Hindawi
International Journal of Intelligent Systems
Volume 2023, Article ID 1587410, 13 pages
https://doi.org/10.1155/2023/1587410

Research Article

Automatic Diagnosis of Different Types of Retinal Vein Occlusion Based on Fundus Images

Cheng Wan,[1] **Rongrong Hua**,[1] **Kunke Li**,[2] **Xiangqian Hong**,[2] **Dong Fang**,[2] **and Weihua Yang**[2]

[1]*College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China*
[2]*Shenzhen Eye Hospital, Jinan University, Shenzhen 518040, China*

Correspondence should be addressed to Weihua Yang; benben0606@139.com

Received 14 October 2022; Revised 8 August 2023; Accepted 31 August 2023; Published 8 September 2023

Academic Editor: Alexander Hošovský

Retinal vein occlusion (RVO) is the second common cause of blindness following diabetic retinopathy. The manual screening of fundus images to detect RVO is time consuming. Deep-learning techniques have been used for screening RVO due to their outstanding performance in many applications. However, unlike other images, medical images have smaller lesions, which require a more elaborate approach. To provide patients with an accurate diagnosis, followed by timely and effective treatment, we developed an intelligent method for automatic RVO screening on fundus images. Swin Transformer learns the hierarchy of low-to high-level features like the convolutional neural network. However, Swin Transformer extracts features from fundus images through attention modules, which pay more attention to the interrelationship between the features and each other. The model is more universal, does not rely entirely on the data itself, and focuses not only on local information but has a diffusion mechanism from local to global. To suppress overfitting, we adopt a regularization strategy, label smoothing, which uses one-hot to add noise to reduce the weight of the categories of true sample labels when calculating the loss function. The choice of different models using a 5-fold cross-validation on our own datasets indicates that Swin Transformer performs better. The accuracy of classifying all datasets is 98.75 ± 0.000, and the accuracy of identifying MRVO, CRVO, BRVO, and normal, using the method proposed in the paper, is 94.49 ± 0.094, 99.98 ± 0.015, 98.88 ± 0.08, and 99.42 ± 0.012, respectively. The method will be useful to diagnose RVO and help decide grade through fundus images, which has the potency to provide patients with further diagnosis and treatment.

1. Introduction

Retinal vein occlusion (RVO) is a common retinal vascular disorder with an incidence of 0.86%–1.63% and increases with age [1, 2]. The results of Chinese epidemiological surveys show that the incidence of RVO in people aged 40–49, 50–59, and 60–69 is 0.3%, 1.3%, and 2.1%, respectively [3]. Fundus images in patients with RVO show retinal vein filling, proximal vascular occlusion, and distal vasodilation. Depending on the area of the lesion, RVO is divided into macular retinal vein occlusion (MRVO) [4], central retinal vein occlusion (CRVO) [5], and branch retinal vein occlusion (BRVO) [6, 7]; BRVO is more common among these [8]. Patients often miss the opportunity for timely and effective treatment because of the mild or unnoticeable symptoms as well as problems, such as time-consuming, laborious manual identification, too strong subjectivity, and no guarantee of accuracy, which may cause permanent and irreversible vision loss [9]. Moreover, lack of specialized ophthalmologists, particularly in remote areas, is an important issue. To solve the above problems, we urgently need to build a screening system model for RVO, which can quickly and accurately identify diseased fundus images and determine the grade of RVO so that patients can grasp the best time for treatment.

Different types of RVO behave somewhat differently in the lesion area. CRVO can form flaming, bleeding, and retinal edema. If CRVO is not treated, there may be significant macular edema or peripheral hard exudation [10, 11]. The lesion area of BRVO is triangular, and the tip

points to the point of obstruction; the retinal vein dilates and twists in the distal vascular distribution area of the obstruction point. The retinal flame-like hemorrhage, retinal edema, and cotton velvet spots become visible. In addition to the manifestation of vascular bleeding [12, 13], MRVO also occurs in the macular area that mostly affects vision. The vascular changes in MRVO are not obvious and primarily seen as bleeding in the macular area, which can easily be confused with a disease that can cause bleeding. We refine RVO into three categories because of the varying specific conditions and vision damage of RVO. Thus, for timely and accurate treatment, specific categories of RVO must be diagnosed.

With the improvement in the computing power of computer and rapid development of machine-learning and deep-learning (DL) algorithms, the development of artificial intelligence (AI) technology has been promoted and applied to various industries. For example, the application of AI technology in the medical field is primarily to guide diagnosis and for treatment plan selection, risk prediction, and reduction of medication errors. Ophthalmic diagnosis relies heavily on imaging examination, and AI methods based on DL can quickly and noninvasively analyze fundus image information. They can identify, locate, and quantify disease characteristics for purposes of screening, diagnosing, grading, and guiding the treatment of diseases. Presently, AI methods based on DL have been widely used in eye diseases such as diabetic retinopathy, glaucoma, and cataract. For example, Takahashi et al. [14] used a modified GoogLeNet DL neural network to grade 4,907 posterior polar photographs with an accuracy rate of 96%. Devalla et al. [15] used an eight-layer CNN that was composed of three convolution layers, three max-pooling layers, and two fully connected layers which can digitally stain an optic disc, retinal pigment epithelium, and choroid and sclera around the optic disc and automatically measure their structural parameters. For all tissues, the dice coefficient, sensitivity, specificity, IU, and accuracy (mean) were 0.84 ± 0.03,0.92 ± 0.03,0.99 ± 0.00, 0.89 ± 0.03, and 0.94 ± 0.02, respectively. Xu et al. [16] proposed to use the global-local convolutional neural network (CNN) automatic classifier to identify and classify a total of 1200 cataract fundus images as normal, mild, moderate, and severe; they used the sharpness of blood vessels and the optic disc as a reference, with an average accuracy of 81.86%, and a deconvolutional neural network to visualize the layer-by-layer characterization of cataracts from the middle layer feature transformation using CNN. In addition, the team of Professor Liu Yizhi of Sun Yat-sen University used DL algorithms to establish a congenital cataract AI diagnosis and treatment platform [17], successfully applied it to the clinic, and showed that the application of AI technology to ophthalmic clinics has excellent prospects.

DL methods have been widely used to diagnose different fundus diseases. For the diagnosis of RVO, Anitha et al. [18] used Kohomen to classify a total of 420 nonvalue-added diabetic retinopathy, CRVO, central serous choroid retinopathy, and central neovascular membrane; the average accuracy, sensitivity, and specificity obtained by using this method were $97.7\% \pm 0.8\%$, 96%, and 98%, respectively. The performance of this method is better; however, histogram equalization, median filtering preprocessing of the images, and texture-based feature extraction made the overall process relatively complex. Nagasato et al. [19] used VGG-16 to classify ultrawide field fundus images, including 237 BRVO fundus images and 176 healthy fundus images, and compared the results; the results showed that the sensitivity, specificity, positive predictive values, negative predictive values, and area under the receiver operating curve (AUC) of BRVO were 94.0%, 97.0%, 96.5%, 93.2%, and 0.976, respectively, using a deep CNN model. But this is only classification recognition of normal fundus images and BRVO fundus images. There have been some studies on RVO; however, only one of the CRVO and BRVO is mentioned, and MRVO is rarely mentioned. In addition, most of the diagnoses of fundus diseases use traditional machine-learning methods and CNN, which have the image feature extraction problem and significantly long network training time. In 2021, Liu et al. [20] proposed Swin Transformer. Swin Transformer achieves the state-of-the-art performance on COCO object detection and ADE20K semantic segmentation, which shows that it performs better in feature extraction. Zhao et al. [21] used Swin Transformer and achieved a mean average precision of 0.934, presenting an efficient and intelligent mutton multipart classification method. Zhao et al. [22] proposed the first Swin Transformer-based mosquito species identification model with a 99.04% accuracy and a 99.16% F1 score. Liu et al. [23] used depthwise separable convolutional Swin Transformer to classify cervical ultrasound lymph-node-level and achieved average accuracy, precision, sensitivity, specificity, and F1 values of the model which were 80.65%, 80.68%, 78.73%, 95.99%, and 79.42%, respectively. Given the present issues, this paper uses the Swin Transformer model combined with the label smoothing method to process the fundus images containing normal, MRVO, CRVO, and BRVO and train a model that can accurately identify diseased fundus images. The main contributions are as follows.

