RESEARCH ARTICLE

MTRA-CNN: A Multi-Scale Transfer Learning Framework for Glaucoma Classification in Retinal Fundus Images

SANLI YI[ID], LINGXIANG ZHOU, LEI MA, AND DANGGUO SHAO

School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

Corresponding authors: Dangguo Shao (20120024@kust.edu.cn) and Lei Ma (lei_ma@kust.edu.cn)

ABSTRACT In the diagnosis of glaucoma based on deep learning, it is more meaningful for ophthalmologist to make a graded diagnosis using fundus images to reflect the degree of disease. Such multi-classification diagnosis tasks require a larger amount of data to enable the neural network to extract more feature information, while it is very difficult in medical field. To address these issues, we propose a novel Multi-Scale Transfer Learning framework (MTRA-CNN) which can make a graded diagnosis effectively based on graded glaucoma dataset with small volume: (1) Aiming at the problem of small data volume, we innovatively apply multi-stage transfer learning technique in glaucoma, in which the advantage of different types of datasets such as ImageNet and Ocular disease intelligent recognition (ODIR) are taken. (2) We innovatively combine the Residual attention(RA) block, a functional module designed specifically for fundus images, with transfer learning techniques and pre-trained it on the ODIR dataset to further improve the network's upper-level feature extraction ability of glaucoma fundus images. (3) To achieve better diagnostic performance, we propose a novel multi-scale transfer learning method by combining the two-stage transfer learning with one-stage transfer learning. To validate our framework, we conducted experiments on the private glaucoma 4 classification dataset, and the results show that the accuracy of our method is 86.8%, compared to the method without any transfer learning, the accuracy of our experiments improved by 19.79%.

INDEX TERMS Multi-scale transfer learning, graded glaucoma diagnosis, deep CNNs, upper-level feature, bottom-level feature.

I. INTRODUCTION

Glaucoma is the second leading cause of blindness worldwide, and accurate diagnosis of the prevalence of glaucoma is crucial [1]. In the field of deep learning-based glaucoma diagnosis, retinal fundus images are convenient and affordable, so most of the current glaucoma diagnosis methods based on it.

In the glaucoma diagnosis, most of the current deep learning-based diagnostic models only perform dichotomous classification of normal eyes and glaucoma, without further diagnosing non-glaucoma, early glaucoma, middle glaucoma, and advanced glaucoma. In fact, it is more helpful

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir[ID].

to ophthalmologist to make an accurate diagnosis of the different grades, which requires more data and stronger network feature extraction capability. However, a common problem in the current field of medical imaging, especially in the field of glaucoma diagnosis, is the difficulty of obtaining large amounts of data. In order to solve these problems, transfer learning techniques are widely utilized and gaining increasing attention in the field of medical imaging.

Transfer learning can be categorized into four strategies: (1) large size dataset: researchers perform pre-training on large size datasets such as ImageNet to improve the bottom-level feature extraction capability. However, many reserchers can only use the weight parameters on public platforms, like [2] and [3]. The high requirement of

computational power limits researchers who want to perform transfer learning on their own. (2) Medium size dataset of different data type: some researchers bagan to perform their transfer learning on medium size dataset, which is smaller than large size datasets like ImageNet but larger than target datasets. However, they found that pre-training on such datasets often led to poor results, i.e., negative transfer phenomena [4], [5]. This is because the size of such dataset is insufficient that makes it difficult to adequately strengthen the bottom-level feature extraction capability; meanwhile, since the image feature types of the pre-training dataset and the target dataset are not the same, not enough upper-level features can be extracted. (3) Medium size dataset of same data type: researchers found that transfer learning on medium size datasets of the same type can effectly enhances the network's ability to extract upper-level features. (4) Two-stage transfer learning: in recent years, researchers have combined above transfer learning methods and proposed two-stage transfer learning algorithms, such as [6], [7], [8], and [9].This technique can extract more effective features and obtain more accurate results.

In the field of glaucoma multi-classification with small sample sizes, it is difficult to achieve satisfactory results only using any one of above strategies. In order to better perform multi-classification of glaucoma, we propose a multi-scale transfer learning framework named MTRA-CNN based on the characteristics of glaucoma fundus data. By using a multi-scale transfer learning method and combining the Residual attention(RA) enhancement block with transfer learning techniques, it can take full advantage of natural images and other fundus images, thus better solving the problem of small number of glaucoma fundus images and complex features. The main contributions of this paper are as follows:

1) To address the issue of limited data, we utilize multistage transfer learning techniques that take advantage of different datasets, such as ImageNet and Ocular disease intelligent recognition (ODIR), to improve the network's ability to extract features.

2) To enhance the network's capacity to extract upper-level features from fundus images, we incorporate the RA block, a specialized functional module designed for this purpose, and pre-train it on the ODIR dataset.

3) We propose a novel multi-scale transfer learning approach that combines both two-stage and one-stage transfer learning to achieve superior diagnostic performance.

4) We have sorted out four classification datasets for glaucoma, which can be used to grade normal eyes and the early, middle and advanced stages of glaucoma.

The rest of this paper is arranged as follows: Section II reviews the related work. In Section III, we introduce the workflow of the proposed MTRA-CNN framework in detail. The experimental setup and comparative evaluation are described in Section IV. Finally, the work of this paper is summarized in Section V.

II. RELATED WORK

A. GLAUCOMA DIAGNOSIS BASED ON DEEP LEARNING

Earlier, most studies of deep learning-based glaucoma diagnosis simply input glaucoma data into a neural network without considering the size of the dataset. For example, Chen et al. designed a simple network, and inputted glaucoma fundus images into it for glaucoma dichotomous classification, with an AUC of 0.887 [10]. Yu et al. proposed a multi-branched deep learning network model to improve the performance of glaucoma classification [11]. Wang et al. aimed at the characteristics of glaucoma fundus data and proposed a network called SCRD-Net, which enhanced the extraction of fundus image nuances and key details, the AUC value reached 0.940 [12]. Later, as transfer learning was increasingly used in the medical field, researchers found that integrating transfer learning techniques into the network would greatly improve the diagnosis of glaucoma. Ahn et al. used a fine-tuned Inception v3 network pre-trained on ImageNet to perform three-classification of normal, early glaucoma and advanced glaucoma on a private dataset, and demonstrated that transfer learning can improve the feature extraction capability of the network and significantly improve diagnostic performance [13]. Singh et al. further validated the powerful performance of transfer learning on multiple glaucoma datasets using multiple leading deep learning classification networks pre-trained on ImageNet [14]. In particular, Bao et al. devised a new SATL transfer learning method and demonstrated that pre-training on the same type of data would significantly improve the diagnostic performance of glaucoma [15].

B. APPLICATION OF TRANSFER LEARNING IN THE IMAGE DOMAIN

In recent years, transfer learning [16] has achieved astonishing results in small-scale image data processing domains. With the development of transfer learning techniques, the implementation scheme of transfer learning has been improved.

In the early days, the most common implementation was to pre-train the network directly on ImageNet. For example, Yosinski et al. pre-trained a simple deep learning network on ImageNet thus verifying the effectiveness of transfer learning using ImageNet [17]. Shaheed et al. used Xception network to pre-train on ImageNet, improving fingerprint recognition accuracy [18]. Subsequently, some improved transfer schemes based on ImageNet datasets have been proposed. For example, Yi et al. find that combining transfer learning with a functional module specifically designed to extract upper-level features can enhance its applicability [19]. Li et al. used three different pre-training methods on ImageNet, and the classification results were significantly improved [20]. Later, the researchers realized that unlike ImageNet with different types of images, using the same type of dataset could better improve the network's ability to extract upper-level features. Meanwhile, researchers found that upper-level feature

TABLE 1. Literature of transfer learning method.

