Survey of Ophthalmology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Survey of Ophthalmology

journal homepage: www.elsevier.com/locate/survophthal

Review article

Multimodal artificial intelligence in retinopathy of prematurity: A comprehensive narrative review

Xinyu Zhao [a], Zhenquan Wu [a], Shirou Wu [a], Kaixuan Cui [a], Wai-Ching Lam [b], Wei-Chi Wu [c], Weihua Yang [a], Baiying Lei [d], Shuqiang Huang [e], Wenbin Wei [f,*], Wei Chi [a,**], Guoming Zhang [a,**]

[a] Shenzhen Eye Hospital, Shenzhen Eye Medical Center, Southern Medical University, Shenzhen, China
[b] Department of Ophthalmology, Vancouver General Hospital, University of British Columbia, Vancouver, Canada
[c] Department of Ophthalmology, Chang Gung Memorial Hospital, College of Medicine and Chang Gung University, Linkou, Taoyuan, Taiwan
[d] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
[e] College of Cyber Security, Jinan University, Guangzhou, China
[f] Beijing Tongren Eye Center, Beijing Key Laboratory of Intraocular Tumor Diagnosis and Treatment, Beijing Ophthalmology & Visual Sciences Key Lab, Medical Artificial Intelligence Research and Verification Key Laboratory of the Ministry of Industry and Information Technology, Beijing Tongren Hospital, Capital Medical University, Beijing, China

ARTICLE INFO

Keywords:
Retinopathy of prematurity
Artificial intelligence
Deep learning
Multimodal imaging
Vascular severity score
Federated learning

ABSTRACT

Retinopathy of prematurity (ROP) remains a leading cause of preventable childhood blindness globally, particularly in regions with limited screening resources. Traditional diagnosis relying on subjective interpretation of fundus images faces challenges of inter-observer variability and manual analytical limitations. Recent advances in artificial intelligence (AI) have demonstrated substantial potential to achieve intelligent ROP management. We present a comprehensive review that synthesizes contemporary advances in AI-driven ROP management and reviewed the AI technologies based on retinal imaging, clinical text data, and smartphone-captured images transform ROP diagnosis, risk prediction, and treatment suggestion. We also illustrate the applications of AI in intelligent diagnosis and treatment of ROP through multimodal imaging. Despite remarkable progress, challenges persist in data heterogeneity, model generalizability, and real-world integration. By mapping technical breakthroughs to unmet clinical needs, we provide actionable insights to accelerate the development of equitable, clinically deployable AI solutions for preventing ROP-related vision loss.

1. Introduction

Retinopathy of prematurity (ROP) is a vision-threatening retinal vascular disorder affecting premature infants, characterized by aberrant neovascularization and fibrovascular proliferation.[18,44,97] The incidence of ROP exhibits a strong inverse correlation with gestational age (GA) and birth weight (BW),[11] a situation further complicated by improvements in neonatal intensive care survival rates, particularly in regions with limited healthcare resources.[36] More than 30,000 newborns worldwide go blind each year due to ROP,[118] underscoring the urgent need for scalable screening and effective intervention systems.

Diagnosing ROP faces significant challenges. Traditional methods rely on subjective ophthalmoscopic examinations or the expert interpretation of digital retinal images. These approaches encounter critical limitations, including substantial inter-observer variability in zoning, staging, and plus disease assessment,[40] persistent shortages of trained specialistsk especially in low-resource settings,[14,40] and an inherent inability to quantify dynamic vascular changes crucial for accurate prognostication.[13] While newer imaging technologies like ultra-widefield (UWF) fundus photographs, optical coherence tomography (OCT), and even smartphone adapters improve accessibility, they generate vast volumes of data that exceed manual analysis capacity,

* Corresponding author.
** Correspondence to: Shenzhen Eye Hospital, Shenzhen Eye Medical Center, Southern Medical University, 18 Zetian Road, Futian District, Shenzhen 518040, China.
E-mail addresses: weiwenbintr@163.com (W. Wei), chiwei@sz-eyes.com (W. Chi), zhangguoming@sz-eyes.com (G. Zhang).

https://doi.org/10.1016/j.survophthal.2026.05.006
Received 30 September 2025; Received in revised form 2 May 2026; Accepted 4 May 2026
Available online 14 May 2026
0039-6257/© 2026 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

underscoring the urgent need for automated objective tools.[41,127]

Artificial intelligence (AI) has emerged to address these challenges by enabling automated, accurate, and scalable ROP assessment. Early studies focused on single-task image-based models for plus disease detection or binary ROP classification,[12,75] while recent research has evolved toward multi-task systems, multi-modal integration, and prediction of disease progression.[64,122] By leveraging diverse clinical and imaging data, AI systems now offer capabilities that extend beyond screening, supporting risk prediction, severity scoring, and treatment suggestion.

