

Received November 24, 2019, accepted December 17, 2019, date of publication December 23, 2019, date of current version January 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2961418

Automatic Diagnosis of Familial Exudative Vitreoretinopathy Using a Fusion Neural Network for Wide-Angle Retinal Images

YU YE¹, JIANBO MAO^{2,3}, LEI LIU¹, SHULIN ZHANG¹, LIJUN SHEN^{2,3}, AND MINGZHAI SUN¹

¹Department of Precision Machinery and Instrumentation, University of Science and Technology of China, Hefei 230022, China

²Eye Hospital of Wenzhou Medical University, Wenzhou 325035, China

³School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou 325035, China

Corresponding authors: Lijun Shen (slj@mail.eye.ac.cn) and Mingzhai Sun (mingzhai@ustc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61836011, and in part by the Fundamental Research Funds for the Central Universities under Grant WK2090090023.

ABSTRACT Familial exudative vitreoretinopathy (FEVR) is a hereditary disorder that can damage the retina. This retinal damage can lead to vision loss and even blindness in the late stages. Thus, early diagnosis and prevention of the disease's progression are critical. The purpose of this study was to develop an automated diagnosis system for FEVR based on combining deep learning and domain knowledge. A transfer learning scheme was designed to train a deep convolutional neural network (DCNN) to provide segmentation of the retinal vessels. Based on this vessel segmentation and prior clinical knowledge, the vascular characteristics, including the retinal avascular area, vessel angle, fractal dimension, branching and density of blood vessels, were automatically evaluated. Finally, the diagnosis of FEVR was achieved by a feature fusion neural network. Our method was evaluated on 300 images with 168 healthy and 132 FEVR images. By combining deep features and handcrafted features (extracted vascular characteristics), the proposed method achieved an average F1-score of 0.95, with excellent accuracy (94.34%) and sensitivity (91.43%); the quadratic weighted κ was 0.88 for the diagnosis of FEVR. We demonstrated the effectiveness and robustness of the proposed method using five-fold cross-validation. The proposed automatic diagnosis system can assist doctors for better judgment and make sense of early diagnosis and prevention of the disease's progression.

INDEX TERMS FEVR, wide-angle retinal images, deep learning, vascular characteristics, fusion.

I. INTRODUCTION

Familial exudative vitreoretinopathy (FEVR) is a progressive and hereditary vitreoretinopathy [1] first reported in 1969 by Criswick and Schepens [2]. The genetics of FEVR are highly heterogeneous, and the expression varies significantly within the same family [3]. Moreover, the presence and severity of symptoms in the left and right eyes can differ [4]. Retinal vessel abnormalities in the temporal periphery are the main pathological features of FEVR, and the detection of these changes requires imaging of the retina with a wider field of view (FOV) than the usual 45° achieved by regular fundus cameras. In clinical practice, wide-angle fundus cameras, such as Optoview (Daytona) are used to achieve the FOV as

wide as 200°, which dramatically expands the region under examination [5]. In combination with fluorescein angiography (FA), wide-angle fundus cameras have become standard tools for diagnosis and monitoring of FEVR. Exudation, neovascularization, retinal fold, ectopic macula, and retinal detachment may also occur with the development of FEVR [6]. A typical image of FEVR (Fig. 2 (a)) shows the avascular zone in the temporal side of the retina with exudation. In addition, blood vessels appear straightened due to the action of vitreoretinal traction. As the disease progresses, the vision of the FEVR patients becomes worse, and patients eventually become blind if the disease is left untreated. Laser photocoagulation [7] is currently the mainstream treatment to prevent the progression of FEVR. Early diagnosis and intervention are important to prevent the progression of the disease and preserve the vision of the patients.

The associate editor coordinating the review of this manuscript and approving it for publication was Chenping Hou¹.

In recent years, deep learning (DL) methods have garnered great attention and have been widely applied in many fields, including image recognition, natural language processing, etc. A deep convolutional neural network (DCNN) is the main component of DL methods. It applies a series of layers, including convolution layer, pooling layer, and fully connected layers, with thousands or even millions of trainable parameters that are continuously updated with a backpropagation algorithm to minimize loss between output and target during the training process. In medicine, deep learning has shown significant advantages in disease diagnosis and lesion segmentation due to its powerful capability for feature extraction. Chudzik *et al.* [8] segmented the microaneurysm, an early lesion of diabetic retinopathy, using a modified U-net and Zheng *et al.* [9] achieved state-of-the-art performance in segmentation of exudates using a conditional generative adversarial network. In terms of disease screening, DL systems have accurately detected age-related macular degeneration (AMD) [10], glaucoma [11], and Alzheimer's disease (from brain MRIs) [12], among others.

One common challenge of applying DL in medical studies is the limitation of annotated data. To alleviate the problem, various training methods, including pseudo-labeling and transfer learning, as well as continuously improved network structures, such as ResNet-18 [13], ResNeXt-50 [14], DenseNet [15], SENet [16], and RU-net [17], have been proposed. Recently, the performance of fusion networks has been shown to be outstanding in semantic segmentation and classification tasks, since useful information is more fully exploited. The flexibility of the deep learning framework allows information to be integrated in various ways. According to the stage of fusion, fusion algorithms can be classified into three approaches: 1) Early fusion integrates multiple sources of information as the input of the network. This simplest fusion approach often cannot fully exploit the complementary nature of the multiple sources of information and may cause information redundancy [18]. Typically, early fusion is most suitable when the input data are similar to each other. 2) Intermediate-level fusion fuses higher-level representations (feature maps) through a series of non-linear mappings. This allows representations to be flexibly fused at different depths, unlike other fusion approaches. However, this approach requires careful design and adequate experimentation with specific tasks to achieve outstanding performance. 3) Late fusion integrates decisions from multiple subnetwork classifiers; thus, this fusion method is also called decision-level fusion [19]. Late fusion methods have been widely used in many fields because these methods can choose the state-of-the-art subnetworks [20], [21]. Of course, decision fusion also has limitations; using only the final vector from classifiers independently will cause loss of spatial information contained in the feature maps, which results in inadequate information use [22]. In summary, all fusion strategies have both advantages and limitations, so the choice of fusion network must be made in the context of the data and task at hand.

The changes in vascular characteristics are critical pathological features of FEVR [4]; therefore, it is crucial to provide accurate segmentation of blood vessels. However, annotated wide-angle retinal images are minimal. To alleviate this problem, we applied the transfer learning scheme to make use of the annotated regular fundus images fully. Furthermore, we adapted the RU-net, which has an outstanding performance in vascular segmentation [17] to segment the vessels in the wide-angle retinal images. Based on the vessel segmentation and prior clinical knowledge, such as the avascular region, we developed a fusion network to classify an input image as either a healthy or FEVR image. This network combines deep features and the handcrafted features (extracted vascular characteristics) to exploit the information of the original image fully. Our system maximized the use of limited data resources and prior clinical knowledge, to achieve an accurate automated diagnosis, a critical goal for disease screening, and the prevention of disease progression.

II. METHOD

Fig. 1 offers a diagram of the proposed algorithm. After pre-processing, we combined the transfer learning scheme and RU-net to achieve accurate segmentation of the blood vessels. Based on the vessel segmentation and prior clinical knowledge, we defined the handcrafted features by evaluating the vascular characteristics. ResNet-18 was used to extract the deep features of the original images and the corresponding blood vessels. Finally, we achieved the diagnosis of FEVR by combining the deep features and handcrafted features.

This study adhered to the tenets of the Declaration of Helsinki [23] and was approved by the institutional review board of the Eye Hospital of Wenzhou Medical University. All subjects provided written informed consent.

A. DATASETS

A total of 150 subjects were recruited at the Eye Hospital of Wenzhou Medical University. The dataset consisted of 300 images (150 patients, 300 eyes), with 168 healthy (84 patients, 168 eyes) and 132 FEVR images (66 patients, 132 eyes). Ultra-wide field retinal imaging facilitated the acquisition of 200 ° panoramic images [5] of the retina, including green laser wavelength images, red laser wavelength images and pseudocolor (two-color) images [5] with a resolution of 3,900×3,072 pixels. The green (red-free) component depicts the retina and its vasculature, whereas the red component highlights deeper structures, such as dilated choroidal vessels. In this paper, we only used the green laser wavelength images, since they provided the best contrast between blood vessels and background; 20 wide-angle images from this dataset are available on GitHub¹ to promote research in this field.