The grading method of retinal vein occlusion based on Swin Transformer is introduced, and there is no need for artificial feature extraction of fundus images in the whole process.

We propose a fine-tuning method to adapt to the pretraining model of retinal fundus images, which is different from the traditional convolutional neural network, in that the model is mainly composed of multihead attention modules, which is more targeted for the extraction of lesion features of fundus images. To prevent the model from predicting labels too confidently when training, we use label smoothing to improve the generalization ability of the model. For the trained model, Grad-CAM was used to visually interpret its lesion area and analyze fundus images with incorrect predictions.

The datasets we used were collected and annotated by the team of professional ophthalmologists, providing strong data support for the entire experiment.

2. Methods

2.1. Data Collection. All data used during the study were from Shenzhen Eye Hospital, and it contained a total of 805 color fundus images after rejecting poor quality, incorrect, and unclear images. These images contained a variety of resolutions from 1024×1024 to 2976×2976, and all were annotated by professional doctors.

Because of the varying resolution sizes and the small number of datasets, the datasets were preprocessed to accommodate model training. The resolution of all fundus images was uniformly modified to 224×224. Then, the images were randomly rotated horizontally, at a given probability, for the purpose of data expansion. Finally, the images were normalized. Image normalization is the centralization of data by de-mean; i.e., the pixels in the image are adjusted to a distribution with an average of 0 and a variance of 1. According to the convex optimization theory and data probability distribution, data centering conforms to the law of data distribution, which makes the network obtain good convergence after training, and it is easier to obtain the generalization effect after training.

The datasets were divided into three parts: training set, validation set, and test set. The training set is a data sample used for model fitting that degrades the error in the training process and learns the weight parameters that can be trained. The validation set is a set of samples left separately during the model training process, which can be used to adjust hyperparameters, such as the learning rate, iteration, batch size, and weights of each part of the loss function, and preliminary evaluate the model. The test set is used to evaluate the generalization ability of the final model but cannot be used as the basis for algorithm-related selection such as tuning parameters and selecting features. A total of 483 fundus images were used as the training set. 161 fundus images were in the validation set, and 161 fundus images were in the test set. The training, validation, and test sets all contained four categories of images that were randomly assigned by setting a random seed. The specific allocation of the dataset is presented in Table 1.

2.2. Model Architecture. In this study, we use the Swin Transformer [20] model for diagnosing different types of RVO. Swin Transformer encodes the original image to obtain pixel features, using a hierarchical construction approach similar to that found in CNN. It uses window multihead self-attention (W-MSA), which can divide feature maps into multiple disjoint regions (windows) and multihead self-attention only within each window; this can reduce the amount of computation; however, it will also isolate information transfer between different windows. Thus, the concept of shifted windows-multihead self-attention (SW-MSA) is proposed, through which information can be transmitted in adjacent windows.

The entire Swin Transformer structure comprises one patch partition, one linear embedding, one layer normalization, one global pooling, one fully connected layer, three patch mergings, and 12 Swin Transformer blocks, as shown in Figure 1(a). First, the image was entered into the patch partition module for chunking; i.e., every 4×4 adjacent pixel is a patch and then flattened in the channel direction. The datasets used in the study were RGB three-channel images, and each pixel has three values of R, G, and B; thus, after flattening, there were 48 channels, and after the patch partition, the image shape was transformed from (H, W, 3) to (H/4, W/4, 48). The linear embedding layer then performed a linear transformation of the channel data for each pixel. Then, different-sized feature maps were constructed through the stages, except for linear embedding in stage 1. The remaining three stages were first downsampled through a patch merging layer, and then, they were all stacked repeatedly with the Swin Transformer block. The block has two structures, as shown in Figure 1(b). They consist of MLP, LN, W-MSA, and SW-MSA. The MLP block consists of fully connections, activation function (GELU), and dropouts. LN is a layer normalization, which can normalize each token. The only difference between the two structures is that one uses a W-MSA structure and the other uses a SW-MSA structure. In addition, the two structures are used in pairs, using a W-MSA structure and then a SW-MSA structure. Because the study is a classification task, the output includes a norm layer, a global pooling layer, and a fully connected layer. The main module of Swin Transformer is multihead self-attention (MSA). Firstly, a_i obtains q^i, k^i, and v_i through the three transformation matrices W^q, W^k, and W^v, respectively. Then, according to the number of heads used h, obtained q^i, k^i, v^i are further divided into h parts. As shown in Figure 2, assuming that $h = 2$, and q^1 is split into $q^{1,1}$ and $q^{1,2}$, then $q^{1,1}$ belongs to head1 and $q^{1,2}$ belongs to head2.The specific calculation process is as follows:

$$\begin{aligned} \text{Multihead}(Q, K, V) &= \text{Concat}\left(\text{head}_1, \ldots, \text{head}_h\right) W^o \\ \text{head}_i &= \text{attention}\left(\text{QW}_i^Q, \text{KW}_i^K, \text{VW}_i^V\right) = \text{softmax}\left(\frac{Q_i W_i^Q K_i W_i^K}{\sqrt{d_k}}\right) V^i W_i^V, \end{aligned} \tag{1}$$

where the projections are parameter matrices $W_i^Q \in R^{d_{\text{model}} \times d_k}$, $W_i^k \in R^{d_{\text{model}} \times d_k}$, $W_i^V \in R^{d_{\text{model}} \times d_k}$, and $W^o \in R^{hd_v \times d_{\text{model}}}$ ($d_k = d_v = (d_{\text{model}}/h)$, d_{model} is the size of the fusion of $q^{i,1}, \ldots, q^{i,j}$).

Experimental hardware configuration is as follows: Intel(R) Core (TM) i7-6700, CPU @ 3.40 GHz, and GPU NVIDIA GeForce RTX 1080. Experimental software configuration is as follows: 64 bit Windows 10 operating system

Table 1: Dataset allocation.

	Train	Valid	Test	Total
Normal	155	52	52	259
MRVO	50	16	16	82
CRVO	82	28	28	138
BRVO	196	65	65	326
Total	483	161	161	805

The dataset is divided into the training set, validation set, and test set.

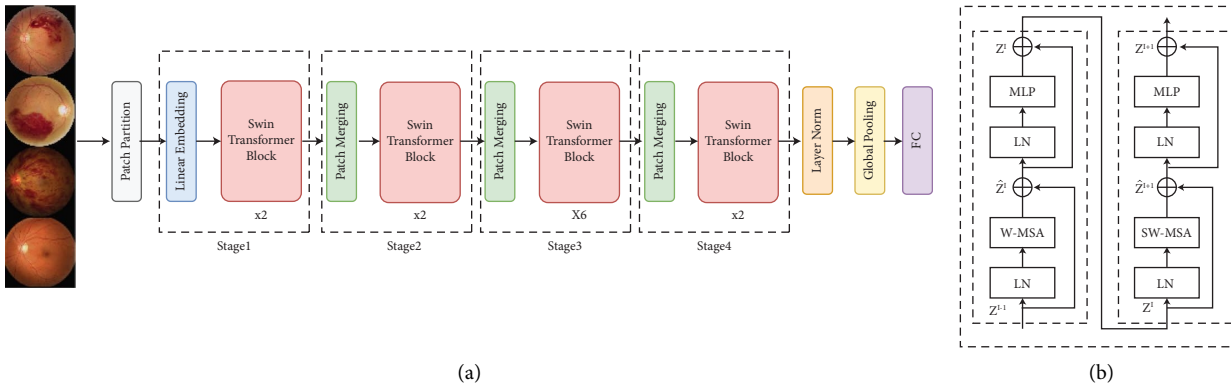

Figure 1: (a) Architecture of Swin Transformer; (b) two successive Swin Transformer blocks.

