Research progress in artificial intelligence assisted diabetic retinopathy diagnosis

Yun-Fang Liu1, Yu-Ke Ji2, Fang-Qin Fei3, Nai-Mei Chen4, Zhen-Tao Zhu4, Xing-Zhen Fei3

1Department of Ophthalmology, First People’s Hospital of Huzhou, Huzhou University, Huzhou 313000, Zhejiang Province, China
2Eye Hospital, Nanjing Medical University, Nanjing 210000, Jiangsu Province, China
3Department of Endocrinology, First People’s Hospital of Huzhou, Huzhou University, Huzhou 313000, Zhejiang Province, China
4Department of Ophthalmology, Huai’an Hospital of Huai’an City, Huai’an 223000, Anhui Province, China

Co-first authors: Yun-Fang Liu and Yu-Ke Ji

Correspondence to: Xing-Zhen Fei. Department of Endocrinology, First People’s Hospital of Huzhou, Huzhou University, Huzhou 313000, Zhejiang Province, China. fxz55998@163.com; Zhen-Tao Zhu. Department of Ophthalmology, Huai’an Hospital of Huai’an City, Huai’an 223000, Anhui Province, China. jshayyzt@163.com

Received: 2023-04-28 Accepted: 2023-06-14

Abstract

- Diabetic retinopathy (DR) is one of the most common retinal vascular diseases and one of the main causes of blindness worldwide. Early detection and treatment can effectively delay vision decline and even blindness in patients with DR. In recent years, artificial intelligence (AI) models constructed by machine learning and deep learning (DL) algorithms have been widely used in ophthalmology research, especially in diagnosing and treating ophthalmic diseases, particularly DR. Regarding DR, AI has mainly been used in its diagnosis, grading, and lesion recognition and segmentation, and good research and application results have been achieved. This study summarizes the research progress in AI models based on machine learning and DL algorithms for DR diagnosis and discusses some limitations and challenges in AI research.

- **KEYWORDS:** diabetic retinopathy; artificial intelligence; machine learning; deep learning; diagnosis; grading; lesions segmentation

DOI:10.18240/ijo.2023.09.05

INTRODUCTION

As a common metabolic disease, diabetes can damage various body tissues and organs[3]. Diabetic retinopathy (DR) can seriously damage patients’ vision and even blindness. As one of the most serious microvascular complications of diabetes, DR has become the main cause of blindness[2-3]. DR is associated with various factors such as the length of the course of diabetes and the control level of blood glucose[4]. Research has shown that in patients who have had a course of diabetes for more than 10y, regardless of age, the fundus has a different course of pathological changes[5]. Another study showed that life expectancy is significantly longer owing to improved socioeconomic conditions, leading to a gradual increase in the incidence of diabetes, approximately 30% of whom develop DR[6]. Strict control blood glucose levels cannot completely prevent DR, but it can significantly reduce the incidence of DR and protect patients’ vision. In addition, timely diagnosis and treatment of patients with DR can effectively protect their visual function.

Artificial intelligence (AI), a branch of computer science, can quickly acquire human knowledge and experience in a short time to simulate the human execution of complex tasks[7]. Machine learning (ML) is a branch of AI in which the machine marks the measured data or features and then enters the sample dataset to train known tags for classification tasks[8]. Deep learning (DL) is a branch of ML and an ML method. Its core is a neural network that can perform classification tasks and extract features[9]. Figure 1 shows the relationship of ML, DL and AI with DR. With the rapid development of AI technology in recent years, AI models based on ML and DL algorithms can automatically diagnose diseases by identifying disease information. In 2016, Google built an intelligent DR diagnostic system based on DR networks. The system’s performance in diagnosing DR was comparable to that of ophthalmologists. This encouraging result has received widespread attention in the industry[10]. The present study directly promotes applied research on global DR and AI and the accelerated development
of global ophthalmic and medical AI. Since 2016, AI has been increasingly used in ophthalmology. Currently, AI has achieved many good results in the study of ophthalmic diseases, such as ocular surface diseases\cite{11,12}, anterior segment diseases\cite{13,14}, retinopathy\cite{15,16,17}, retinopathy of prematurity\cite{18,19}, glaucoma\cite{20,21,22}, age-related macular degeneration\cite{23,24}, and retinal vein occlusion\cite{25,26}. In this review, we summarize the research progress of AI in DR diagnosis in the past few years and discuss some limitations and challenges encountered in AI research.