We emphasize how recent AI advances in ROP management by critically comparing existing approaches and clarifying their strengths and limitations within clinical and diagnostic contexts. We summarize AI models, highlight barriers such as dataset heterogeneity, single-center bias, difficulties integrating with hospital information systems, and discuss practical pathways for equitable deployment of AI-based ROP screening, drawing on telemedicine experiences and health-economic evaluations. Ultimately, by synthesizing existing AI research in ROP, this review outlines its developmental trajectory, current state, and future directions, and offers insights to facilitate the translation of intelligent technologies into real-world clinical practice. To clearly delineate the novelty and value of the present review, Table 1 provides a concise comparison of recent review studies on AI applications in ROP, highlighting differences in scope, data modalities, and emerging topics. The multimodal data, main tasks of AI models, and clinical applications discussed are concisely illustrated in Figure 1.

2. Results

2.1. Multimodal data sources and quality control

2.1.1. Fundus photographs

Wide-field color fundus photography (CFP), currently the most commonly used imaging modality for objectively documenting the condition of ROP in infants, plays a vital role in the development and validation of AI models. These images are primarily captured using devices such as the RetCam 3 and RetCam Envision, which share highly similar imaging principles and acquisition angles.[121] In China, there are also several domestically produced devices, such as the ROP Screening System and the Nautilus Fundus Imaging System.[132] With a viewing-field of approximately 130°, these systems image both the posterior pole and the peripheral retina to ensure comprehensive coverage of lesion areas. At present, research centers worldwide use RetCam imaging data to annotate ROP lesions in terms of zone, stage, and plus disease, serving as the foundational dataset for training and external validation of AI models.

In recent years, Mao and colleagues[72] used the Optos 200Tx UWF scanning laser ophthalmoscope (non-contact mode) to image infants with ROP. This device successfully captured color and fundus fluorescein angiography (FFA) images from infants at various stages of ROP, including aggressive ROP (A-ROP), without the need for anesthesia or direct ocular contact. Jayanna and coworkers[54] conducted a retrospective analysis of CFP using the Optos 200Tx in a tertiary eye care outpatient setting. Their study demonstrated the effectiveness of CFP in capturing detailed images of the peripheral retina, which is crucial for accurate ROP diagnosis. The Optos 200Tx offers a viewing-field ranging from 180° to 200°, providing superior visualization of peripheral retinal lesions compared to traditional devices. [121]

Although most AI models for ROP screening have been developed using wide-field RetCam images, differences in imaging devices, particularly in fields of view, raise important concerns regarding model generalizability in real-world clinical deployment. To address this issue, a recent external validation study evaluated a RetCam-trained AI model using retinal images acquired with the Phoenix ICON camera (NeoLight, Scottsdale, AZ), which provides a smaller field of view. Despite this difference, the algorithm achieved comparable performance, demonstrating 100% sensitivity for TR-ROP at the infant level, suggesting that well-trained models may retain diagnostic robustness across imaging platforms.[28]

2.1.2. FFA images

In infants, FFA images are typically obtained by switching a dedicated fundus camera to a FFA camera. Various infant fundus cameras, both domestic and international, are equipped with this function. The orientation of the captured images is the same as that of CFPs, covering a 130° viewing-field.[33] In recent years, some ophthalmologists have begun attempting to use the angiography function of Optos 200Tx designed for adults in infants, achieving similarly excellent image quality. Mao and coworkers[72] employed the Optos 200Tx to perform ultra-wide field (UWF)-FFA, which enables a more comprehensive visualization of peripheral ROP lesions. This enhances the ability to identify and monitor disease progression, making it a valuable tool in ROP screening and follow-up.