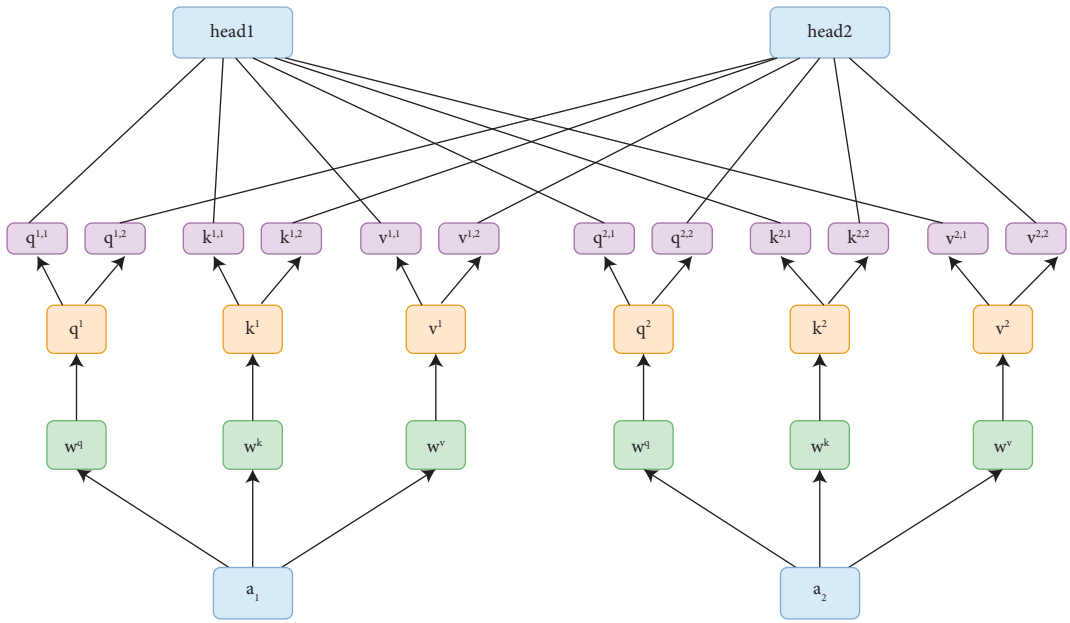

Figure 2: Multihead self-attention module.

and PyCharm Community Edition 2021.3, Python 3.6.13. Because there are too few samples for training, overfitting is easy; i.e., the model performs better on the training samples, but the generalization effect on the test set is unsatisfactory. To alleviate this phenomenon, we use different methods to enhance the dataset, primarily including random horizontal and vertical flipping and arbitrary direction rotation [24]. During training, the loss function takes cross-entropy loss, using AdamW as the optimizer, and weight decay is set to 0.05; we use a transfer-learning approach that uses ImageNet-based pretrained weights. After loading the pretrained [25] weights, we froze all the parameters, except the last layer, which not only improved the training speed but also alleviated the model overfitting phenomenon. In addition, to further alleviate this, the label smoothing method was used in the process of training the model. The entire training process iterated a total of 100 epochs, and the initial learning rate was set to 0.0001.

For comparison, deep CNNs, VGG-16, VGG-19, MobileNet-v2, ResNet-18, ResNet-50, WP-CNN-105, and DenseNet-121 were also used during the study. VGG [26] is a classic classification network. The ResNet [27] series model was proposed in 2015. Because its structure uses residual structures, the image features acquired at the shallow level of the network are superimposed into the deep network. This can alleviate the problem of network degradation (as the number of layers of the network deepens, the effect will deteriorate) and make it perform well in image classification; it won first place in the classification task in the 2015 ImageNet competition. In contrast to the previous deep CNN, the ResNet series network is not just a simple stack of convolutional layers, which is one of the reasons for choosing the ResNet [28] series network. MobileNet-v2 [29] was selected as the network for comparison because its network structure model is small and its computation speed is fast. Guo et al. [30] used MobileNet-v2 to predict different fundus diseases, which performed well. WP-CNN-105 [31] performs well on the grading of diabetic retinopathy. DenseNet-121 [32] is a densely connected network that extracts image features well. To verify generalization, we used 5-fold cross-validation for different models, as shown in Figure 3. Because of the similar hardware configuration and software configuration of the comparison network, all models adopt a transfer-learning [33] approach during training. The total training epochs of different models are 100. The selected optimizer is Adam. The loss function is cross-entropy loss, and the initialization learning rate is set to 0.00001.

2.3. Label Smoothing. In regression problems or classification problems, the loss function is used to measure the difference between the predicted value and true value, and the resulting difference is also called "punishment." For regression models using neural networks, the most commonly used loss function is the mean squared error, where y_i is the true value of the i_{th} data. y_i' is the predicted value of the i_{th} data, and n is the total number of data:

$$L(y, y') = \frac{\sum_i (y_i - y_i')^2}{n}. \tag{2}$$

For classification models, because the final output results in a predicted probability for each class, the output layer generally uses the sigmoid function (for binary classification) or the softmax function (for multiclassification), and the loss function uses a combination of cross-entropy methods. The final output of both the sigmoid function and the softmax function is a value between 0 and 1 and can be used to represent the probability of the classification. The cross-entropy loss function is similar to entropy in information theory:

$$L(y, y') = -\sum_i y_i \log(y_i'). \tag{3}$$

The closer the predicted value is to the true value, the smaller is the loss function of cross-entropy. However, the use of one-hot encoding has the disadvantage of making the network heavily rely on training samples during training, resulting in poor performance robustness. Artificially reducing the probability of the correct value of the sample label and increasing the probability of the wrong value (such as changing the label in the above example to [0.1, 0.9]) can help train the model, further improve the prediction ability, and avoid extreme cases. This method is called label smoothing, and in many classification models, it performs better than the original. In 2015, Szegedy et al. [34] proposed label smoothing regularization (LSR), and LSR achieves the desired goal of preventing the largest logit from becoming much larger than all others. LSR would result in a large cross-entropy loss.

2.4. Statistical Analysis. This is a multiclassification problem; thus, two evaluation criteria are selected during the evaluation process. The first is to convert a multiclass problem into a binary classification problem and then evaluate the model using the evaluation criteria sensitivity (SE), specificity (SP), precision (P), accuracy (AC), F1 score, and AUC in the binary classification [35]. The second is to calculate the kappa coefficient as the multiclassification evaluation criteria.

The classification categories in the study include four types: normal, MRVO, CRVO, and BRVO. When using the biclass evaluation criteria, positive and negative samples should first be determined. Taking BRVO as an example, when calculating its relevant evaluation indicators, BRVO is regarded as positive samples, and all other categories are automatically regarded as negative samples. True negative (TN) is a predicted negative sample but actually a negative sample; false positive (FP) is a predicted positive sample but actually a negative sample; true positive (TP) is a predicted positive sample but actually a positive sample; false negative (FN) is a predicted negative sample but actually a positive sample. After understanding all concepts, we introduce the SE, SP, P, AC, and F1 values of the evaluation criteria.