In the present study, we divided all images into two datasets: one training and one independent test dataset. The training dataset was used to train the DCNN to find the

¹<https://github.com/FundusResearch/WARI>

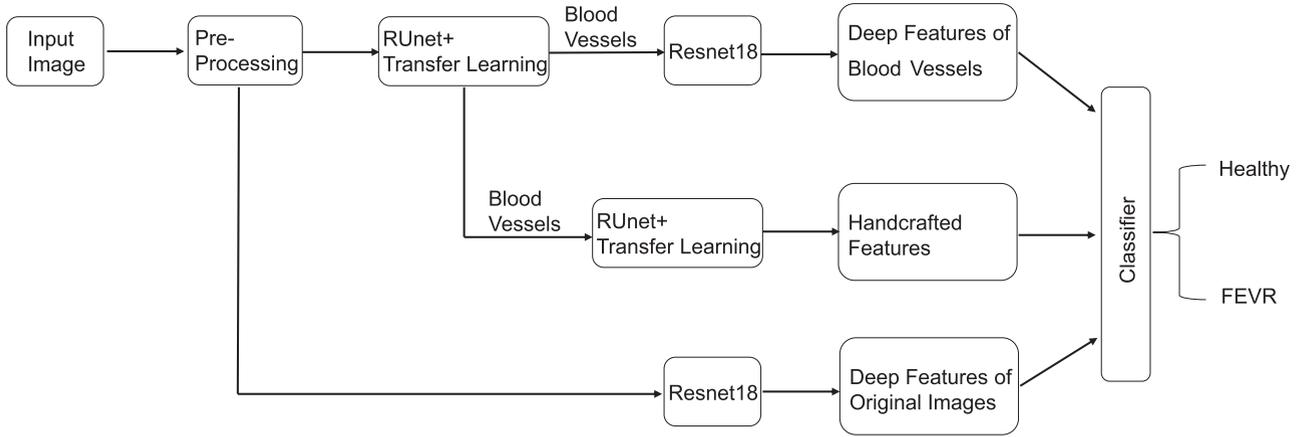


FIGURE 1. Block diagram for the proposed algorithm.

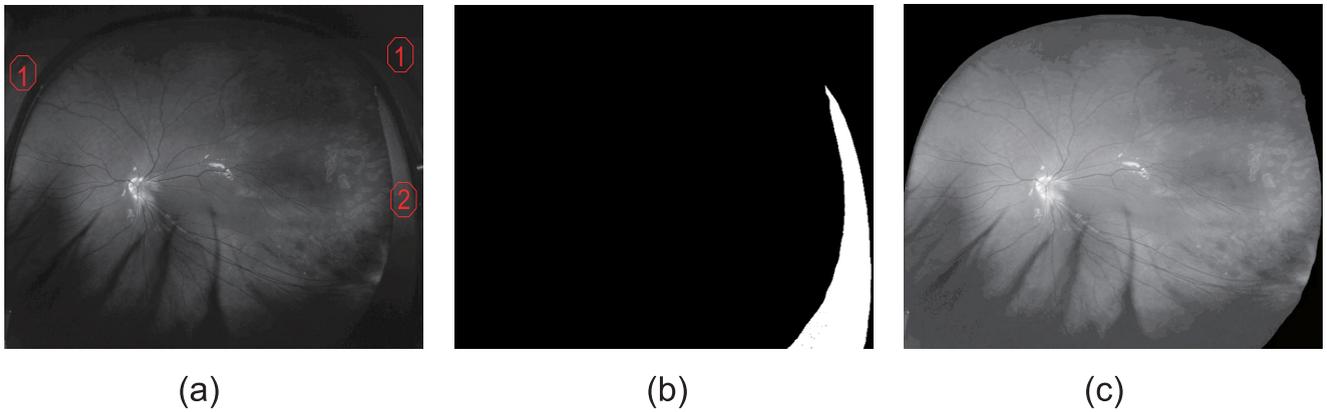


FIGURE 2. Pre-processing of wide-angle retinal images. (a) The original image. Areas ① and ② in Fig. 2a represent the nonuniform and irregular edges of the camera, respectively. (b) The corresponding irregular edge of the camera. (c) The processed image.

optimal model, and the independent test dataset was used to further test the performance of the best model.

B. IMAGE PRE-PROCESSING

The pre-processing of wide-angle retinal images involved region cropping with a mask, removing the irregular edges of the camera with the DCNN, gamma correction [24], histogram equalization [25], and histogram matching [26], as shown in Fig. 2 and the following mathematical expressions.

$$f(I) = A \cdot I^\gamma \cdot (1 - mask1) \cdot (1 - mask2) \tag{1}$$

$$S_k = T(r_k) = (L - 1) \cdot \sum_{j=0}^k (P_r(r_j)) = \frac{L - 1}{MN} \sum_{j=0}^k (n_j) \tag{2}$$

$$S = T(r) = (L - 1) \cdot \int_0^r P_r(w) dw$$

$$G(z) = (L - 1) \cdot \int_0^z P_z(t) dt = S \tag{3}$$

We provide an original image, the corresponding irregular edges of the camera, and the processed image in Fig. 2. We cropped the original image with a mask to exclude extraneous and nonuniform parts of the camera (area① in Fig. 2a). Gamma correction and histogram equalization were used

to adjust the brightness and contrast of images. Histogram matching was adopted to reduce the difference between different images and we chose a clear image with high contrast and the distribution of gray value matches the Gaussian distribution as the reference image. The presence of irregular edges (area② in Fig. 2a) around the image can affect the detection of the avascular area. We manually labeled the irregular edges of 50 images. Then a U-net [27] was used to segment the irregular edges. As shown in Fig. 2(c), we successfully solved the interference of the edges of the device, and the visibility of the blood vessels is higher in the processed image.

Eq. (1) describes the process of region cropping with mask1, removing the irregular edges of the device with the mask (mask2) provided by DCNN and gamma correction, where I and $f(I)$ represent the original image and the processed image, respectively, and the parameters A and γ were set as 1.2 and 0.75. Eq. (2) describes the histogram equalization, where S_k and r_k represent the target pixel value and raw pixel value, respectively. We set L to 255; the parameters M and N represent the width and height of the image. $P_r(r_j)$ represent the ratio of the number of pixels in the current gray level to the total pixels and n_j represent the number of pixels in the current gray level. The third equation describes the histogram matching adopted in our study;

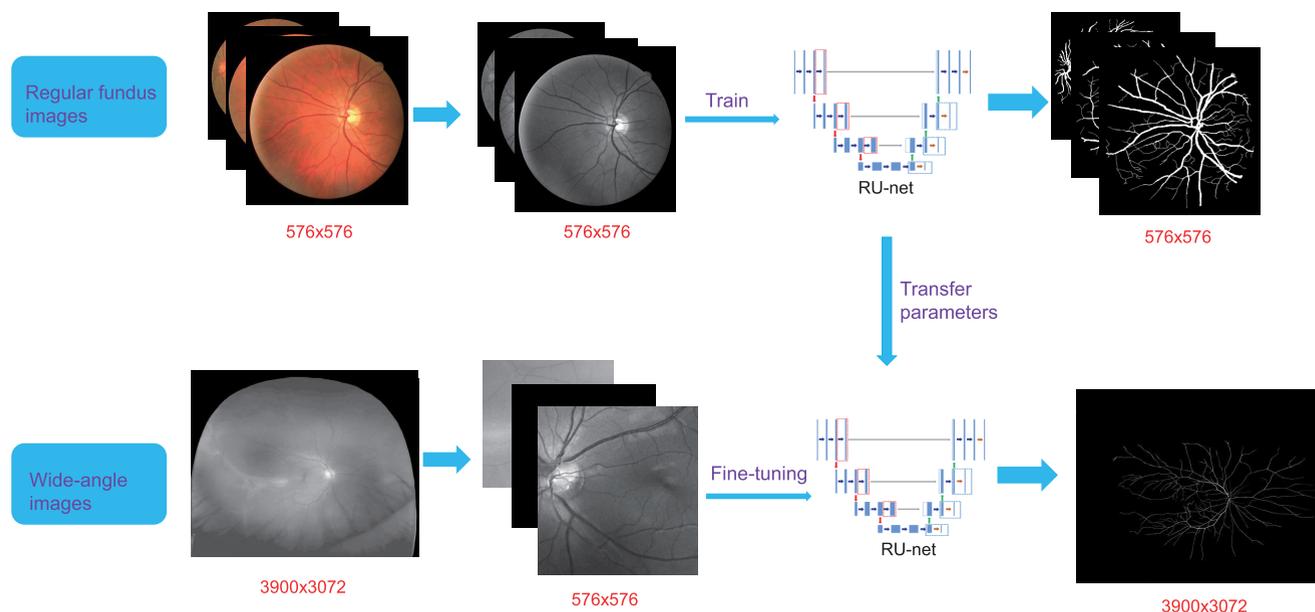


FIGURE 3. Transfer learning using regular retina images to improve the network performance on the wide-angle images. We first trained the RU-net with the green channel image of a regular fundus photograph and then refined the network with the patches of the wide-angle images.

