



Intelligent Diagnosis of Multiple Peripheral Retinal Lesions in Ultra-widefield Fundus Images Based on Deep Learning

Tong Wang · Guoliang Liao · Lin Chen · Yan Zhuang ·
Sibo Zhou · Qiongzen Yuan · Lin Han · Shanshan Wu ·
Ke Chen · Binjian Wang · Junyu Mi · Yunxia Gao · Jiangli Lin ·
Ming Zhang

Received: November 9, 2022 / Accepted: January 5, 2023 / Published online: January 24, 2023
© The Author(s) 2023

ABSTRACT

Introduction: Compared with traditional fundus examination techniques, ultra-widefield fundus (UWF) images provide 200° panoramic images of the retina, which allows better detection of peripheral retinal lesions. The advent of UWF provides effective solutions only for detection but still lacks efficient diagnostic capabilities. This study proposed a retinal lesion

detection model to automatically locate and identify six relatively typical and high-incidence peripheral retinal lesions from UWF images which will enable early screening and rapid diagnosis.

Methods: A total of 24,602 augmented ultra-widefield fundus images with labels corresponding to 6 peripheral retinal lesions and normal manifestation labelled by 5 ophthalmologists were included in this study. An object detection model named You Only Look Once X (YOLOX) was modified and trained to locate and classify the six peripheral retinal lesions including rhegmatogenous retinal detachment (RRD), retinal breaks (RB), white without pressure (WWOP), cystic retinal tuft (CRT), lattice degeneration (LD), and paving-stone degeneration (PSD). We applied coordinate attention block and generalized intersection over union (GIOU) loss to YOLOX and evaluated it for accuracy, sensitivity, specificity, precision, *F1* score, and average precision (AP). This model was able to show the exact location and saliency map of the retinal lesions detected by the model thus contributing to efficient screening and diagnosis.

Results: The model reached an average accuracy of 96.64%, sensitivity of 87.97%, specificity of 98.04%, precision of 87.01%, *F1* score of 87.39%, and mAP of 86.03% on test dataset 1 including 248 UWF images and reached an average accuracy of 95.04%, sensitivity of 83.90%, specificity of 96.70%, precision of

Tong Wang, Guoliang Liao have contributed equally.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40123-023-00651-x>.

T. Wang · L. Chen · Q. Yuan · S. Wu · B. Wang ·
Y. Gao · M. Zhang (✉)
Department of Ophthalmology, West China
Hospital, No. 37 Guoxue Lane, Wuhou District,
Chengdu 610041, Sichuan, People's Republic of
China
e-mail: zhangmingscu0905@163.com;
zhangming@scu.edu.cn

T. Wang · Q. Yuan · S. Wu · B. Wang
West China School of Medicine, Sichuan University,
Chengdu 610041, Sichuan, People's Republic of
China

G. Liao · Y. Zhuang · S. Zhou · L. Han · K. Chen ·
J. Mi · J. Lin (✉)
College of Biomedical Engineering, Sichuan
University, No. 24, Section 1, South 1st Ring Road,
Wuhou District, Chengdu 610065, Sichuan,
People's Republic of China
e-mail: linjlsu@163.com; linjiangli@scu.edu.cn

78.73%, *F1* score of 81.96%, and *mAP* of 80.59% on external test dataset 2 including 586 UWF images, showing this system performs well in distinguishing the six peripheral retinal lesions. **Conclusion:** Focusing on peripheral retinal lesions, this work proposed a deep learning model, which automatically recognized multiple peripheral retinal lesions from UWF images and localized exact positions of lesions. Therefore, it has certain potential for early screening and intelligent diagnosis of peripheral retinal lesions.

Keywords: Peripheral retinal lesion; Ultra-widefield fundus; Deep learning; Object detection; You Only Look Once X

Key Summary Points

Compared with traditional fundus examination techniques, ultra-widefield fundus (UWF) images provide a 200° panoramic image of the retina, which allows better detection of peripheral retinal lesions.

This study proposed the hypothesis that deep learning can be applied to screening of peripheral retinal lesions based on ultra-widefield fundus images. On test dataset 1 with 248 images, the model performed well in distinguishing and locating six kinds of peripheral retinal lesions with average accuracy of 96.64%, sensitivity of 87.97%, specificity of 98.04%, precision of 87.01%, *F1* score of 87.39%, and *mAP* of 86.03%.

This result supports the initial hypothesis and indicates certain potential of the deep learning model in assisting doctors in improving diagnosis efficiency based on ultra-widefield fundus images.

INTRODUCTION

It is reported that over 43 million people are blind and 550 million suffer from vision

impairment [1, 2]. Retinal lesions may be one of the main causes. Peripheral retinal lesions, as a typical part of retinal lesions, have the characteristics of high concealment and risk and can cause irreversible visual loss without early treatment.

There is clinical relevance among the selected lesions, but the urgency of ophthalmic assessment and treatment required for different kinds of peripheral retinal lesions are quite different. Paving-stone degeneration (PSD) indicates degeneration of retinal neurosensory layers only and may require no clinical intervention, but some cases reported relevance between PSD and extensive macular atrophy routines, so routine follow-up is still recommended [3]. White without pressure (WWOP) also seems not urgent when peripheral vascular abnormalities are observed, and WWOP is thought to be associated with retinal break (RB) and rhegmatogenous retinal detachment (RRD) [4]. Cystic retinal tuft (CRT) is believed to not only develop into a retinal break, but may also be a finding that reflects the pathogenesis of RRD [5]. Lattice degeneration (LD) can cause RB and RRD due to fibrosis of the nearby vitreous body, and clinical association between LD and RRD is frequently reported [6] and is especially common among Asian patients [7]. RB can lead to RRD in a short time since a retinal break allows the ingress of fluid from the vitreous cavity to the subretinal space, resulting in retinal separation [8]. Lesions like PSD may require no intervention, while RB requires laser treatment and RRD needs urgent surgical intervention in most cases. Since different peripheral retinal lesions require variable interventions, it is important for the diagnosis of peripheral retinal lesions.

However, traditional fundus examination techniques, such as ocular color Doppler ultrasound, color fundus photography, and optical coherence tomography (OCT), have long had difficulty detecting peripheral retinal lesions due to the limitation of view field and image quality. Therefore, peripheral retinal lesions often quietly develop into serious conditions like retinoschisis and RRD, leading to significant vision loss before being detected. Ultra-widefield fundus (UWF) imaging system is a

relatively new technique that can detect 200° of the retinal region, thus being able to show much more information than traditional fundus examination techniques, making it possible to scan the peripheral retina quickly and non-invasively. However, because of the complexity of UWF images, the interpretation requires high professionalism and is time-consuming; currently, this work can only be done by experienced and professional ophthalmologists, which makes it difficult for more people to benefit from early screening for peripheral retinal lesions [9]. The lack of automatation, reliability, and accurate identification techniques is still an obstacle to applying the UWF imaging system to large-scale screening in clinical practice.

