. Sun, S.-I. Pao, K.-H. Huang, C.-Y. Wei, K.-F. Lin, and P.-N. Chen, "Generative adversarial network-based deep learning approach in classification of retinal conditions with optical coherence tomography images," *Graefe's Arch. Clin. Experim. Ophthalmol.*, vol. 261, no. 5, pp. 1399–1412, May 2023.

[47] V. Das, S. Dandapat, and P. K. Bora, "Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images," *Biomed. Signal Process. Control*, vol. 54, pp. 101605–101615, Sep. 2019, Art. no. 101605. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809419301867

[48] A. C. Mandal and A. Phatak, "Optimizing deep learning based retinal diseases classification on optical coherence tomography scans," in *Proc. Eur. Conf. Biomed. Opt.*, Aug. 2023, p. 70.

AO ZHANG received the master's degree from the Chongqing University of Science and Technology, China, in 2022. She is currently with the Chongqing Institute of Technology. She has academic experience in teaching, curriculum planning, and ensuring the quality of the teaching process. Her research interests include computer-aided medical diagnostics and nonlinear system control.

XIN QIAN received the bachelor's degree in computer science from Jinggangshan University, China, in 2021. He is currently pursuing the master's degree with the Chongqing University of Science and Technology. His research interests include machine learning and medical image analysis.

CHENGCHENG XU received the bachelor's degree in statistics from Hefei Normal University, China, in 2022. She is currently pursuing the master's degree with the Guangxi University of Science and Technology. Her research interests include statistics of industry and medical data analysis.

JIE ZHANG received the bachelor's degree in information security from the School of Communication Engineering, Xidian University, in June 2007. He is currently the Director of the Department of Network Engineering and Information Security, School of Computer and Internet of Things, Chongqing Institute of Technology. His main research interests include information security, cloud computing, and big data.