The variables r and z represent the gray level of the input image and the output image, respectively; $P_r(w)$ and $P_z(t)$ are corresponding probability density functions. The T and G are the gray level conversion function and $L = 255$ in our study.

C. VASCULAR SEGMENTATION

Although there are publicly available datasets of regular fundus images (45° FOV) with manually annotated ground truth, such as the DRIVE dataset [28] and the E-ophtha dataset [29], there are few such resources for wide-angle retinal images. To alleviate the problem, we applied the transfer learning method [8] to train an RU-net [17] to provide segmentation of the retinal vessels. RU-net is a refined DCNN based on U-NET and has demonstrated outstanding performance in vascular segmentation. As shown in Fig. 3, we first trained the RU-net with 380 regular fundus images from publicly available datasets. The network could initially detect almost ninety percent of blood vessels in the wide-angle images, albeit with a high false-positive rate and an inability to distinguish blood vessels and eyelashes. We fine-tuned the network with 50 wide-angle retinal images manually labeled by two ophthalmologists from Wenzhou Medical University. The original resolution of the wide-angle retinal images was 3,900×3,072 pixels. we cropped the original images into 576×576 patches with a stride of 100 pixels, and we selected the patches of blood vessels and background with an unbalanced sampling ratio of 3:1. The patch-based training method successfully solved the problem of computer memory overflow during training with the whole images. In this way, we achieved a precise vascular segmentation of the wide-angle retinal images. Segmentation of the optic disc was achieved using similar procedures.

D. VASCULAR CHARACTERIZATION

Based on the pathological features of FEVR, we evaluated the vascular characteristics, including the retinal avascular area, vessel angle, fractal dimension, vascular branches, and density of blood vessels. In the process of the vascular characterization, we calculated various metrics for all vascular pixels without having to consider which blood vessel each pixel belongs to. So we did not need to apply vascular tracking.

To obtain a binary image of the blood vessels and optic discs, an adaptive threshold method was adopted to the output of the segmentation network described above. Then, we fitted a circle with the optic disc and calculated its center and diameter. Since the vascular abnormalities of FEVR appear on the temporal side of the retina, we cropped sixteen times the diameter of optic disc on the temporal side of images as the region of interest (ROI), which covered all the main blood vessels. The area inside the red circle in Fig. 8 shows the ROI we selected; we focused on the evaluation of vascular characteristics in ROIs in this study.

We detected the avascular area by searching the boundary of blood vessels as shown in Fig. 4b, which is the important characteristic of FEVR, especially in its early stages, and is the key to distinguish between FEVR and other similar diseases [30].

Traction in the temporal side of the retina causes the straightening of the blood vessels. We used the change of vessel angle within ROI to reflect this traction. In the process of angle measurement, we connected the center of the macula and optic disc as the reference axis (red line with an arrow in Fig. 4a). For each point, we fitted a line with five points around it and then calculated the angle between the line with the reference axis (blue sector in Fig. 4a). As shown in Fig. 5,

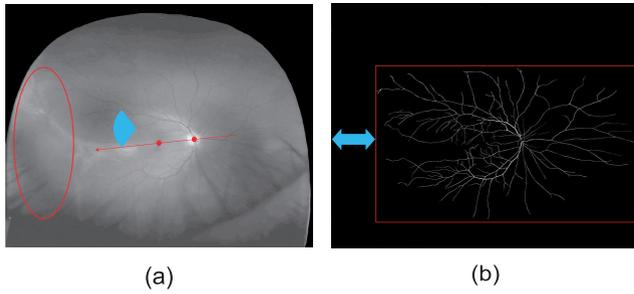


FIGURE 4. Evaluation of the retinal avascular area and vessel angle. (a) A typical image of FEVR with the avascular area on the temporal side of the retina (the red line with an arrow represents the reference axis and the blue sector represents the angle between the blood vessels and the reference axis). (b) The boundary of blood vessels and the detected avascular area.

the color of each point in the pseudocolor map represents the absolute value of the angle.

Fractal dimension was defined as a pattern of expanding symmetry [31]; the more complex graphics correspond to a larger value. First, we obtained the skeletonized blood vessels within the ROI. Then, we determined the fractal dimension using the standard box-counting method [32]–[34].

The number of vascular branches and crossing points demonstrates the geometric complexity of the blood vessels. Based on the skeletonized blood vessels, we detected all crossing points by counting the number of pixels in a connected domain and broke the skeletonized blood vessels into branches [35].

To evaluate the density of blood vessels, we removed the avascular area around the binary blood vessel images and calculated the ratio of blood vessel points to the entire area of the map.

E. BASELINE OF THE FUSION NETWORK

For ease of understanding, we described the baselines of the fusion network in this section as shown in Fig. 6. We used the original image or extracted blood vessels as a single input. Then, we adopted ResNet-18 [13] except for the last layer as the feature extractor to extract the deep feature of the input and got the original feature (the deep feature of the original image) and blood vessels feature (the deep feature of the blood vessels) for all but the last two layers. In the previous section, we evaluated vascular characteristics, which are important for disease diagnosis in a clinical setting. We presented these vascular characteristics (the retinal avascular area detected, the mean of vessel angle, the standard deviation (SD) of vessel angle, fractal dimensions, number of vascular branches, density of blood vessels) as a feature vector with a length of 6 and obtained the handcrafted feature. As a result, we have three different features as the basis for diagnosis in each case: the original feature, blood vessels feature and handcrafted feature. Finally, the fully connected layer and rectified linear unit (ReLU) were used to map the features to the final decision in a nonlinear manner.

F. FEVR CLASSIFICATION-FUSION NETWORKS

In this section, we describe the development of a fusion neural network by combining deep features and the handcrafted feature. The proposed method aimed to maximally utilize all available data and prior clinical knowledge to train an effective and robust deep learning model.

The details of the architecture of the fusion network for FEVR classification was shown in Fig. 7. We used the trained RU-net to segment the blood vessels of the original image. Based on the vessel segmentation, we evaluated the vascular characteristics as described in Section II-D. These were

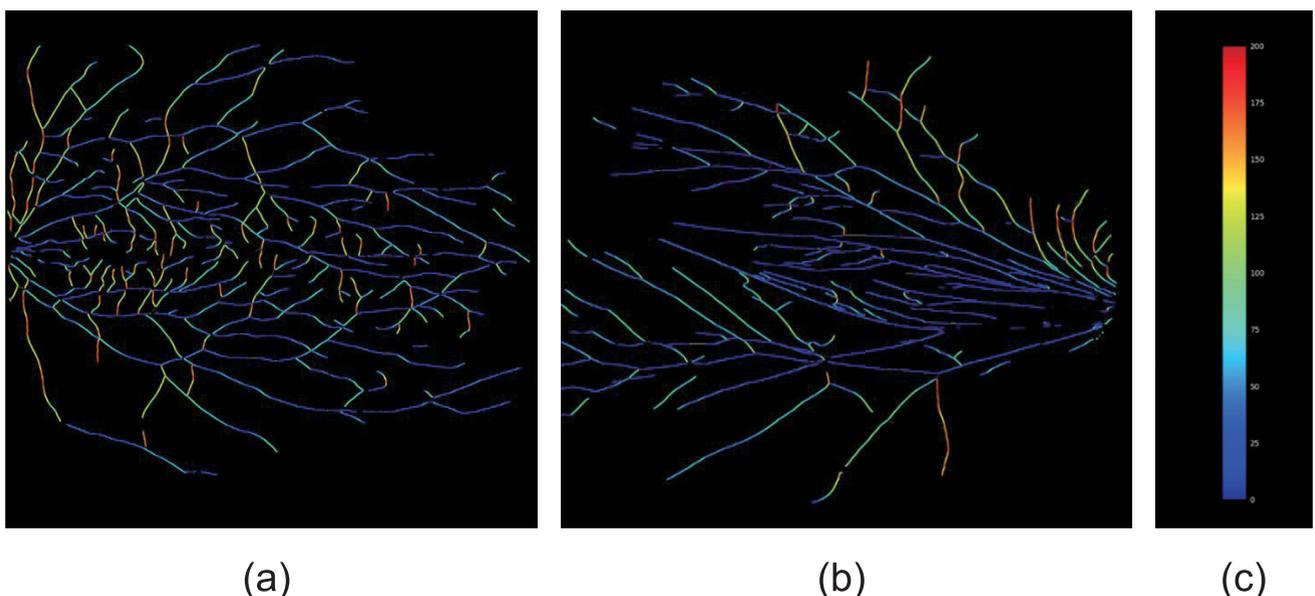


FIGURE 5. Angle pseudocolor map within the ROI. (a) The angle pseudocolor map of a healthy eye. (b) The angle pseudocolor map of FEVR. (c) Colorbar: blue represents an angle value of 0°, green, an angle value of 45°, and red, an angle value of 90°.