2.1.3. OCT images

OCT has emerged as a powerful cross-sectional imaging modality that complements conventional fundus imaging and FFA in the assessment of ROP.[41,79] OCT provides depth-resolved, micron-level visualization of retinal layers, enabling clinicians to inspect the 3-dimensional anatomy of the vascular-avascular junction and neurovascular tissue.[113]

Table1
comparison of recent review studies on AI in ROP.

Review	Year	Scope of tasks	Data modalities	Emerging topics	Coverage of studies
Bao et al.[8]	2021	Screening, detection	Retinal images	Telemedicine, DL, ML	Up to 2020
Tan et al.[104]	2022	Screening	Retinal images	Teleophthalmology	Up to 2021
Bai et al.[6]	2022	Screening, detecting	Demographic data,retinal images	DL	Up to 2021
Nakayama et al.[76]	2023	Screening, severity grading, detection, prediction	Retinal images	DL	Up to 2022
Hoyek et al.[46]	2023	Screening, detection,biomarker discovery	Retinal images, blood biomarkers, multiomics data	Biomarkers, DL, ML	Up to 2023
Shah et al.[98]	2023	Detection, risk prediction, screening	Retinal images, demographic data	ML, DL, VSS, telemedicine	Up to 2023
Ramanathan et al.[91]	2023	Image quality, detection, staging	Retinal images	DL, VSS	Up to 2022
Kearns et al.[59]	2025	Screening, detection	Clinical data,retinal images	DL, telemedicine	Up to 2025
Jafarizadeh et al.[53]	2025	Screening, detection, staging, treatment, prognostication	Retinal images,demographic data	ML, DL, XAI, telemedicine	Up to 2024
Marra et al.[73]	2025	Screening, treatment, management	Clinical data, retinal images	DL, telemedicine	Up to 2025
Our review	2025	Detection, prediction, severity assessment, treatment, management	Multimodal	VSS, telemedicine, economic evaluation, LLM, XAI, FL, DL, ML	Up to 2025

AI: artificial intelligence, ROP: ROP: retinopathy of prematurity, DL: deep learning, ML: machine learning, VSS: vascular severity scoring, XAI: explainable artificial intelligence, LLM: large language models, FL: federated learning.

