Heliyon 10 (2024) e33108

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Automated interpretation of retinal vein occlusion based on fundus fluorescein angiography images using deep learning: A retrospective, multi-center study

Shenyu Huang [a,1], Kai Jin [a,1], Zhiyuan Gao [a,1], Boyuan Yang [b], Xin Shi [a], Jingxin Zhou [a], Andrzej Grzybowski [c,**], Maciej Gawecki [d,e], Juan Ye [a,*]

[a] *Eye Center, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Zhejiang Provincial Key Laboratory of Ophthalmology, Zhejiang Provincial Clinical Research Center for Eye Diseases, Zhejiang Provincial Engineering Institute on Eye Diseases, Hangzhou, Zhejiang, China*
[b] *Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA*
[c] *Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland*
[d] *Department of Ophthalmology of Specialist Hospital in Chojnice, Lesna 10, 89-600, Chojnice, Poland*
[e] *Dobry Wzrok Ophthalmological Clinic, Zabi Kruk 10, 80-402, Gdańsk, Poland*

ARTICLE INFO

Keywords:
Retinal vein occlusion
Deep learning
Fundus fluorescein angiography
Multi-label
Multi-center

ABSTRACT

Purpose: Fundus fluorescein angiography (FFA) is the gold standard for retinal vein occlusion (RVO) diagnosis. This study aims to develop a deep learning-based system to diagnose and classify RVO using FFA images, addressing the challenges of time-consuming and variable interpretations by ophthalmologists.

Methods: 4028 FFA images of 467 eyes from 463 patients were collected and annotated. Three convolutional neural networks (CNN) models (ResNet50, VGG19, InceptionV3) were trained to generate the label of image quality, eye, location, phase, lesions, diagnosis, and macular involvement. The performance of the models was evaluated by accuracy, precision, recall, F-1 score, the area under the curve, confusion matrix, human-machine comparison, and Clinical validation on three external data sets.

Results: The InceptionV3 model outperformed ResNet50 and VGG19 in labeling and interpreting FFA images for RVO diagnosis, achieving 77.63%–96.45% accuracy for basic information labels and 81.72%–96.45% for RVO-relevant labels. The comparison between the best CNN and ophthalmologists showed up to 19% accuracy improvement with the inceptionV3.

Conclusion: This study developed a deep learning model capable of automatically multi-label and multi-classification of FFA images for RVO diagnosis. The proposed system is anticipated to serve as a new tool for diagnosing RVO in places short of medical resources.

1. Introduction

Retinal vein occlusion (RVO) is defined as a blockage of retinal venous circulation. Predominantly, RVO affects the central or hemi-

* Corresponding author.
** Corresponding author.
E-mail addresses: ae.grzybowski@gmail.com (A. Grzybowski), yejuan@zju.edu.cn (J. Ye).
[1] These authors contributed equally to this work and share the first authorship.

https://doi.org/10.1016/j.heliyon.2024.e33108
Received 30 November 2023; Received in revised form 13 June 2024; Accepted 14 June 2024
Available online 19 June 2024

central veins of the retina [1]. Based on the obstruction site, RVO can be divided into two main types [2]: branch retinal vein occlusion (BRVO) and central retinal vein occlusion (CRVO). Clinically, RVO is usually characterized by retinal/intraretinal hemorrhages, edema, cotton-wool spots, venous dilations, and tortuosity [3]. As the leading cause of visual acuity loss in developed countries, approximately 16 million people suffer from RVO worldwide and BRVO is about four times more prevalent [4]. The sudden, painless, unilateral loss of vision is a common presentation in RVO patients, and the severity of vision loss is determined by the level of retinal involvement and macular perfusion status.

The diagnosis and management of retinal vein occlusion have evolved significantly over the last three decades [5]. However, it is still crucial to assess retinal ischemia at baseline and during follow-up examinations. Besides clinical examination, ocular multimodal imaging is used to diagnose RVO. Color fundus photography (CFP) captures high-resolution images of the retina and surrounding tissues in color, allowing for the detection of any abnormalities. Optical coherence tomography (OCT) is commonly used to assess macular edema [6], while OCT angiography (OCTA) facilitated a distinct evaluation of the superficial and deep capillary network in RVO [7]. Fundus fluorescein angiography (FFA) is the gold standard for diagnosing RVO [8] as it can evaluate retinal perfusion status, neovascularization (NV), peripheral retinal vascular leakage and macular edema. Advanced imaging techniques have contributed to a better understanding of the disease's pathophysiology, and the treatment options have expanded beyond the traditional laser therapy. In the majority of cases, therapeutic agents like anti-vascular endothelial growth factors (anti-VEGF) and steroid injections are currently favored.

With the development of computer technology, artificial intelligence (AI) has shown significant progress in image-based medical fields such as radiology, dermatology, pathology, and ophthalmology. AI is mostly used in ophthalmology to aid in the diagnosis and treatment of diseases [9] such as diabetic retinopathy (DR) [10], glaucoma [11], and age-related macular degeneration [12]. It is highly probable that AI, machine learning (ML) and deep learning (DL) will assume a pivotal role in the clinical practice of ophthalmology, with potential ramifications for the screening, diagnosis, and monitoring of the primary causes of visual impairment in