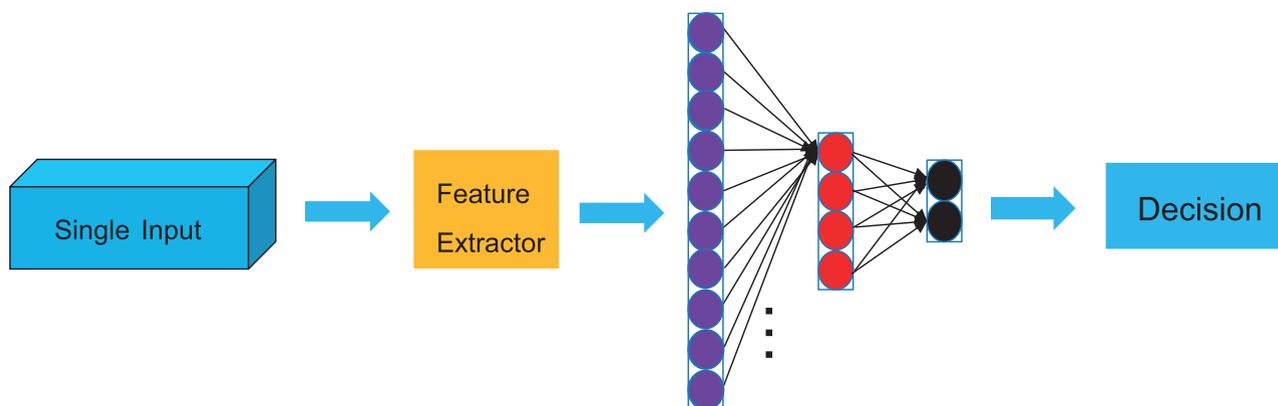


FIGURE 6. Structure of the baseline of the fusion network.

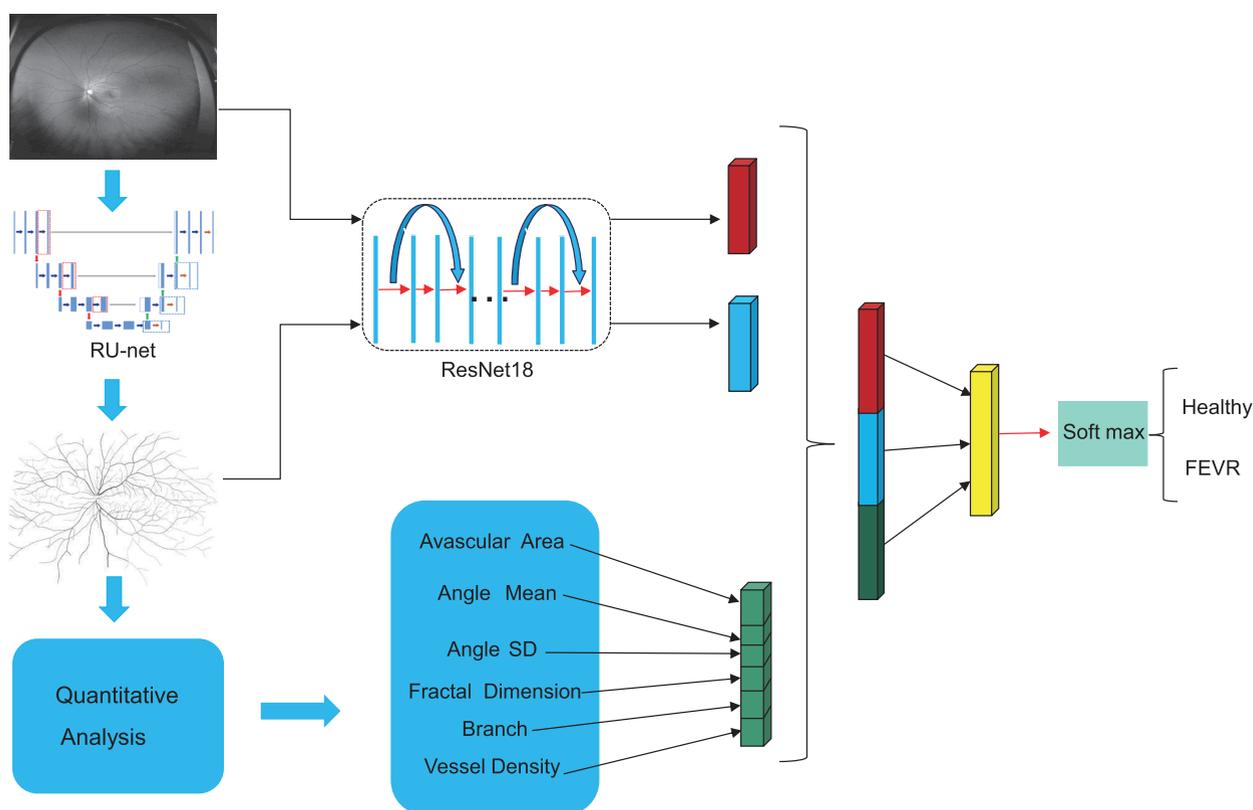


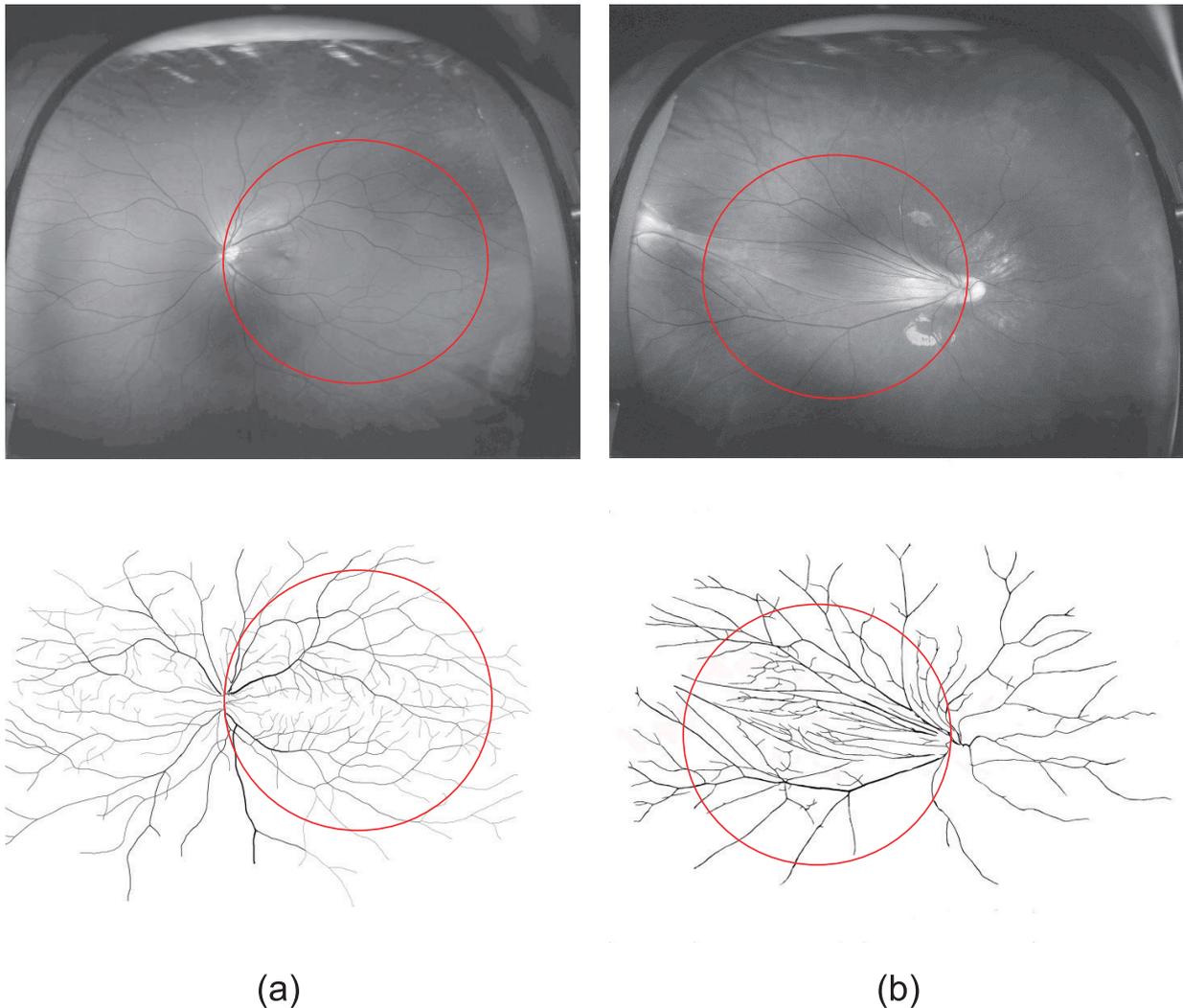
FIGURE 7. Architecture of the proposed fusion classification model.