**DIABETIC RETINOPATHY**

DR, as a serious diabetic microvascular complication, is also a common blinding ophthalmopathy worldwide\cite{27,28}. Currently, the pathogenesis of DR is not fully clarified; however, the root cause is glucose metabolism disorder\cite{29,30}. Regarding its clinical manifestations, patients with DR often show no obvious symptoms in the early stages; however, with the development of the disease, patients with DR experience varying degrees of vision loss and blindness\cite{31,32,33}. Fundus manifestations of DR include microhemangioma, bleeding spots, rigid exudation, retinal angiopathy, and optic neuropathy\cite{34,35}. In severe cases, vitreous hemorrhage, retinal neovascularization, macular edema, tractional retinal detachment, neovascular glaucoma, and other serious complications occur, causing serious damage to patients’ visual acuity\cite{36}. According to the severity of DR, it is divided into two categories: non-proliferative DR (NPDR) and proliferative DR (PDR). The core difference between PDR and NPDR is the formation of retinal neovascularization, which breaks through the internal limiting membrane\cite{37,38}. Serious complications, such as tractional retinal detachment, vitreous hemorrhage, diabetic macular edema (DME) and neovascular glaucoma, can easily occur in patients with PDR. In clinical settings, the main treatment methods for DR include: 1) strict control blood glucose levels and diabetic complications, such as hypertension, hyperlipidemia, and nephropathy; 2) retinal laser photocoagulation; 3) surgical treatment, such as vitrectomy and intraocular photocoagulation; 4) anti-vascular endothelial growth factor drugs inhibiting neovascularization\cite{39,40}.

**RESEARCH OF ARTIFICIAL INTELLIGENCE IN DIABETIC RETINOPATHY**

Herein, we introduce the applications of AI in DR cases over the past five years. To focus on the output performance of the AI model because it is associated with whether the AI model can be applied to clinical diagnosis and treatment; therefore, we did discuss much on other aspects of AI research.

**Research of AI in DR Diagnosis (not Involving DR Grading Diagnosis)** Because late-stage DR seriously damages vision and leads to irreversible blindness, timely diagnosis and treatment for patients with DR are crucial. Currently, DR is primarily diagnosed by professional ophthalmologists based on fundus changes and related ophthalmological examination results, thus requiring many high-level ophthalmologists to diagnose DR. However, in many remote areas, economically backward areas, and areas lacking medical resources, the number of high-level ophthalmologists is very low, making it difficult to meet patients’ needs and indicating the need for other methods for DR diagnosis. In recent years, AI-assisted diagnosis of DR has been widely studied, and many computer-aided diagnostic systems that are helpful in the clinical diagnosis of DR have been constructed.

**Research of AI-assisted diagnosis of DR based on color fundus photographs** Arsalan et al\cite{41} proposed a Vess-Net network model that can differentiate DR from hypertensive retinopathy based on convolutional neural networks. Three public datasets (DRIVE, CHASE-DB1, and STARE) were used for verification to evaluate the accuracy of Vess-Net model for diagnosing DR. The model’s sensitivity, specificity, area under curve (AUC), and accuracy were respectively 0.80, 0.98, 0.98, and 0.96 for the DRIVE dataset; 0.82, 0.98, 0.98, and 0.97 for the CHASE-DB1 dataset; and 0.85, 0.99, 0.99, and 0.97 on STARE dataset. Cen et al\cite{42} jointly used InceptionV3, Xception, InceptionResNet-V2, ResNet, and other DL algorithms to build an AI system that could automatically diagnose 39 fundus retinal diseases, including DR. They compared the system’s diagnostic results with that of five ophthalmologists. Finally, the system exhibited superior performance in diagnosing DR, with a sensitivity, specificity, and specificity of 0.94, 0.98, and 0.98, respectively. The diagnostic performance was similar to that of the ophthalmologists.