Deep learning (DL) has developed rapidly in recent years and has been widely applied to medical imaging interpretation and intelligent diagnosis, which shows potential for the recognition of thyroid nodules [10], skin cancer [11], COVID-19 [12], and so on. Owing to its high efficiency and accuracy, DL contributes to efficient early screening and diagnosis.

Based on the traditional color fundus photography technique, deep learning has accomplished automated classification of eye diseases [13–16], including retinal lesion of prematurity [17], myopic maculopathy [18], neovascularization [19], and diabetic retinopathy [20, 21]. Studies utilizing a DL model based on UWF images also perform well in the classification of many diseases including diabetic retinopathy (DR), retinal vein occlusion (RVO), age-related macular degeneration (AMD), and so on [22–28]. However, most studies cannot show the specific location of the retinal lesion. Meanwhile, there are very few studies focusing on the peripheral retina.

According to the circumstances above, we proposed a hypothesis: an object detection model, as one branch of DL, could be applied to distinguishing and locating peripheral retinal lesions from UWF images. This study established a dataset with UWF images and modified an object detection model, aiming at distinguishing and locating six different peripheral retinal lesions including WWOP, PSD, CRT, LD, RB, and RRD.

Overall, this work aims to evaluate the performance of the object detection model on UWF images not only to locate peripheral retinal lesions but also to prove that the object detection model can be applied to UWF images.

METHODS

This study adhered to the tenets of the Declaration of Helsinki and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines [29] and was approved by the ethical review committee of the Chinese Center for Disease Control and Prevention, Beijing, China (ChiCTR2100053526), and all the images were desensitized.

A total of > 100,000 original UWF images were taken by an Optos Daytona (P200T) ophthalmoscope from patients undergoing UWF imaging photography at the Department of Ophthalmology, West China Hospital, from January 2016 to December 2021. The inclusion criteria for this study were as follows: (1) images presenting with at least one kind of listed peripheral retinal lesion including WWOP, PSD, LD, CRT, RB, and RRD; (2) patients > 18 years old; (3) no history of previous laser treatment.

Due to the great influence of low-quality images on the DL model performance [30], we established exclusion criteria for this study: (1) blurred images; (2) over-coverage ratio of the eyelid > 30%; (3) abnormal color of the whole

Table 1 Exclusion criteria for images

Exclusion reason	Criteria
Blurred image	Image lacks boundary information and sharpness and does not show meaningful details
Over-coverage of eyelid	Eyelid covering part is > 30% of the full picture, and the image does not show enough information
Abnormal color	Whole image only shows red or green

image. Details of exclusion criteria are shown in Table 1 and Fig. 1.

Since peripheral retinal lesions could hardly be seen by traditional non-invasive examination and were rarely diagnosed without obvious symptoms, we could not directly obtain peripheral retinopathy UWF images with corresponding labels from the clinical imaging system. Moreover, establishment of a DL model strongly depends on precise and reliable data support. Therefore, we invited five ophthalmologists (WT, CL, WSS, YQZ, WBJ) with qualifications and > 2 years' clinical experience and two specialists (ZM and GYX) with > 10 years' clinical experience from West China Hospital, Sichuan University, to establish the dataset including qualified images with corresponding labels.

As shown in Fig. 2, after excluding unqualified images, the remaining UWF images were distributed to five ophthalmologists, who labeled them with an open-source tool called LabelImg. Five ophthalmologists were divided into two groups (WT, YQZ in group A and CL, WBJ, WSS in group B), and the same dataset was labeled independently by both groups. Each image was labeled according to what kind of peripheral retinal lesion it presented and where it was located. Criteria for consistent labeling of one image were: (1) category of lesions is the same; (2) number of rectangles surrounding corresponding lesions is the same; (3) intersection over union (IOU) of corresponding rectangles is > 90%.

Consistent labels on one image were adapted directly; if there was divergence, the image was transferred to specialists (ZM and GYX) for final

judgment. If disagreement remained, consensus on the final judgment was reached after discussion between two specialists. Since the UWF imaging system uses red and green lasers to synthesize the dual-channel pseudo-color image [9], the ophthalmologists first read the red channel of the image, then read the green channel, and finally read the pseudo-color image for the final diagnosis (Fig. 2b). Table 2 showed features of all target retinal lesions in the whole labeling process, which were referred to EyeWiki from the American Academy of Ophthalmology, while normal image was defined as one UWF image without any sign of retinal lesion or disease.

After all the above procedures, we obtained a database consisting of 2622 UWF images with corresponding labels in XML format respectively. Ninety percent of the former database was selected randomly as the training dataset (with 2374 images) and the remaining 10% as test dataset 1 (with 248 images). Test dataset 1 was used for testing only and not utilized during the training procedure.

The quantity and quality of the database have a great influence on the DL model, so we used an open-source tool named Albumentations to randomly augment the training dataset by horizontal flipping, vertical flipping, transposition, scaling, rotation, brightness variation, and contrast variation. For the training dataset, we applied horizontal and vertical flip for all the images, and we adjusted 60% of the images' brightness and contrast randomly. For CRT, WWOP, PSD, and LD, we also applied random scaling and rotating to augment them again. Finally, to rebalance the dataset, we applied

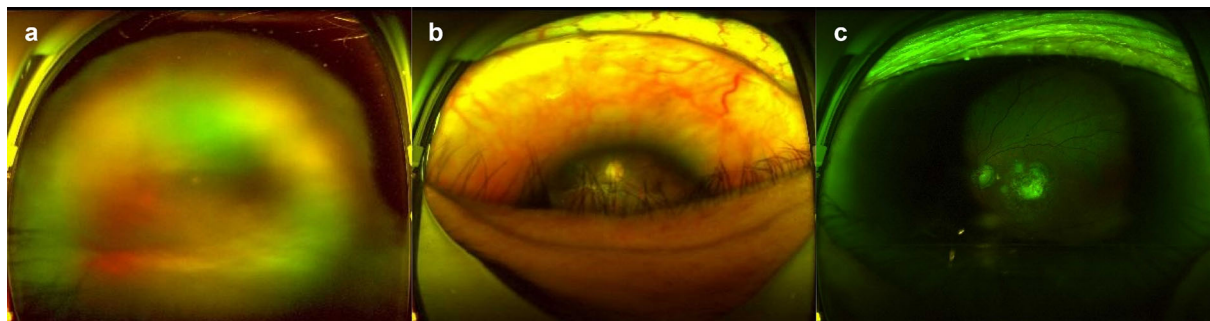


Fig. 1 Typical low-quality examples. **a** Blurred image; **b** over-coverage of eyelid; **c** abnormal color