Study	Transfer learning method	Field of research
[17]	Simple deep learning network on ImageNet	Transfer learning generalizability test for multiple datasets in the image classification domain
[18]	Xception network pretrain on ImageNet	Fingerprint recognition
[19]	Combining the Xception network pre-trained on imagenet with the designed function block	Diagnosis of rectal cancer
[20]	Different pre-training methods on ImageNet	Diabetic retinopathy fundus image classification
[21]	Pre-trained deep learning networks on the same type of dataset	Pediatric pneumonia diagnosis
[7]	Two-stage transfer learning strategy	Central serous chorioretinopathy (CSC) images classification.
[22]	Multi-stage transfer learning approach	Vssential tremor of voice (ETV) and abductor and adductor spasmodic dysphonia (ABSD and ADSD)
[8]	Make full use of Natural images (i.e., ImageNet) and same type medical images to achieve two-stage transfer learning	Lung segmentation
[9]	Make full use of Natural images (i.e., ImageNet) and same type medical images to achieve two-stage transfer learning	Skin lesions segmentation
[23]	Use cell line images serve as an intermediary domain between the natural domain (exemplified by ImageNet) and the specialized medical domain of mammography	Mammographic Breast Mass Classification

transfer from different types of datasets is prone to negative transfer phenomena [4], [5]. Based on these views, many researchers have adopted the scheme of transferring from similar datasets. For example, Zheng et al. pre-trained deep learning networks on the same type of dataset, thus effectively improving the classification effect of pediatric pneumonia [21]. This strategy of using same type of dataset, although better able to extract upper-level features from images, can not provide enough bottom-level features

due to its small volume compared to the different types of datasets which is usually huge, e.g. ImageNet. To further enhance the feature extraction capability of the network based on small datasets, multi-stage transfer learning has been focused. For example, Yoo et al. proposed a two-stage transfer learning strategy for the characteristics of Central serous chorioretinopathy(CSC) images, and achieved better classification results [7]. Yao et al. further utilized a multi-stage transfer learning approach to address the insufficient datasets of vssential tremor of voice (ETV) and abductor and adductor spasmodic dysphonia (ABSD and ADSD) [22]. Liu et al. made full use of natural images (i.e., ImageNet) and medical images of the same type (i.e., LIDC-IDRI), and proposed a two-stage transfer learning method for lung segmentation [8]. Karri et al. used a similar method for segmentation of skin lesions with significant improvement [9]. Gelan et al. mitigates the difficulty associated with procuring a substantial volume of annotated mammogram training data by leveraging an extensive repository of microscopic images from cancer cell lines. These cell line images serve as an intermediary domain for learning, bridging the gap between the natural domain (exemplified by ImageNet) and the specialized medical domain of mammography [23].

Although deep learning has achieved many results in the domain of glaucoma. As far as we know, there are no models of deep learning that consider the integration of multi-stage transfer learning methods into glaucoma diagnosis. The fusion of multi-stage transfer learning methods with improved deep networks would better improve the diagnosis of glaucoma.

III. METHOD
In this paper, to perform glaucoma diagnosis more accurately we propose a deep learning framework called MTRA-CNN, which can solve the problems of small data volume and four classifications for glaucoma diagnosis. The main ideas of the framework are as follows:

We adopt a multi-stage transfer learning strategy. At the first stage, we use the natural image dataset ImageNet to pre-train the neural network to enhance the extraction ability of the bottom-level features of the network. At the second stage, we further train the pre-trained network on the fundus dataset ODIR to enhance the network's extraction ability of the upper-level features of fundus images.

To further improve the capabilities of feature extraction for fundus images of glaucoma, we incorporated the RA (Residual attention) block into MTRA-CNN. The RA block is a functional module proposed for fundus images and can better extract fundus image features [24]. Since this module is more suitable for the extraction of upper-level features of fundus images, we combine the RA block with the pre-trained neural network based on the first stage, and further train the combined network on the ODIR fundus dataset, which together constitute the multi-scale transfer learning framework.

The above transfer learning strategy is shown in the following Equations:

$$ResW_1 = \{ResNet : S_1 \rightarrow S_2\} \tag{1}$$

$$ResW_2 = \{ResW_1 : S_2 \rightarrow T\} \tag{2}$$

$$RAW_1 = \{RA : S_2 \rightarrow T\} \tag{3}$$

$$MTRA\_W = ResW_2 \cup RAW_1 \tag{4}$$

where S_1 denotes the first source domain, which is used for pre-training in this paper, i.e., the ImageNet. S_2 denotes the second source domain, which is used for pre-training in this paper, i.e., the ODIR dataset. $ResW_1$ denotes the ResNet50 weights transferred from the ImageNet dataset to the ODIR dataset. $ResW_2$ denotes on the basis of $ResW_1$, the weights of ResNet50 transferred from the ODIR dataset to the target dataset,.i.e. the glaucoma 4 classification dataset. RAW_1 denotes the weights of the RA functional blocks used in this paper were transferred from the ODIR dataset to the target dataset i.e. the glaucoma 4 classification dataset. $MTRA\_W$ denotes the final network weights used in this paper, combining A and B, applied jointly to the glaucoma 4 classification task.

The MTRA-CNN framework based on multi-scale transfer learning is shown in Figure 4, which consists of four main steps: (1) Pre-trains ResNet50 on the ImageNet to obtain weighted1 ResNet50. (2) The RA functional module is added after the weighted1 ResNet50 to form a new RA-ResNet50 network, then it is pre-trained on the ODIR dataset to obtain the weighted2 RA-ResNet50. (3) The sample dataset of the target task, i.e., glaucoma 4 classification data, is input to the weighted2 RA-ResNet50 for final training to obtain the feature maps reflecting the glaucoma 4 classification data. (4) The feature maps are fed into the classifier to obtain the final glaucoma 4 classification diagnosis.

The main theories concerned in the framework will be detailed later, including A: an overview of the datasets used in our experiments; B: the first-stage transfer learning; C: Second-stage transfer learning based on ODIR dataset; D: the procedure for training the data; E: the classifier of glaucoma.

A. DATASETS

In the experiments of this paper, we used four datasets, including: the Glaucoma fundus 4 classification dataset for glaucoma diagnosis; the ImageNet dataset and ODIR fundus dataset for the multi-stage transfer learning; and the Chest X-Ray Image (Pneumonia) dataset for the validation of negative transfer. Details are as follows:

1) Glaucoma fundus 4 classification dataset. This dataset was approved by the First Affiliated Hospital of Kunming Medical University and the Chinese Tertiary Care Center. These data were determined consecutively by reviewing our hospital's pathology database from January 2015 to March 2021.Patient inclusion criteria were as follows: (a) confirmed diagnosis of glaucoma pathology; (b) preoperative examination of fundus images; (c) definition of the patient's glaucoma severity.The exclusion criteria for patients were:

TABLE 2. Data distribution of Glaucoma fundus 4 classification dataset.

	Normal	Early	Middle	Advanced
Fundus images	492	297	173	306
Augmented	492	492	492	492

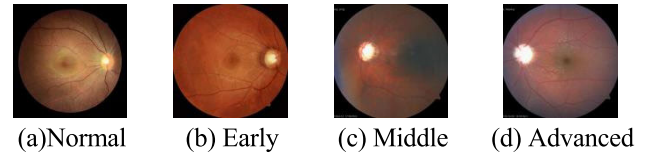| (a)Normal | (b) Early | (c) Middle | (d) Advanced |

FIGURE 1. Fundus images of different stage Glaucoma: (a) Normal; (b) Early; (c) Mild; (d) Advanced.