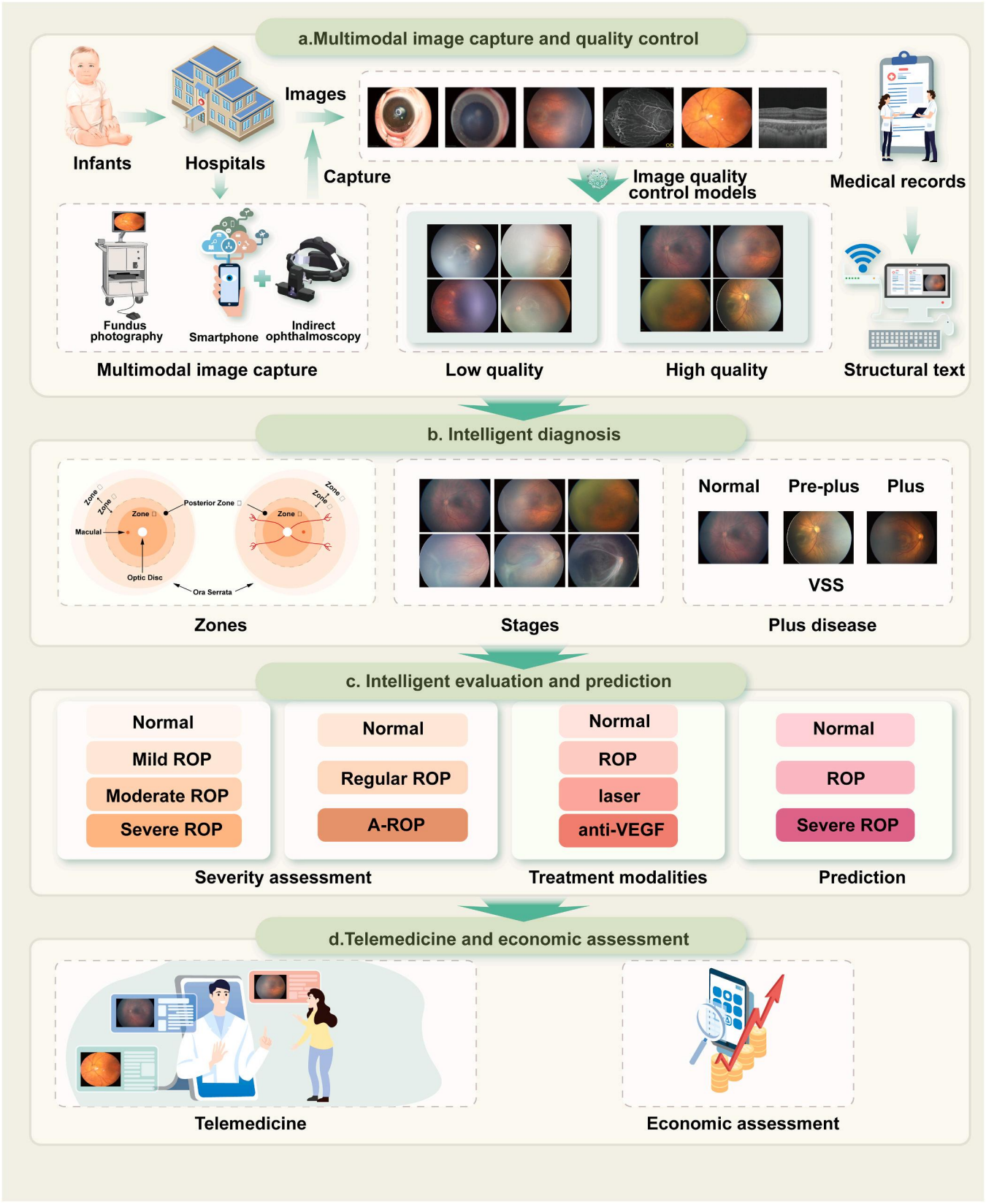


Fig. 1. Concise illustration of the multimodal AI workflow in ROP, showing the integration of clinical data and imaging for diagnosis, treatment suggestion, and prognosis prediction. ROP: retinopathy of prematurity; VSS: vascular severity scoring; A-ROP: aggressive ROP.

Burt and coworkers[13] employed an investigational 800 kHz swept-source UWF-OCT, which could offer an objective, continuous biomarker that outperforms traditional ordinal staging. It indicates that OCT adds depth and volume dimensions to ROP assessment, paving the way for precision medicine in infantile retinal disease.

2.1.4. Smartphone-captured images

Fundus images captured using smartphone in combination with a binocular indirect ophthalmoscopy have gained attention in recent years due to their simplicity, low cost, and accessibility. This technique has been gradually adopted in certain countries and regions. Developing AI models based on these images represents a promising direction for future research and application in countries and regions with limited medical resources.[82,127]

2.1.5. Clinical text data

Beyond imaging data, clinical text data, including demographic information, physiological metrics, and longitudinal clinical records, serve as another critical and valuable input for developing intelligent ROP models, offering complementary insights into disease risk and progression. The Postnatal Growth and ROP Study (G-ROP) study developed screening criteria based on large-scale, multi-center clinical text data (including BW, GA, and postnatal weight gain) from 7483 cases. This criterion achieved 100% sensitivity in predicting type 1 ROP while reducing unnecessary examinations by 32.5%.[10] Poppe's study[87] constructed a model using physiological text data from the first postnatal month (such as SpO/FiO_2ratio and heart rate skewness). It achieved a balanced accuracy of 0.81 in predicting treatment-requiring ROP (TR-ROP), verifying the effectiveness of clinical text data in ROP prediction. In addition to static text data, a historical cohort study from Iran applied a random survival forest to longitudinal clinical text data from 338 preterm infants to model the time to development of ROP. The model achieved a concordance index of 0.878 and identified key nonlinear predictors. This work highlights the value of longitudinal clinical text data for capturing the dynamic and time-dependent nature of ROP risk.[105]