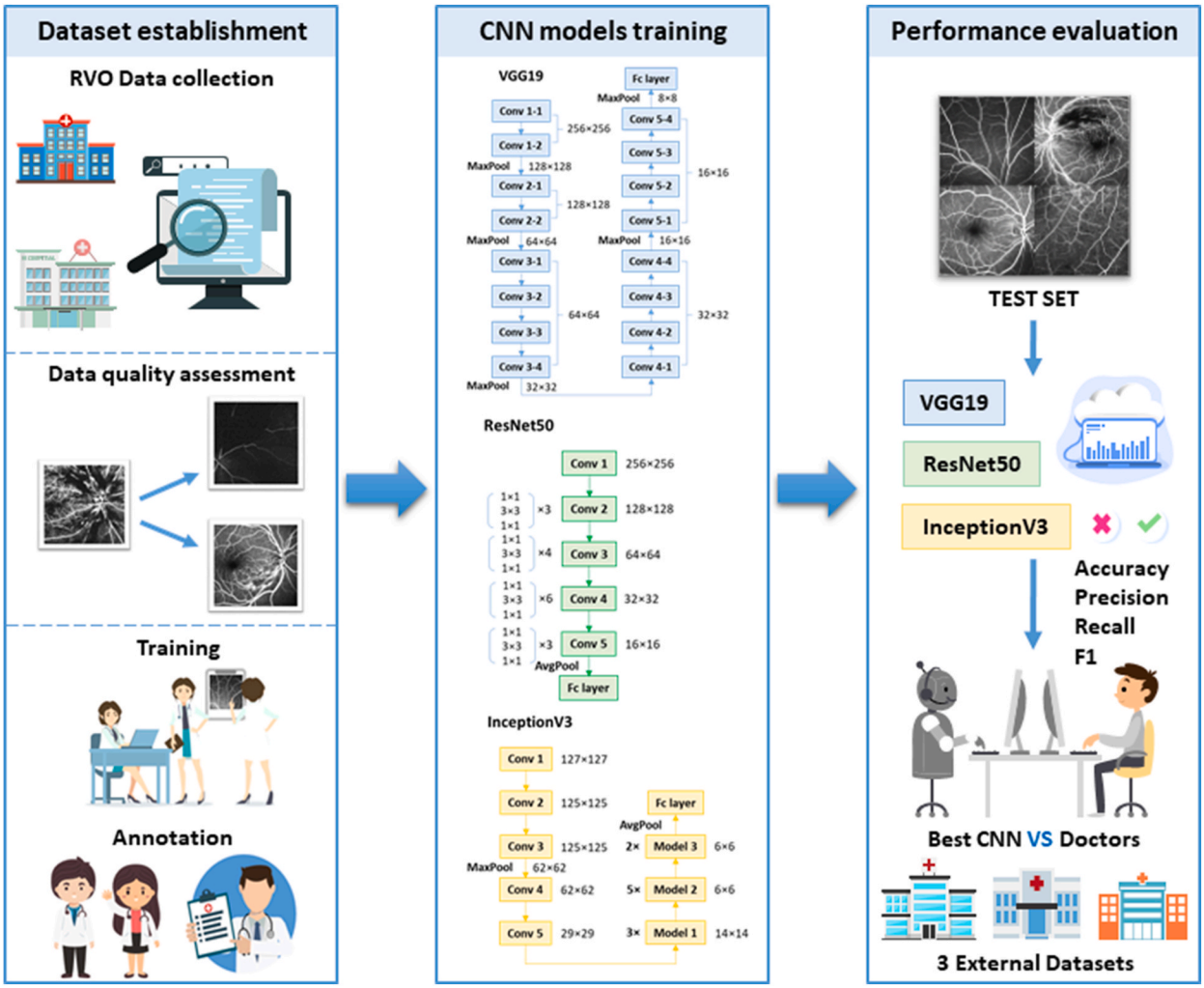

Fig. 1. Illustration of overall study. CNN, convolutional neural network; RVO, Retinal Vein Occlusion.

the context of aging populations worldwide [13]. As for RVO, some studies have used deep learning models to classify and diagnose RVO through fundus images [14–16], but there is still a need for further research on the use of AI to interpret and analyze FFA images for the diagnosis and classification of RVO. FFA is a commonly used medical imaging method that can visually display the situation of retinal vessels, including the location and degree of retinal vein occlusion. However, traditional FFA image analysis methods require manual judgment by doctors, which is easily influenced by subjective factors, and the analysis efficiency is relatively low. Therefore, by integrating deep learning algorithms into the workflow of interpreting FFA images, the costs and time of analysis can be reduced, as well as eliminate human subjectivity in producing objective and reproducible results through quality metrics based on quantitative analysis [17]. Our team had developed a deep learning model for multi-label and multi-classification of FFA images in diabetic retinopathy patients [18].

On the basis of the previous work, the objective of this study is to create an intelligent system based on deep learning and explore its potential for labeling and interpreting FFA images for the diagnosis and categorization of RVO. The novelty of this study lies in its utilization of deep learning, which has not been previously applied in multi-label and multi-classification of FFA images for RVO diagnosis. This novel application allows for automated interpretation of FFA images, which will be the theoretical basis for precise diagnosis of RVO, setting our work apart from conventional studies.

2. Methods

The current study was approved by the ethics committee of the Second Affiliated Hospital, Zhejiang University School of Medicine (No. Y2020-1027) and the Second Affiliated Hospital of Xi'an Jiaotong University had adhered to the principles of the Declaration of Helsinki. Informed consent was obtained from each research subject before study participation. The overall design and methodology of this study is illustrated in Fig. 1.

2.1. Data collection

From 2016 August to 2021 July 4028 FFA images of 467 eyes from 463 patients were collected from the Eye Center at the Second Affiliated Hospital of Zhejiang University and the Department of Ophthalmology at the Second Affiliated Hospital of Xi'an Jiaotong University. The FFA was conducted with the tabletop HRA-II system at 30° (Heidelberg, Germany), at 768x768 pixels.

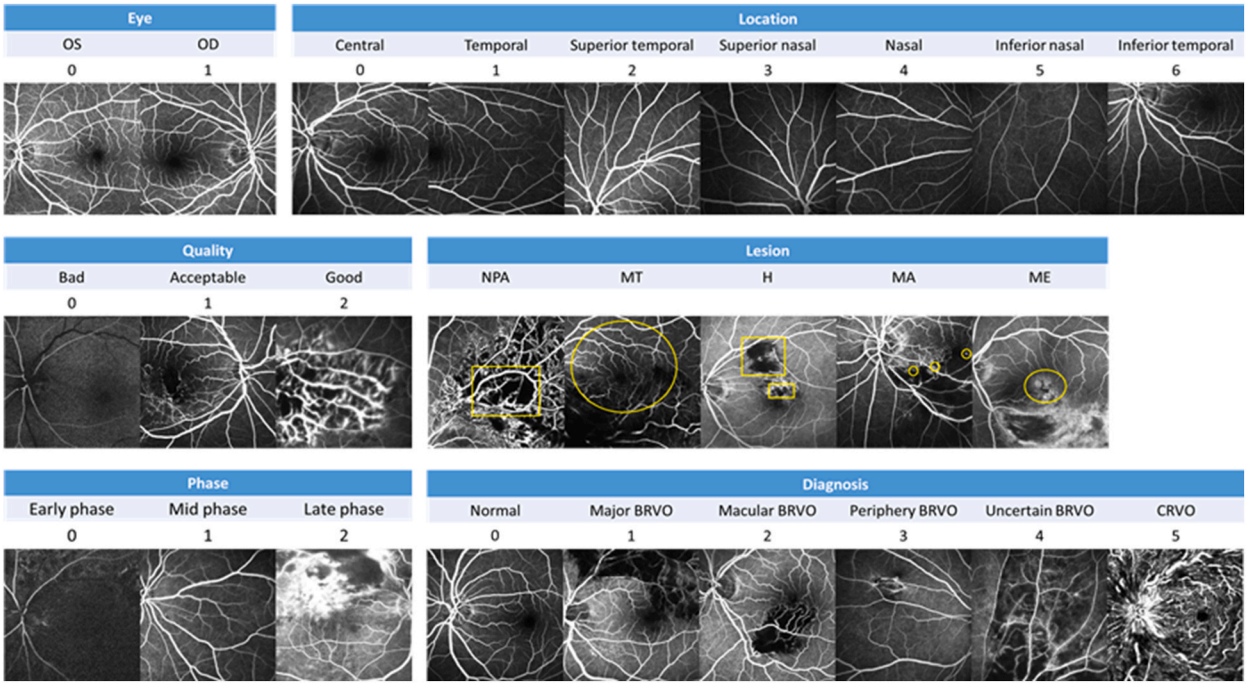

Fig. 2. Examples of representative FFA images. The laterality of eye is either left eye (OS) or right eye (OD). The imaging location is labeled as central, temporal, superior temporal, superior nasal, nasal, inferior nasal and inferior temporal. The image quality is divided into bad, acceptable or good. The imaging phase is divided into early phase, mid-phase or late phase. The lesions include non-perfusion areas (NPA), macular tortuosity (MT), hemorrhage (H), microaneurysms (MA) and macular edema (ME). The diagnosis of image is labeled as normal, major branch retinal vein occlusion (BRVO), macular BRVO, periphery BRVO, uncertain BRVO and central retinal vein occlusion (CRVO).

2.2. Annotation

After thoroughly studying a detailed document of the guidelines for labeling, three ophthalmologists (SH, XS, and JZ) independently annotated FFA images using a straightforward yet effective approach. The images were opened and reviewed using standard image viewing software available on most operating systems, which allowed for clear visualization of the FFA images. For the annotation process itself, we utilized Microsoft Excel, a tool familiar to most researchers, which enabled structured and systematic labeling of the images. In the Excel sheets, we created specific columns for each category relevant to our study.