(a) lesion-related treatment prior to examination, including surgery, chemotherapy, radiotherapy, and targeted drug therapy; (b) previous history of ocular or systemic disease affecting the visual field or optic nerve; and (c) substandard image quality. Pathological data of all patients were annotated by two ophthalmologists. This study was approved by the First Affiliated Hospital of Kunming Medical University and followed the principles of the Declaration of Helsinki. In this paper, we used this dataset for 4 classification of glaucoma diagnosis.

This dataset includes 492 normal eyes, 297 early glaucoma, 173 mild glaucoma, and 306 advanced glaucoma. Such distribution is not balance, which would cause the results be biased in favor of the category with larger amount of data, thus affecting its training result. Whereas a data balanced dataset would not have this problem. Therefore, we equalize each category of data to a fixed number 492 by preprocessing the data with random combinations of flip, rotation, translation and scale adjustment. The detailed distribution of the data in each category is shown in Table 2. The fundus image for each category is shown in Figure 1.

2) ImageNet [25]. ImageNet is a database contains over 14 million annotated high-resolution images, including 1,000 different image categories with approximately 1,000 images per category. It is widely used in the field of computer vision because it contains enough data and rich annotation information to improve the generalization ability and accuracy of the trained model. In this paper, the weights of networks pre-trained on ImageNet, can be called directly from the keras library function.

3) Ocular disease intelligent recognition (ODIR) dataset(https:// odir2019.grand-challenge.org/.) [26]. This dataset contains more than 10,000 fundus images for multiple eye disease categories. The diseases are classified into eight categories which are normal, diabetes, glaucoma, cataract, AMD, hypertension and myopia and others. In this paper, we use this dataset to validate the proposed multi-scale transfer learning algorithm. Because it has the problem of unbalanced data categories, we took the same preprocessing operation described above to equalize each category of fundus disease to 4,000 samples, for a total of 32,000 fundus images. The specific

TABLE 3. Data distribution of ODIR fundus data.

	Normal	Diabetes	Glaucoma	Cataract	AMD	Hypertension	Myopia	Others
Fundus images	3994	1970	264	386	379	145	312	1681
Augmented	4000	4000	4000	4000	4000	4000	4000	4000

details of the data are shown in Table 3. The data examples of different diseases are shown in Figure 2.

4) Chest X-Ray Image (Pneumonia) (https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia) [27]. The dataset is composed of chest X-ray images of children aged 1 to 5 years from Guangzhou Women and Children Medical Center. This dataset contains 4273 pneumonia images, 1583 normal images, and a total of 5856 chest X-ray images. We use it to verify that when the amount of data in the source domain is insufficient and the image types are different from the target image, it leads to negative transfer phenomenon. The data examples are shown in Figure 3.

B. FIRST-STAGE TRANSFER LEARNING BASED ON ImageNet

In order to improve the bottom-level feature extraction capability of the network, in this study we pre-train ResNet50 on the ImageNet. ResNet50 is designed to mitigate the problem of gradient disappearance in the network, so it can achieve better results in applications of small datasets such as glaucoma compared to other deeper networks. The structure of ResNet50 is shown in Figure 5, which is composed of four residual convolution blocks, i.e. Residual Block1, Residual Block2, Residual Block3, and Residual Block4. Each residual block is composed of different convolution layers and residual connections. The ImageNet contains 14 million annotated images covering 1,000 categories, which allows us to efficiently tune the convolutional layer training parameters and initialize the weights of the model, thus enhancing the network's ability to extract the bottom-level features with powerful generalization capabilities, which can also be applied to the processing of medical images such as glaucoma.

C. SECOND-STAGE TRANSFER LEARNING BASED ON ODIR DATASET

1) FORMATION OF RA-RESNET50 NETWORK

To further improve the network's ability to extract features of fundus images, we added the RA block, a functional module, after the pre-trained ResNet50 network, i.e., weighted1 ResNet50, to form the RA-ResNet50 network. RA block was initially used for a diabetic retinopathy classification task based on fundus images, due to the ability to effectively extract the upper-level features of fundus images, it can also achieve excellent results in 4 classification of glaucoma based on fundus images. Therefore, we incorporated the RA block into the proposed multi-scale transfer learning framework to enhance the 4 classification diagnosis of glaucoma.The
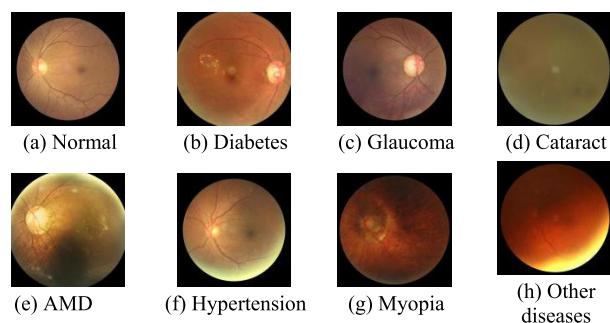

(a) Normal	(b) Diabetes	(c) Glaucoma	(d) Cataract
(e) AMD	(f) Hypertension	(g) Myopia	(h) Other diseases

FIGURE 2. Fundus images of different diseases in ODIR dataset: (a) Normal; (b) Diabetes; (c) Glaucoma; (d) Cataract; (e) AMD; (f) Hypertension; (g) Myopia; (h) Other diseases.
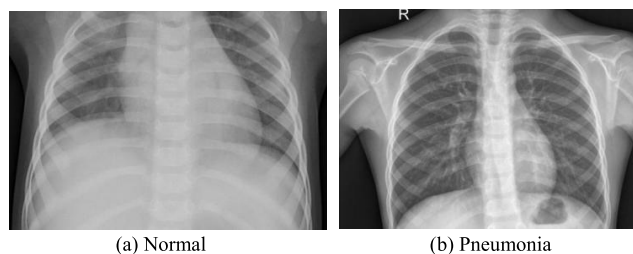

(a) Normal	(b) Pneumonia

FIGURE 3. X-Ray Images of Normal and Pneumonia: (a) Normal; (b) Pneumonia.

structure of the RA block is shown in Figure 6, which integrates the depth-separable convolution, residual mechanism and attention mechanism.

#### 2) SECOND STAGE TRANSFER LEARNING

ResNet50 network based on the ODIR dataset. Since the ResNet50 network was transferred twice, while the RA block was only transferred once, we named this technique multi-scale transfer learning. The fundus image data in ODIR dataset and the target dataset belong to the same type of data, so the weighted2 RA-ResNet50, which is obtained in the second stage transfer learning on ODIR dataset, has stronger upper-level feature extraction ability, i.e., the feature extraction ability of fundus image data.