2.1.6. Standardization of image quality

Image quality standardization is a core prerequisite for the accurate application of fundus imaging in intelligent diagnosis. Its main goal is to minimize the impact of data heterogeneity on diagnostic model performance by enforcing strict quality control, which uses quantitative indicators to screen qualified images and applies targeted enhancement to reduce the influence of low-quality data.[37,110] The core lies in establishing multi-dimensional and interpretable evaluation standards.[108] Image quality control and enhancement models have been investigated in adult multimodal fundus images and have achieved good performance.[56,119] Meanwhile, similar models specifically designed for infant fundus images, such as the DeepQuality, have also been developed[66]. DeepQuality accurately evaluates integrity, illumination, and clarity, scores overall image quality, and performs targeted enhancement. When applied to over 2 million real-world images, DeepQuality revealed that 58.3% had quality defects, with significant variation across regions and hospital types. Targeted enhancement of these substandard images helped improve clinicians' diagnostic performance. Furthermore, integrating DeepQuality with AI diagnostic models further increased ROP detection accuracy, providing valuable insights for future image-based disease screening systems.[66] In addition, a DL–based infant fundus image quality feedback system (DLIF-QFS) was developed to automatically assess overall image quality and identify common acquisition errors during infant fundus photography. The DLIF-QFS demonstrated robust performance in distinguishing poor, adequate, and high quality images, with area under the receiver operating curve (AUC) of 0.802, 0.691, and 0.926, respectively. Importantly, the system improved diagnostic accuracy and expert consistency in clinical evaluation, highlighting the value of quality feedback and error identification in standardizing infant fundus imaging and supporting efficient ROP screening.[117]

High-quality, standardized modal data serve as the foundation upon which AI models are developed, and their availability directly influences performance in ROP diagnosis.

2.2. AI technology in ROP diagnosis

This section systematically reviews the progress of AI technologies in ROP diagnosis along data modalities: starting with single imaging modalities, then clinical text, followed by low-cost smartphone video, and finally discussing multimodal fusion. To evaluate the model's capacity for detecting ROP, various metrics, including accuracy, sensitivity, specificity, and the AUC, were extracted and presented in Table 2.

2.2.1. Image-based AI technology in ROP diagnosis

Application in ROP zoning. Zone identification is a foundational step in ROP assessment, as accurate zoning guides clinical decision-making for screening frequency, intervention timing, and treatment modalities. Recent advances in AI have enabled automated zoning in fundus images, demonstrating promising performance across different imaging systems and clinical scenarios.

Zhao and colleagues[133] first developed a convolutional neural networks (CNN)-based framework to automatically identify Zone I from RetCam images. Using 2849 images with optic disc and macula, zone I identification accuracy reached 91%, with 0.8-second per-image processing; however, it can only identify Zone I in images that contain both the optic disc and macular center, which limits its ability to precisely locate and assess the peripheral lesions in clinical settings. Agrawal and coworkers[1] used 2 lightweight U-Nets to segment optic disc and vessels, infer macula, and delineate Zones I–III from RetCam images in 30 ms with 98% accuracy. These advancements highlight significant progress toward more accurate and efficient automated zoning in ROP diagnosis.

Application in ROP Staging. ROP staging, which categorizes the lesion severity, is critical for determining the urgency of intervention and predicting disease progression. Among AI applications in ROP, staging models represent the core focus of automated diagnosis efforts due to their clinical importance and relatively well-defined classification criteria.

Huang and coworkers[49] and Li and coworkers[67] were all focused on the early stage identification. Huang trained a CNN on 11,372 RetCam images to separate "no ROP" from stages 1–2 with an accuracy reached 92.2% and AUC 0.96, confirming strong early-screening value. Li focused on stages 1–3 classification using 18,827 RetCam images. U-Net segmentation followed by DenseNet classification yielded stage-wise sensitivities of 90.2%, 92.8%, and 91.8%, with specificities of 97.7%, 98.7%, and 99.3%. This study demonstrated high sensitivity and specificity, but did not include stages 4–5 of ROP. Compared with this single network, Peng and coworkers[86] adopted a 3-stream framework to cover all 5 ROP stages. Analyzing 635 images, it achieved a weighted recall of 0.91 and a precision of 0.91; however, stage 1 was prone to misdiagnosis, mainly due to limited training data and subtle clinical differences from stage 2 in this study. Tong's group's study, based on a large dataset, demonstrated excellent performance in ROP staging.[109] They developed a 101-layer CNN (ResNet) and a faster region-based CNN (Faster-RCNN), which achieved 95.7% accuracy for ROP stage classification. Another study improved both internal and external performance by combining images from different populations and cameras. Chen and coworkers[20] pooled 10,894 images from North American RetCam and Nepali 3nethra cameras (Forus Health, Bengaluru, India). The model reached AUC 0.99 on North American data and 0.97 on Nepali data, showing clear cross-device benefit.