'Quality' includes good, acceptable, or bad. FFA images which are difficult to distinguish lesions or locations were defined as bad quality and excluded. 'Eye' was divided into left (OS) and right (OD). 'Location' means different areas of the retina. 'Phase' includes early/mid/late phase. The early phase means the arteriovenous phase (30s to 1min), the mid-phase means the venous phase, and the late phase means the post-venous phase (5–10min). As for the lesions, non-perfusion area (NPA) means filling defects, macular tortuosity (MT) refers to the presence of abnormally curved blood vessels in the macula region, 'hemorrhage' represents blocked fluorescence, microaneurysms (MA) refers to tiny dot or clustered fluorescence and macular edema (ME) means the accumulation of fluid in the macula region. Then, students were then tested on a pre-annotated dataset. The diagnosis of FFA image includes normal, major branch retinal vein occlusion (BRVO), macular BRVO, periphery BRVO, uncertain BRVO, and central retinal vein occlusion (CRVO). Macular involvement is also annotated, ensuring a comprehensive assessment of each image. Fig. 2 provides an example of how the annotations appear in our Excel setup. To ensure consistency and accuracy in the annotation process, medical students were trained using a pre-annotated dataset. Each image was labeled by two medical students and then reviewed by an ophthalmologist. The images were evenly distributed among the annotators, and any discrepancies in annotations were resolved by the ophthalmologist, who cast the deciding vote.

2.3. Preprocessing, multi-label and multi-classification model construction

First, all images are resized to 256 × 256 pixels. Then these images are rescaled into the range [−1, 1] with a mean of 0.5 and standard deviation of 0.5 as a conventional method for better model performance. We randomly divide images into training and testing sets with a ratio of 4:1. All images for each eye is kept either in the training data set or in the internal testing dataset.

Multi-classification is a generalization of binary classification problems where a label could have more than two results. While multi-label typically refers to classification problems where multiple labels are targeted at the same time. The convolutional neural networks (CNNs) models have a good capacity for solving these image classification problems. The most commonly utilized deep learning architectures for medical image processing are CNN or specialized CNN frameworks [19] like AlexNet, VGG, inception, and ResNet. In this study, we choose three common models, VGG19 [20], Resnet50 [21], and InceptionV3 [22], to compare their capabilities. The CNN models have been widely used in medical image processing tasks. Ronneberger achieved good segmentation performance on biomedical images [23]. Kisilev used multi-task CNN for the detection and semantic description of lesions in diagnostic images [24]. Ayyachamy built a medical image retrieval framework based on CNN [25].

We use the following setup across three models. The batch size is 64. The optimizer is Adam optimizer with an initial learning rate of 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The learning rate decays exponentially with a rate of 0.95. We choose Cross Entropy Loss as the loss function. The model is trained with Nvidia V100 with 500 epochs.

2.4. Human-machine comparison

To evaluate the performance of machine learning in comparison to human judgment, a sample of 100 images from the test set was randomly selected. In order to provide a comprehensive assessment, three ophthalmologists who were not involved in the earlier labeling process were recruited to assign 11 labels to each image. These labels included factors such as image quality, laterality of the eye, location, Phase, the presence of lesions (NPA, MT, hemorrhage, MA, and ME), diagnosis, and macular involvement. The labels from both the physicians and the best CNN were compared, with the gold standard being derived from the annotation section.

2.5. Clinical validation on three external data sets

This study incorporated three distinct external data sets for analysis. The first data set, designated as the Poland data set, was procured from the Department of Ophthalmology at the University of Warmia and Mazury in Olsztyn, Poland. It was comprised of 196 FFA images (conducted with ZEISS VISUCAM 532) obtained from 16 patients diagnosed with BRVO. The second data set, referred to as the Linfen data set, was assembled from 59 FFA images captured by Topcon during examinations of 22 RVO patients at Linfen People's Hospital in Shanxi, China. The third data set, known as the Ningbo data set, was collected from the Affiliated People's Hospital of Ningbo University in Zhejiang, China, consisting of 98 FFA images (conducted with Heidelberg) sourced from 12 RVO patients. To assess the efficacy of our deep learning model, data extracted from external data sets were subjected to preprocessing and labeling procedures before testing.

Table 1
Statistics of the RVO dataset.

Labels	Quality		Eye		Location					Phase			Lesion										Diagnosis				Macular involvement	
													NPA		MT		H		MA		ME							
	0	1	0	1	0	1	2	3	4	0	1	2	0	1	0	1	0	1	0	1	0	1	0	1	2	3	0	1
Training set (80 %)	1230	1736	1536	1430	1315	386	535	324	406	245	2001	720	823	2143	2479	487	814	2152	2093	873	2801	165	714	853	484	915	1955	1011
Test set (20 %)	288	445	356	377	324	107	134	73	95	54	494	185	202	531	589	144	193	540	483	250	687	46	172	237	153	171	465	268

NPA, non-perfusion areas; MT, macular tortuosity; H, hemorrhage; MA, microaneurysms; ME, macular edema.

3. Results

3.1. Information of dataset and annotation

329 FFA images in this study were ineligible for labeling due to their inadequate image quality. The remaining 3699 images were relabeled as acceptable (0) or good (1). Acceptable images are those that have unsatisfactory image quality with blurs or high noise, but lesions can still be identified. Table 1 presents the annotation of the entire dataset. The laterality of the eye is distributed evenly. In order to balance the dataset, we merged location 2 (superior temporal) and 3 (superior nasal) into superior and relabeled them as 2 (superior). Location 4 is relabeled as 3, location 5 (inferior nasal) and 6 (inferior temporal) are combined as 4 (inferior). Location information reveals that 44.3% of images are focused on the central area. Mid-phase images account for 67.5%, whereas early phase images only account for 8.1%. Regarding lesion distribution, hemorrhage (H) is the most frequently observed (72.8%) among these images, followed by NPA (72.3%). Only 30.4% of images have MA and less than 23% of images were recognized with lesions MT and ME. For the diagnosis label, we adjust diagnosis 2 (macular BRVO), 3 (periphery BRVO), and 4 (uncertain BRVO) into new label 2 (not major BRVO), diagnosis 5 (CRVO) was relabeled as 3. After the adjustment, nearly 30% of images were diagnosed major BRVO or CRVO, normal was 24% and only 17.2% was not major BRVO. Approximately one-third of all images noticed macular involvement.

3.2. Performance of models

Model performance metrics, such as precision, sensitivity/recall, specificity, accuracy, F1-measure, and ROC curves, allow us to determine the best CNN model [26]. As shown in Table 2, three CNNs have the ability to perform 11 classification tasks concurrently. InceptionV3 emerged as the superior model in all aspects among the CNNs, and ResNet50 was close to matching its performance. The accuracy of 'Quality' and 'Phase' tasks was moderate. InceptionV3 outperformed the others for the 'Eye' task with an accuracy of 96.45%. InceptionV3 and ResNet50 had higher accuracy (over 90%) than VGG19 for the 'Location' task. While the three CNNs performed evenly well for the 'NPA', 'H' and 'ME' task, the best accuracy achieved for 'MT' or 'MA' was lower than 85%. For 'Diagnosis' task, InceptionV3 showed best accuracy as 89.22%. For 'Macular involvement' task, Inception V3 and ResNet50 shared same accuracy of 86.63%.

Fig. 3 displays the area under the curve (AUC) columns for each task. InceptionV3 outperformed the other two CNNs. It has better

Table 2
The performance of three CNNs on nine tasks.