D. GLAUCOMA DATA TRAINING

After completing the multi-scale transfer learning pre-training, we input the glaucoma sample data into the pre-trained weighted2 RA-ResNet50 for training to obtain feature maps,which reflect the 4 classification information of glaucoma. Then, we input it into the classifier to achieve 4 classification of glaucoma.
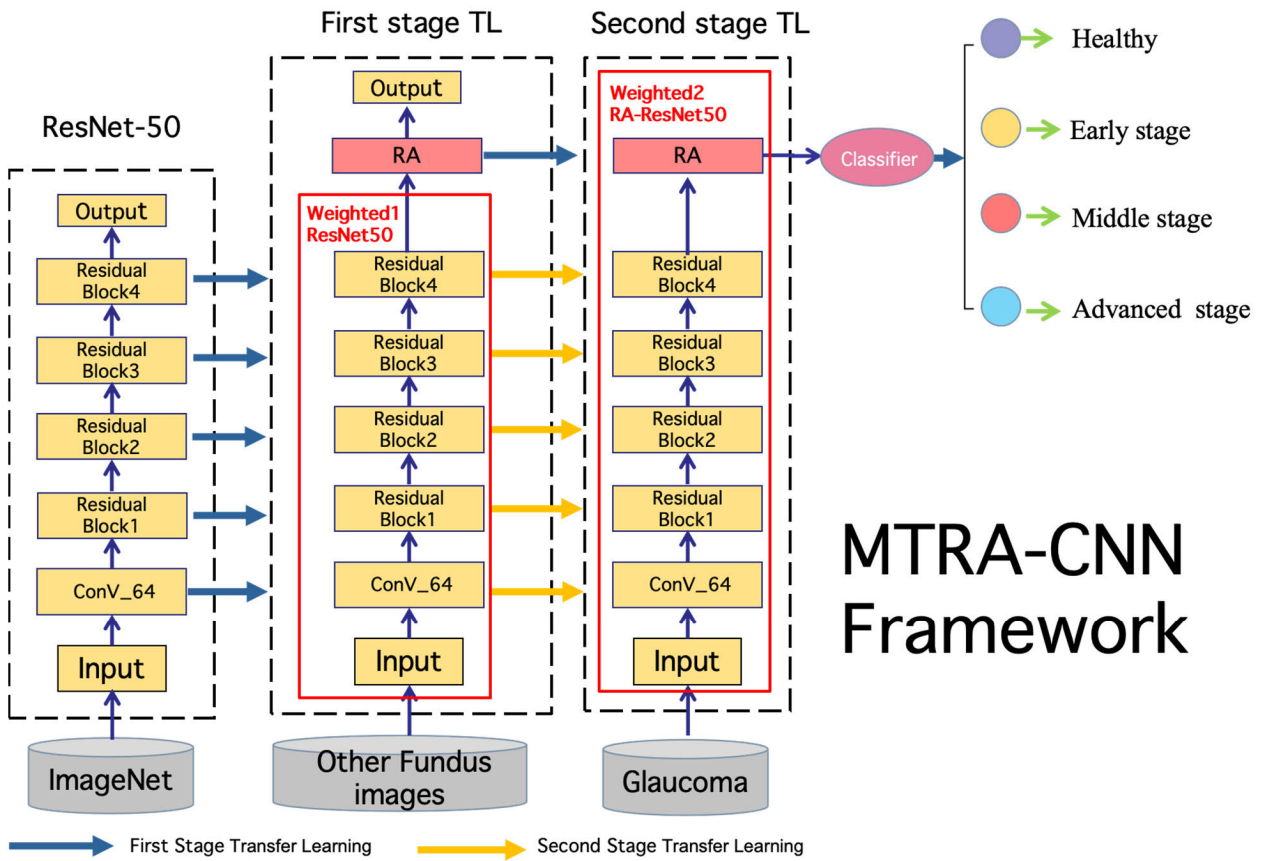
E. CLASSIFIER

In our MTRA-CNN framework, the structure of the classifier used to achieve 4 classification of glaucoma is shown in Figure 7. The inputs of the classifier are feature maps, which are obtained from glaucoma sample dataset trained on weighted2 RA-ResNet50. Our classifier consists of a global average pooling (GAP) layer, a BN layer, a fully connected
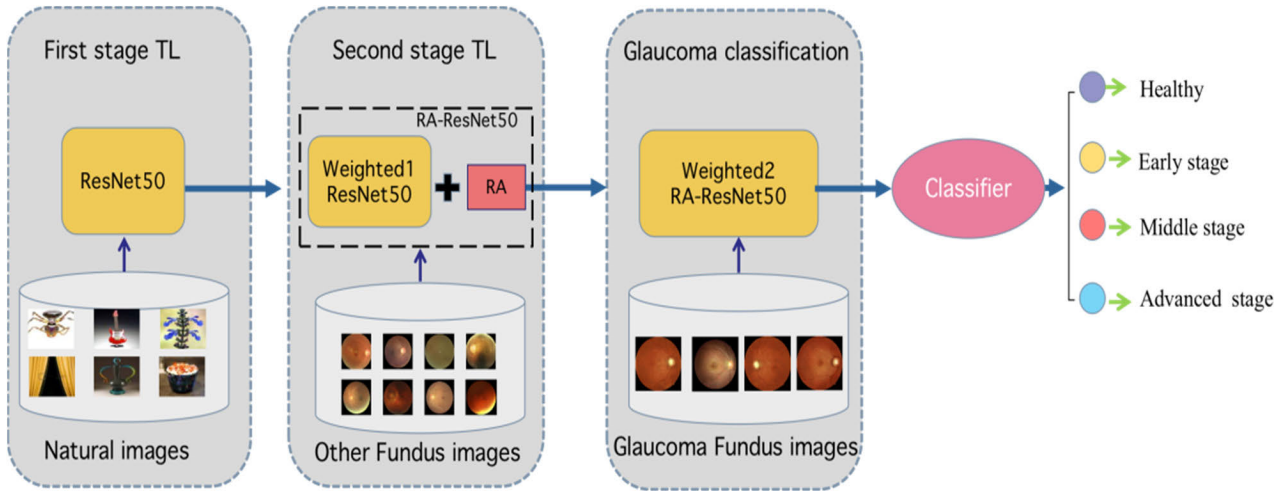
(a) Overall architecture of the MTRA-CNN framework.

(b) Detailed structure of the MTRA-CNN framework .

FIGURE 4. Framework of the proposed MTRA-CNN: (a) Overall architecture of the MTRA-CNN framework; (b) Detailed structure of the MTRA-CNN framework.

layer(the nodes for the filters of the layer set as 128) and a Softmax layer. the GAP layer reduces the spatial resolution, the BN layer enhances the convergence of the network, and the Softmax layer is used for the fundus image classification
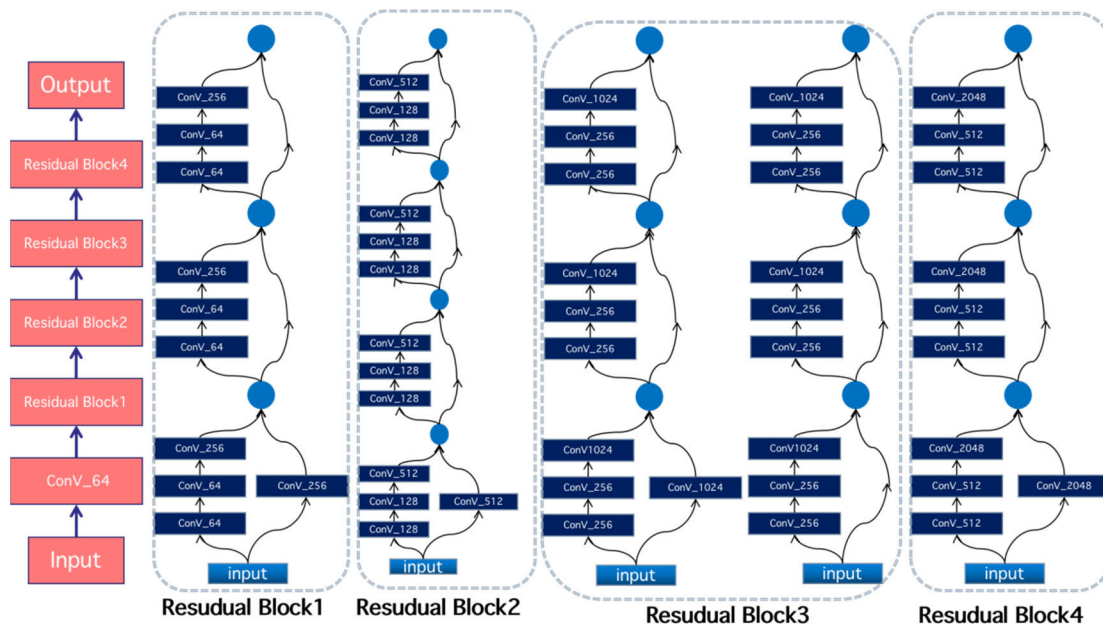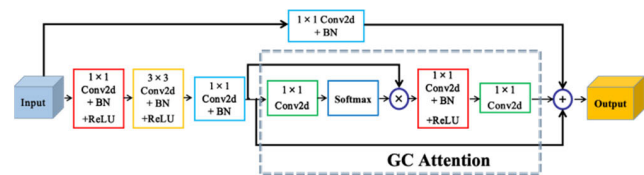
FIGURE 5. ResNet50 architecture.