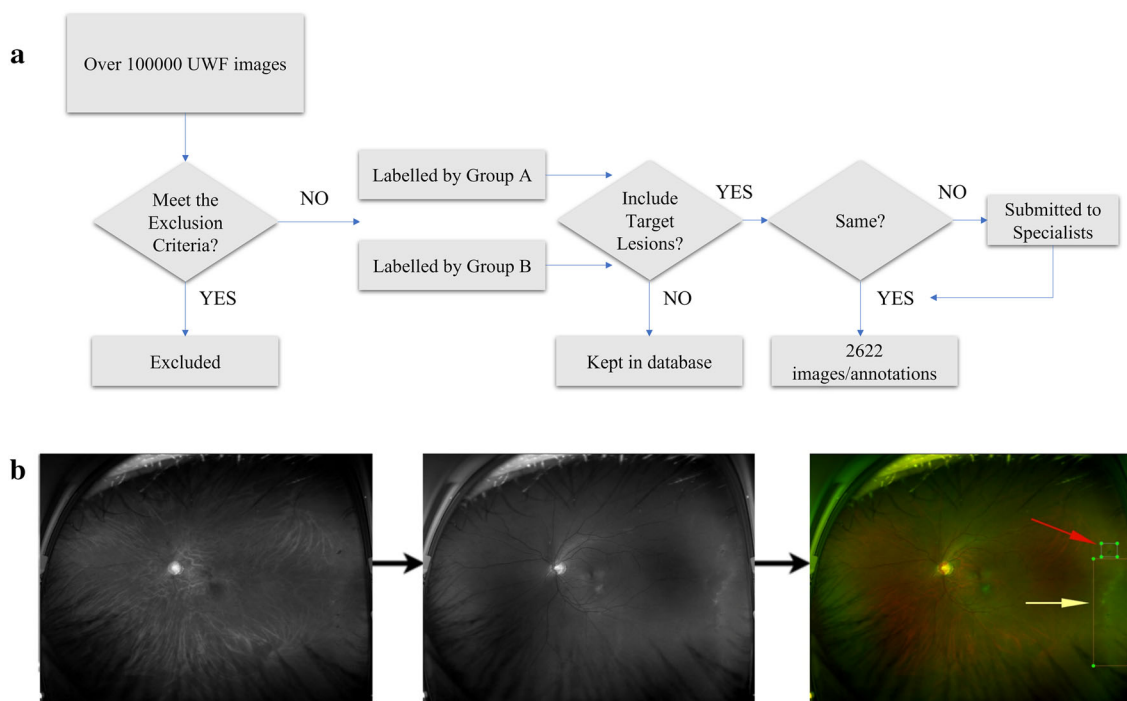


Fig. 2 Labeling process for the dataset used in this study. (a) Establishment of the dataset. (b) Labeling process using LabelImg. As shown above, the retinal lesion area is surrounded by a box (indicated by red and yellow arrows)

transposition. The augmented dataset consists of 24,602 images and 22,601 retinal lesion areas.

To further evaluate the DL model, external test dataset 2 was collected separately from West China Hospital, composed of 122 normal UWF images and 464 UWF images with 600 parts of peripheral retinal lesions. Neither test dataset 1 nor 2 was used for training or augmented, and they were never used in the DL model during the training procedure. Details of the datasets are shown in Table 3.

The details of UWF images are complex because there can be multiple different retinal lesion areas in one UWF image, and the same retinal lesion may also emerge in different parts of the retina, resulting in multiple retinal lesions and multiple retinal lesion areas in one UWF image, like a multi-instance multi-label (MIML, multiple instances and multiple labels in one single image) problem in the research area of machine learning [31, 32]. Under such circumstances, a traditional classification network may not work; meanwhile, the simple use

of a multi-label, multi-instance classification network may have little effect and lack practicability for medical images [33, 34].

To choose our baseline, based on test dataset 1, we tested some state-of-the-art (SOTA) algorithms including SSD, RetinaNet, CenterNet, YOLOv5, Fcos, and YOLOX, and we finally chose YOLOX as the baseline.

In this study, the improved YOLOX, a SOTA object detection model, was innovatively used to locate and classify peripheral retinal lesions. The retinal lesions detected are surrounded by rectangle boxes so that we could see exactly where the retinal lesion areas are.

YOLOX is an improved implementation based on YOLOv3, which is one of the SOTA algorithms in the research area of object detection and has high precision and speed. Compared with DarkNet-19 used in YOLOv2, YOLOv3 adopts DarkNet-53 as the feature extraction network (backbone) and adds a residual network [35], which alleviates the gradient disappearance and explosion problem caused by increasing depth in the deep neural

Table 2 Description of target retinal lesions

Retinal lesion	Features
White without pressure	<p>Definition: distinctive white appearance of the peripheral retina without indentation and without mechanical stimulus</p> <p>Characteristic: whiter than the retina with pressure and the choroidal markings are almost obscured, found in post-equatorial region at the base of the vitreous and ora serrata, whiteness further accentuated with scleral depression, margins are sharply demarcated from normal retina</p>
Paving-stone degeneration	<p>Definition: lesions show multiple rounded, punched-out areas of choroidal and retinal atrophy</p> <p>Characteristic: yellow-white in color due to sclera being partly visible through the atrophic choroid, discrete margins that may be pigmented, may become confluent. Lesions located between ora and equator with size of one to several disc diameters</p>
Lattice degeneration	<p>Definition: retinal thinning with fibrosis and vitreous liquefaction over the lesion</p> <p>Characteristic: oval or linear pattern of lesions, may be one lesion or multiple, may have yellow deposits, pigment, atrophic holes, or retinal breaks</p>
Cystic retinal tuft	<p>Definition: area of retinal degeneration caused by attachment to and pulling of the vitreous</p> <p>Characteristics: circular, cotton ball-like structure adjacent to a dark background</p>
Retinal break	<p>Definition: isolated tear due to a defect in the sensory retina from the traction of the vitreous after acute posterior vitreous detachment</p> <p>Characteristic: either retinal tear or retinal hole, retinal holes due to atrophic changes in the neurosensory retina</p>
Rhegmatogenous retinal detachment	<p>Definition: subretinal fluid accumulates between the neurosensory retina and the retinal pigment epithelium with occurrence of a break in the retina allowing vitreous to directly enter the subretinal space</p> <p>Characteristic: a rhegmatogenous retinal detachment has a corrugated appearance and undulates with eye movements</p>

network. Furthermore, YOLOv3 draws on the experience of the idea of FPN (pyramid feature map), and multi-scale features are used for object detection [36]. Deconvolution is added to extract features, which has greatly improved the detection of small objects.