Tasks	CNNs	Accuracy	Precision	Recall	F1
Quality	VGG19	0.7381	0.7963	0.7640	0.7798
	ResNet50	0.7722	0.8233	0.7955	0.8091
	InceptionV3	0.7763	0.8486	0.7685	0.8066
Eye	VGG19	0.8104	0.8459	0.7719	0.8072
	ResNet50	0.9304	0.9429	0.9204	0.9315
	InceptionV3	0.9645	0.9606	0.9708	0.9657
Location	VGG19	0.7899	0.7373	0.7335	0.7348
	ResNet50	0.9168	0.9123	0.8910	0.8976
	InceptionV3	0.9318	0.9278	0.9145	0.9175
Phase	VGG19	0.7776	0.6648	0.6343	0.6478
	ResNet50	0.8581	0.7961	0.7737	0.7842
	InceptionV3	0.8649	0.8007	0.8058	0.8030
NPA	VGG19	0.9059	0.9278	0.9435	0.9356
	ResNet50	0.9372	0.9417	0.9736	0.9574
	InceptionV3	0.9372	0.9385	0.9774	0.9576
MT	VGG19	0.7967	0.4797	0.4097	0.4419
	ResNet50	0.8377	0.6404	0.3958	0.4893
	InceptionV3	0.8349	0.6575	0.3333	0.4424
H	VGG19	0.9304	0.9374	0.9704	0.9536
	ResNet50	0.9291	0.9453	0.9593	0.9522
	InceptionV3	0.9441	0.9495	0.9759	0.9626
MA	VGG19	0.7080	0.5957	0.4480	0.5114
	ResNet50	0.7844	0.7644	0.5320	0.6274
	InceptionV3	0.8172	0.7959	0.6240	0.6996
ME	VGG19	0.9454	0.5938	0.4130	0.4872
	ResNet50	0.9618	0.8750	0.4565	0.6000
	InceptionV3	0.9645	0.8571	0.5217	0.6486
Diagnosis	VGG19	0.7572	0.7490	0.7466	0.7430
	ResNet50	0.8554	0.8549	0.8459	0.8430
	InceptionV3	0.8922	0.8955	0.8826	0.8849
Macular involvement	VGG19	0.8145	0.7500	0.7388	0.7444
	ResNet50	0.8663	0.8102	0.8284	0.8192
	InceptionV3	0.8663	0.8244	0.8060	0.8151

NPA, non-perfusion areas; MT, macular tortuosity; H, hemorrhage; MA, microaneurysms; ME, macular edema.

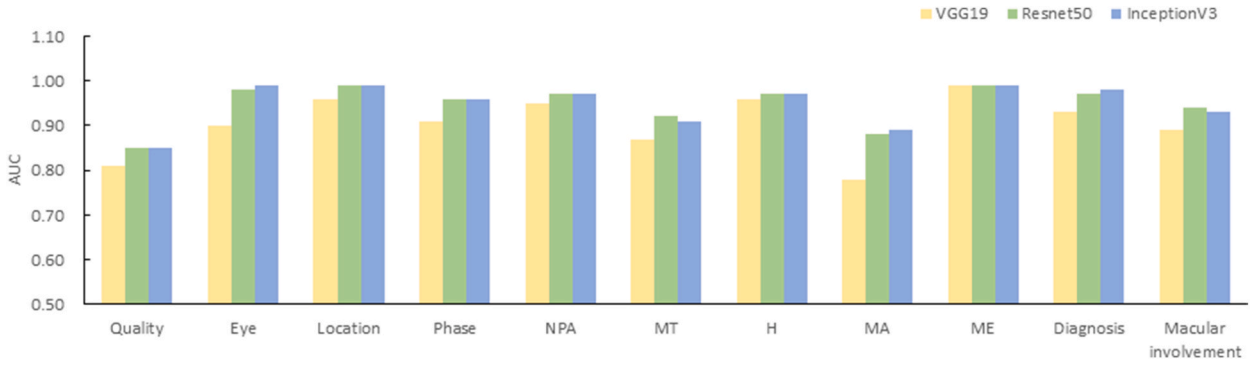

Fig. 3. AUC for each task of three CNNs. AUC, area under the curve; CNN, convolutional neural network; NPA, non-perfusion areas; MT, macular tortuosity; H, hemorrhage; MA, microaneurysms; ME, macular edema.

or similar AUC value in all tasks except 'MT' and 'macular involvement'. AUC for image quality and 'MA' was less than 0.90, while AUC for other tasks exceed 0.90 which indicate our model has strong capability.

The confusion matrices for InceptionV3 are presented in Fig. 4, which includes all tasks. For the 'quality' task, all performance indicators, which are accuracy, precision, recall and F1 score, exceed 76%. It shows a balanced predicting power. For the 'location' task, as we merge several locations into one to create a more balanced dataset, all performance indicators exceed 91%. For 'laterality of eye', the model showed an outstanding performance, as all indicators exceed 96%, suggesting our model is a solid classifier. In the 'phase' task, images with a middle phase are better classified than the early phase and late phase. Regarding lesion detection, NPA and

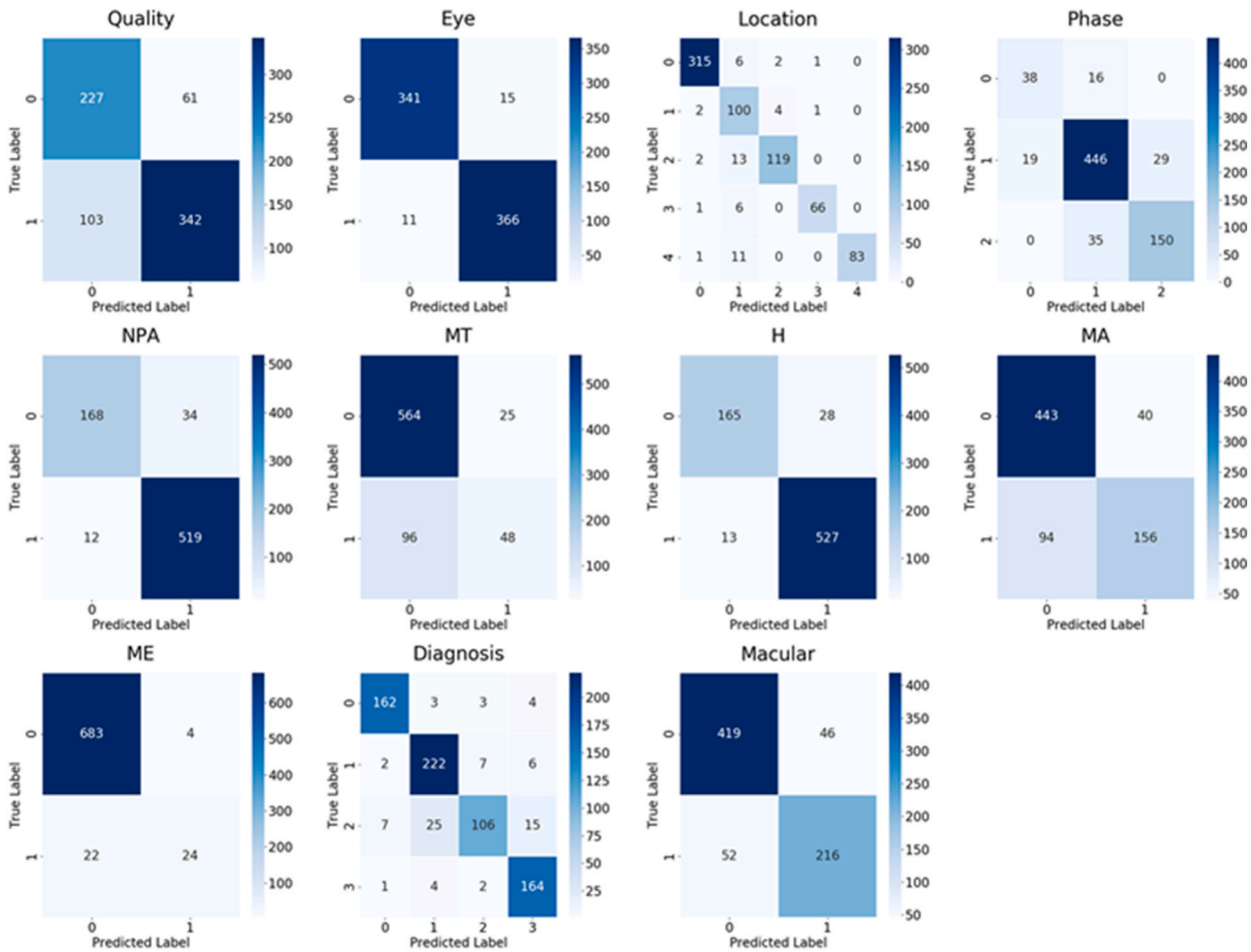

Fig. 4. Confusion matrix for each task of InceptionV3 model.