FIGURE 6. The structure of the RA block.

FIGURE 7. The structure of the classifier.

task, which finally completes the diagnosis of normal, early, middle, and advanced glaucoma.

IV. EXPERIMENTS RESULTS AND DISCUSSION

A. IMPLEMENTATION DETAILS

All experiments were conducted on a computer equipped with Intel(R) Core(tm) i9-10900 k cpu@3.7 GHz, 64.0 GB RAM, NVIDIA GEFORCE GTX 3080 and 64-bit Windows 10 operating system, compiler is Pycharm 3.6.0, and all experiments are based on the TensorFlow 2.4.0 framework. All classification experiments in this paper use ten-fold cross-validation, where 90% of the data are trained in each fold and the results are obtained by testing on the remaining 10% of the data. The parameters of our model are set as follows: the size of the image input to the network is 224×224, the optimizer of the network is set to ADAM, the learning rate is set to 0.001, the loss function is set to cross-entropy, the batch_size is set to 16, and the epoch is set to 50.

Our experiments including: comparison of different CNN networks, ablation experiment of RA block, and comparison of different transfer learning strategies. All the transfer learning strategies are presented in Figure 8, and their detailed description are in sub-section C.(3) and C.(4).
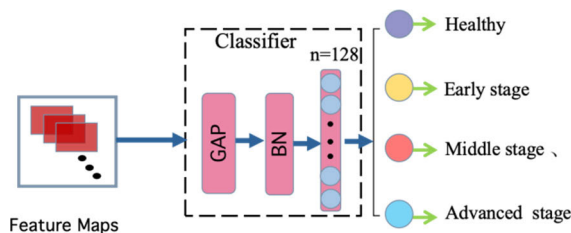
B. EVALUATION METRICS

To present the results of the above diagnostic model, the following evaluation methods were used. We compared Accuracy, Precision, F1-Score, Recall, and Area Under the Subject Operating Characteristic (ROC) curve (AUC). Accuracy is the ratio of the number of samples correctly predicted by the classifier to the total number of samples. It measures the overall performance of the classifier. The Precision rate indicates the proportion of samples that the classifier predicts correctly out of all the samples that it predicts as positive examples. It measures how accurately the classifier predicts as positive cases. Recall measures the proportion of samples that the classifier predicts as positive examples out of all true positive examples. It measures how well the classifier checks for true positive examples and reflects the sensitivity of the classifier. The F1-score is the reconciled mean of precision and recall, combining these two metrics.The F1-score allows for a comprehensive assessment of the model. The AUC value represents the area under the ROC curve of the classifier, which is used to measure the overall performance of the classification model.The closer the AUC value is to 1, the

better the performance of the classifier, which is able to better distinguish between positive and negative cases.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (5)$$

$$Precision = \frac{TP}{TP + FP} \qquad (6)$$

$$Recall = \frac{TP}{TP + FN} \qquad (7)$$

$$F1 - Score = \frac{2 * TP}{(2 * TP) + FP + FN}, \qquad (8)$$

C. EXPERIMENT RESULTS

1) PROBLEMS OF DIFFERENT CNNS WITHOUT TRANSFER LEARNING

In the grade diagnosis of glaucoma, the performance of the commonly used CNNs are usually not satisfactory due to the small size of the dataset and the need for four classifications. The problem is exemplified by comparison experiments on several commonly used CNNs such as MobileNetV2 [28], EfficientNetB4 [29], VGG19 [30], DenseNet169 [31] and ResNet50 [32].

Table 4 shows the classification results of each deep learning network in the glaucoma 4 classification dataset used in this paper. From the results, it can be seen that (1) direct use of deep learning networks could not achieve satisfactory results on the glaucoma 4 classification dataset. (2) Directly using too lightweight networks like MobileNetV2 cannot converge on small datasets like glaucoma 4 classification, while directly using too deep networks like DenseNet169 will not work well due to overfitting. (3) Among all networks, ResNet50 achieved the best classification results, which indicated that ResNet50 was the most suitable among these deep learning networks for a 4 classification diagnosis of glaucoma.

2) ABLATION EXPERIMENT OF RA BLOCK

To verify the performance of the RA block, we pre-trained each network on ImageNet and then compared the diagnostic effects of the network before and after adding the RA module.

From the results, it can be seen: (1) the performance of each network has been greatly improved after transfer learning is applied, by comparing Table 4 and Table 5. (2) the performance of each network achieved some improvement with the addition of RA block, due to the improved attention to the lesion region by RA block. (3) The ResNet+RA achieved the best classification results among all networks. It reached 83.75% in Accuracy; 84.14% in Precision; 83.72% in Recall; 83.78% in F1-score; and 89% in AUC.

3) ABLATION EXPERIMENT OF THE MULTI-SCALE TRANSFER LEARNING

In order to explore and validate the effectiveness of each transfer learning step of our proposed multi-scale transfer learning framework, we conducted ablation experiments in this section, and adopt five transfer learning (i.e. TL)

strategies of Figure 8: (1) No TL: the glaucoma images are directly input to each network to perform glaucoma 4 classification. (2) TL Strategy1: each network is pre-trained on the ODIR dataset first, and then glaucoma images are input to the pre-trained network for glaucoma grading. (3) TL Strategy2: each network is pre-trained on the ImageNet first, and then glaucoma images are input to the pre-trained network for glaucoma grading. (4) TL Strategy3: each network is first pre-trained on the ImageNet, and subsequently pre-trained on the ODIR dataset in the second stage, at last the glaucoma images are input to the network undergoing the two stages of pre-training for glaucoma grading. (5) TL Strategy4: is the proposed MTRA-CNN framework in this paper. It indicates that the network is first pre-trained on ImageNet, then RA block is added after the pre-trained network, followed by the second stage of pre-training on ODIR dataset, and finally the glaucoma images are input to the pre-trained RA-ResNet for glaucoma grading.

Table 6 lists the results of the ablation experiments for each transfer learning step on the glaucoma 4 classification dataset, from which we can draw the following conclusions: (1)The accuracy of both TL strategy 1 and TL strategy 2 is higher than that of No TL, indicating that pre-training either on the same type of dataset or on different types of datasets with huge volumes and multiple categories enhances the network performance. Where TL strategy 1 enhances the network's ability to extract upper-level features of images and TL strategy 2 enhances the network's ability to extract bottom-level features of images. (2) The accuracies TL Strategy3 are higher than that of TL Strategy1 and TL Strategy2, which indicates that multi-stage transfer learning on the ImageNet and ODIR datasets can further enhance the feature extraction capability of the model and thus further increase the diagnostic performance. (3) TL strategy 4, which is the MTRA-CNN framework proposed in this paper, performs best, indicating that each module of this algorithm can improve the performance and their combinations can achieve satisfactory results, among which the best results are achieved by using ResNet50 as the backbone network.