The ridge is the key morphological hallmark marking the transition in ROP stages. Accurate segmentation of the ridge not only determines staging precision, but also provides AI models with quantifiable morphologic features. Traditional studies treated the ridge only as part

Table 2
Characteristics of AI Studies for Diagnosis, Prediction, Assessment, and Treatment of ROP.

Author (Country, Year)	Purpose	Algorithm	Algorithm Type	Imaging	Number of Image	Dataset type	Performance	
ROP Diagnosis								
Image-based AI technology								
Zoning								
	Zhao et al. (China, 2019)[133]	Automatic identification of zone I	DCNN	Classification	RetCam	2849 (training: validation: test = 20:3:1)	Private	Accuracy = 91%
	Agrawal et al. (India, 2021)[1]	Detect all ROP zones (I, II, III) automatically	U. Net, Circle Hough Transform	Classification, Segmentation	RetCam, Neo	3000 (training: validation: test = 70:15:15)	Private	Accuracy = 98%
Staging								
	Burt et al. (USA, 2024)[13]	Automated quantification of ROP stage	U-Net	Segmentation	Ultrawidefield OCT	3347 (training: validation = 3:1)	Private	DSC = 0.61 ± 0.13
	Huang et al. (Taiwan, 2021)[49]	Stages I–II ROP classification	CNN	Segmentation	RetCam	11,372 (training: validation = 9:1) 244 for testing	Private	For stage 1 ROP: Sensitivity = 91.8%, Specificity = 94.5%; For stage 2 ROP: Sensitivity = 89.8%, Specificity = 99.0%
	Li & Liu (China, 2022)[67]	Automatic identification of Stages I–III	U-Net, DenseNet	Segmentation	RetCam	18,827 (training: validation: test = 28:7:1)	Private	For stage 1 ROP: Sensitivity = 90.2%, Specificity = 97.7%; For stage 2 ROP: Sensitivity = 92.8%, Specificity = 98.7%; For stage 3 ROP: Sensitivity = 91.8%, Specificity = 99.3%
	Peng et al. (China, 2021)[86]	ROP staging classification	ResNet18, DenseNet121	Classification	RetCam	3147 (training: validation: test = 3:1:1)	Private	Precision = 0.91, Accuracy = 98.3%
	Tong et al. (China, 2020)[109]	5-level ROP staging classification	ResNet-101, Faster R-CNN	Classification	RetCam	36,231 (training: validation = 9:1)	Private	Accuracy = 95.7%
	Chen et al. (USA/Nepal, 2021)[20]	ROP stage(I–III) detection	CNN	Classification	RetCam,3nethra	10,894 (training: test = 9:1)	Private	In North America: AUROC = 0.99, Sensitivity = 94%, Specificity = 96%; In Nepal:AUROC = 0.97, Sensitivity = 73%, Specificity = 99%
	Kurup et al. (India, 2018)[62]	Early ROP detection via demarcation line/ridge	k-means Clustering	Segmentation	RetCam	33	Private	Sensitivity = 100%, Specificity = 92%
	Jia et al. (China, 2025)[55]	ROP lesion segmentation	SeBSNet	Segmentation	RetCam	350 (training: validation: test = 7:2:1)	Private	ROC_AUC = 98.8%, Dice scores = 66.9%
Plus Disease								
	Brown et al. (USA, 2018)[12]	Automated diagnose plus disease in ROP	CNN	Classification	RetCam	5511 (training: validation = 4:1)	Private	AUC = 0.98, Sensitivity = 93%, Specificity = 94%
	Mao et al. (China, 2020)[71]	Automated diagnosis plus disease in ROP	DCNN	Classification, Segmentation	RetCam	6224 (training: test = 12:1)	Private	AUC = 0.93, Sensitivity = 95.1%, Specificity = 97.8%
	Deng et al. (China, 2023)[27]	Investigate retinal vascular morphological characteristics	U-Net	Segmentation	Fundus photography	Not reported	Private	Sensitivity = 72%, Specificity = 99%