the handcrafted features (green cuboid elements in Fig. 7). A ResNet-18 [13] was adopted as the feature extractor to extract the deep features of the original image (red cuboid element) as shown in Fig. 6 and the corresponding segmented blood vessels (blue cuboid element). The handcrafted features and deep features were concatenated before applying the fully connected layer. In this way, we successfully integrated the vascular characteristics into the network. The proposed fusion network has fully exploited the information of the original image, blood vessels, and prior clinical knowledge to achieve accurate classification.

To demonstrate the importance of the three features (original feature, blood vessels feature and handcrafted features), we designed a series of ablation structures and classification methods to get the optimal model. Based on the proposed architecture (shown in Fig. 7), we removed the network parameters corresponding to each feature; we used only: 1) the original and handcrafted features without the blood vessels feature; 2) the blood vessels and handcrafted features without the original feature; 3) the original and blood vessels features without the handcrafted features. For purposes of comparison, we also tried to classify only using the

TABLE 1. Performance of the vascular segmentation of our method with transfer learning (RU-net + transfer learning) with that of an RU-net trained only with 50 wide-angle retinal images.

	Accuracy	Sensitivity	Specificity	Precision	F1-score
RU-net	97.81%	70.23%	99.32%	65.34%	67.69%
RU-net + transfer learning	98.86%	72.21%	99.34%	66.69%	69.34%

**FIGURE 8.** Segmentation results of blood vessels for wide-angle retinal images. (a) Segmented blood vessels of a healthy eye. (b) Segmented blood vessels of an eye with FEVR.

handcrafted feature by the fully connected layers and support vector machine (SVM).

To get the best model for the diagnosis of FEVR, we compared the proposed fusion network with the current competitive classification networks, including ResNeXt-50 [14], DenseNet [15], and SENet [16].

All proposed networks were implemented in Python 3.5 using Pytorch as the backend. All experiments were performed on a cluster with 10 GeForce GTX 1080Ti GPUs. During the training process, we adopted data augmentation for all images, including rotating with a random angle, flipping randomly, affine transformation,

adding Gaussian noise, and normalizing by subtracting the mean value of the image followed by dividing the corresponding standard deviation. We set the batch size of 8, and Adam was adopted as the optimizer with the learning rate of 0.005.

III. EXPERIMENTAL RESULTS

A. VASCULAR SEGMENTATION OF THE WIDE-ANGLE RETINAL IMAGES

Based on the transfer learning scheme [8] and RU-net [17], we made full use of 380 fundus images from public datasets (DRIVE dataset, and E-ophtha dataset, etc.) and

TABLE 2. Evaluation of the performance of various ablation architectures. The inputs of the networks were: 1) original feature, blood vessels feature and handcrafted feature. 2) original feature and handcrafted feature. 3) blood vessels feature and handcrafted feature. 4) original feature and blood vessels feature. 5) handcrafted feature via two fully connected layers. 6) handcrafted feature via SVM.

	Quadratic weighted κ	Accuracy	Sensitivity	Specificity	Precision	F1-score
Proposed method	88.48% $\pm 2.47\%$	94.34% $\pm 2.18\%$	91.43% $\pm 3.07\%$	99.12% $\pm 2.01\%$	99.68% $\pm 1.01\%$	95.01% $\pm 2.22\%$
Original feature + Handcrafted feature	86.58% $\pm 4.35\%$	93.48% $\pm 3.95\%$	90.71% $\pm 4.07\%$	96.42% $\pm 2.31\%$	97.85% $\pm 1.63\%$	94.44% $\pm 1.95\%$
Blood vessels feature + Handcrafted feature	85.53% $\pm 4.13\%$	92.93% $\pm 2.08\%$	90.18% $\pm 4.49\%$	96.26% $\pm 2.59\%$	97.45% $\pm 1.86\%$	93.92% $\pm 1.89\%$
Original feature + Blood vessels feature	84.09% $\pm 4.02\%$	92.37% $\pm 2.87\%$	90.39% $\pm 4.06\%$	95.94% $\pm 2.95\%$	97.06% $\pm 1.38\%$	93.06% $\pm 2.37\%$
Handcrafted feature+ FC	82.27% $\pm 3.78\%$	90.31% $\pm 3.04\%$	87.49% $\pm 4.05\%$	96.21% $\pm 2.37\%$	96.68% $\pm 1.72\%$	91.94% $\pm 2.71\%$
Handcrafted feature+ SVM	79.46% $\pm 4.03\%$	88.22% $\pm 3.88\%$	81.74% $\pm 4.17\%$	95.62% $\pm 2.96\%$	93.95% $\pm 1.39\%$	91.45% $\pm 2.38\%$

TABLE 3. Comparison of the proposed method with state-of-the-art classification networks. We used the original images as the input of the networks (SENet, ResNeXt-50, DenseNet) and the quadratic weighted κ , accuracy, sensitivity, specificity, precision and F1 score were adopted to evaluate the performance of various models.

	Quadratic weighted κ	Accuracy	Sensitivity	Specificity	Precision	F1-score
Proposed method	88.48% $\pm 2.47\%$	94.34% $\pm 2.18\%$	91.43% $\pm 3.07\%$	99.12% $\pm 2.01\%$	99.68% $\pm 1.01\%$	95.01% $\pm 2.22\%$
SENet [16]	84.51% $\pm 4.77\%$	92.38% $\pm 2.85\%$	88.22% $\pm 4.88\%$	96.82% $\pm 3.04\%$	97.54% $\pm 1.74\%$	93.63% $\pm 2.71\%$
ResNeXt-50 [14]	82.44% $\pm 5.16\%$	91.30% $\pm 3.58\%$	85.72% $\pm 5.67\%$	97.37% $\pm 3.78\%$	97.34% $\pm 1.92\%$	92.32% $\pm 2.39\%$
DenseNet [15]	82.10% $\pm 5.12\%$	91.27% $\pm 3.83\%$	85.39% $\pm 5.04\%$	94.44% $\pm 3.96\%$	96.15% $\pm 1.58\%$	92.59% $\pm 2.51\%$

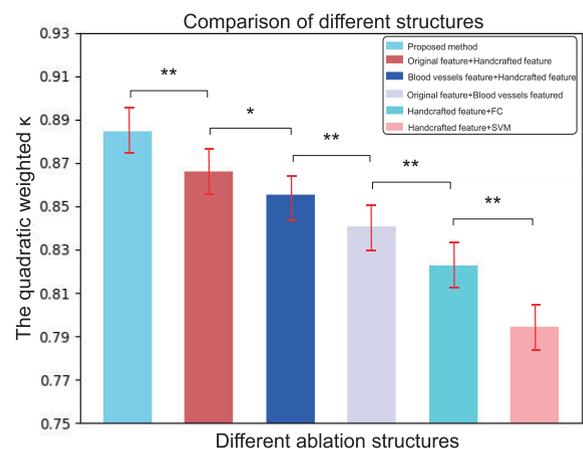
50 wide-angle retinal images manually annotated by the ophthalmologists to perform the segmentation of blood vessels.