4) EXPERIMENT OF NEGATIVE TRANSFER LEARNING

We conducted negative transfer experiments to verify the significance of using datasets of the same type during pre-training of upper-level features on fundus images. In this experiment, two transfer learning strategies of Figure 8 are adopted, which are detailed as: (1) TL Strategy5: each network is first pre-trained on Chest X-Ray Image (Pneumonia) dataset, and subsequently the glaucoma images are input to the pre-trained network for glaucoma classification. (2) TL Strategy6: each network is first pre-trained on the ImageNet, and subsequently pre-trained on the Chest X-Ray Image (Pneumonia) dataset at the second stage, the glaucoma images were then input to a network that underwent two stages of pre-training to perform glaucoma grading.

Table 7 and Table 8 show the results of negative transfer experiments, from which we can draw the following
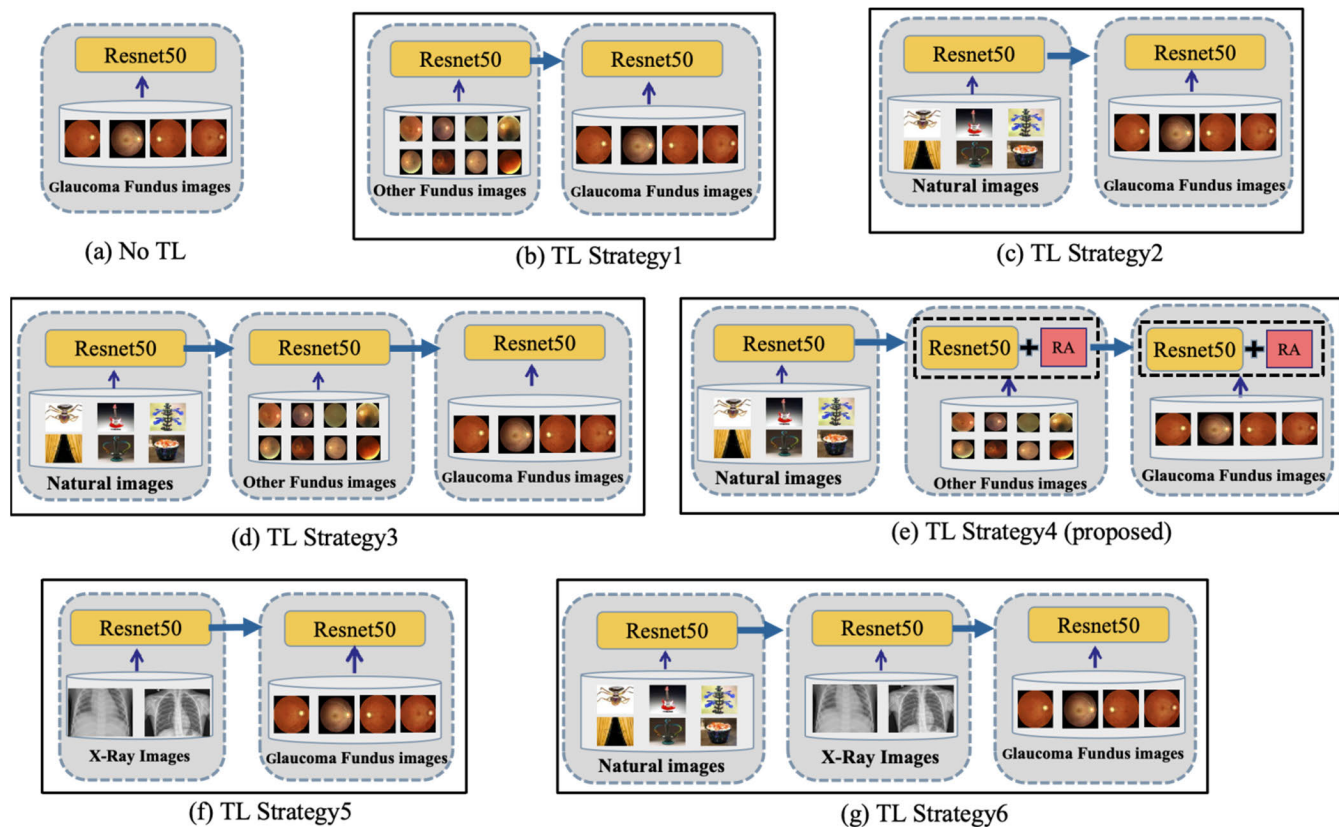
FIGURE 8. Flow charts of various transfer learning strategies: (a) No TL; (b) TL Strategy1; (c) TL Strategy2; (d) TL Strategy3; (e) TL Strategy4; (f) TL Strategy5; (g) TL Strategy6.

TABLE 4. Comparison of different CNN networks.

Model	Accuracy(%)	Precision(%)	Recall(%)	F1 -score(%)	AUC(%)
MobileNetV2	26.39	6.96	25.00	10.89	50
EfficientNetB4	56.85	56.15	56.31	56.23	71
VGG19	61.97	66.72	61.67	64.10	75
DenseNet169	64.97	67.73	64.62	66.14	76
ResNet50	67.01	67.01	66.41	66.71	78

TABLE 5. Comparison of CNNs with and without RA block.

Model	Accuracy(%)	Precision(%)	Recall(%)	F1- score(%)	AUC(%)
MobileNetV2	78.68	78.43	78.21	78.32	86
EfficientNetB4	79.18	81.88	78.80	80.31	86
VGG19	76.65	76.37	76.19	76.28	84
DenseNet169	81.21	82.08	80.78	81.42	87
ResNet50	82.23	83.25	81.99	82.62	88
MobileNetV2+RA	80.71	81.42	80.08	80.74	87
EfficientNetB4+RA	80.71	80.63	80.22	80.42	87
VGG19+RA	77.66	78.78	77.73	78.25	85
DenseNet169+RA	82.74	83.42	82.77	83.09	88
ResNet50+RA	83.75	84.14	83.72	83.93	89

TABLE 6. Comparison of CNNs using different transfer learning strategies.

	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
No TL	MobileNetV2	26.39	6.96	25.00	10.89	50
	EfficientNetB4	56.85	56.15	56.31	56.23	71
	VGG19	61.97	66.72	61.67	64.10	75
	DenseNet169	64.97	67.73	64.62	66.14	76
	ResNet50	67.01	67.01	66.41	66.71	78
TL Strategy1	MobileNetV2	68.02	67.25	67.81	67.53	79
	EfficientNetB4	72.08	72.46	71.28	71.87	81
	VGG19	71.07	71.69	70.73	71.21	81
	DenseNet169	74.11	74.01	73.64	73.82	83
	ResNet50	74.62	78.63	74.00	76.24	83
TL Strategy2	MobileNetV2	78.68	78.43	78.21	78.32	86
	EfficientNetB4	79.18	81.88	78.80	80.31	86
	VGG19	76.65	76.37	76.19	76.28	84
	DenseNet169	81.21	82.08	80.78	81.42	87
	ResNet50	82.23	83.25	81.99	82.62	88
TL Strategy3	MobileNetV2	80.71	80.63	80.22	80.42	87
	EfficientNetB4	81.21	83.08	81.12	82.09	87
	VGG19	77.66	78.78	77.73	78.25	85
	DenseNet169	83.25	84.38	83.04	83.70	89
	ResNet50	84.77	85.04	84.18	84.61	90
TL Strategy4 (Proposed)	MobileNetV2	82.23	82.57	81.98	82.27	88
	EfficientNetB4	83.75	84.78	83.70	84.24	89
	VGG19	79.69	82.26	79.58	80.90	86
	DenseNet169	85.78	86.04	85.29	85.66	91
	ResNet50	86.80	86.76	86.55	86.65	91

TABLE 7. Comparison of first negative transfer learning strategy.