SE demonstrates the number of samples predicted to be positive in the total of all actual positive samples, and the higher the SE, the stronger is the ability to recognize the positive samples:

$$\text{SE} = \frac{\text{TP}}{(\text{TP} + \text{FN})}. \tag{4}$$

SP is how many of the samples predicted to be negative are negative samples, indicating the ability to recognize negative samples:

$$\text{SP} = \frac{\text{TN}}{(\text{TN} + \text{FP})}. \tag{5}$$

P is the probability that total samples are correctly judged to be positive for all judged positive samples:

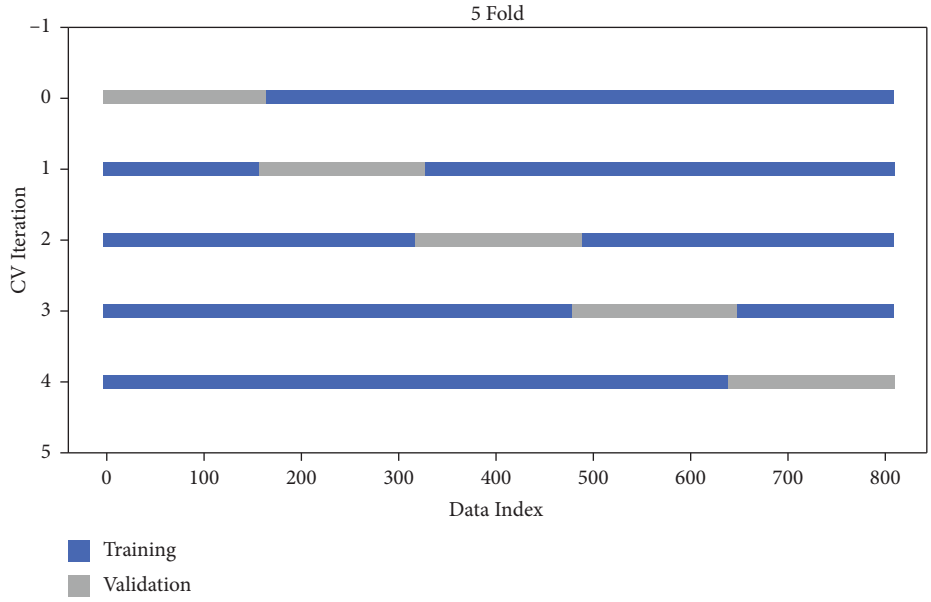

Figure 3: The process of 5-fold cross-validation.

$$P = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FP})}. \tag{6}$$

AC is the ratio of the correct number that is predicted by the model to the whole, and the higher the accuracy rate, the better is the model:

$$\mathrm{Ac} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}. \tag{7}$$

The F1 value is the harmonization value of P and SE, and to balance P and SE evaluation, simply increasing one side does not increase the F1 value:

$$F1 = \frac{2 * P * \mathrm{SE}}{\mathrm{SE} + P}. \tag{8}$$

AUC is defined as the area under the ROC curve enclosed by the coordinate axis. The ROC curve is above the line of $y = x$, so the value range of AUC is [0.5, 1.0]. The closer the AUC is to 1.0, the higher is the accuracy of the model prediction.

The kappa [36] coefficient is a method of assessing consistency in statistics, and its value range is [0, 1]. When evaluating a multiclassification model, the higher the Kappa coefficient, the higher the classification accuracy of the model. The kappa coefficient is calculated as follows:

$$\begin{aligned} k &= \frac{p_0 - p_e}{1 - p_e}, \\ p_e &= \frac{a_1.b_1 + a_2.b_2 + \cdots + a_c.b_c}{n.n}, \end{aligned} \tag{9}$$

where p_0 represents the classification accuracy, a_i represents the number of true samples of class i, and b_i represents the number of samples predicted in class i.

2.5. Diagnosis Visualization. To intuitively analyze the influence of each region on the fundus image classification results, we determine whether the lesion area concerned by the classification model is consistent with the medically identified one, i.e., determine whether the model has learned the correct features or information so that the model recognition results can be better analyzed. We use gradient-weighted class activation mapping (Grad-CAM) to visualize Swin Transformer.

We used heat maps drawn by Grad-CAM [37] to visualize the areas of interest of the model. When the network is forward propagated, it obtains the feature layer and the predicted value. The feature layer is the result of the network extracting the features of the original image. The deeper the feature layer, the richer the semantic information. Because the feature layer contains semantic information for all classes, it is necessary to backpropagate the predicted value to obtain the gradient information of the feature layer. The gradient information represents the contribution of each element in the feature layer to the predicted value of a certain category. Finally, a weighted sum can be performed to obtain a heat map through ReLU, in which the larger the contribution area, the warmer the color. The specific implementation of Grad-CAM is shown in (10) and (11):

$$L^{C}_{\mathrm{Grad-CAM}} = \mathrm{ReLU}\left(\sum_k \alpha^c_k A^k\right), \tag{10}$$

$$\alpha^c_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A^k_{\mathrm{ij}}}, \tag{11}$$

where A represents a feature layer, k represents the k_{th} channel in feature layer A, c represents category c, A^k represents the data of channel k in feature layer A, α^c_k

represents the weight for A^k, y^c represents the score predicted by the network for category c, and A^k_{ij} represents the data of feature layer A at coordinate i and j in channel k, and Z is equal to the width of the feature layer x height.

3. Results

The Swin Transformer model is used in the study, and other models used for comparison were trained and verified with 483 and 161 fundus images, respectively. All models used label smoothing to avoid overfitting. The validation set is used to adjust hyperparameters and make a preliminary assessment of the model. Thus, we use the accuracy of the validation set to make a preliminary evaluation of the model, and the results can be obtained according to the experimental details in Section 2.2, as shown in Table 2. Ablation of the label smoothing approach on the task is reported in Table 3. Swin Transformer with label smoothing outperforms the counterpart by 0.5%, which indicates the effectiveness of using label smoothing. The analysis of different networks' parameters is shown in Table 4.

The trained Swin Transformer and comparable models predicted a test set containing 161 fundus images, and the model could successfully identify fundus images as normal, MRVO, CRVO, and BRVO. The specifics of each category identification are shown in the confusion matrix of Figure 4—the horizontal axis represents the true label and the vertical axis represents the predicted label. In 161 fundus images used for testing, Swin Transformer only identified two images wrongly.

To compare the identification performance of different models, MRVO, CRVO, BRVO, and normal through the test set and the specific results are shown in Tables 5–8, respectively. For the identification of MRVO fundus images, the results in Table 5 show that the Swin Transformer model performs better on most indicators and that the SE, SP, P, and F1 values are higher than those of comparable models and reach more than 90%; specifically, the SP value is $99.98\% \pm 0.017$. For the identification of CRVO, the results are shown in Table 6, DenseNet-121 performs best in precision, and the indicators AC, SE, SP, and F1 acquired by Swin Transformer had better performance than those of comparable models. For the identification of BRVO, the results are shown in Table 7, the AC, SE, SP, P, and F1 values of Swin Transformer were $98.88\% \pm 0.080$, $98.55\% \pm 0.056$, $99.04\% \pm 0.041$, $98.56\% \pm 0.066$, and $98.56\% \pm 0.068$, respectively, and it is evident that the values of each index reached more than 95%, which shows better performance than those models used for comparison, except MobileNet-v2 in sensitivity. All models performed well in detecting normal fundus images, and the results are shown in Table 8. All things considered, Swin Transformer performed better in grading RVO. Figure 5 shows the AUC values of the Swin Transformer model in identifying MRVO, CRVO, BRVO, and normal of 0.9991, 1.0000, 0.9992, and 1.0000, respectively.

In addition to converting a multiclassification problem into multiple binary classification problems and using the indicators of binary classification to evaluate the model [38], we also used the evaluation indicator that can directly implement the multiclassification model, namely, the kappa coefficient. The kappa coefficient is a method of evaluating data consistency in statistics; the higher the value of the kappa coefficient, the better the model performs in multiclassification. The kappa coefficient obtained by different models by predicting the test set is presented in Table 9, and it is evident that the kappa coefficient of Swin Transformer is higher than those of other models.

We use Grad-CAM to visualize different models in identifying areas of concern, especially when identifying different types of RVO, to demonstrate the interpretability of the model [39]. As shown in Figure 6, from up to down are BRVO, CRVO, and MRVO, respectively. The first line is the original fundus images, and the other lines are the corresponding heat maps of different models; the warmer the color of the heat map, the greater the role it plays in the classification process. From the heat map, we can see that the lesion area of MRVO is mainly distributed near the macula. The bleeding spots in the lesion area of CRVO are irregular in shape and diffusely flame-like. The lesion area of BRVO is more concentrated than that of CRVO. Comparing the heat maps obtained from fewer models, we found that the Swin Transformer model extracted more accurate lesion features when grading retinal vein occlusion.