H detection show a good performance. MT, MA and ME has a relatively low recall value indicates that the model tends to misclassify the positive case, as only 33% MT positive case, 62% MA positive case and 52% ME positive case are identified. As for macular involvement, the model shows an acceptable performance with four indicators above 80%. At last, for the overall diagnosis task, the model has four indicators over 88%. It shows our model has a strong capability on disease diagnosis.

3.3. Human-machine comparison

The result of the human-machine comparison was presented in Fig. 5. We randomly select 100 examples from the test dataset, then compare the predicted labels among the best network (InceptionV3) and ophthalmologists. Doctor 1 and doctor 3 are junior ophthalmologists, and doctor 2 is a senior ophthalmologist. If a label is the same as the ground-truth label, it is shown in green. Otherwise, it is shown in red. With the exception of eye, phase 'MT' and 'MA' labels, InceptionV3 achieved higher accuracies compared to human ophthalmologists across all other labeling categories. The highest accuracy achieved by InceptionV3 was 98%, while the lowest was 77%. For the 'MA' label, InceptionV3 achieved an accuracy of only 77%, which was lower than that of all three human ophthalmologists (78%, 84% and 80%). However, accuracy of InceptionV3 in predicting 'location' was 94%, only slightly higher than that of the highest-performing physician Doctor 2, whose accuracy was 93%.

3.4. External data sets

Our top-performing model, InceptionV3, was utilized to evaluate the three external datasets. The resulting performance metrics, including accuracy, precision, recall, and F1 score, are presented in Table 3. During the external testing using InceptionV3, the accuracy of the model on the quality, eye, location, NPA, H, and ME labels were observed to be similar or higher than the internal test set. Among the three external datasets, the Ningbo dataset had the highest accuracy in most of the tasks. InceptionV3's performance on the Poland dataset showed significantly lower accuracy in the phase (22.67%) and diagnosis (38.95%) tasks compared to the other two datasets. The accuracy of identifying MA lesions is very low in all three datasets, not exceeding 40%. Considering there was no images with ME in Linfen and Ningbo datasets, we excluded DL results for ME on these.

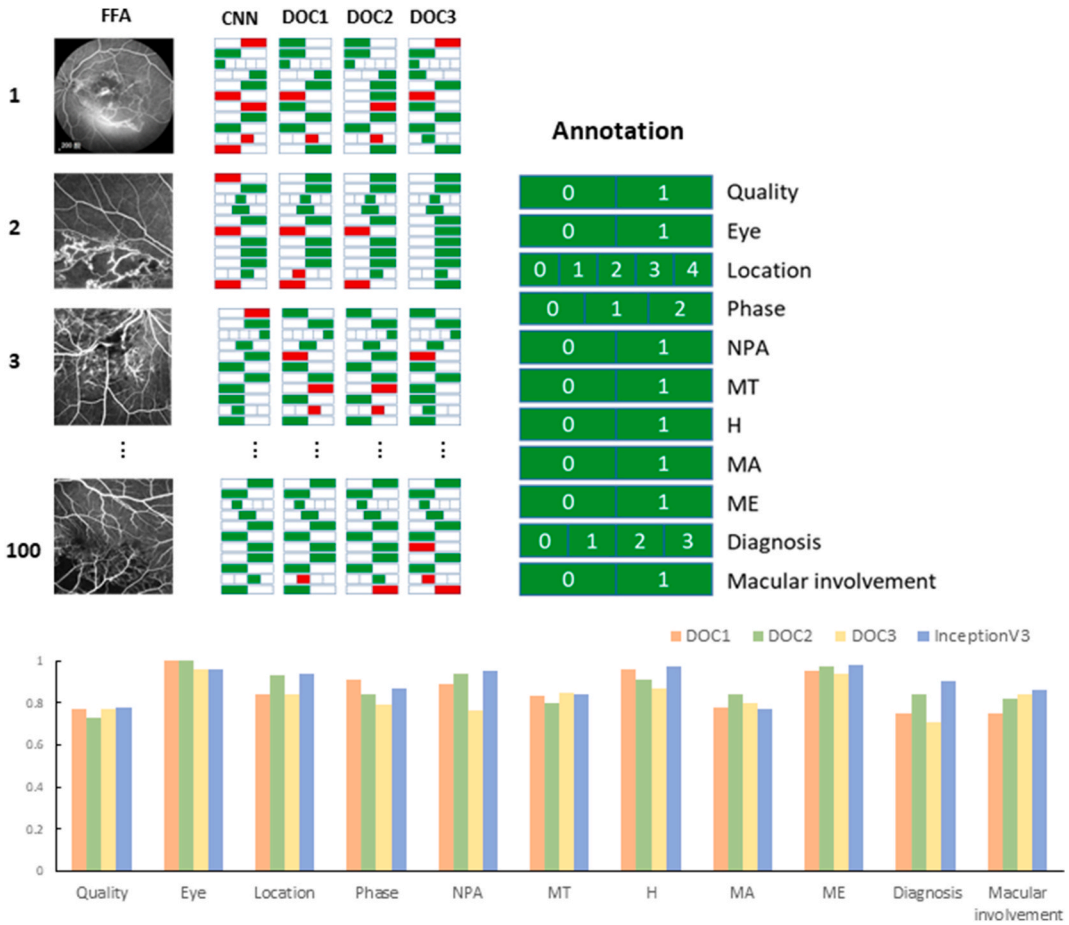

Fig. 5. Comparison between doctors and CNN.

Table 3
The performance of InceptionV3 for three external datasets.

Tasks	Hospitals	Accuracy	Precision	Recall	F1
Quality	Poland	0.8081	0.2857	0.0667	0.1081
	Linfen	0.9231	1.0000	0.2000	0.3333
	Ningbo	0.7634	0.7000	0.4667	0.5600
Eye	Poland	0.9419	0.9811	0.8525	0.9123
	Linfen	0.9231	0.8333	1.0000	0.9091
	Ningbo	0.9140	0.8500	0.9444	0.8947
Location	Poland	0.7849	0.6767	0.6497	0.6037
	Linfen	0.9808	0.5000	0.5000	0.5000
	Ningbo	0.8387	0.7553	0.7983	0.7472
Phase	Poland	0.2267	0.4526	0.4047	0.2258
	Linfen	0.5385	0.5833	0.5775	0.5327
	Ningbo	0.7204	0.7193	0.7144	0.6833
NPA	Poland	0.7326	0.9524	0.7092	0.8130
	Linfen	0.8269	1.0000	0.8269	0.9053
	Ningbo	0.9570	1.0000	0.9512	0.9750
MT	Poland	0.4419	1.0000	0.0400	0.0769
	Linfen	0.6731	1.0000	0.0556	0.1053
	Ningbo	0.5484	1.0000	0.1600	0.2759
H	Poland	0.8198	1.0000	0.7862	0.8803
	Linfen	0.9423	1.0000	0.9423	0.9703
	Ningbo	0.9032	0.9726	0.9103	0.9404
MA	Poland	0.1919	1.0000	0.0142	0.0280
	Linfen	0.3462	0.0000	0.0000	0.0000
	Ningbo	0.3441	1.0000	0.1029	0.1867
ME[a]	Poland	0.9535	0.0000	0.0000	0.0000
	Linfen				
	Ningbo				
Diagnosis	Poland	0.3895	0.2853	0.3849	0.2992
	Linfen	0.8462	0.6883	0.6755	0.6804
	Ningbo	0.8280	0.8041	0.8250	0.7951
Macular involvement	Poland	0.7267	0.8842	0.7000	0.7814
	Linfen	0.7308	1.0000	0.6667	0.8000
	Ningbo	0.7742	0.8889	0.6531	0.7529