	Model	Accuracy (%)	Precision (%)	Recall (%)	F1- score (%)	AUC (%)
No TL	MobileNetV2	26.39	6.96	25.00	10.89	50
	EfficientNetB4	56.85	56.15	56.31	56.23	71
	VGG19	61.97	66.72	61.67	64.10	75
	DenseNet169	64.97	67.73	64.62	66.14	76
	ResNet50	67.01	67.01	66.41	66.71	78
TL Strategy5	MobileNetV2	26.39	6.96	25.00	10.89	50
	EfficientNetB4	48.73	59.15	49.10	53.66	66
	VGG19	53.30	56.33	53.10	54.67	70
	DenseNet169	55.84	58.77	55.51	57.09	71
	ResNet50	59.39	61.07	58.15	59.57	73

conclusions: (1) The performance of TL Strategy5 is lower than that of No TL, which indicates that pre-training upper-level features on small-scale images of different types will reduce the performance of the model and lead to the negative transfer phenomenon. (2) The performance of TL Strategy6 is lower than that of TL Strategy2, which indicates that pre-training on small-scale dataset different from the target dataset at the second-stage of transfer learning, not only fails to enhance the network's ability to extract upper-level features of the target data but also weakens the effect of pre-training at the first stage.

5) COMPARISON WITH OTHER STATE-OF-THE-ART METHODS

Table 9 shows the comparison of our proposed framework with the two existing methods in other papers, among them OZER CAN designed new deep learning network named self-onns for the characteristics of glaucoma fundus image data. Law Kumar Singh pre-trained several existing deep learning networks on imagenet to diagnose glaucoma using one stage transfer learning. The results show that, our Accuracy reached 86.80%, which was 2.54% higher than the method proposed by Law Kumar Singh, which is the highest

TABLE 8. Comparison of second negative transfer learning strategy.

	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 -score (%)	AUC (%)
TL Strategy2	MobileNetV2	78.68	78.43	78.21	78.32	86
	EfficientNetB4	79.18	81.88	78.80	80.31	86
	VGG19	76.65	76.37	76.19	76.28	84
	DenseNet169	81.21	82.08	80.78	81.42	87
	ResNet50	82.23	83.25	81.99	82.62	88
TL Strategy6	MobileNetV2	73.09	72.74	72.96	72.85	82
	EfficientNetB4	75.13	84.39	75.31	79.59	83
	VGG19	69.54	68.57	68.79	68.68	80
	DenseNet169	77.66	78.78	77.73	78.25	85
	ResNet50	79.18	81.88	78.80	80.31	86

TABLE 9. Comparison with other state-of-the-art methods.

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1- score (%)	AUC (%)
OZER CAN [33]	79.18	81.88	78.80	80.31	86
Law Kumar Singh [14]	84.26	84.72	84.32	84.52	89
Proposed	**86.80**	**86.76**	**86.55**	**86.65**	**91**
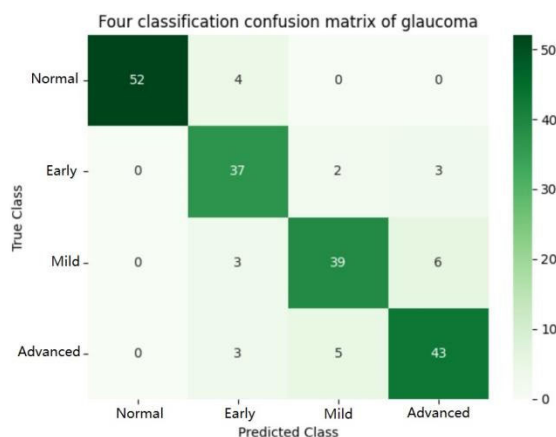
FIGURE 9. The confusion matrix of the MTRA-CNN method.

FIGURE 10. The ROC curve of the MTRA-CNN method: class0 denotes normal eyes; class1 denotes early glaucoma; class2 denotes middle glaucoma; class3 denotes advanced glaucoma.

of the others. In terms of Precision, it reached 86.76%, which was 2.04% higher than the method proposed by Law Kumar Singh. In terms of Recall, it reached 86.55%, 2.23% higher than the method proposed by Law Kumar Singh, which is the highest of the others. In terms of F1-score, it reached 86.67%, 2.31% higher than the method proposed by Law Kumar Singh, which is the highest of the others. In terms of AUC, it reached 91%, 2% higher than the method proposed by Law Kumar Singh, which is the highest of the others. Since our multi-scale transfer learning framework combines data features and transfer learning methods, whereas [33] only uses data characteristics and [14] only uses transfer learning methods, our framework obtains the best results. Figures 9, 10 and 11 show plots of the confusion matrix, ROC curves and
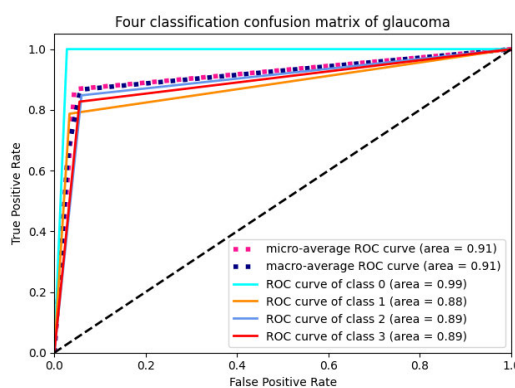
the training process.Thus, overall, the model proposed in this paper has the best performance in glaucoma 4 classification.

V. CONCLUSION
In this paper, we propose a new automatic glaucoma classification framework, MTRA-CNN, and to verify the effectiveness of the algorithm, four experiments are performed: (1) in the first experiment, the commonly used CNNs were compared for the graded diagnosis of glaucoma, and the results showed that ResNet50 was the most appropriate. (2) in the second experiment, we performed ablation expriment of RA block on the pre-trained CNNS, the results demonstrate that RA block can enhance the perfermence of each pretrained CNNS. (3) In the third experiment, different transfer learning
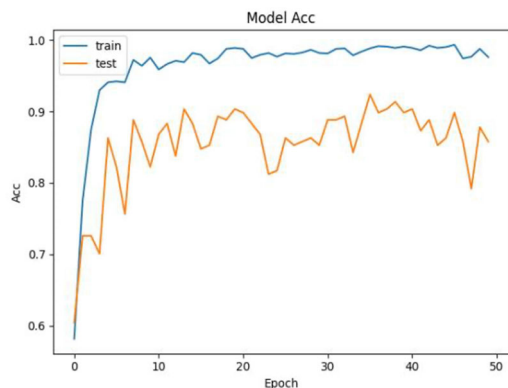
FIGURE 11. Change in accuracy with epoch increase of the MTRA-CNN method.

stratages were compared, the results demonstrate that the multi-scale transfer learning stratage proposed in this paper performed best. (4) At last, we conducted a negative transfer learning experiment. The results show that pre-training on different types of small-scale datasets cannot improve the upper-level feature extraction ability, leading to the negative transfer phenomenon

The innovative points of this paper are as follows: (1) we propose a multi-scale transfer learning framework, MTRA-CNN, which combines the two stage transfer learning with the one stage transfer learning and makes full use of multiple datasets for transfer learning, including ImageNet, ODIR, and other datasets. This transfer learning method can improve the network's ability to extract features at different levels, thus improving the performance of glaucoma diagnosis. (2) We connect RA blocks after the network pre-trained at the first transfer learning stage, and combine it with transfer learning techniques, then pre-trained the combined network on ODIR dataset to further improve the network's ability to extract glaucoma fundus image features. (3) Aiming at the characteristics of glaucoma fundus data, we have compiled a 4 classification dataset for glaucoma and are able to classify normal eyes as well as early, middle and advanced stages of glaucoma. The results show that our accuracy reaches 86.8%, which is higher than other transfer learning strategies, proving the effectiveness and superiority of multi-scale transfer learning strategies.