Original YOLOs are mostly anchor-based, which means the researchers need to design the anchors beforehand, and the model may lack generalization when the distribution of the test dataset is much different from the training dataset.

The architecture of YOLOX is shown in Fig. 3. Researchers turned the detector to an anchor-free method, which means the users no longer need to design the anchors and the model becomes more robust. The use of other advanced detection techniques, such as CSPNet, Mosaic augmentation, decoupled head [37], and SiLU [38], makes YOLOX a high-performance detector. We applied the coordinate attention block [39] to YOLOX and replaced the loss function with GIOU [40] improve it.

Table 3 Distribution of the retinal lesions in all datasets

Dataset	Normal	WWOP	PSD	CRT	LD	RB	RRD	Total
Training dataset (original)								
Image	389	404	159	97	473	595	306	2374
Retinal lesion	0	423	217	106	547	670	391	2354
Training dataset (augmented)								
Image	4835	3984	2251	2213	2281	4789	4641	24,602
Retinal lesion	0	4159	3099	2526	2616	5414	4787	22,601
Test dataset 1								
Image	40	47	29	21	50	57	26	248
Retinal lesion	0	55	41	26	69	68	28	287
Test dataset 2								
Image	122	76	63	54	91	138	52	586
Retinal lesion	0	94	76	73	112	180	65	600

WWOP white without pressure, *PSD* paving-stone degeneration, *CRT* cystic retinal tuft, *LD* lattice degeneration, *RB* retinal break, *RRD* rhegmatogenous retinal detachment

All the experiments were accomplished using PyTorch-1.10.2 and CUDA-11.3, and the model was trained by a GTX 3060 (12 GB).

First, we loaded YOLOX-l's pretrained weight (COCO2017) and trained it with the input size of (640, 640, 3) for 50 epochs with the backbone frozen. Then, we unfroze the backbone and trained it for another 150 epochs. When we froze the backbone, we trained it with the batch size of 16. After it was unfrozen, the batch size changed to 8.

While training, we applied Mosaic for further data augmentation. Mosaic makes the dataset more various, which is undoubtedly helpful for the model, but it also makes the image completely different from the real one, so we shut it down during the last 30% epochs.

In this study, we used true positive (TP), false positive (FP), and false negative (FN) to calculate sensitivity, precision, *F1* score, and average precision (AP) [41]. True negative (TN) was not defined in the research area of object detection so we manually calculated true negative (TN) and specificity by the following formula.

Finally, we evaluated the model with accuracy of the formula $(ACC) = (TP + TN)/(TP + TN + FP + FN)$, sensitivity $(SE) = TP/(TP + FN)$, specificity $(SP) = TN/(TN + FP)$, precision $(PR) = TP/(TP + FP)$, *F1* score $= 2 \times (PR \times SE)/(PR + SE)$, and average precision (AP). AP is actually the area under the P–R (precision-recall) curve, and the area is equivalent to the integral of the equation of the curve. The computational formula for AP and mean AP (mAP) is listed in Supplementary Fig. 1.

In the database, some images presented plural parts of retinal lesions; we counted 1 TP result only when the YOLOX model accurately located all lesions with $IOU > 0.5$. Figure 4 illustrates a prediction example of our model: One TP result was counted for RB since two RB lesions were all detected and surrounded by the yellow rectangle. One FN result was counted for RRD since RRD was missed by the model. This image showed no sign of other peripheral retinal lesion but RB and RRD, so we additionally counted one TN result for normal and each other peripheral retinal lesion. One FP result was counted when there was no actual lesion

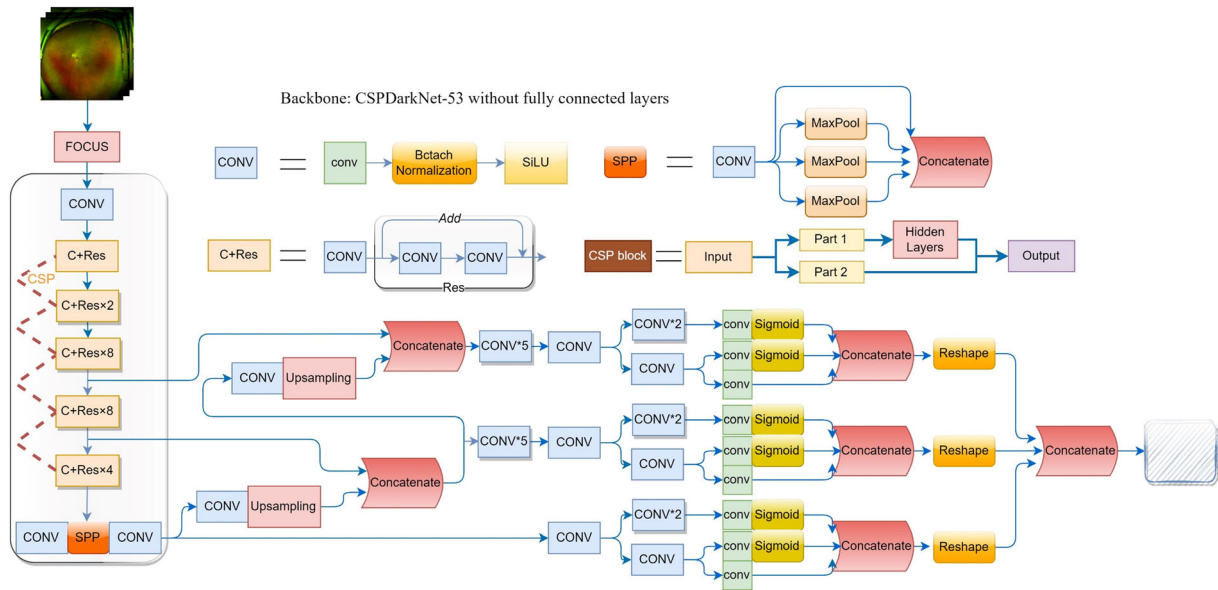


Fig. 3 The architecture of YOLOX with the backbone CSPDarkNet-53, which means it has 53 convolutional layers, and CSP block is added. Three effective features are used for prediction and detection