NPA, non-perfusion areas; MT, macular tortuosity; H, hemorrhage; MA, microaneurysms; ME, macular edema.
[a] : no ME was found in Linfen or Ningbo dataset.

4. Discussions

Our study aimed to explore the feasibility of creating a deep learning model for RVO classification and diagnosis from FFA images. Due to the shortage of fundus specialists, RVO screening is challenging in areas with limited medical resources. The findings from our study indicate that the deep learning model exhibits high accuracy, sensitivity, and specificity for RVO diagnosis and classification. Additionally, it facilitates large-scale screening for ocular diseases in underserved and rural regions.

4.1. Clinical significance of each label

In this study, four kinds of labels (quality, eye, location, phase) describing image basic information were included, three kinds of labels (lesion, diagnosis, macular involvement) related to RVO were chosen. Image quality is critical in clinical practice, as ophthalmologists rely on imaging data to evaluate diseases, monitor disease progression, and determine the most effective treatment options for their patients. As a result, assessing image quality is a crucial preliminary step in order to ensure that the diagnosis, patient management, and treatment decisions are not compromised, delayed, or impropriate [27]. Image quality is usually labeled as bad or good [18]. However, in our study, we introduced a new label between bad and good, which we defined as 'acceptable'. Images with poor quality but still identifiable location, phase, and lesion are considered as acceptable. This label is more clinically relevant because if the key information required for diagnosis can be identified from the image, the diagnosis can continue even if the image is not very clear. 'Eye' label which means laterality of the eye, is a fundamental but critical piece of information in any ophthalmic report. We selected the label 'location' because RVO classification based on localization [28]. By labeling location, we can determine the exact distribution of the lesion in the fundus of the eye, which can lead to better diagnosis and treatment. The interpretation of FFA is highly dependent on the 'phase'. Vascular leakage is shown in the early or mid-phase of FFA and filling of cystoid spaces in the late phase [29]. Considering the main manifestations of RVO in FFA, our study included 5 labels for 'lesion', which are: include non-perfusion areas (NPA), macular tortuosity (MT), hemorrhage (H), microaneurysms (MA) and macular edema (ME). Neovascularization is associated with a large non-perfusion area (NPA) on FFA. Both CRVO and BRVO can be broadly classified into ischemic and non-ischemic types based on the area of capillary non-perfusion[34], and this distinction is useful for clinical management. However, the absence of quantitative NPA measurements is a notable limitation of this study. Future research should prioritize the development of standardized

protocols for the precise measurement of NPA to improve the diagnostic and therapeutic approaches in RVO. MA can be effectively detected using FFA, but the presence of leakage can sometimes hinder their detection [30]. MT is a risk factor for the development of macular edema. Hemorrhage is a common finding on FFA in patients with RVO, which may indicate that the retinal vein is obstructed, causing rupture of the retinal vessels and resulting in blood leakage into the retinal tissue. This blood leakage can result in retinal tissue swelling and damage, which can lead to vision loss. Therefore, hemorrhage is one of the signs that RVO patients need to be treated early. ME is a common complication of central retinal vein occlusion (CRVO). ME is characterized by the accumulation of fluid in the macula, which can cause visual impairment or even blindness. Anti-vascular endothelial growth factor (VEGF) therapy has been shown to be effective in reducing ME and improving visual acuity in patients with CRVO-associated ME [1]. The 'diagnosis' and 'macular involvement' label will help in selecting an appropriate treatment plan. While visual acuity is indeed a primary criterion for treatment decisions, it is important to note that the overall management strategy must also consider the type of RVO and specific pathological changes such as macular edema, which is a leading cause of vision loss in these patients [35]. While systemic treatments for RVO are generally consistent, the ophthalmological strategies for managing CRVO differ significantly from those applied to BRVO [36], indicating the necessity of tailored ophthalmic interventions based on the diagnosis and detailed ocular findings. The inclusion of macular tortuosity (MT), hemorrhage (H), microaneurysms (MA), and macular edema (ME) in our evaluation, though not traditional markers for occlusion severity, offers a more comprehensive assessment of the retinal condition. Macular Tortuosity (MT) and Hemorrhage (H) are often reflective of underlying venous pressure changes and can indicate worsening vascular integrity. Microaneurysms (MA) signify localized areas of capillary weakness and can precede more severe vascular complications. Most critically, Macular Edema (ME) directly impacts visual acuity and guides the urgency of therapeutic interventions. The assessment of ME is crucial as it influences both prognosis and treatment options available to patients. Collecting data on these parameters may also provide valuable insights for future research, potentially leading to revised or new criteria for occlusion severity based on empirical evidence.

4.2. Model performance

From Table 2 and Fig. 3, it can be concluded that InceptionV3 is the best model for the multi-label and multi-classification tasks in terms of accuracy, F-1 score and AUC. Meanwhile, it has the fastest training time comparing with other two networks. The superiority of the network structure of Inception may make it suitable for complex classification tasks. From Fig. 4, we can try to understand the performance of the model. The quality of the image is relatively subjective task which leads to diversified performance results. In our study, we noted that while the task of determining the laterality of the eye is relatively simple for trained clinicians, achieving comparable accuracy with a neural network presents unique challenges, especially in peripheral images where visual cues might be less distinct. We acknowledged that the current model's performance in determining laterality does not yet match the near-perfect accuracy that clinicians achieve. We analyzed the model's performance limitations and are implementing enhancements to improve its accuracy. Future developments will focus more on complex diagnostic tasks where AI can provide significant value beyond human capabilities.

We specifically compared the performance of InceptionV3 model with the widely-used ResNet18 model, as reported in prior multilabel and multiclassification deep learning model studies [14,37]. For tasks requiring precise localization and detailed area analysis, such as localization and lesions, InceptionV3 achieved accuracies of 93.18% in localization, 93.72% in NPA and 94.41% in hemorrhage, significantly outperforming ResNet18, which achieved 80.79%, 64.10% and 63.37% respectively.

The recall value of MT, MA and ME are relatively low comparing with other indictors. For ME, we only have 46 positive samples in test set, which makes the result less convincing. For MT and MA, the model has difficulty predicting positive samples. It could be the result of image artifacts such as image noise or blurriness. Moreover, in late phase images, the leakage makes it difficult to recognize MA or MT. It's worth noting that doctors also struggle in these tasks. In clinical practice, doctors often rely on the combination of FFA images from different phases to make judgments, while deep learning models only analyze a single image at a time.