We compared the performance of our method (RU-net+transfer learning) with that of an RU-net trained only with 50 wide-angle retinal images. The result demonstrated that transferring learning did substantially improve the performance of the network, and our method achieved accurate segmentation of blood vessels (accuracy of 0.98, the sensitivity of 0.71; Table 1), noticeably better than that achieved by the network trained with 50 wide-angle images.

Fig. 8 displays the vascular segmentation result for the wide-angle retinal images. We successfully resolved the issue of interference from the edges of the device and of uneven illumination by pre-processing. We could distinguish the blood vessels and eyelashes very well, with few false positives in the segmented blood vessels. Not only the main blood vessels but also thin vessels at the periphery were well-segmented (Fig. 8 (c,d)).

B. ABLATION EXPERIMENTS TO COMPARE DIFFERENT FUSION NETWORKS

In this section, we compare our network (Fig. 7) with other fusion structures and classification methods through a series

**FIGURE 9.** The result of the paired T-test of the quadratic weighted κ for different ablation structures. We compared the proposed fusion network with the networks that lack any feature to verify whether each feature (original feature, blood vessels feature and handcrafted feature) improves the performance of the network (*: $p < 0.05$, **: $p < 0.01$).

of ablation experiments (as described in Section II-F) to obtain the optimal model.

We compared the proposed architecture with the performance of networks without any features to verify

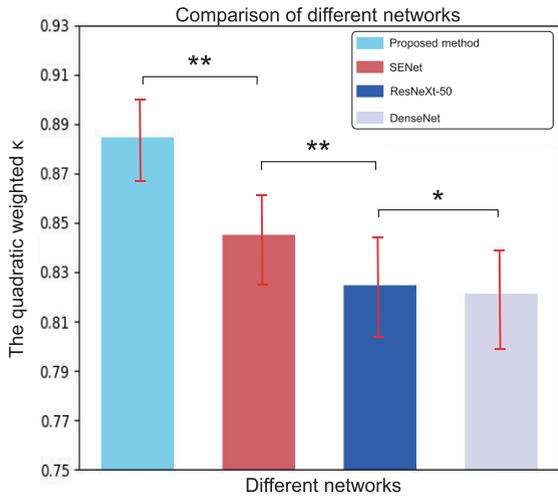


FIGURE 10. Comparison of the quadratic weighted κ for the state-of-the-art classification networks and the proposed method by paired T-test (*: $p < 0.05$, **: $p < 0.01$).

whether each feature (original, blood vessels, and handcrafted features) improves the performance of the network. The following measures were used to assess the performance of each architecture: quadratic weighted κ , accuracy, sensitivity, specificity, precision, and F1 score. The mean and standard deviation (SD) of obtained metrics are provided in Table 2.

The results demonstrate that the proposed fusion network not only achieved excellent performance, with the quadratic weighted κ of 0.88 but also outperformed other

ablation structures, with a significant difference (Fig. 9). In other words, every feature in our fusion structure contributes to the diagnosis of FEVR.

C. COMPARISON OF THE PROPOSED METHOD WITH THE STATE-OF-THE-ART NETWORKS

The performance of our proposed method and current state-of-the-art classification networks is described in Table 3. We can see that the proposed method outperformed other models in the diagnosis of FEVR; the difference is significant, as shown in Fig. 10.

D. OPTIMAL MODEL-OUR FUSION NETWORK

We performed five-fold cross-validation to evaluate the robustness of our algorithm. All the images were divided into 5 groups. We used four groups as training data whereas one group was used as validation data. The right and the left eyes of the same patients were assigned to the same group to ensure complete independence of the training and testing data. The receiver operating characteristics (ROC) curve was used to assess the performance of our algorithm, and the area under the receiver operator curve (AUC) to characterize the level of classification (Fig. 11(a)). Our algorithm showed excellent performance, with an average AUC of 0.97.

Furthermore, we selected an additional 50 images (22 FEVR and 28 healthy) as a separate test dataset. As shown in Fig. 11(b), we achieved excellent performance, with the quadratic weighted κ of 0.87 and the accuracy of 0.94. All FEVR patients were correctly diagnosed, and only three

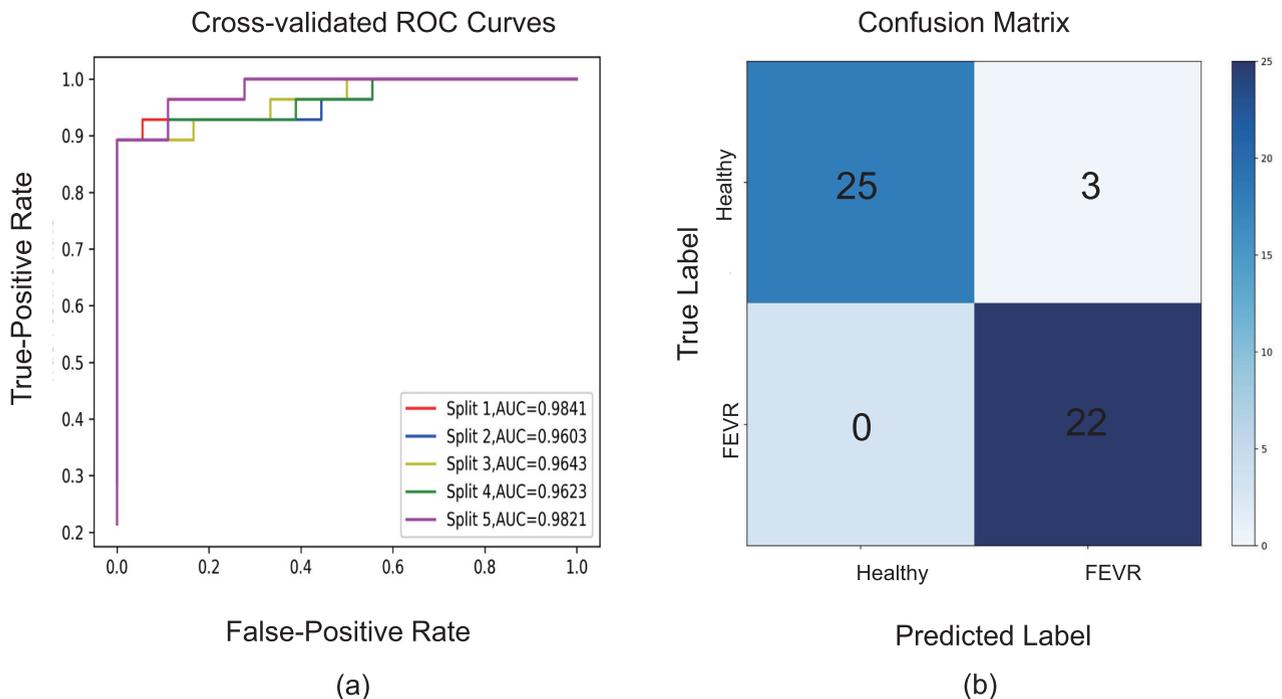


FIGURE 11. Performance of the optimal model-our fusion network. (a) ROC curves of five-fold cross-validation experiments with the corresponding AUC. (b) Confusion matrix of our fusion network’s performance on the separate test dataset.

normal people were classified as FEVR as a result of small pupils and noise at the edge of the images.

IV. DISCUSSION

In this paper, we proposed a system for the automatic diagnosis of FEVR that combines deep learning and prior clinical knowledge. By evaluating vascular characteristics (the retinal avascular area, vessel angle, fractal dimension, branch and vessel density) and presenting these vascular characteristics as a feature vector, we successfully integrated prior clinical knowledge into our fusion network, greatly improving its performance. The fusion approach adopted made up for the difference between different information resources and achieved an accurate diagnosis. The automatic diagnosis system can assist doctors in more accurate diagnosis and support early screening for FEVR and prevention of the disease's progression.