(*continued on next page*)

Table 2 (*continued*)

Author (Country, Year)	Purpose	Algorithm	Algorithm Type	Imaging	Number of Image	Dataset type	Performance
ROP Diagnosis							
Image-based AI technology							
Zoning							
	Bai et al. (Australia, 2023)[7]	Automated diagnose plus disease in ROP	ROP.AI	Classification	RetCam	8052	Private	AUC = 0.75, Sensitivity = 84%, Specificity = 43%
Text-based AI technology							
	Takeda et al. (Japan, 2025)[103]	ROP occurrence using clinical data	Decision Tree, Random Forest, Gradient-Boosted Tree, Neural Network, Naive Bayes	Classification	-	-	Private	AUC-ROC = 0.93–0.94, Accuracy = 90.1–90.6%, Sensitivity = 94.6–95.7%, Specificity = 66.0–73.6%
	Eckert et al. (Brazil, 2012)[31]	Develop a ROPscore in ROP diagnosis	ROPscore	Classification	-	-	Private	AUC = 0.77 and 0.88
	Ricard et al. (USA, 2017)[95]	Develop a clinical screening tool for ROP	Logistic regression	Classification	-	-	Private	Sensitivity = 95%
Smartphone video-based AI technology							
	Young et al. (India, 2023)[127]	Assess efficacy of SBFI systems for ROP screening	ResNet18	Classification	SBFI (MII Retcam or MIO)	1522 (training: validation: test = 7:1:2)	Private	Diagnose TR-ROP: Sensitivity =100%, Specificity = 58.6%; Diagnose RW-ROP: Sensitivity = 80.0%, Specificity = 59.3%
	Ortiz et al. (Mexico/ Argentina, 2025)[82]	ROP screening using smartphone videos	Machine Learning	Classification	Smartphone+ magnifier	524 videos (training: validation: test = 4:1:40)	Private	Sensitivity = 93.3%
Multi-modal fusion							
	Legocki et al. (USA, 2023)[64]	Multivariate models to diagnose early RW-ROP	XGBoost	Classification	Handheld OCT	167	Private	AUC = 0.83, Sensitivity = 91.7%, Specificity = 77.8%
ROP Prediction							
	Chen et al. (China, 2024)[21]	Establish an AI algorithm for ROP risk prediction	Logistic regression, Decision tree, Multi-layer perceptron	Classification	-	-	Private	Predicting ROP:AUC = 0.81; Predicting TR-ROP:AUC = 0.83
	Shoeibi et al. (Iran, 2025)[100]	Design models for identifying TR-ROP	Machine Learning	Classification	-	-	Private	AUC = 0.69–0.87, Accuracy = 61–96%
	Foote et al. (the USA, 2025)[47]	Develop a model to predict TR-ROP	Machine Learning	Classification	-	-	Private	AUROC = 0.92–0.96
	Huang et al. (Taiwan, 2019)[50]	Prediction of Prediction of visual outcomes	Artificial Neural Network	Classification	RetCam	60 (training: validation = 4:1)	Private	In the IVI group: RMSE for SE = 0.131; In the laser group: RMSE for SE = 0.104
	He et al. (China, 2025)[43]	Predict TR-ROP	Machine Learning	Classification	RetCam	1705	Private	AUC = 0.78–0.85
	Lu et al. (China, 2023)[70]	retrospective validation predictive algorithms in two Chinese medical centers	G-ROP, CO-ROP, Alex-ROP, ROPscore	Classification	RetCam	Not reported	Private	G. ROP:Sensitivity = 96.0%, Specificity = 35.0%;Co-ROP: Sensitivity = 96.0%, Specificity = 64.3%; Alex-ROP:Sensitivity = 56.0%, Specificity = 61.4%; ROPscore: Sensitivity = 91.3%,

(*continued on next page*)

Table 2 (*continued*)