To create a DR detection model, Sundaram et al\cite{43} constructed an AI model using an ensemble convolutional neural network and trained and tested it using three public datasets: DIARETDB0, STARE, and DIARETDB1. Finally, the AI model’s accuracy for detecting DR was 0.99, and its detection performance was excellent. Dai et al\cite{44} established a DeepDR system based on ResNet. The system can automatically detect DR by identifying pathological changes, such as
microaneurysms, hard exudation, and retinal hemorrhage, on color fundus photographs. They used 675 569 images to train and verify the DeepDR system. The AUC of DR detected by the system was 0.91. In addition, Lu et al.\(^\text{[44]}\) created a computer DR detection system, DLS, based on convolutional neural networks. The system’s specificity, sensitivity, and AUC for detecting DR were 0.96, 0.90, and 0.98, respectively. Nazir et al.\(^\text{[45]}\) proposed an AI model based on DenseNet-100, which can detect DR by identifying DR lesions such as retinal hemorrhage, microaneurysms, exudates. In addition, they trained and validated the model using the public datasets APTOS-2019 and IDRiD. Finally, the model’s accuracy was 0.97 in the APTOS-2019 dataset and 0.98 in the IDRiD datasets. Similarly, Li et al.\(^\text{[46]}\) constructed an intelligent diagnostic AI model based on the Inception-V4 network to automatically detect DR. They collected 8739 color fundus photographs and used the Messidor-2 dataset to evaluate the AI model. In addition, they compared the AI model’s diagnostic results with that of ophthalmologists. The final results showed that the model’s sensitivity was 0.92, specificity was 0.96, and AUC was 0.99, indicating better performance than that of ophthalmologists. Zhang et al.\(^\text{[47]}\) developed an automatic DR detection and diagnosis model based on the Inception-V3 network, which was trained and verified using 88 702 color fundus photographs. After verification, the model’s sensitivity was 0.92, specificity was 0.90, and AUC was 0.97. Bhuiyan et al.\(^\text{[48]}\) used three DL algorithms (Xception, Inception-V3, Inception-Resnet-V2) to construct an AI model to screen DR. They collected 90 450 images to train and test the model. Finally, the model’s sensitivity was 0.98, specificity was 0.99, and AUC was 0.99 for DR screening. Similarly, Wang et al.\(^\text{[49]}\) constructed DR screening model based on the framework of the Inception-V3 network. In this study, they used 33 115 images to train, test and evaluate the AI model. Additionally, they compared the model’s diagnostic results with that of ophthalmologists. Finally, the model’s AUC was 0.95, sensitivity was 0.97, specificity was 0.88, and accuracy was 0.87. The model’s diagnostic performance of was similar to that of the ophthalmologists, and the time required by the model was less than that of the ophthalmologists. To achieve automatic screening for DR, Hassan et al.\(^\text{[50]}\) combined three DL algorithms (VGG16, ResNet-50, and U-Net) to construct a DR screening model. They used 1840 color fundus photographs to train the model and evaluated it using an externally validated dataset. After verification, the model’s performance was good, with a diagnostic accuracy of 0.99. Islam et al.\(^\text{[51]}\) proposed an AI model using supervised contrastive learning to assist in DR screening. They used the APTOS 2019 datasets to evaluate the model’s performance. After testing, the model’s AUC was 0.99 and accuracy was 0.98 for diagnosing DR. To create AI models that could automatically screen for DR, Khalifa et al.\(^\text{[52]}\) constructed multiple AI models using six DL algorithms: AlexNet, ResNet18, SqueezeNet, GoogleNet, VGG16, and VGG19. They used the APTOS 2019 dataset to train and validate AI models and compared the diagnostic performances of various AI models. After verification, the AlexNet model’s accuracy was the highest (0.98). Similarly, Tariq et al.\(^\text{[53]}\) constructed multiple AI diagnostic models based on the five DL algorithms and compared the models. They collected 1440 color fundus photographs to train and test the models. Finally, the ResNeXt-50 model’s diagnostic performance was the best, with an accuracy of 0.98 (Table 1).