Table 2: Average valid accuracy of different models.

Models	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
VGG-16	91.25	90.04	91.95	92.25	92.30	91.56
VGG-19	92.45	92.40	93.01	92.10	91.98	92.39
MobileNet-v2	93.75	93.80	93.65	93.82	93.70	93.74
ReaNet-18	94.98	94.95	95.01	95.02	95.08	95.01
ResNet-50	96.27	96.25	96.30	96.26	96.27	96.27
WP-CNN-105	95.60	95.75	95.50	95.63	95.63	95.62
DenseNet-121	97.05	97.13	97.10	97.08	97.09	97.09
Swin Transformer	98.20	98.25	98.23	98.27	98.25	**98.25**

Bold represents the best performance.

Table 3: Ablation study on label smoothing.

	Swin Transformer	Swin Transformer with label smoothing
Accuracy	98.25	98.75

Table 4: The analysis of different networks' parameters.

	Prams (million)
VGG-16	138.26
VGG-19	142.67
MobileNet-v2	3.51
ResNet-18	11.69
ResNet-50	25.56
WP-CNN-105	13.27
DenseNet-121	7.98
Swin Transformer	28.27

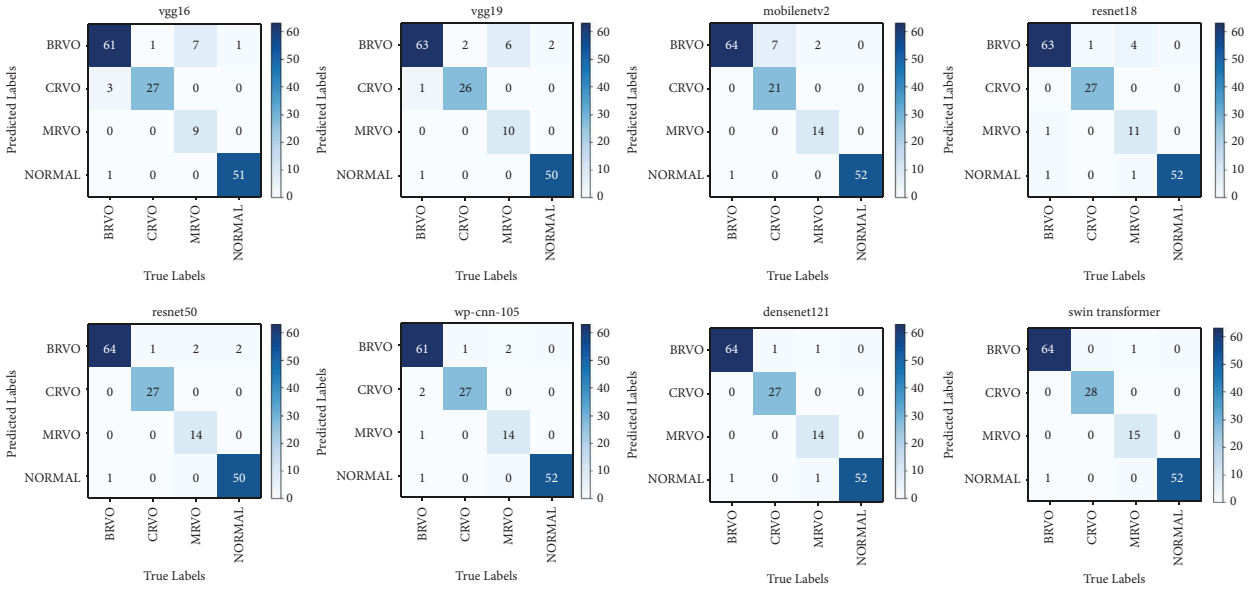

Figure 4: Confusion matrix obtained by different models predicting the test set.

Table 5: Accuracy, sensitivity, specificity, precision, and F1 value for identifying MRVO by different models.

Models	Accuracy ± std	Sensitivity ± std	Specificity ± std	Precision ± std	F1 ± std
VGG-16	95.71 ± 0.026	56.24 ± 0.047	99.91 ± 0.091	99.94 ± 0.056	72.05 ± 0.052
VGG-19	96.33 ± 0.035	62.53 ± 0.032	99.93 ± 0.067	99.92 ± 0.076	76.93 ± 0.065
MobileNet-v2	**98.85 ± 0.053**	87.52 ± 0.042	99.95 ± 0.048	99.91 ± 0.088	93.33 ± 0.035
ResNet-18	96.36 ± 0.064	68.84 ± 0.048	99.39 ± 0.095	91.75 ± 0.054	78.63 ± 0.036
ResNet-50	98.82 ± 0.047	87.55 ± 0.056	99.96 ± 0.036	99.96 ± 0.039	93.32 ± 0.028
WP-CNN-105	98.14 ± 0.041	87.57 ± 0.075	99.38 ± 0.086	99.35 ± 0.064	90.32 ± 0.035
DenseNet-121	98.80 ± 0.032	87.59 ± 0.092	99.97 ± 0.029	99.89 ± 0.110	93.34 ± 0.042
Swin Transformer	94.49 ± 0.094	**93.89 ± 0.095**	**99.98 ± 0.017**	**99.97 ± 0.026**	**96.81 ± 0.084**

Bold represents the best performance.

Table 6: Accuracy, sensitivity, specificity, precision, and F1 value for identifying CRVO by different models.

Models	Accuracy ± std	Sensitivity ± std	Specificity ± std	Precision ± std	F1 ± std
VGG-16	97.53 ± 0.055	96.46 ± 0.061	97.71 ± 0.015	90.16 ± 0.064	93.12 ± 0.031
VGG-19	98.12 ± 0.023	92.93 ± 0.032	99.99 ± 0.001	96.32 ± 0.027	94.66 ± 0.082
MobileNet-v2	95.73 ± 0.031	75.00 ± 0.0001	99.98 ± 0.015	99.92 ± 0.072	85.72 ± 0.044
ResNet-18	99.42 ± 0.025	96.42 ± 0.026	99.95 ± 0.048	99.81 ± 0.089	98.23 ± 0.034
ResNet-50	99.45 ± 0.052	87.55 ± 0.053	98.57 ± 0.083	99.91 ± 0.085	98.26 ± 0.063
WP-CNN-105	99.81 ± 0.017	96.47 ± 0.074	99.86 ± 0.065	93.64 ± 0.042	94.72 ± 0.028
DenseNet-121	99.47 ± 0.074	96.41 ± 0.028	99.75 ± 0.057	**99.95 ± 0.043**	98.27 ± 0.075
Swin Transformer	**99.98 ± 0.015**	**99.97 ± 0.016**	**99.99 ± 0.006**	99.73 ± 0.062	**99.99 ± 0.006**

Bold represents the best performance.

Table 7: Accuracy, sensitivity, specificity, precision, and F1 value for identifying BRVO by different models.

Models	Accuracy ± std	Sensitivity ± std	Specificity ± std	Precision ± std	F1 ± std
VGG-16	91.93 ± 0.031	93.84 ± 0.044	90.62 ± 0.025	87.13 ± 0.034	90.32 ± 0.035
VGG-19	92.55 ± 0.055	96.96 ± 0.064	99.22 ± 0.023	86.33 ± 0.032	91.30 ± 0.031
MobileNet-v2	93.84 ± 0.052	**98.57 ± 0.075**	90.67 ± 0.072	87.74 ± 0.045	92.81 ± 0.014
ResNet-18	95.77 ± 0.071	96.95 ± 0.059	94.88 ± 0.085	92.68 ± 0.081	94.70 ± 0.011
ResNet-50	96.33 ± 0.034	98.53 ± 0.032	94.83 ± 0.036	92.87 ± 0.072	95.65 ± 0.056
WP-CNN-105	95.74 ± 0.041	93.83 ± 0.033	96.96 ± 0.062	95.35 ± 0.057	94.54 ± 0.043
DenseNet-121	98.16 ± 0.062	98.54 ± 0.049	97.94 ± 0.051	97.06 ± 0.064	97.78 ± 0.082
Swin Transformer	**98.88 ± 0.080**	98.55 ± 0.056	**99.04 ± 0.041**	**98.56 ± 0.066**	**98.56 ± 0.068**

Bold represents the best performance.