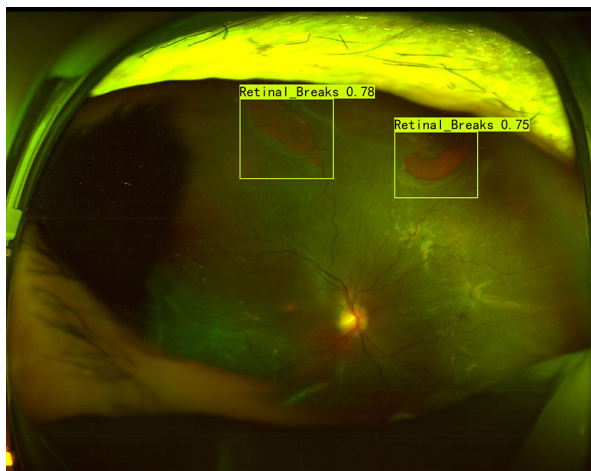


Fig. 4 Detection example of our model. As shown in the picture, two retinal breaks were indicated by yellow rectangles

but identified as one lesion by the model. If any lesion was missed or identified in the wrong category of retinal lesion by the model, the image was counted as a FN result.

RESULTS

This study contained 2374 original images for training data; of those, 1985 UWF images

showed peripheral lesions (mean age 44.92 ± 22.46 years, 930 images from men and 1055 images from women), and 389 showed normal UWF images (mean age 40.52 ± 15.89 years, 142 images from men and 247 images from women). In our preliminary studies, we compared the performance of several SOTA models to choose our baseline, and all the models were trained with 24,602 augmented images. To test the models' abilities for detecting target retinal lesions, we used test dataset 1 (208 original images with peripheral lesions, mean age 42.17 ± 18.09 years, 92 images from men and 116 images from women; 40 normal UWF images, mean age 38.30 ± 12.48 years, 22 images from men and 18 images from women).

We built CenterNet, SSD, RetinaNet, Fcos, YOLOv5, original YOLOX, and modified YOLOX with the same dataset and hyperparameters (batch size, input size, learning rate, etc.) and trained them for 150 epochs without extra data augmentation or training skills. We tested their performance with test dataset 1, and Table 4 shows the results of the experiments.

As shown in Table 4, YOLOX performed better than the other models, so we chose it as

Table 4 Comparison between some SOTA algorithms based on test dataset 1 (%)

	CenterNet	SSD	RetinaNet	Fcos	YOLOv5-6.1	YOLOX	Ours
Recall	61.92	56.86	66.38	79.90	70.50	80.47	81.59
Precision	74.83	81.34	77.03	76.79	83.02	80.02	81.54
<i>F1</i>	65.33	64.83	69.83	77.50	76.17	80.00	81.33
mAP	75.44	73.24	78.19	81.20	80.79	84.08	85.91

the baseline, and after modifying the architecture of the network, it improved.

The YOLOX model can distinguish whether the UWF image is normal, classify categories of the six peripheral retinal lesions (WWOP, PSD, CRT, LD, RB, RRD), and locate where the retinal lesion is with average accuracy, sensitivity, specificity, precision, *F1* score, and mAP of 96.64%, 87.97%, 98.04%, 87.01%, 87.39%, and 86.03%, respectively. More details about the performance of the model are shown in Table 4, indicating that the model is effective for the identification and localization of the peripheral retinal lesions based on the test dataset 1.

For further evaluating stability of the YOLOX model's performance, we prepared another dataset called test dataset 2 based on another bunch of original UWF images separately collected from West China Hospital between January to April of 2022. This set included 586 UWF images (464 with peripheral retinal lesions and 122 normal) and had never been augmented or used for a training procedure. Based on test dataset 2, the model achieved average accuracy, sensitivity, specificity, precision, *F1* score, and mAP of 95.04%, 83.90%, 96.70%, 78.73%, 81.96%, and 80.59%, respectively. Details are shown in Table 5.

DISCUSSION

After being trained with 24,602 augmented UWF images with 22,601 parts of peripheral retinal lesion, the modified YOLOX model achieved average accuracy of 96.64%, sensitivity of 87.97%, specificity of 98.04%, precision of 87.01%, *F1* score of 87.39%, and mAP of 86.03% based on test dataset 1. To ensure the generalizability, we tested the model with test dataset

2, and it achieved average accuracy of 95.04%, sensitivity of 83.90%, specificity of 96.70%, precision of 78.73%, *F1* score of 81.96%, and mAP of 80.59%. One of the innovative aspects of this study is that we focus on peripheral retinal lesions, which were less concerned because of limited examination techniques in the past but could also cause vision loss even in more silent ways. The second aspect is that we conducted an object detection so that the DL model could not only identify what the retinal lesion was but also locate the exact position of the retinal lesion.

DL models have received significant attention and interest over the past few decades, but the models lack interpretation, especially in healthcare applications. Essentially, interpretability is one of the major obstacles in the implementation of DL. DL model can extract features from original images and make numerous calculations to identify what the images are. Figure 5a, c, e, g, i, k shows some examples, and the lesions detected are surrounded by boxes with different colors. Figure 5b, d, f, h, j, l shows the corresponding saliency maps and may help to explain how our model located peripheral retinal lesions, which improves the interpretability of our study. The saliency maps partially ripped a crack in the black box of the DL model, but were still far from tearing that black box down.

Recent studies based on DL have shown some stimulating achievements for the identification of eye diseases. Most focus on polar retinal disease such as DR [23], AMD [24], RVO [26], and neovascularization [19]; some also pay attention to diseases in the anterior part like cataract and angle-closure glaucoma [13]. However, these studies generally applied