Since FEVR was first reported in 1969 by Criswick and Schepens [2], many medical experts have studied this disease due to its potential to cause serious vision loss or even blindness. Most of this research has focused on pathogenic mechanisms, including genetic defects [36] and genotype-phenotypic correlations [37], [38], and on treatment methods [39], [40]. Wide-field fluorescein angiography has become the gold standard for doctors in the diagnosis and monitoring of FEVR patients [41]. Unlike these prior studies, the aim of our study was to develop an end-to-end system dedicated to automatic diagnosis to support the early prevention of disease progression. A major benefit of our method is that it is completely non-invasive, without the need for fluorescein angiography.

Early-stage FEVR is characterized by avascularity and anomalous vasculature in the temporal quadrants of the peripheral retina [3]. In clinical practice, the detection of early-stage FEVR requires rich clinical experience and knowledge. Based on the vascular segmentation (Fig. 8), we detected the retinal avascular area by searching for the boundary of blood vessels. We also evaluated typical pathological features, including the vessel angle, fractal dimensions, branching, and density of blood vessels; all of these vascular characteristics are meaningful in the diagnosis of the disease. By combining deep features with the handcrafted feature (vascular characteristics), our method achieved a precise diagnosis for FEVR, including patients in the early stages of the disease. For many patients with early-stage FEVR, we can detect an obvious avascular area on the temporal side of the retina, as shown in Fig. 2 (a). However, in some patients with more advanced FEVR with blood vessels pulled to the temporal side of the retina, we detected no avascular area, as shown in Fig. 8 (b). Thus, it is inaccurate to diagnose only based on the avascular area; we must assess all pathological characteristics and available information to do so accurately.

Vascular segmentation is a crucial step that can greatly affect the accuracy of the image analysis and classification. Compared with regular fundus images (45° FOV), the vascular distribution of the wide-angle images is much wider,

with a smaller width; meanwhile, the problem of uneven illumination and noise is more serious. As a result, the vascular segmentation of wide-angle images is more challenging in the case of the few wide-angle images with manually annotated ground truth. Based on the transfer learning scheme, we made full use of the publicly available datasets of regular fundus images.

As discussed in [42], the main technical problem of transfer learning is how to reduce the shifts in data distributions across domains. In our case, wide-angle retinal imaging facilitated the acquisition of 200-degree panoramic images [5] of the retina. A regular fundus camera mostly has a 45-degree of a field of view (FOV), which is about the middle area of the wide-angle image. Therefore, the shifts in data distribution of the regular and wide-angle retinal images are minimum. Furthermore, during training, we cropped the wide-angle images into patches of the same size as regular fundus images, applied histogram equalization [25], and histogram matching [26] to reduce the differences between the wide-angle images and the regular images. As a result, pre-trained network on the regular fundus images significantly improves the network performance on wide-angle retinal images.

In our study, we developed a fusion architecture to maximally utilize all the available resources, including image information and prior clinical knowledge. We adopted ResNet-18 [13] as a basic network unit to extract the deep features of the original images and blood vessels; this basic network achieved good classification performance with few parameters. The retinal vessel abnormalities in the temporal periphery are typical pathological features of FEVR. The peripheral retinal avascularity and varying degrees of macular ectopia and tractional retinal detachment lead to changes in blood vessel patterns. We used the handcrafted feature obtained from the evaluation of blood vessels to reflect abnormalities of the retinal vessels. In a sense, the handcrafted feature belongs to the high-level feature, and it is hard to learn directly from images using the DCNN, especially with a small sample size. Through the evaluation of vascular characteristics and presenting the vascular characteristics in a vector, we successfully integrated the vascular characteristics into our network. As a result of the obvious difference between the original image, blood vessels, and handcrafted feature, we adopted feature-level fusion by concatenating all feature vectors rather than the data-level fusion [19] which is performed by integrating multiple sources of information as the input of the network. Finally, the fully connected layers and rectified linear units (ReLU) were used to integrate the joint representation (fused feature) through the non-linear mapping. In this way, we made full use of the data resources available to us, while at the same time successfully integrating prior clinical knowledge into our network, which significantly improved the performance of classification. We applied information fusion by merging the deep features with handcrafted features. Image fusion was not applied in our case since it may fuse the tiny vessel pixels with background pixels. We carried out a series of

ablation experiments and demonstrated that the proposed fusion approach achieved the best performance (Fig. 9). In addition, the proposed method achieved excellent performance on the separate test dataset, with a quadratic weighted κ of 0.87.

We analyze the proposed system in terms of time and memory usage. During the training process, it takes 240 minutes to train the RU-net and 153 to train our fusion classification model on a GeForce GTX 1080Ti GPU, with 1.86M and 2.64M trainable parameters, respectively. It takes only about one second to test a real-time image with our end-to-end automated diagnostic system. In the process of testing a real-time clinical image, we continuously perform four steps: pre-processing, vessel segmentation, vascular characterization, and classification; the pre-processing takes about 150 milliseconds, the segmentation of blood vessels about 300 milliseconds, vascular characterization about 100 milliseconds, and the classification task about 420 milliseconds.

In conclusion, we proposed an end-to-end automatic diagnosis system for FEVR based on deep learning and prior clinical knowledge. Our automatic diagnosis system provides important advice to doctors for better diagnosis and is quite meaningful for early screening and prevention of disease progression. Moreover, the proposed method is completely non-invasive and does not require fluorescein angiography.

ACKNOWLEDGMENT

(Yu Ye and Jianbo Mao contributed equally to this work.)