Table 8: Accuracy, sensitivity, specificity, precision, and F1 value for identifying normal by different models.

Models	Accuracy ± std	Sensitivity ± std	Specificity ± std	Precision ± std	F1 ± std
VGG-16	98.81 ± 0.028	98.11 ± 0.041	99.12 ± 0.039	98.12 ± 0.035	98.24 ± 0.037
VGG-19	98.12 ± 0.031	96.21 ± 0.032	99.13 ± 0.021	98.00 ± 0.004	97.15 ± 0.052
MobileNet-v2	99.43 ± 0.026	99.97 ± 0.028	99.10 ± 0.054	98.14 ± 0.012	99.05 ± 0.055
ResNet-18	98.84 ± 0.041	99.96 ± 0.035	98.21 ± 0.045	96.32 ± 0.038	98.12 ± 0.032
ResNet-50	98.12 ± 0.035	96.21 ± 0.046	99.11 ± 0.048	98.02 ± 0.025	97.16 ± 0.064
WP-CNN-105	**99.44 ± 0.021**	99.98 ± 0.014	**99.18 ± 0.085**	98.13 ± 0.021	99.02 ± 0.023
DenseNet-121	98.81 ± 0.048	99.95 ± 0.037	98.24 ± 0.014	96.34 ± 0.042	98.18 ± 0.078
Swin Transformer	99.42 ± 0.012	**99.99 ± 0.0001**	99.12 ± 0.031	**98.19 ± 0.065**	**99.19 ± 0.085**

Bold represents the best performance.

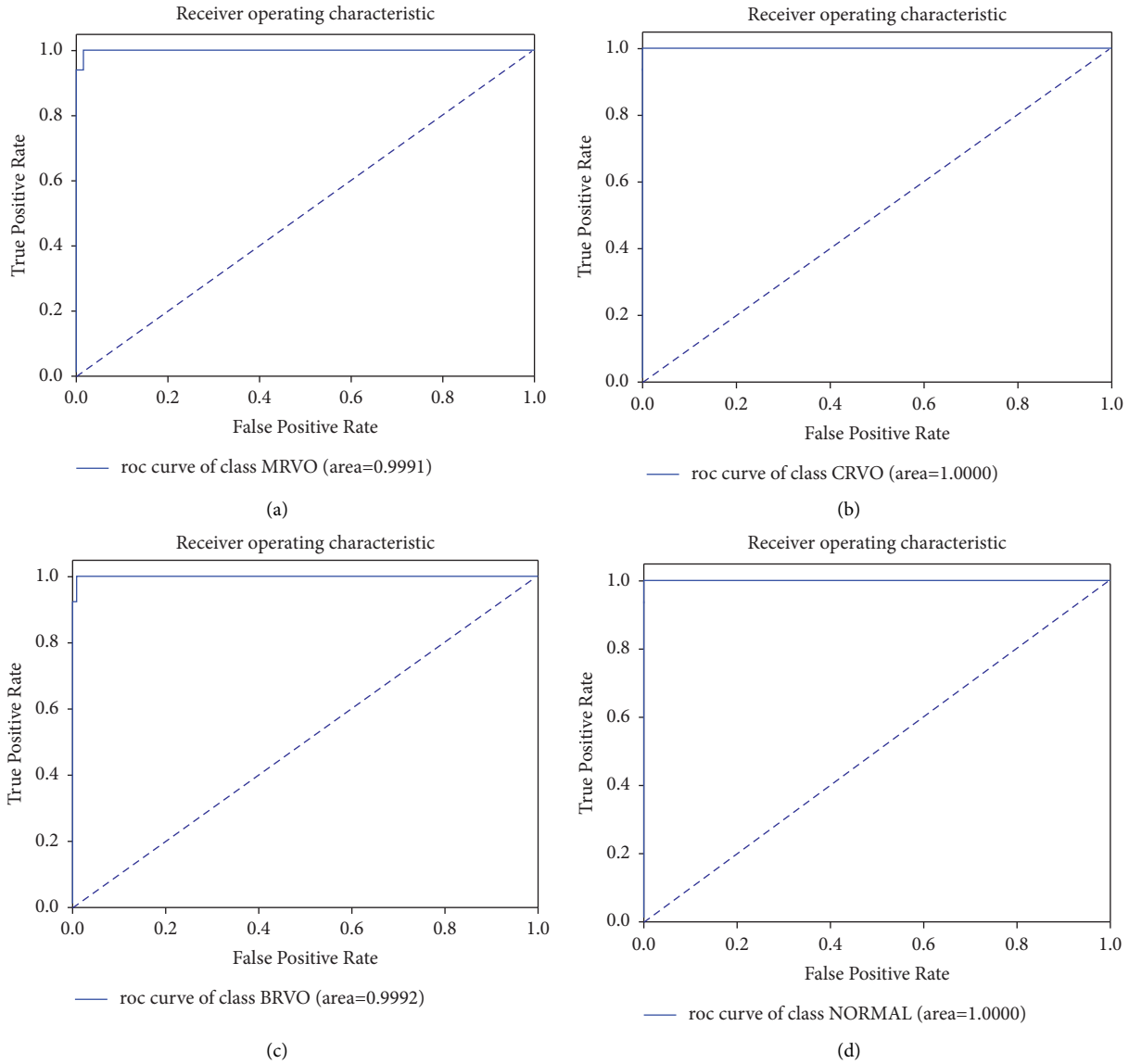

Figure 5: (a–d): ROC curves of MRVO, CRVO, BRVO, and normal, respectively.

4. Discussion

We use the Swin Transformer model in this study to intelligently diagnose fundus images containing normal, MRVO, CRVO, and BRVO. Presently, the combination of image processing and medical images is becoming highly extensive [40, 41] because image processing plays a significant role in medical diagnosis [42, 43]. In this study, for intelligent diagnosis of RVO, we mainly attempted to use Swin Transformer, VGG-16, VGG-19, MobileNetV2,

Table 9: Kappa coefficients for different models on the test set.

Models	Kappa
VGG-16	0.8816 ± 0.0012
VGG-19	0.8898 ± 0.0014
MobileNet-v2	0.9086 ± 0.0008
ResNet-18	0.9274 ± 0.0006
ResNet-50	0.9457 ± 0.0010
WP-CNN-105	0.9372 ± 0.0011
DenseNet-121	0.9639 ± 0.0004
Swin Transformer	**0.9820 ± 0.0002**

Bold represents the best performance.

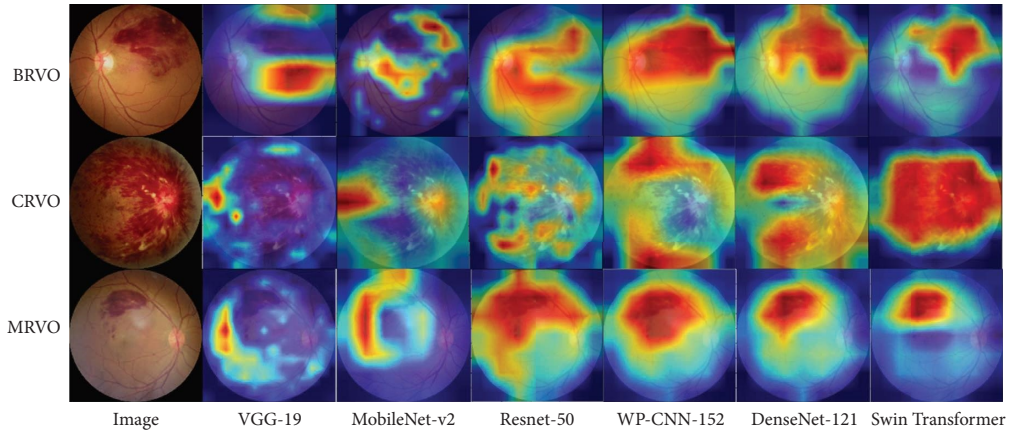

Figure 6: The original fundus images of BRVO, CRVO, MRVO, and corresponding heat maps of different models.