Table 5 Evaluation indexes for detecting each retinal lesion with YOLOX

	ACC (95% CI)	SE (95% CI)	SP (95% CI)	PR (95% CI)	F1(95% CI)	AP (95%CI)
Test dataset 1						
Normal	97.96 (97.22–98.69)	91.99 (88.21–95.77)	98.93 (98.32–99.55)	93.36 (89.25–97.47)	92.66 (89.83–95.48)	N/A
WWOP	96.88 (96.02–97.75)	90.38 (87.54–93.22)	98.24 (97.19–99.29)	91.50 (86.17–96.83)	90.91 (87.53–94.29)	90.92 (88.79–93.05)
PSD	97.37 (96.52–98.21)	86.35 (81.47–91.24)	98.73 (98.42–99.13)	88.83 (83.80–93.85)	87.51 (84.23–90.80)	87.86 (82.78–92.94)
CRT	95.91 (94.74–97.09)	81.94 (78.40–85.48)	97.20 (96.43–97.97)	70.67 (65.67–75.68)	75.85 (72.49–79.20)	74.63 (68.10–81.16)
LD	96.02 (94.83–97.22)	82.75 (80.01–85.50)	96.65 (95.69–97.62)	85.60 (82.40–88.80)	84.13 (82.59–85.67)	83.31 (78. 98–87.64)
RB	93.92 (92.30–95.55)	89.43 (86.30–92.56)	97.67 (96.51–98.84)	90.98 (87.05–94.91)	90.19 (87.11–93.26)	89.85 (86.48–93.22)
RRD	98.33 (97.79–98.88)	92.98 (89.05–96.90)	98.82 (98.24–99.39)	88.10 (82.61–93.59)	90.45 (86.26–94.64)	89.62 (88.35–90.89)
Average	96.64	87.97	98.04	87.01	87.39	86.03
Test dataset 2						
Normal	97.37 (96.17–98.5)	92.24 (89.32–95.15)	98.47 (97.33–99.61)	93.26 (90.04–96.48)	92.73 (90.64–94.82)	N/A
WWOP	94.49 (93.12–95.86)	85.67 (81.70–89.64)	95.92 (94.47–97.37)	77.97 (73.81–82.13)	81.63 (78.00–85.25)	81.32 (78.81–83.83)
PSD	94.34 (92.30–96.38)	79.62 (73.33–80.52)	96.57 (95.27–97.87)	74.16 (70.28–78.03)	75.49 (72.64–78.34)	76.29 (73.35–79.23)
CRT	93.87 (91.62–96.12)	74.85 (70.39–79.32)	95.87 (94.07–97.66)	65.88 (60.96–70.80)	70.04 (66.71–73.36)	70.46 (66.31–74.61)
LD	94.49 (93.50–95.48)	84.53 (80.31–88.76)	96.22 (95.31–97.14)	79.79 (76.19–83.40)	82.08 (78.86–85.30)	82.65 (80.69–84.61)
RB	92.93 (91.03–94.83)	85.04 (81.42–88.66)	95.20 (92.82–97.58)	84.08 (78.53–89.63)	84.53 (80.97–88.08)	84.97 (79.9889.96)
RRD	97.82 (96.69–98.49)	88.06 (81.81–94.31)	98.69 (97.88–99.49)	86.40 (83.20–89.60)	87.20 (83.17–91.23)	87.84 (84.49–91.19)
Average	95.04	83.90	96.70	78.73	81.96	80.59

WWOP white without pressure, *PSD* paving-stone degeneration, *CRT* cystic retinal tuft, *LD* lattice degeneration, *RB* retinal break, *RRD* rhegmatogenous retinal detachment, *ACC* accuracy, *SE* sensitivity, *SP* specificity, *PR* precision, *AP* average precision

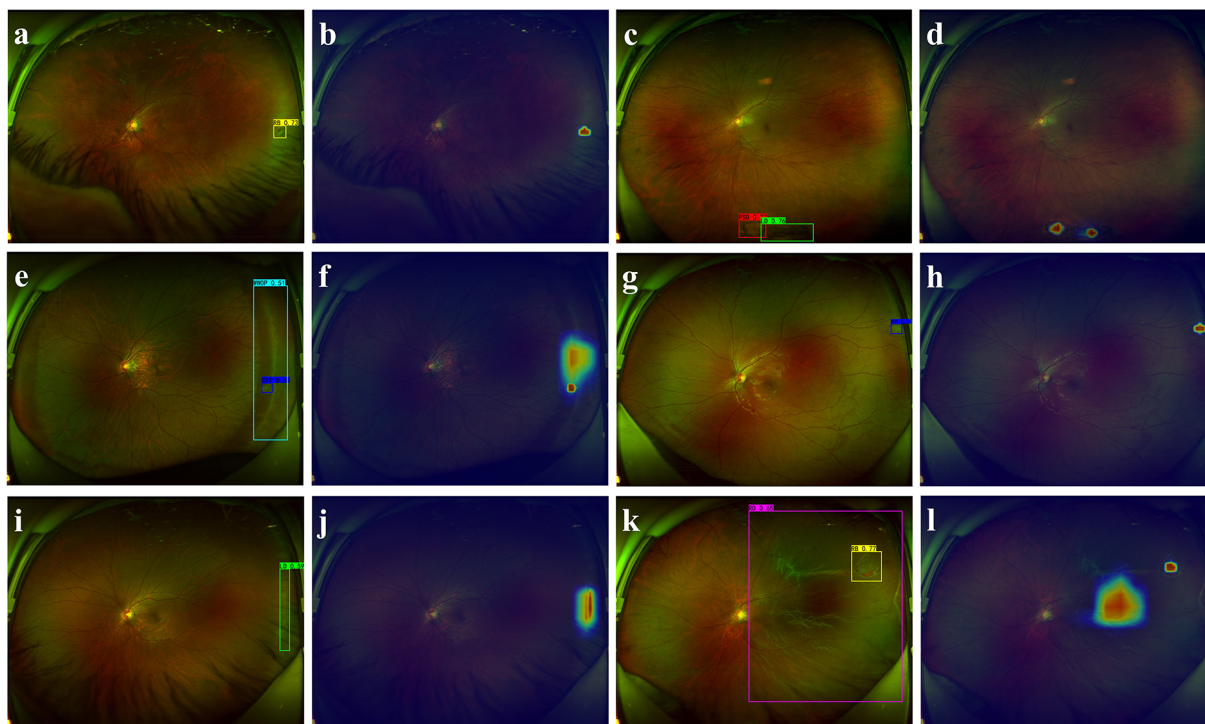


Fig. 5 Examples of peripheral retinal lesions located by YOLOX. **a** Retinal break; **c** paving-stone degeneration and lattice degeneration; **e** white without pressure and cystic retinal tuft; **g** cystic retinal tuft; **i** lattice degeneration;

k retinal detachment and retinal break; **b, d, f, h, j, l** corresponding saliency maps

traditional classification networks, which were only able to evaluate the existence of eye diseases; they could not locate the specific position of the diseases, and the system may not work if there are multiple diseases in one single image. Different from traditional classification networks, this study applied an object detection model so that the retinal lesion areas could be located directly and accurately, being able to help ophthalmologists make rapid diagnoses. By outputting the saliency map and exact position of retinal lesion areas with clear borders, this study may also help to improve the clinical interpretability in another way.