**Research of AI-assisted diagnosis of DR based on other ophthalmic imaging data** To build an auxiliary diagnostic system that can diagnose DR automatically, Abitbol et al.\(^\text{[54]}\) proposed an AI model using the DenseNet121 and convolutional neural networks and used 224 ultra-widefield color fundus photographs (UWF-CFP) for training and verification. Their model can be used for the differential diagnosis of many types of retinal vascular diseases, such as DR. The model’s AUC was 0.90, accuracy was 0.85, and specificity was 0.92 for diagnosing DR, indicating superior diagnostic performance. Sun et al.\(^\text{[55]}\) created an AI diagnostic model that can diagnose many types of fundus diseases, such as DR, using three DL algorithms: EfficientNet-B7, DenseNet, and ResNet-101. They used 4574 ultra-widefield images (UWFIs) to train and test the model. Finally, the model’s specificity was 0.99 and accuracy was 0.99 for diagnosing DR. Additionally, they compared the accuracy of the AI model for diagnosing DR with that of ophthalmologists and the model’s accuracy was comparable to that of ophthalmologists. Elgafi et al.\(^\text{[56]}\) constructed a computer diagnostic system that can automatically detect DR based on convolutional neural networks. This system can automatically segment the retinal layer on optical coherence tomography (OCT) images and extract pathological features for DR detection. They collected three-dimensional (3D)-OCT images of 188 participants for training and verification. Finally, the system’s accuracy for detecting DR was 0.97. Similarly, Ryu et al.\(^\text{[57]}\) created an AI model based on the ResNet-101 network to assist in diagnosing DR and used 301 optical coherence tomography angiography (OCTA) images to evaluate the AI model. Finally, the accuracy was 0.91–0.98, sensitivity was 0.86–0.97, specificity was 0.94–0.99, AUC was 0.92–0.98 (Table 1). An accurate diagnosis is important for the treatment of patients with DR. However, in many poor and remote areas, owing to the lack of professional ophthalmologists, many patients with DR cannot receive an accurate diagnosis; therefore, missing the best time for treatment leads to serious visual impairment and even blindness. Based on the above results, AI showed
superior performance in the automatic diagnosis of DR and the potential to diagnose DR. This would greatly reduce the demand for professional ophthalmologists in poor and remote areas, improve the efficiency of diagnosing DR, and reduce DR-induced visual function impairment.

**Research of AI in DR Grading Diagnosis**

To accurately reflect the severity of fundus lesions in patients with DR and facilitate the adoption of corresponding treatment methods, doctors must grade patients with DR accurately. The International Ophthalmology Congress established a new grading standard that divides DR into five grades. This can be used to more accurately evaluate the degree of fundus lesions in patients with DR. Table 2 shows the data for grades 0–4 DR, and Figure 2 shows their fundus images. However, these DR grading tasks must be performed by professional ophthalmologists, which consumes the clinician’s time and
increases the doctor’s work pressure. AI is widely used in DR classification, and good results have been achieved.

**Research of AI-assisted grading diagnosis of DR based on color fundus photographs**  
Sugeno et al.\(^6\) created an AI model that could automatically classify DR based on the EfficientNet-B3. The model divides DR into five levels (grades 0-4). They used the APTOS 2019 dataset to evaluate the performance of the AI model, and the final results showed that the DR classification model’s accuracy was as high as 0.98. Similarly, Zhang et al.\(^6\) constructed an AI hierarchical model by ResNet-34 and Inception-V3 networks, which can automatically divide DR into five levels: no DR, mild DR, moderate DR, severe DR, PDR. In addition, they collected 1089 color fundus photographs to evaluate the performance of the AI model. Finally, the DR classification model’s sensitivity was 0.93, specificity was 0.93; the negative predictive values (NPV) were 0.95, whereas the positive predictive values (PPV) were 0.91. Xu et al.\(^6\) constructed a DR grading model based on ResNet-50. The model classifies DR by extracting the features of DR lesions, such as retinal hemorrhage, exudate, and neovascularization, from fundus images. They collected 15 988 color fundus photographs to train and test the model. The final result is that the DR grading model’s accuracy was 0.97, sensitivity was 0.97, and specificity was 0.98 (Table 3).