ResNet-18, ResNet-50, WP-CNN-105, and DenseNet-121 to screen out fundus images with normal, MRVO, CRVO, and BRVO and then used Grad-CAM to visualize the retinopathy areas of different types of RVO, while demonstrating the efficiency of Swin Transformer in the classification process.

To diagnose RVO, a classification method of fundus images based on the Swin Transformer model was used, which divides fundus images into four categories: normal, MRVO, CRVO, and BRVO. When using Swin Transformer, we primarily focus on two aspects, the high accuracy of the model in the training process and the strong generalization ability of the model in the application process; i.e., the model must accurately identify the specific category of RVO through fundus images that have not been trained. This specific research has the following aspects: the dataset used in this study has a total of 805 fundus images, of which 483 are used to train the network; the dataset is preprocessed, such as image cropping and data expansion. Although data expansion has been performed, the overall amount of data is still very small, which is not enough to fully train the network; thus, the fine-tuning method of transfer learning is used to speed up the training speed. To further improve the recognition accuracy and generalization ability of the model, we also use the label smoothing method and 5-fold cross validation. For Swin Transformer, training on fundus images containing different categories is primarily a process of learning characteristics regarding the different categories of fundus images. Furthermore, we used the Grad-CAM method to visualize the lesion area with a heat map of the predicted image, and the visualization results show that it is consistent with the medically determined lesion area. Clinically, there is no known golden rule for the diagnosis of different types of RVO; thus, the diagnosis process will be affected by subjective cognition and the experience of ophthalmologists [44, 45]. Compared with ophthalmologists' diagnosis, the model proposed in this paper is not affected by subjective factors in the process of diagnosis, and it performs well in generalization. The visualization results in Figure 6 show that the Swin Transformer model focuses on the lesion area and is negligibly affected by the background when diagnosing RVO through fundus images and indicate that the Swin Transformer model is interpretable.

Our research has the following advantages: First, the Swin Transformer model used can automatically diagnose RVO through fundus images, and its diagnostic accuracy is higher than that of comparable models. In addition, it can process datasets automatically and efficiently without manual assistance. Second, our model extracts and predicts the morphological features of fundus images, which are not easily affected by subjective cognition and experience. As long as the diagnostic criteria are given, the prediction results of the model will always be consistent with the given diagnostic criteria. Third, we can not only diagnose RVO but also accurately judge its specific type, which has an important clinical significance in real life.

This study also has some limitations. First, the distribution of different categories is uneven. In the future, we

need larger datasets, a more balanced sample distribution, and more training data to improve the classification accuracy of the model. Second, the Swin Transformer model should be applied to evaluate the fundus images of different ethnic groups [46] to verify the robustness in diagnosing RVO. In this paper, because of the limitations of the dataset, our research objects are all Asians. Third, because of the differences of symptoms in different situations [47, 48], age, gender, family genetic history, and other factors should be considered when building the model to improve the accuracy of the algorithm for different categories of RVO. Finally, for the lesion area, the analysis of fundus images with wavelet transforms and the extraction of spectral characteristics of bleeding point areas and edges can further improve efficiency [49].

Abbreviations

RVO:	Retinal vein occlusion
MRVO:	Macular retinal vein occlusion
CRVO:	Central retinal vein occlusion
BRVO:	Branch retinal vein occlusion
DL:	Deep learning
AI:	Artificial intelligence
CNN:	Convolutional neural network
AUC:	Area under the receiver operating curve
W-MSA:	Windows multihead self-attention
SW-MSA:	Shifted windows-multihead self-attention
SE:	Sensitivity
SP:	Specificity
P:	Precision
AC:	Accuracy
TN:	True negative
FP:	False positive
TP:	True positive
FN:	False negative
Grad-CAM:	Gradient-weighted class activation mapping.

Data Availability

All data used during the study are available from the corresponding author upon request.

Disclosure

The preprint "Automatic Diagnosis of Different Types of Retinal Vein Occlusion Based on Fundus Images" [49] of this article is already published.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

CW, RH, and WY acquired, analyzed, and discussed the data and drafted the manuscript. RH designed the research and drafted the manuscript. CW and WY acquired the clinical information and revised the manuscript. All the authors have contributed to the manuscript and approved the submitted version.

Acknowledgments

This work was supported by the Shenzhen Fund for Guangdong Provincial High-Level Clinical Key Specialties (SZGSP014) and the Sanming Project of Medicine in Shenzhen (SZSM202011015).

References

[1] L. L. Lim, N. Cheung, J. J. Wang et al., "Prevalence and risk factors of retinal vein occlusion in an Asian population," *British Journal of Ophthalmology*, vol. 92, no. 10, pp. 1316–1319, 2008.

[2] Y. N. Yan, Y. X. Wang, Y. Yang et al., "10-year fundus tessellation progression and retinal vein occlusion," *International Journal of Ophthalmology*, vol. 11, no. 7, pp. 1192–1197, 2018.

[3] W. Liu, L. Xu, and J. B. Jonas, "Vein occlusion in Chinese subjects," *Ophthalmology*, vol. 114, no. 9, pp. 1795-1796, 2007.

[4] H. F. Qin, C. Y. Zhang, D. W. Luo et al., "Anti-VEGF reduces inflammatory features in macular edema secondary to retinal vein occlusion," *International Journal of Ophthalmology*, vol. 15, no. 8, pp. 1296–1304, 2022.

[5] Y. Deng, Q. W. Zhong, A. Q. Zhang et al., "Microvascular changes after conbercept therapy in central retinal vein occlusion analyzed by optical coherence tomography angiography," *International Journal of Ophthalmology*, vol. 12, no. 5, pp. 802–808, 2019.

[6] J. Y. Yang, B. You, Q. Wang, S. Y. Chan, J. B. Jonas, and W. B. Wei, "Retinal vessel oxygen saturation in healthy subjects and early branch retinal vein occlusion," *International Journal of Ophthalmology*, vol. 10, no. 2, pp. 267–270, 2017.

[7] M. Coban-Karatas, R. Altan-Yaycioglu, B. Ulas, S. Sizmaz, H. Canan, and C. Sariturk, "Choroidal thickness measurements with optical coherence tomography in branch retinal vein occlusion," *International Journal of Ophthalmology*, vol. 9, no. 5, pp. 725–729, 2016.

[8] S. Polizzi, F. Barca, T. Caporossi, G. Virgili, and S. Rizzo, "Branch retinal vein occlusion following cataract surgery," *Clinical and Experimental Optometry*, vol. 101, no. 1, pp. 135-136, 2018.

[9] P. Song, Y. Xu, M. Zha, Y. Zhang, and I. Rudan, "Global epidemiology of retinal vein occlusion: a systematic review and meta-analysis of prevalence, incidence, and risk factors," *Journal of Global Health*, vol. 9, no. 1, Article ID 010427, 2019.

[10] Y. H. Wang, P. Zhang, L. Chen et al., "Correlation between obstructive sleep apnea and central retinal vein occlusion," *International Journal of Ophthalmology*, vol. 12, no. 10, pp. 1634–1636, 2019.

[11] D. Călugăru and M. Călugăru, "Comment on "Microvascular changes after conbercept therapy in central retinal vein occlusion analyzed by optical coherence tomography angiography"," *International Journal of Ophthalmology*, vol. 13, no. 5, pp. 848–850, 2020.

[12] N. E. Lee, H. M. Kang, J. H. Choi, H. J. Koh, and S. C. Lee, "Sectoral changes of the peripapillary choroidal thickness in patients with unilateral branch retinal vein occlusion," *International Journal of Ophthalmology*, vol. 12, no. 3, pp. 472–479, 2019.