The results from the human-machine comparison suggest that InceptionV3 outperformed the three ophthalmologists in terms of accuracy for most of the labels. However, it should be noted that the doctors' accuracy for the 'quality' and 'diagnosis' were relatively lower than the model. These findings highlight the potential of deep learning models in assisting medical professionals in image analysis tasks, but also emphasize the need for careful evaluation and continued refinement to ensure optimal performance. The inferior performance of doctors compared to deep learning networks may be attributed to the subjective nature of doctors' judgment criteria for image quality and the need for doctors to integrate additional patient information or more images for diagnosis. Our findings from external datasets suggest that InceptionV3 has promising performance for detecting various features in FFA images, but its accuracy may be influenced by specific FFA dataset characteristics and the complexity of certain tasks.

In this study, we explored the discriminative performance of neural networks (NNs) in diagnosing RVO including the assessment of laterality and macular involvement. While the neural network exhibited promising results, we recognize that from a clinical standpoint, the practical demand for defining laterality or macular involvement may be limited. This feedback highlights a crucial aspect of our research and suggests that the focus on such parameters, while scientifically valid, might not translate directly into clinical necessity.

4.3. Limitations and future directions

In recent years, deep learning (DL) has shown remarkable success in ophthalmology [31], particularly in disease screening and diagnosis. The potential benefits of DL include high working efficiency and low cost of human labour. DL models have been previously

developed for detecting various fundus diseases [32], assessing image quality [17] and distinguishing retinal lesions in FFA images [33]. Our study aimed to develop a standardized DL model to automatically analyze FFA images of RVO, highlighting the need for improved data diversity and dynamic analysis in future models. The model was developed using over 4000 FFA images with meaningful and clinically relevant labels. The network structure was relatively simple, making it accessible and promising for clinical application. Our model can help improve the diagnostic ability of junior ophthalmologists and homogenize the level of medical care in different areas, though further validation and adaptation in diverse clinical settings are needed. Moreover, our work may pave the way for future FFA report generation, as a clinical FFA report generally consists of the labels we proposed in our study.

Deep learning has made significant progress in medical image processing, but there are still challenges and limitations. One of the challenges is the reliability and interpretability of deep learning algorithms. The "black-box" nature of deep learning algorithms makes it difficult to explain their decision-making processes and results. In addition, deep learning algorithms face challenges related to data collection and privacy protection. Obtaining large amounts of high-quality medical image data and protecting patient privacy are important challenges for the application of deep learning algorithms in medical image processing. The interpretability of deep learning (DL) systems is crucial in healthcare, as it is not only the quantitative algorithmic performance that matters, but also the underlying features used to classify diseases that can improve physician acceptance [31]. However, it is important to note that the concept of 'interpretability' may differ between healthcare professionals and machine learning experts. Despite the potential for interpretable algorithms to be more easily accepted by ophthalmologists, further research is needed to determine whether this leads to measurable improvements in clinical effectiveness for patients. Despite the challenges and limitations, the development of deep learning models in medical image processing has opened up new possibilities for medical research and clinical practice. To further improve the accuracy and reliability of deep learning algorithms for the interpretation of FFA images of RVO, integrating with advanced technologies such as multimodal image data, adaptive learning, and weakly supervised learning could be promising directions. These technologies can potentially enhance the diagnostic and treatment outcomes of RVO.

Besides, our method does have some limitations. Currently, our training dataset consists of FFA images from only two centers, which limits its diversity in terms of ethnic backgrounds and imaging equipment used. Additionally, the type of lesion is restricted. To mitigate these constraints and enhance the robustness of our model, future research should aim to broaden the dataset to encompass a wider array of ophthalmologic diseases and incorporate images from a more diverse range of demographic and geographic sources. Furthermore, our methodology involved labeling separate images rather than the complete series of FFA images from a single patient, which restricts our ability to analyze temporal changes and dynamic information inherent in FFA sequences. Future developments could focus on creating methodologies that can process these dynamic aspects to provide more comprehensive diagnostic insights. Another limitation arises from the traditional FFA's relatively limited field of view, which can lead to inaccuracies in defining the ischemic range and subsequently affect therapeutic decision-making. The advent of ultra-wide-angle fundus fluorescein angiography (UWFFA) presents an opportunity to overcome this limitation. We recommend that future studies incorporate UWFFA to enhance the observation of the peripheral retina, potentially improving the accuracy and efficacy of FFA-based diagnostics.

In conclusion, while our study provides a promising framework for automated interpretation and clinical evaluation of FFA images of RVO, there is significant scope for further refinement. Future research should not only focus on expanding the dataset and improving the methodology but also explore the integration of advanced imaging technologies like UWFFA to broaden the clinical utility and diagnostic accuracy of our approach.

5. Conclusion

In this study, a deep learning model was developed for the automated interpretation of FFA images for RVO diagnosis. The model is highly consistent with clinical diagnosis in a series of labels and effectively identifies the presence of RVO, classifying it according to the site of occlusion. While the proposed system promises to mitigate the impact of medical resource shortages by providing a new tool for precise RVO diagnosis, it also presents several challenges that must be addressed. The accuracy of the model is currently limited by the diversity and size of the training datasets. Looking forward, there is a critical need to expand our datasets to include a broader spectrum of retinal diseases and to source data from a wider variety of centers. This expansion will not only improve the model's robustness but also enhance its applicability across different demographic groups. Additionally, future research should explore the development of automated FFA report generation, which could further streamline the diagnostic process and support clinicians in delivering timely and accurate patient care. Furthermore, we aim to integrate more comprehensive diagnostic features, such as assessing disease progression and response to treatment, which could profoundly impact clinical decision-making and patient management. These endeavors will pave the way for more sophisticated diagnostic tools in ophthalmology, potentially transforming patient outcomes in RVO and beyond.

Funding

This work has been financially supported by National Natural Science Foundation of China (grant number 82201195), Natural Science Foundation of Zhejiang Province (LQ21H120002), Medical and Health Science and Technology Program of Zhejiang Province (grant number 2021RC064), Clinical Medical Research Center for Eye Diseases of Zhejiang Province (grant number 2021E50007), National Natural Science Foundation Regional Innovation and Development Joint Fund (U20A20386), National Natural Science Foundation of China (82330032).

Authorship

All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Compliance with ethics guidelines

This study adhered to the tenets of the Declaration of Helsinki, and the protocol is obtained from the Ethics Committee of the Second Affiliated Hospital, Zhejiang University School of Medicine (No. Y2020-1027).

Data availability statement

The datasets generated during and/or analyzed during the current study are available from corresponding authors on reasonable request.

CRediT authorship contribution statement

Shenyu Huang: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Kai Jin:** Writing – review & editing, Formal analysis. **Zhiyuan Gao:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Boyuan Yang:** Writing – review & editing, Software, Methodology, Formal analysis. **Xin Shi:** Methodology, Formal analysis, Data curation. **Jingxin Zhou:** Methodology, Formal analysis, Data curation. **Andrzej Grzybowski:** Writing – review & editing, Supervision, Conceptualization. **Maciej Gawecki:** Writing – review & editing, Supervision, Conceptualization. **Juan Ye:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the Second Affiliated Hospital of Xi'an Jiaotong University (Shanxi, China), the Department of Ophthalmology at the University of Warmia and Mazury (Olsztyn, Poland), Linfen People's Hospital (Shanxi, China), the Affiliated People's Hospital of Ningbo University (Zhejiang, China) for their contribution in dataset construction.