Zhang et al.\(^6\) used convolutional neural and graph convolutional networks to construct a deep graph correlation network. The network system can automatically grade DR without needing ophthalmologists to professionally label fundus images. In addition, the network system can divide DR into no DR, mild DR, moderate DR, severe DR, and PDR based on the severity of the DR. The EyePACS-1 and Messidor-2 datasets were used to evaluate the performance of the network system. Finally, the system’s accuracy was 0.90, sensitivity was 0.88, and specificity was 0.91 for the EyePACS-1 dataset, whereas they were 0.91, 0.90, and 0.93 for the Messidor-2 dataset. To achieve an automatic grading of DR, Alyoubi et al.\(^6\) created a DR grading model based on the convolution neural network 299, convolution neural network 512, and EfficientNetB0 networks. The model divides DR into five levels: no DR, mild DR, moderate DR, severe DR, and PDR. They used color fundus photographs from two public datasets (DDR and APTOS Kaggle 2019) to evaluate the AI model, and the final result was the accuracy of 0.89 and 0.84 for the DDR and APTOS Kaggle 2019 datasets, respectively. Dai et al.\(^6\) constructed a DeepDR system by ResNet, which can automatically grade DR by detecting DR lesions on color fundus photographs. They used 87 5705 color fundus photographs to train and test the system. Finally, the result was the system’s AUC for mild DR, moderate DR, severe DR, and PDR were 0.90, 0.94, 0.95, and 0.97, respectively. Liu et al.\(^6\) constructed a DR automatic grading model based on various convolution neural networks and used three hybrid model structures to improve the grading performance of the AI model. They evaluated the model’s grading performance using three public datasets (EyePACS, APTOS, and DeepDR). Finally, the model’s accuracy was 0.86, sensitivity was 0.99, specificity was 0.75, and precision was 0.91.

The accurate and reliable classification of DR plays an important role in preventing DR-induced visual impairment and blindness. Erçiyas and Barışçı\(^6\) used a convolution neural network, transfer learning, and attention mechanisms to construct a DR classification model. This model classifies DR by detecting and classifying DR lesions on color fundus photographs. After testing, the model’s accuracy and AUC were 0.99 and 1.00, respectively, for the Kaggle dataset. Katz et al.\(^6\) created an AI classification model by the W-net. This model divides patients with DR into DR and DME groups by identifying the DR and DME markers on color fundus photographs. They used 6981 color fundus photographs to train and test the AI model, and the final results showed that the model’s classification accuracy was 0.99. Yaqoob et al.\(^6\) used the ResNet-50 network and random forest algorithm to construct a DME classification model. They used Messidor-2 datasets to train and test the model. Additionally, they compared the proposed model with six state-of-the-art network models. After testing, the AI model’s accuracy was 0.96, and its performance was significantly better than those of the six most advanced network models (Table 3).

**Research of AI-assisted grading diagnosis of DR based on other ophthalmic imaging data**  
Wang et al.\(^6\) established an AI model that could automatically grade DR using a convolution neural network and ultra-widefield fluorescein angiography (UWFA) images. The model graded DR by analyzing the ischemia and deep fistula indices on UWFA images and divided DR into three grades: normal, NPDR, and PDR. After verification, the model’s accuracy for DR classification was 0.89. Zang et al.\(^6\) constructed a DR classification model based on a 3D convolution neural network. This model divides DR into three categories: non-referrable DR (nrDR), referable DR (rDR), and vision-threatening DR (vtDR). They collected 355 OCT images to test the model. Finally, the model’s AUC was 0.96 and 0.83 for the rDR and vtDR classifications. Tang et al.\(^6\) constructed a DL network system using ResNet-18 and ResNet-34 to classify DR and DME. In addition, to evaluate the performance of the DL system, they used 100 727 OCT images for training and verification. The final result was that the system’s AUC was >0.89, and its performance was excellent. Asif et al.\(^6\) constructed a ResNet50-based model and applied it to DME classification. A total of 84452 OCT images were collected.
AI-assisted DR diagnosis