This study also has limitations: First, the UWF images we used to train the model were two-dimensional images, lacking stereoscopic qualities rather than three-dimensional images, thus making the identification of proposed lesions such as cystic retinal tuft challenging. Second, only UWF images were included in this study, lacking assistance from different

dimensions including medical history and other physical examinations. Thus, this study could not achieve multi-dimensional evaluation, which limited the accuracy in realistic scenes. Third, images in this study were taken from one hospital, and most images come from only Asians, which lowers the ethnic representation of the study. Besides, the UWF image cannot cover 100% of the retina, and the green-red pseudo-color image differs from the real look of the fundus, so some peripheral retinal lesions may still be missed. Moreover, the sensitivity of the model is relatively low, but the specificity is high, which suggests that the model tends to produce false-negative results, which is not advantageous in large-scale screening. Future studies will focus more on improving sensitivity while maintaining high specificity and accuracy of the DL models.

Focusing on peripheral retinal lesions, we successfully established a modified YOLOX model based on UWF images. This work

provides an efficient solution for the localization and identification of peripheral retinal lesions, and the results indicated that the proposed model showed high efficiency in intelligent diagnosis for six peripheral retinal lesions.

Main contributions and innovations are as follows: (1) this study developed an object detection dataset including six peripheral retinal lesions based on UWF images; (2) the object detection model YOLOX was applied to locate and identify the peripheral retinal lesions, being able to show the exact locations of the retinal lesions, which has not been frequently reported in other studies, and showed high efficiency; (3) this study focused on the six peripheral retinal lesions that are high risk but sometimes ignored because of limited examination techniques.

Currently, the interpretation of UWF images can only be completed by experienced ophthalmologists after years of training. There is a huge gap between the large group of patients and the relatively scarce fundus disease specialists, which hinders the popularity of the UWF imaging system and makes it difficult for patients to obtain convenient fundus examination and supporting interpretation services. How to diagnose fundus diseases with high efficiency and low cost in basic medical units has been a concern and discussed widely. According to a review based on AI and UWF images, object detection technology has hardly been applied to fundus images, which means there is still research space for lesion detection on UWF images [42]. Meanwhile, although some meaningful diseases were included for study and the results were inspiring, there are still many neglected diseases, and different diseases may be identified as abnormal but not classified as well [43–45]. Focusing on peripheral retinal lesions, we applied an object detection model to UWF images, verifying its certain potential for the identification and location of retinal lesions. With the increase of quantity and quality of data, more diseases can be involved in the future study, which may contribute more to early screening and clinical diagnosis of retinal lesions. In this way, ophthalmologists can pay more attention to patients with severe conditions, and clinical efficiency might be greatly improved.

In future studies, a larger dataset including more retinal lesions should be used to improve and evaluate our method. Furthermore, multi-center studies may be conducted to promote the applicability of this study. Many kinds of images and information could be involved, which may contribute to the intelligent diagnosis of retinal lesions, making this method practical in clinical settings.

CONCLUSION

In summary, our method showed certain potential for intelligent identification and localization of six peripheral retinal lesions based on UWF images. The results verified that the model had relatively high efficiency and accuracy and preliminarily approached ophthalmologists' diagnosis level [46], indicating that the model may contribute to clinical diagnosis in the future.

ACKNOWLEDGEMENTS

Funding. This study was supported by grants from the Project of National Key Research and Development Program of China (grant no. 2018YFC1106103, 2021YFB3802100). The journal's Rapid Service Fee is also funded by grants from the Project of National Key Research and Development Program of China.

Author Contributions. Conception and design of the study (ZM, LJJ, WT, and LGL); data collection and image labeling (WT, CL, YQZ, WSS, WBJ, GYX); data analysis (LGL, WT, ZY, ZSB, HL, CK, and MJY); writing of the manuscript and preparation of figures (WT, LGL, ZY); supervision (ZM and LJJ); review and discussion of the results (all authors); edition and revision of the manuscript (all authors).

Disclosures. Tong Wang, Guoliang Liao, Lin Chen, Yan Zhuang, Sibozhou, Qiongzhen Yuan, Lin Han, Shanshan Wu¹, Ke Chen, Binjian Wang, Junyu Mi, Yunxia Gao, Jiangli Lin, and Ming Zhang have nothing to disclose.

Compliance with Ethics Guidelines. This study was approved by the ethical review committee of the Chinese Center for Disease Control and Prevention (Beijing, China) with the number ChiCTR2100053526. This study is a retrospective study, and written consent to participate is not applicable. All images in study were desensitized with consent for publication. This study adhered to the tenets of the Declaration of Helsinki and the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis guidelines.

Data Availability. The code of the deep learning model used in this study is available through the link: https://github.com/JohnLeoXJTU/PerLesionDet_for_Oph-The. The specific subset of images used in this study is available upon reasonable request from the corresponding authors.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Flaxman SR, Bourne R, Resnikoff S, Ackland P, Braithwaite T, Cicinelli MV, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health*. 2017;5(12):e1221–34. [https://doi.org/10.1016/S2214-109X\(17\)30393-5](https://doi.org/10.1016/S2214-109X(17)30393-5).
2. Bourne R, Steinmetz JD, Flaxman S, Briant PS, Taylor HR, Resnikoff S, et al. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. *Lancet Glob Health*. 2021;9(2):e130–43. [https://doi.org/10.1016/S2214-109X\(20\)30425-3](https://doi.org/10.1016/S2214-109X(20)30425-3).
3. Vilela MAP, Parodi MB. Extensive macular atrophy with pseudodrusen-like: case series and review published online ahead of print, 2022 May 23]. *Eur J Ophthalmol*. 2022. <https://doi.org/10.1177/11206721221102269>.
4. Orlin A, Fattoo A, Ehrlich J, D'Amico DJ, Chan RP, Kiss S. Ultra-widefield fluorescein angiography of white without pressure. *Clin Ophthalmol*. 2013;7:959–64. <https://doi.org/10.2147/OPHT.S43450>.
5. Nishitsuka K, Nishi K, Namba H, Kaneko Y, Yamashita H. Peripheral Cystoid Degeneration Finding using intraoperative optical coherence tomography in rhegmatogenous retinal detachment. *Clin Ophthalmol*. 2021;15:1183–7. <https://doi.org/10.2147/OPHT.S306623>.
6. Sengillo JD, Smiddy WE, Yannuzzi NA, Flynn Jr HW. Giant retinal tears: long-term outcomes of fellow eyes. *Ophthalm Surg Lasers Imaging Retina*. 2022;53(11):619–25.
7. Banerjee AC, Davis D, Charteris D. Ethnic variation in rhegmatogenous retinal detachments. *Eye*. 2015;29(6):803–7.
8. Steel D. Retinal detachment. *BMJ Clin Evid*. 2014;2014:0710.
9. Nagiel A, Lalane RA, Sadda SR, Schwartz SD. Ultra-widefield fundus imaging: a review of clinical applications and future trends. *Retina*. 2016;36(4):660–78.
10. Zhou S, Qiu Y, Han L, Liao G, Zhuang Y, Ma B, et al. A lightweight network for automatic thyroid nodules location and recognition with high speed and accuracy in ultrasound images. *J X ray Sci Technol*. 2022;30(5):967–81.
11. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–8.
12. Zhuang Y, Rahman MF, Wen Y, Pokojovy M, McCaffrey P, Vo A, et al. An interpretable multi-task system for clinically applicable COVID-19 diagnosis using CXR. *J Xray Sci Technol*. 2022;30(5):847–62.

13. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167–75.
14. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. *Prog Retin Eye Res*. 2018;67:1–29. <https://doi.org/10.1016/j.preteyeres.2018.07.004>.
15. Cao C, Liu F, Tan H, Song D, Shu W, Li W et al. Deep learning and its applications in biomedicine. *Biomed Genom Proteom Bioinform*. 2018;16(1):17–32.
16. Caixinha M, Nunes S. Machine learning techniques in clinical vision sciences. *Curr Eye Res*. 2017;42(1):1–15.
17. Tong Y, Lu W, Deng Q, Chen C, Shen Y. Automated identification of retinopathy of prematurity by image-based deep learning. *Eye Vis*. 2020;7(1):40.
18. Li J, Wang L, Gao Y, Liang Q, Chen L, Sun X et al. Automated detection of myopic maculopathy from color fundus photographs using deep convolutional neural networks. *Eye Vis*. 2022;9(1):1–12.
19. Cai SPFUM. Deep learning detection of sea fan neovascularization from ultra-widefield color fundus photographs of patients with sickle cell hemoglobinopathy. *JAMA Ophthalmol*. 2021;139(2):206–13.
20. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402.
21. Voets M, Møllersen K, Bongo LA. Reproduction study using public data of: development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *PLoS ONE*. 2019;14(6): e217541.
22. Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen C, et al. A deep learning system for identifying lattice degeneration and retinal breaks using ultra-widefield fundus images. *Ann Transl Med*. 2019;7(22):618.
23. Nagasawa T, Tabuchi H, Masumoto H, Enno H, Niki M, Ohara Z, et al. Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naïve proliferative diabetic retinopathy. *Int Ophthalmol*. 2019;39(10):2153–9.
24. Matsuba S, Tabuchi H, Ohsugi H, Enno H, Ishitobi N, Masumoto H, et al. Accuracy of ultra-wide-field fundus ophthalmoscopy-assisted deep learning, a machine-learning technology, for detecting age-related macular degeneration. *Int Ophthalmol*. 2019;39(6):1269–75.
25. Xie H, Zeng X, Lei H, Du J, Wang J, Zhang G, et al. Cross-attention multi-branch network for fundus diseases classification using SLO images. *Med Image Anal*. 2021;71: 102031.
26. Nagasato D, Tabuchi H, Ohsugi H, Masumoto H, Enno H, Ishitobi N et al. Deep neural network-based method for detecting central retinal vein occlusion using ultrawide-field fundus ophthalmoscopy. *J Ophthalmol*. 2018;2018:1875431.
27. Nagasato D, Tabuchi H, Ohsugi H, Masumoto H, Enno H, Ishitobi N, et al. Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion. *Int J Ophthalmol*. 2019;12(1):94–9.
28. Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen C, et al. Development and evaluation of a deep learning system for screening retinal hemorrhage based on ultra-widefield fundus images. *Transl Vis Sci Technol*. 2020;9(2):3.
29. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *Eur J Clin Invest*. 2015;45(2):204–14.
30. Khan SM, Liu X, Nath S, Korot E, Faes L, Wagner SK, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *The Lancet Digit Health*. 2021;3(1):e51–66.
31. Zhou Z, Zhang M, Huang S, Li Y. Multi-instance multi-label learning. *Artif Intell*. 2012;176(1):2291–320. <https://doi.org/10.1016/j.artint.2011.10.002>.
32. Wang W, Zhou Z. Learnability of multi-instance multi-label learning. *Chin Sci Bull*. 2012;57(19):2488–91.
33. Mercan C, Aksoy S, Mercan E. Multi-Instance multi-label learning for multi-class classification of whole slide breast histopathology images. *IEEE Trans Med Imaging*. 2018;37(1):316–25.
34. Humpire-Mamani GE, Arnaud Arindra AS, van Ginneken B, Colin J. Organ detection in thorax abdomen CT using multi-label convolutional neural networks. *Phys Med Biol*. 2018;63(8):85003.
35. Kaiming H, Xiangyu Z, Shaoqing R, Jian S. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; p. 770–8. <https://doi.org/10.1109/CVPR.2016.90>.

36. Pan H, Chen G, Jiang J. Adaptively dense feature pyramid network for object detection. *IEEE Access*. 2019;7:81132–44. <https://doi.org/10.1109/access.2019.2922511>.
37. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell*. 2020;42(2):318–27. <https://doi.org/10.1109/TPAMI.2018.2858826>.
38. Elfwing S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw*. 2018;107:3–11.
39. Qibin H, Daquan Z, Jiashi F. Coordinate attention for efficient mobile network design. *Neural Netw*. 2018;107:3–11.
40. Seyed HR, Nathan T, JunYoung G, Amir S, Ian DR, Silvio S. Generalized intersection over union: a metric and a loss for bounding box regression. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019;2019: 658–66.
41. Trevelyan R. Sensitivity, specificity, and predictive values: foundations, pliabilitys, and pitfalls in research and practice. *Front Public Health*. 2017;5: 307.
42. Yang J, Fong S, Wang H, Hu Q, Lin C, Huang S, et al. Artificial intelligence in ophthalmopathy and ultra-wide field image: a survey. *Expert Syst Appl*. 2021;182: 115068.
43. Tan T, Ting DSW, Wong TY, Sim DA. Deep learning for identification of peripheral retinal degeneration using ultra-wide-field fundus images: is it sufficient for clinical translation? *Ann Transl Med*. 2020;8(10):611.
44. Ragab M, Aljedaibi WH, Nahhas AF, Alzahrani IR. Computer aided diagnosis of diabetic retinopathy grading using spiking neural network. *Comput Electr Eng*. 2022;101: 108014.
45. Ragab M, AL-Ghamdi ASA, Fakieh B, Choudhry H, Mansour RF, Koundal D. Prediction of diabetes through retinal images using deep neural network. *Comput Intel Neurosc*. 2022;2022:7887908.
46. Cen L, Ji J, Lin J, Ju S, Lin H, Li T et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nat Commun*. 2021;12(1):4828.