[13] M. Zhao, C. Zhang, X. M. Chen, Y. Teng, T. W. Shi, and F. Liu, "Comparison of intravitreal injection of conbercept and triamcinolone acetonide for macular edema secondary to branch retinal vein occlusion," *International Journal of Ophthalmology*, vol. 13, no. 11, pp. 1765–1772, 2020.

[14] H. Takahashi, H. Tampo, Y. Arai, Y. Inoue, and H. Kawashima, "Applying artificial intelligence to disease staging: deep learning for improved staging of diabetic retinopathy," *PLoS One*, vol. 12, no. 6, Article ID e0179790, 2017.

[15] S. K. Devalla, K. S. Chin, J. M. Mari et al., "A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head," *Investigative Ophthalmology & Visual Science*, vol. 59, no. 1, pp. 63–74, 2018.

[16] X. Xu, L. Zhang, J. Li, Y. Guan, and L. Zhang, "A hybrid global-local representation CNN model for automatic cataract grading," *IEEE J Biomed Health Inform*, vol. 24, no. 2, pp. 556–567, 2020.

[17] E. Long, H. Lin, Z. Liu et al., "An artifical intelligence platform for the multihospital collaborative management of congenital cataracts," *Nature Biomedical Engineering*, vol. 1, no. 2, pp. 1–8, 2017.

[18] J. Anitha, C. K. Vijila, A. I. Selvakumar, A. Indumathy, and D. Jude Hemanth, "Automated multi-level pathology identification techniques for abnormal retinal images using artificial neural networks," *British Journal of Ophthalmology*, vol. 96, no. 2, pp. 220–223, 2012.

[19] D. Nagasato, H. Tabuchi, H. Ohsugi et al., "Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion," *International Journal of Ophthalmology*, vol. 12, no. 1, pp. 94–99, 2019.

[20] Z. Liu, Y. Lin, Y. Cao et al., "Swin Transformer: hierarchical vision transformer using shifted windows," 2021, https://arxiv.org/abs/2103.14030.

[21] S. Zhao, Z. Bai, S. Wang, and Y. Gu, "Research on automatic classification and detection of mutton multi-parts based on swin-transformer," *Foods*, vol. 12, no. 8, p. 1642, 2023.

[22] D. Z. Zhao, X. K. Wang, T. Zhao et al., "A Swin Transformer-based model for mosquito species identification," *Scientific Reports*, vol. 12, no. 1, Article ID 18664, 2022.

[23] Y. Liu, J. Zhao, Q. Luo, C. Shen, R. Wang, and X. Ding, "Automated classification of cervical lymph-node-level from ultrasound using depthwise separable convolutional swin transformer," *Computers in Biology and Medicine*, vol. 148, Article ID 105821, 2022.

[24] G. Zhang, B. Sun, Z. Zhang, J. Pan, W. Yang, and Y. Liu, "Multi-model domain adaptation for diabetic retinopathy classification," *Frontiers in Physiology*, vol. 13, Article ID 918929, 2022.

[25] Q. Chen, S. Lin, B. S. Liu et al., "Artificial intelligence can assist with diagnosing retinal vein occlusion," *International Journal of Ophthalmology*, vol. 14, no. 12, pp. 1895–1902, 2021.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recongination," 2015, https://arxiv.org/abs/1409.1556.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *ArVix. Preprint posted online on December*, vol. 10, 2015.

[28] Y. Huang, L. Lin, P. Cheng, J. Lyu, R. Tam, and X. Tang, "Identifying the Key components in ResNet-50 for diabetic retinopathy grading from fundus images: a systematic investigation," *Diagnostics*, vol. 13, no. 10, p. 1664, 2023.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," 2018, https://arxiv.org/abs/1801.04381.

[30] C. Guo, M. Yu, and J. Li, "Prediction of different eye diseases based on fundus photography via deep transfer learning," *Journal of Clinical Medicine*, vol. 10, no. 23, p. 5481, 2021.

[31] Y. P. Liu, Z. Li, C. Xu, J. Li, and R. Liang, "Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network," *Artificial Intelligence in Medicine*, vol. 99, Article ID 101694, 2019.

[32] G. Huang, Z. Liu, and L. van der Maaten, "Densely Connected Convolutional Networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.

[33] A. Usman, A. Muhammad, A. M. Martinez-Enriquez, and A. Muhammad, "Classification of diabetic retinopathy and retinal vein occlusion in human eye fundus images by transfer learning," *Advances in Intelligent Systems and Computing*, Springer, New York, NY, USA, pp. 642–653, 2020.

[34] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, June 2015.

[35] Z. Y. Liu, B. Li, S. Xia, and Y. X. Chen, "Analysis of choroidal morphology and comparison of imaging findings of subtypes of polypoidal choroidal vasculopathy: a new classification system," *International Journal of Ophthalmology*, vol. 13, no. 5, pp. 731–736, 2020.

[36] C. Wan, H. Li, G. F. Cao, Q. Jiang, and W. H. Yang, "An artificial intelligent risk classification method of high myopia based on fundus images," *Journal of Clinical Medicine*, vol. 10, no. 19, p. 4488, 2021.

[37] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and B D. Grad-Cam, "Visual explanations from deep networks via gradient-based localization," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, Venice, October 2017.

[38] W. Mi, J. Li, Y. Guo et al., "Deep learning-based multi-class classification of breast digital pathology images," *Cancer Management and Research*, vol. 13, pp. 4605–4617, 2021.

[39] B. Babenko, A. Mitani, I. Traynis et al., "Detection of signs of disease in external photographs of the eyes via deep learning," *Nature Biomedical Engineering*, vol. 6, no. 12, pp. 1370–1383, 2022.

[40] K. Cao, J. Xu, and W. Q. Zhao, "Artificial intelligence on diabetic retinopathy diagnosis: an automatic classification method based on grey level co-occurrence matrix and naive Bayesian model," *International Journal of Ophthalmology*, vol. 12, no. 7, pp. 1158–1162, 2019.

[41] M. Ucar, H. B. Cakmak, and B. Sen, "A statistical approach to classification of keratoconus," *International Journal of Ophthalmology*, vol. 9, no. 9, pp. 1355–1357, 2016.

[42] G. Zhang, B. Sun, Z. Chen et al., "Diabetic retinopathy grading by deep graph correlation network on retinal images without manual annotations," *Frontiers of Medicine*, vol. 9, Article ID 872214, 2022.

[43] H. Zhang, K. Niu, Y. Xiong, W. Yang, Z. He, and H. Song, "Automatic cataract grading methods based on deep learning," *Computer Methods and Programs in Biomedicine*, vol. 182, Article ID 104978, 2019.

[44] B. Mo, H. Y. Zhou, X. Jiao, and F. Zhang, "Evaluation of hyperreflective foci as a prognostic factor of visual outcome in retinal vein occlusion," *International Journal of Ophthalmology*, vol. 10, no. 4, pp. 605–612, 2017.

[45] A. H. Bayat, A. Çakır, B. Erden, S. Bölükbaşı, and Ş G. Şehirli, "Comment on "Evaluation of hyperreflective foci as

a prognostic factor of visual outcome in retinal vein occlusion"," *International Journal of Ophthalmology*, vol. 11, no. 5, p. 898, 2018.

[46] L. P. Cen, J. Ji, J. W. Lin et al., "Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks," *Nature Communications*, vol. 12, no. 1, p. 4828, 2021.

[47] D. Călugăru and M. Călugăru, "Intraocular pressure modifications in patients with acute central/hemicentral retinal vein occlusions," *International Journal of Ophthalmology*, vol. 14, no. 6, pp. 931–935, 2021.

[48] W. J. Dong, C. Q. Li, Y. Q. Shu et al., "Altered brain network centrality in patients with retinal vein occlusion: a resting-state fMRI study," *International Journal of Ophthalmology*, vol. 14, no. 11, pp. 1741–1747, 2021.

[49] C. Wan, R. Hua, K. Li, X. Hong, D. Fang, and W. Yang, *Automatic Diagnosis of Different Types of Retinal Vein Occlusion Based on Fundus Images*, JMIR Publications Inc, Toronto, Canada, 2022.