REFERENCES

- [1] D. Shukla, J. Singh, G. Sudheer, M. Soman, R. K. John, K. Ramasamy, and N. Perumalsamy, "Familial exudative vitreoretinopathy (FEVR). Clinical profile and management," *Indian J. Ophthalmol.*, vol. 51, no. 4, p. 323, 2003.
- [2] V. Criswick and C. Schepens, "Familial exudative vitreoretinopathy," *Amer. J. Ophthalmol.*, vol. 68, no. 4, pp. 578–594, 1969.
- [3] Z. Tauqeer and Y. Yonekawa, "Familial exudative vitreoretinopathy: Pathophysiology, diagnosis, and management," *Asia-Pacific J. Ophthalmol.*, vol. 7, no. 3, pp. 176–182, 2018.
- [4] J. Z. Kartzchner and M. E. Hartnett, "Familial exudative vitreoretinopathy presentation as persistent fetal vasculature," *Amer. J. Ophthalmol. Case Rep.*, vol. 6, pp. 15–17, Jun. 2017.
- [5] A. Nagiel, R. A. Lalane, S. R. Sadda, and S. D. Schwartz, "Ultra-widefield fundus imaging: A review of clinical applications and future trends," *Retina*, vol. 36, no. 4, pp. 660–678, 2016.
- [6] X. Chen, "Development of retinal and choroidal vasculature and peripheral retina," in *Handbook of Pediatric Retinal Oct and the Eye-Brain Connection E-Book*. Amsterdam, The Netherlands: Elsevier, 2019, p. 63.
- [7] M. Takahashi, T. Yokoi, S. Katagiri, T. Yoshida-Uemura, S. Nishina, and N. Azuma, "Surgical treatments for fibrous tissue extending to the posterior retina in eyes with familial exudative vitreoretinopathy," *Jpn. J. Ophthalmol.*, vol. 62, no. 1, pp. 63–67, 2018.
- [8] P. Chudzick, S. Majumdar, F. Calivá, B. Al-Diri, and A. Hunter, "Microaneurysm detection using fully convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 158, pp. 185–192, May 2018.
- [9] R. Zheng, L. Liu, S. Zhang, C. Zheng, F. Bunyak, R. Xu, B. Li, and M. Sun, "Detection of exudates in fundus photographs with imbalanced learning using conditional generative adversarial network," *Biomed. Opt. Express*, vol. 9, no. 10, pp. 4863–4878, Oct. 2018.
- [10] P. M. Burlina, N. Joshi, M. Pekala, K. Pacheco, D. Freund, and N. Bressler, "Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks," *JAMA Ophthalmol.*, vol. 135, no. 11, pp. 1170–1176, Nov. 2017.
- [11] H. Muhammad, T. J. Fuchs, N. C. De, C. M. De, D. M. Blumberg, J. M. Liebmann, R. Ritch, and D. C. Hood, "Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects," *J. Glaucoma*, vol. 26, no. 12, pp. 1086–1094, 2017.
- [12] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 1015–1018.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [14] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1492–1500.
- [15] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.
- [16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [17] S. Zhang, R. Zheng, Y. Luo, X. Wang, J. Mao, C. J. Roberts, and M. Sun, "Simultaneous arteriole and venule segmentation of dual-modal fundus images using a multi-task cascade network," *IEEE Access*, vol. 7, pp. 57561–57573, 2019.
- [18] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 96–108, Nov. 2017.
- [19] Y. Xie, Y. Xia, J. Zhang, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT," *Inf. Fusion*, vol. 42, pp. 102–110, Jul. 2018.
- [20] Y. Zhu, Z. Lan, S. Newsam, and A. Hauptmann, "Hidden two-stream convolutional networks for action recognition," in *Proc. Asian Conf. Comput. Vis. Cham, Switzerland: Springer*, 2018, pp. 363–378.
- [21] D. Wu, L. Pigou, P.-J. Kindermans, N. D.-H. Le, L. Shao, J. Dambre, and J.-M. Odobez, "Deep dynamic neural networks for multimodal gesture segmentation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1583–1597, Aug. 2016.
- [22] Y.-R. Cho, S. Shin, S.-H. Yim, K. Kong, H.-W. Cho, and W.-J. Song, "Multistage fusion with dissimilarity regularization for SAR/IR target recognition," *IEEE Access*, vol. 7, pp. 728–740, 2018.
- [23] World Medical Association, "World medical association declaration of helsinki. ethical principles for medical research involving human subjects," *Bull. World Health Org.*, vol. 79, no. 4, p. 373, 2001.
- [24] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, "Automatic contrast-limited adaptive histogram equalization with dual gamma correction," *IEEE Access*, vol. 6, pp. 11782–11792, 2018.
- [25] H. Masumoto, H. Tabuchi, S. Nakakura, H. Ohsugi, H. Enno, N. Ishitobi, E. Ohsugi, and Y. Mitamura, "Accuracy of a deep convolutional neural network in detection of retinitis pigmentosa on ultrawide-field images," *PeerJ*, vol. 7, p. e6900, May 2019.
- [26] J. Zhang, G. Chen, and Z. Jia, "An image stitching algorithm based on histogram matching and sift algorithm," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 4, 2017, Art. no. 1754006.
- [27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent. Cham, Switzerland: Springer*, 2015, pp. 234–241.
- [28] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "BDD100k: A diverse driving video database with scalable annotation tooling," 2018, *arXiv:1805.04687*. [Online]. Available: <https://arxiv.org/abs/1805.04687>
- [29] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, 2017.
- [30] J. Lyu, Q. Zhang, S.-Y. Wang, Y.-Y. Chen, Y. Xu, and P.-Q. Zhao, "Ultra-wide-field scanning laser ophthalmoscopy assists in the clinical detection and evaluation of asymptomatic early-stage familial exudative vitreoretinopathy," *Graefe's Arch. Clin. Experim. Ophthalmol.*, vol. 255, no. 1, pp. 39–47, 2017.
- [31] Z. Oztas, C. Akkin, S. Nalcaci, O. Ilim, and F. Afrashi, "Branch retinal vein occlusion: The importance of the topographical distribution of retinal vessels among risk factors," *Eye*, vol. 31, no. 5, p. 726, 2017.
- [32] N. T. Miloevic, "Fractal analysis of two dimensional images: Parameters of the space-filling and shape," in *Proc. IEEE 20th Int. Conf. Control Syst. Comput. Sci.*, May 2015, pp. 539–544.

[33] N. Milošević, “The morphology of the brain neurons: Box-counting method in quantitative analysis of 2D image,” in *The Fractal Geometry of the Brain*. New York, NY, USA: Springer, 2016, pp. 109–126.

[34] N. Rajković, B. Krstošić, and N. Milošević, “Box-counting method of 2d neuronal image: Method modification and quantitative analysis demonstrated on images from the monkey and human brain,” *Comput. Math. Methods Med.*, vol. 2017, May 2017, Art. no. 8967902.

[35] S. B. Wang, P. Mitchell, G. Liew, T. Y. Wong, K. Phan, A. Thiagalingam, N. Joachim, G. Burlutsky, and B. Gopinath, “A spectrum of retinal vasculature measures and coronary artery disease,” *Atherosclerosis*, vol. 268, pp. 215–224, Jan. 2018.

[36] X.-Y. Huang, H. Zhuang, J.-H. Wu, J.-K. Li, F.-Y. Hu, Y. Zheng, L. C. A. M. Tellier, S.-H. Zhang, F.-J. Gao, and J.-G. Zhang, “Targeted next-generation sequencing analysis identifies novel mutations in families with severe familial exudative vitreoretinopathy,” *Mol. Vis.*, vol. 23, p. 605, Aug. 2017.

[37] J. M. Robitaille, B. Zheng, K. Wallace, M. J. Beis, C. Tatlidil, J. Yang, T. G. Sheidow, L. Siebert, A. V. Levin, and W.-C. Lam, “The role of frizzled-4 mutations in familial exudative vitreoretinopathy and coats disease,” *Brit. J. Ophthalmol.*, vol. 95, no. 4, pp. 574–579, 2011.

[38] G. Iarossi, M. Bertelli, P. E. Maltese, E. Gusson, G. Marchini, A. Bruson, S. Benedetti, S. Volpetti, G. Catena, and L. Buzzonetti, “Genotype-phenotype characterization of novel variants in six italian patients with familial exudative vitreoretinopathy,” *J. Ophthalmol.*, vol. 2017, Jul. 2017, Art. no. 3080245.

[39] T. Yamane, T. Yokoi, Y. Nakayama, S. Nishina, and N. Azuma, “Surgical outcomes of progressive tractional retinal detachment associated with familial exudative vitreoretinopathy,” *Amer. J. Ophthalmol.*, vol. 158, no. 5, pp. 1049–1055, 2014.

[40] A. H. Kashani, K. T. Brown, E. Chang, K. A. Drenser, A. Capone, and M. T. Trese, “Diversity of retinal vascular anomalies in patients with familial exudative vitreoretinopathy,” *Ophthalmology*, vol. 121, no. 11, pp. 2220–2227, 2014.

[41] J. S. Chang, T. G. Murray, D. J. Hess, B. J. Fallas, and A. M. Berrocal, “Widefield fluorescein angiography based laser treatment in pediatric retinal disease,” *Vis. Pan-Amer. Pan-Amer. J. Ophthalmol.*, vol. 16, no. 1, pp. 20–23, 2017.

[42] M. Long, H. Zhu, J. Wang, and M. I. Jordan, “Deep transfer learning with joint adaptation networks,” in *Proc. 34th Int. Conf. Mach. Learn. (JMLR)*, vol. 70, 2017, pp. 2208–2217.



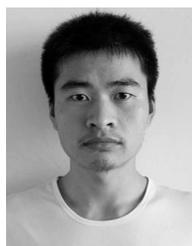
YU YE is currently pursuing the master’s degree with the Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China (USTC), Hefei, China. His major research interests include machine learning, deep learning neural network, and medical image processing and their application in assisted medical diagnosis.



JIANBO MAO is currently a Physician with the Department of Ophthalmology, Eyesight Hospital, Wenzhou Medical University. The medical direction of him is retinopathy of prematurity, macular degeneration, diabetic retinopathy, fundus laser, and angiography.



LEI LIU received the Ph.D. degree from the Department of Electronic Engineering and Information Science, University of Science and Technology of China (USTC), Hefei, China, in 2019. He is currently a Postdoctoral Researcher with the Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China (USTC). His current research interests include deep learning, statistical learning, evolutionary computation, computer vision, and medical image processing and their applications in industry.



SHULIN ZHANG is currently pursuing the Ph.D. degree with the Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China (USTC), Hefei, China. His major research interests include deep learning, intelligent information processing, computer vision, and medical image processing and their applications in industry.



LIJUN SHEN is currently the Vice President of the Department of Ophthalmology, Eyesight Hospital, Wenzhou Medical University, major in vitreoretinal diseases and pediatric vitreoretinal diseases. She is skilled in the surgical treatment of macular diseases, diabetic retinopathy, high myopia fundus diseases, and complex retinal detachment. She is managing two National Science and Technology Major Project and several national projects. She has published more than 100 articles on SCI, such as Ophthalmology, Retina, and other authoritative journals.



MINGZHAI SUN received the Ph.D. degree in physics from the University of Missouri, USA, in 2008. He is currently a Professor with the Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China (USTC), Hefei, China. His research interests include optical super-resolution microscopy, biomedical device development, and bioimage informatics, particularly on medical image analysis and deep learning-based algorithm development.

...