Since all multi-stage transfer learning strategies require multiple pre-training, this results in high computational costs. In the future, we plan to maintain the performance benefits of multi-stage transfer learning while simplifying processes and reducing complexity.

USE OF AI TOOLS DECLARATION
The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

CONFLICTS OF INTEREST
The authors declare no conflicts of interest.

DATA AVAILABILITY
Data underlying the results presented in this paper are available in Dataset 1, [23].

Data underlying the results presented in this paper are available in Dataset 2, [24]. (https:// odir2019.grand-challenge.org/.)

Data underlying the results presented in this paper are available in Dataset 3, [25].(https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia)

REFERENCES
[1] M. Aamir, M. Irfan, T. Ali, G. Ali, A. Shaf, S. A. Saeed, A. Al-Beshri, T. Alasbali, and M. H. Mahnashi, "An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification," *Diagnostics*, vol. 10, no. 8, p. 602, Aug. 2020.
[2] Accessed: May 1, 2023. [Online]. Available: https://storage.googleapis.com/tensorflow/keras-applications/nasnet
[3] Accessed: May 1, 2023. [Online]. Available: https://download.pytorch.org/models
[4] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[5] J. Plested and T. Gedeon, "Deep transfer learning for image classification: A survey," 2022, *arXiv:2205.09904*.
[6] J. Liu, F. Guo, H. Gao, Z. Huang, Y. Zhang, and H. Zhou, "Image classification method on class imbalance datasets using multi-scale CNN and two-stage transfer learning," *Neural Comput. Appl.*, vol. 33, no. 21, pp. 14179–14197, Nov. 2021.
[7] T. K. Yoo, S. H. Kim, M. Kim, C. S. Lee, S. H. Byeon, S. S. Kim, J. Yeo, and E. Y. Choi, "DeepPDT-Net: Predicting the outcome of photodynamic therapy for chronic central serous chorioretinopathy using two-stage multimodal transfer learning," *Sci. Rep.*, vol. 12, no. 1, p. 18689, Nov. 2022.
[8] J. Liu, B. Dong, S. Wang, H. Cui, D.-P. Fan, J. Ma, and G. Chen, "COVID-19 lung infection segmentation with a novel two-stage cross-domain transfer learning framework," *Med. Image Anal.*, vol. 74, Dec. 2021, Art. no. 102205.
[9] M. Karri, C. S. R. Annavarapu, and U. R. Acharya, "Skin lesion segmentation using two-phase cross-domain transfer learning framework," *Comput. Methods Programs Biomed.*, vol. 231, Apr. 2023, Art. no. 107408.
[10] X. Chen, Y. Xu, D. W. Kee Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 715–718.
[11] S. Yu, H.-Y. Zhou, K. Ma, C. Bian, C. Chu, H. Liu, and Y. Zheng, "Difficulty-aware glaucoma classification with multi-rater consensus modeling," in *Proc. 23rd Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Lima, Peru. Springer, Oct. 2020, pp. 741–750.
[12] H. Wang, J. Hu, and J. Zhang, "SCRD-Net: A deep convolutional neural network model for glaucoma detection in retina tomography," *Complexity*, vol. 2021, pp. 1–11, Apr. 2021.
[13] J. M. Ahn, S. Kim, K.-S. Ahn, S.-H. Cho, K. B. Lee, and U. S. Kim, "A deep learning model for the detection of both advanced and early glaucoma using fundus photography," *PLoS ONE*, vol. 13, no. 11, Nov. 2018, Art. no. e0207982.
[14] L. K. Singh, Pooja, H. Garg, and M. Khanna, "Deep learning system applicability for rapid glaucoma prediction from fundus images across various data sets," *Evolving Syst.*, vol. 13, no. 6, pp. 807–836, Dec. 2022.
[15] Y. Bao, J. Wang, T. Li, L. Wang, J. Xu, J. Ye, and D. Qian, "Self-adaptive transfer learning for multicenter glaucoma classification in fundus retina images," in *Proc. 8th Int. Workshop Ophthalmic Med. Image Anal. (OMIA), Held Conjunct With MICCAI*, Strasbourg, France. Springer, Sep. 2021, pp. 129–138.
[16] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.
[17] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[18] K. Shaheed, A. Mao, I. Qureshi, M. Kumar, S. Hussain, I. Ullah, and X. Zhang, "DS-CNN: A pre-trained xception model based on depth-wise separable convolutional neural network for finger vein recognition," *Expert Syst. Appl.*, vol. 191, Apr. 2022, Art. no. 116288.

[19] S. Yi, Y. Wei, X. Luo, and D. Chen, "Diagnosis of rectal cancer based on the xception-MS network," *Phys. Med. Biol.*, vol. 67, no. 19, Oct. 2022, Art. no. 195002.

[20] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–11.

[21] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 104964.

[22] Y. Yao, M. Powell, J. White, J. Feng, Q. Fu, P. Zhang, and D. C. Schmidt, "A multi-stage transfer learning strategy for diagnosing a class of rare laryngeal movement disorders," *Comput. Biol. Med.*, vol. 166, Nov. 2023, Art. no. 107534.

[23] G. Ayana, J. Park, and S.-W. Choe, "Patchless multi-stage transfer learning for improved mammographic breast mass classification," *Cancers*, vol. 14, no. 5, p. 1280, Mar. 2022.

[24] S.-L. Yi, X.-L. Yang, T.-W. Wang, F.-R. She, X. Xiong, and J.-F. He, "Diabetic retinopathy diagnosis based on RA-EfficientNet," *Appl. Sci.*, vol. 11, no. 22, p. 11035, Nov. 2021.

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[26] (2019). *Peking University International Competition on Ocular Disease Intelligent Recognition (ODIR-2019)*. Accessed: Mar. 29, 2022. [Online]. Available: https://odir2019.grand-challenge.org/

[27] 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.

[28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[29] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[32] 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.

[33] O. C. Devecioglu, J. Malik, T. Ince, S. Kiranyaz, E. Atalay, and M. Gabbouj, "Real-time glaucoma detection from digital fundus images using self-ONNs," *IEEE Access*, vol. 9, pp. 140031–140041, 2021, doi: 10.1109/ACCESS.2021.3118102.

SANLI YI was born in Yueyang, Hunan, in 1977. She received the Ph.D. degree from Central South University, in 2011. Her current research interest includes medical image processing investigate.

LINGXIANG ZHOU was born in 1998. He received the B.S. degree from the Zhejiang University of Technology, in 2020. He is currently pursuing the master's degree with the Kunming University of Science and Technology. His research interests include medical image processing and transfer learning.

LEI MA was born in Kunming, Yunnan, in 1978. He received the master's degree from Monash University, Australia, in 2004. His current research interest includes medical image processing investigate.

DANGGUO SHAO was born in 1979. He received the Ph.D. degree from Sichuan University, in 2012. His current research interest includes medical image processing Investigate.

• • •