Table 3 Summary of AI research in DR grading diagnosis

<table>
<thead>
<tr>
<th>Research</th>
<th>Grading</th>
<th>Image type</th>
<th>Dataset</th>
<th>AI model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sugeno et al[48]</td>
<td>No DR, mild DR, moderate DR, severe DR, PDR</td>
<td>Fundus image</td>
<td>APTOS 2019 dataset</td>
<td>EfficientNet-B3</td>
</tr>
<tr>
<td>Zhang et al[49]</td>
<td>No DR, mild DR, moderate DR, severe DR, PDR</td>
<td>Fundus image</td>
<td>1089 images</td>
<td>ResNet-34, Inception-V3</td>
</tr>
<tr>
<td>Xu et al[50]</td>
<td>No DR, mild DR, moderate DR, severe DR, PDR</td>
<td>Fundus image</td>
<td>15988 images</td>
<td>ResNet-50</td>
</tr>
<tr>
<td>Zhang et al[51]</td>
<td>No DR, mild DR, moderate DR, severe DR, PDR</td>
<td>Fundus image</td>
<td>EyePACS-1 dataset, Messidor-2 dataset</td>
<td>Convolution neural network, Graph convolutional network</td>
</tr>
<tr>
<td>Alyoubi et al[52]</td>
<td>No DR, mild DR, moderate DR, severe DR, PDR</td>
<td>Fundus image</td>
<td>DRR dataset, APTOS Kaggle 2019 dataset</td>
<td>Convolution neural network 299, Convolution neural network 512, EfficientNet-B0</td>
</tr>
<tr>
<td>Dai et al[53]</td>
<td>No DR, mild DR, moderate DR, severe DR, PDR</td>
<td>Fundus image</td>
<td>875705 images</td>
<td>ResNet</td>
</tr>
<tr>
<td>Liu et al[54]</td>
<td>No DR, mild DR, moderate DR, severe DR, PDR</td>
<td>Fundus image</td>
<td>EyePACS dataset, APTOS dataset, DeepDR dataset</td>
<td>EfficientNet-B4, EfficientNet-B5, NASNet-Large, Xception, InceptionResNetV2</td>
</tr>
<tr>
<td>Erçiyas and Barigo[55]</td>
<td>Classification</td>
<td>Fundus image</td>
<td>Kaggle dataset</td>
<td>Convolution neural network, Transfer learning, Attention mechanism</td>
</tr>
<tr>
<td>Katz et al[56]</td>
<td>Classification</td>
<td>Fundus image</td>
<td>6981 images</td>
<td>W-net</td>
</tr>
<tr>
<td>Yaqoob et al[57]</td>
<td>No referable DME, referable DME</td>
<td>Fundus image</td>
<td>Messidor-2 dataset</td>
<td>ResNet-50, Random Forest</td>
</tr>
<tr>
<td>Wang et al[58]</td>
<td>Normal, NPDR, PDR</td>
<td>UWFA</td>
<td>399 images</td>
<td>Convolution neural network</td>
</tr>
<tr>
<td>Tang et al[59]</td>
<td>Classification</td>
<td>OCT</td>
<td>355 individuals</td>
<td>Convolution neural network</td>
</tr>
<tr>
<td>Tang et al[60]</td>
<td>Classification</td>
<td>OCT</td>
<td>100727 images</td>
<td>ResNet-18, ResNet-34</td>
</tr>
<tr>
<td>Asif et al[61]</td>
<td>Classification</td>
<td>OCT</td>
<td>84452 images</td>
<td>ResNet-50</td>
</tr>
</tbody>
</table>


to train and validate the model. After validation, the model’s accuracy was 0.99 (Table 3).

Accurate DR classification is essential for the treatment of patients with DR. However, it takes time and effort for professional doctors to complete these tasks. The above studies show that AI performs well in completing DR grading diagnoses and can automatically complete DR grading, saving clinicians’ time, greatly reducing doctors’ workload, providing great help to clinicians, and portraying important clinical significance.