References

[1] L. Nicholson, S.J. Talks, W. Amoaku, K. Talks, S. Sivaprasad, Retinal vein occlusion (RVO) guideline: executive summary, Eye (Lond). 36 (5) (2022) 909–912.
[2] I.U. Scott, P.A. Campochiaro, N.J. Newman, V. Biousse, Retinal vascular occlusions, Lancet 396 (10266) (2020) 1927–1940.
[3] T.Y. Wong, I.U. Scott, Clinical practice. Retinal-vein occlusion, N. Engl. J. Med. 363 (22) (2010) 2135–2144.
[4] S. Rogers, R.L. McIntosh, N. Cheung, et al., The prevalence of retinal vein occlusion: pooled data from population studies from the United States, Europe, Asia, and Australia, Ophthalmology 117 (2) (2010) 313–319 e1.
[5] F. Romano, F. Lamanna, P.H. Gabrielle, et al., Update on retinal vein occlusion, Asia Pac J Ophthalmol (Phila). 12 (2) (2023) 196–210.
[6] I. Voo, E.C. Mavrofrides, C.A. Puliafito, Clinical applications of optical coherence tomography for the diagnosis and management of macular diseases, Ophthalmol Clin North Am 17 (1) (2004) 21–31.
[7] N. Ben Abdesslem, S. Haddar, A. Mahjoub, et al., Retinal vein occlusions: an OCT- Angiography analysis, Tunis. Med. 99 (5) (2021) 538–543.
[8] W. An, J. Han, Research progress of UWFFA and OCTA in retinal vein occlusion: a review, Eur. J. Ophthalmol. 31 (6) (2021) 2850–2855.
[9] K. Jin, J. Ye, Artificial intelligence and deep learning in ophthalmology: current status and future perspectives, Advances in Ophthalmology Practice and Research 2 (3) (2022) 100078.
[10] Z. Gao, K. Jin, Y. Yan, et al., End-to-end diabetic retinopathy grading based on fundus fluorescein angiography images using deep learning, Graefes Arch. Clin. Exp. Ophthalmol. 260 (5) (2022) 1663–1673.
[11] X. Huang, K. Jin, J. Zhu, et al., A structure-related Fine-grained deep learning system with diversity data for universal glaucoma visual field grading, Front. Med. 9 (2022) 832920.
[12] Y. Yan, K. Jin, Z. Gao, et al., Attention-based deep learning system for automated diagnoses of age-related macular degeneration in optical coherence tomography images, Med. Phys. 48 (9) (2021) 4926–4934.
[13] D.S.W. Ting, L. Peng, A.V. Varadarajan, et al., Deep learning in ophthalmology: the technical and clinical considerations, Prog. Retin. Eye Res. 72 (2019) 100759.
[14] W. Xu, Z. Yan, N. Chen, et al., Development and application of an intelligent diagnosis system for retinal vein occlusion based on deep learning, Dis. Markers 2022 (2022) 4988256.
[15] D. Nagasato, H. Tabuchi, H. Ohsugi, et al., Deep neural network-based method for detecting central retinal vein occlusion using ultrawide-field fundus ophthalmoscopy, J Ophthalmol 2018 (2018) 1875431.
[16] D. Nagasato, H. Tabuchi, H. Ohsugi, et al., Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion, Int. J. Ophthalmol. 12 (1) (2019) 94–99.
[17] M. Konig, P. Seebock, B.S. Gerendas, et al., Quality assessment of colour fundus and fluorescein angiography images using deep learning, Br. J. Ophthalmol. 108 (1) (2022) 98–104.
[18] Gao Z., Pan X., Shao J., et al., Automatic interpretation and clinical evaluation for fundus fluorescein angiography images of diabetic retinopathy patients by deep learning, Br. J. Ophthalmol. 107(12)(2022)1852-1858.
[19] S. Suganyadevi, V. Seethalakshmi, K. Balasamy, A review on deep learning in medical image analysis, Int J Multimed Inf Retr 11 (1) (2022) 19–38.

[20] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv:1409.556, https://ui.adsabs.harvard.edu/abs/2014arXiv1409.1556S. (Accessed 1 September 2014).
[21] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, 2015 arXiv:1512.03385, https://ui.adsabs.harvard.edu/abs/2015arXiv151203385H. (Accessed 1 December 2015).
[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Ieee. Rethinking the inception architecture for computer vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826. Seattle, WA.
[23] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional Networks for Biomedical Image Segmentation, Springer International Publishing, Cham, 2015, pp. 234–241.
[24] P. Kisilev, E. Sason, E. Barkan, S. Hashoul, Medical image description using multi-task-loss CNN, in: *2nd International Workshop On Deep Learning in Medical Image Analysis (DLMIA)/1st International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis (LABELS)*. 10008, 2016, pp. 121–129. Athens, GREECE.
[25] S. Ayyachamy, V. Alex, M. Khened, G. Krishnamurthi, Medical image retrieval using ResNet-18, in: Conference on Medical Imaging - Imaging Informatics for Healthcare, Research, and Applications, 2019, 10954. San Diego, CA.
[26] M. Togacar, B. Ergen, Z. Comert, COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches, Comput. Biol. Med. 121 (2020) 103805.
[27] X. Wang, S. Zhang, X. Liang, C. Zheng, J. Zheng, M. Sun, A cnn-based retinal image quality assessment system for teleophthalmology, J. Mech. Med. Biol. 19 (5) (2019) 1950030.
[28] M.J. Turczynska, P. Krajewski, J.E. Brydak-Godowska, Wide-field fluorescein angiography in the diagnosis and management of retinal vein occlusion: a retrospective single-center study, Med Sci Monit 27 (2021) e927782.
[29] U. Schmidt-Erfurth, J. Garcia-Arumi, B.S. Gerendas, et al., Guidelines for the management of retinal vein occlusion by the European society of retina specialists (EURETINA), Ophthalmologica 242 (3) (2019) 123–162.
[30] D.V. Weinberg, A.E. Wahle, M.S. Ip, et al., Score study report 12: development of venous collaterals in the score study, Retina 33 (2) (2013) 287–295.
[31] D.S.W. Ting, L.R. Pasquale, L. Peng, et al., Artificial intelligence and deep learning in ophthalmology, Br. J. Ophthalmol. 103 (2) (2019) 167–175.
[32] A.D. Moraru, D. Costin, R.L. Moraru, D.C. Branisteanu, Artificial intelligence and deep learning in ophthalmology - present and future, Exp. Ther. Med. 20 (4) (2020) 3469–3473.
[33] D. Holomcik, P. Seebock, B.S. Gerendas, et al., Segmentation of macular neovascularization and leakage in fluorescein angiography images in neovascular age-related macular degeneration using deep learning, Eye (Lond) 37 (7) (2022) 1439–1444.
[34] B. Tan, Y.C. Sim, J. Chua, D. Yusufi, D. Wong, A.P. Yow, C. Chin, A.C.S. Tan, C.C.A. Sng, R. Agrawal, L. Gopal, R. Sim, G. Tan, E. Lamoureux, L. Schmetterer, Developing a normative database for retinal perfusion using optical coherence tomography angiography, Biomed. Opt. Express 12 (7) (2021) 4032–4045.
[35] EE Cornish, SL Zagora, K Spooner, S Fraser-Bell, Management of macular oedema due to retinal vein occlusion: An evidence-based systematic review and meta-analysis, Clin. Exp. Ophthalmol. 51 (4) (2023) 313–338.
[36] W. Xu, Z. Yan, N. Chen, Y. Luo, Y. Ji, M. Wang, Z. Zhang, Development and application of an intelligent diagnosis system for retinal vein occlusion based on deep learning, Dis. Markers 2022 (2022) 4988256.
[37] Z. Gao, X. Pan, J. Shao, X. Jiang, Z. Su, K. Jin, J. Ye, Automatic interpretation and clinical evaluation for fundus fluorescein angiography images of diabetic retinopathy patients by deep learning, Br. J. Ophthalmol. 107 (12) (2023) 1852–1858.