Research of AI in DR Lesions Segmentation

Research of AI-assisted DR lesion segmentation based on color fundus photographs The grading and classification of DR are very important for its diagnosis, detection, and treatment; however, DR lesion segmentation is also of great significance in diagnosing DR. Because professional ophthalmologists complete DR diagnosis by identifying DR lesions, accurate identification and segmentation of DR lesions also play an important role in the diagnosis and treatment of DR. Wan et al[72] constructed an EAD-Net network by U-Net network, which can automatically segment the pathological changes of DR, such as microaneurysm, retinal hemorrhage, and exudates, on color fundus photographs to assist DR diagnosis. They evaluated the network’s performance using a public dataset: e_ophtha_EX dataset, and the final result was that the sensitivity was 0.93, specificity was 1.00, and accuracy was 1.00. Similarly, Alyoubi et al[63] constructed an AI model that could automatically segments DR lesions by the CNN512 and YOLOv3 networks and evaluated the model’s performance using a public dataset: DDR dataset. This model can automatically segment microaneurysms, hemorrhages, hard exudates, soft exudates, and other DR lesions on color fundus photographs. Finally, the result was that the model’s accuracy was 0.89, sensitivity was 0.89, and specificity was 0.97. To accurately segment DR lesions on color fundus photographs, Zheng et al[73] constructed a DR lesion segmentation model based on the U-Net network and used a conditional generative adversarial network to improve the model’s performance. To evaluate the model’s performance, they tested it using four public datasets (E_ophtha_EX, DiaReTDB1, HEI-MED, and Messidor). The final results were that the model’s accuracy in the four datasets was 0.95, 0.92, 0.89, and 0.90, respectively (Table 4).

Research of AI-assisted DR lesion segmentation based on other ophthalmic imaging data Hu et al[74] proposed an AI model through the ResNet50 network, which can diagnose macular edema caused by diseases such as DR by segmenting pathological features such as subretinal fluid (SRF) and pigment epithelial detachment (PED) in OCT images. In the study, they used the OCT dataset (from the “AI Challenger” platform) and 3DIRCADb dataset to train and evaluate the performance of the model. Additionally, they also compared the model with other algorithms. After testing, the model has higher accuracy of lesion segmentation, and the accuracy of lesion segmentation was better than other algorithms (Table 4).
Accurate segmentation and identification of DR lesions are very important for diagnosing DR. AI-assisted segmentation of DR lesions can greatly help doctors diagnose DR, which saves diagnosis time, improves diagnosis efficiency, and greatly reduces the doctors’ work pressure. In addition, the accurate segmentation of DR lesions is beneficial for the grading and screening diagnosis of DR.

**LIMITATIONS AND CHALLENGES**

The above research show that AI is widely used in DR cases, and many good results have been achieved, especially in the image processing and recognition of DR. Furthermore, many AI models in the present study showed good performance in diagnosing, screening, detecting, and grading of DR; however, there are also many limitations and challenges in AI research, which can hinder further research on AI in DR cases and clinical application of AI. We have listed some limitations and challenges encountered using AI in DR cases.

**Dataset Size** [58-74] The performance of an AI model is closely associated with the dataset size. The larger the dataset, the more accurate the AI model. However, in many studies, the dataset used for AI model training and verification is small, which will affect the reliability of AI models performance [77]. Therefore, the dataset used for AI research should contain a sufficient sample size.

**Accuracy of Manual Labeling of DR Images** [78-79] The accuracy of the DR image annotation has an important impact on the performance of AI models. In many studies, images in the datasets were annotated by professional doctors [80]. In addition, when annotating DR images, multiple experts must annotate the same image several times to ensure the accuracy of manual labeling.

**Quality of the DR Image in the Dataset** [81-82] In the AI research, the quality of DR images in datasets has a vital impact on the results. The higher the quality of the DR image, the higher the accuracy of the AI model. However, image quality is affected by many factors, such as the advanced level of equipment, operator proficiency, and degree of patient cooperation [83]. Therefore, high-quality images should be selected in AI research to eliminate the influence of image quality on research results.

**Clinical Verification of the AI Models** [84-86] In many studies, the AI model performed well on the external verification set; however, this is the only result obtained under the research conditions. Because the difference between “clinical conditions” and “research conditions”, a series of problems are often encountered in the practical application of the superior AI model, affecting the accuracy of the AI model.

**Clinicians’ Acceptance of AI Models** [87-88] AI is a branch of computer science, but for clinicians, it is beyond the professional scope of clinicians, which will lead to clinicians’ lack of professional knowledge related to AI, prone to the “black box phenomenon”, resulting low acceptance of AI, thus hindering AI application [89].

**Protection of Patient Medical Data and Privacy** AI research requires the use of patient medical data and privacy, and the leakage of these important information can have serious consequences. The protection of patient medical data and privacy will have a significant impact on the clinical application of AI.

**Accountability for Medical Accidents** AI is not flawless, it can also make mistakes, and even cause bigger problems. Who should bear legal responsibility for medical accidents caused by AI? This important issue also requires the establishment of clear ethical norms and legal regulations. In conclusion, because DR is the most common retinal vascular disease and the main cause of blindness in the working-age population, it has always been a focus of attention and research. Late-onset DR can lead to irreversible visual impairment and blindness. However, timely diagnosis and effective treatment of DR can control the condition in most patients, thereby reducing and avoiding the occurrence of blindness. Therefore, early diagnosis and treatment are important in patients with DR. In recent years, AI has developed rapidly, promoting its application in DR cases, particularly in the diagnosis, grading, and lesion segmentation of DR. AI research is increasingly extensive. AI can process the fundus images of DR to achieve the diagnosis, grading, and even lesion segmentation of DR. This can not only facilitate the screening of DR patients, but also help to determine the treatment plan for patients. This review reports the significant progress made in the research on AI models based on DR fundus image-processing technology. The DR intelligent diagnostic system can effectively improve the current situation of insufficient medical resources in poor and backward areas, effectively alleviate...
AI-assisted DR diagnosis

the tension of ophthalmologists caused by DR screening, effectively help clinicians complete DR grading to determine the best treatment plan for patients, effectively complete lesion segmentation, and help doctors diagnose DR quickly and accurately to reduce their workload. The limitations and challenges in AI research may hinder the further application and development of AI in diagnosing and treating DR; however, with the continuous progress in AI technology, these limitations and challenges will be gradually overcome, enabling AI to provide better help for the clinical diagnosis and treatment of DR.

ACKNOWLEDGEMENTS

Authors’ contributions: Liu YF and Ji YK wrote the manuscript. Fei FQ and Chen NM reviewed the manuscript. Fei XZ and Zhu ZT designed the research and collected the literature.

Foundation: Supported by Huzhou Science and Technology Planning Program (No.2019GY13).

Conflicts of Interest: Liu YF, None; Ji YK, None; Fei FQ, None; Chen NM, None; Zhu ZT, None; Fei XZ, None.

REFERENCES


38 Relhan N, Jr Flynn HW. The Early Treatment Diabetic Retinopathy Study historical review and relevance to today’s management of diabetic macular edema. *Curr Opin Ophthalmol* 2017;28(3):205-212.


40 Wang RY, Zuo GX, Li KK, Li WT, Xuan ZQ, Han YZ, Yang WH. Systematic bibliometric and visualized analysis of research hotspots and trends on the application of artificial intelligence in diabetic retinopathy. *Front Endocrinol* 2022;13:1036426.


AI-assisted DR diagnosis


82 Ma D, Pasquale LR, Girard MJA, Leung CKS, Jia YL, Sarunic MV, Sappington RM, Chan KC. Reverse translation of artificial intelligence in glaucoma: connecting basic science with clinical applications. *Front Ophthalmol* 2023;2:1057896.

83 Dioa JL, Chen XX, Sheng Y, Li J, Chen YQ, He LF, Chen SN, Mou P, Ma XY, Wei RL. Research progress and application of artificial