Author (Country, Year)	Purpose	Algorithm	Algorithm Type	Imaging	Number of Image	Dataset type	Performance
ROP Diagnosis							
Image-based AI technology							
Zoning							
							Specificity = 62.4%
	Sun et al. (China, 2021)[102]	Compare ROPScore and CHOP ROP for early prediction of ROP	Logistic regression	Classification	-	-	Private	ROPScore:Sensitivity = 55.8%, Specificity = 77.8%; CHOP-ROP: Sensitivity = 100%, Specificity = 21.4%
	Wu et al. (China, 2022)[122]	Predict ROP occurrence and severity	OC-Net and SE-Net	Classification	RetCam	7033	Private	AUC = 0.90, Accuracy = 52.8%, Sensitivity = 100%, Specificity = 37.8%
Severity assessment and treatment suggestion							
	Coyner et al. (Portland, 2021)[24]	Prediction of Severe ROP	i-ROP	Classification	RetCam	Not reported	Private	Sensitivity = 100%, Specificity = 48.9% and 80.8%
	Yenice et al. (Turkey, 2024)[61]	Automated detection of Type 1 ROP, Type 2 ROP and A-ROP	RegNetY002	Classification	RetCam	634 (training: test = 9:1)	Private	Detect A-ROP:AUC = 0.91, Accuracy = 91%, Sensitivity = 92%
	Wang et al. (China, 2018)[116]	Automated ROP screening using deep neural networks	Id-Net and Gr-Net	Classification	RetCam	25,884	Private	Id-Net:Sensitivity = 96.6%, Specificity = 99.3%; Gr-Net : Sensitivity = 88.5%, Specificity = 92.3%
	Yang et al. (China, 2024)[125]	Interpretable AI system for assessing ROP severity level	CNN	Classification	RetCam	6100 (training: validation = 4:1)	Private	AUC = 0.95, F1 = 0.76
	Liu et al. (China, 2023)[68]	AI system for screening and treatment modalities	ResNet-18, DenseNet-121	Classification	RetCam	24,495 (training: validation: test = 7:1:2)	Private	Accuracy = 86.3%, Sensitivity = 70.6%, Specificity = 94.1%
	Xie et al. (China, 2023)[124]	Automated diagnosis of AP-ROP	Knowledge Distillation	Classification	RetCam	Not reported	Private	Accuracy = 96%
	Zhang et al. (China, 2022)[131]	Automatic diagnosis for AP-ROP	CNN	Classification	RetCam	12,230 (training: validation: test = 3:1:2)	Private	Accuracy = 93.09%, Sensitivity = 87.70%, Specificity = 95.40%
	Lepore et al. (Italy, 2020)[65]	ROP management driven by fluorescein angiography	CNN	Classification	RetCam	835 (training: test = 9:1)	Private	AUC = 0.91, Accuracy = 0.88
VSS quantification							
	Campbell et al. (USA, 2021)[16]	Applications of AI for ROP screening	U-Net	Classification	RetCam	4175	Public	AUC = 0.98, Sensitivity = 100%, Specificity = 78%
	Bellsmith et al. (USA, 2020)[9]	Clinical and quantitative imaging features of AP-ROP	i-ROP	Classification	RetCam	5945	Private	VSS = 8.79 of AP-ROP
	Eilts et al. (German, 2023)[32]	Assessment of ROP regression and reactivation	i-ROP	Classification	RetCam	1046	Private	VSS = 2.7 after treatment
	Coyner et al. (Portland, 2024)[25]	Verify the AI-generated VSS improved plus disease diagnosis	i-ROP	Classification	RetCam	750	Private	AUROC = 0.94–0.98
	Sharafi et al. (USA, 2024)[99]	Automated diagnosis of plus disease	SVM	Classification	RetCam	76	Private	Accuracy = 0.86 ± 0.01
	Campbell et al. (USA, 2022)[15]	Validation of a VSS against expert diagnosis	i-ROP	Classification	RetCam	58	Public	Weighted kappa = 0.67, Pearson CC = 0.88
Telemedicine and economic evaluation							
	Greenwald et al. (USA, 2020)[39]	Evaluation of AI-based telemedicine	i. ROP	Classification	RetCam	613	Private	AUC = 0.99, Sensitivity = 100%, Specificity = 90%

(*continued on next page*)

Table 2 (*continued*)

Author (Country, Year)	Purpose	Algorithm	Algorithm Type	Imaging	Number of Image	Dataset type	Performance	
ROP Diagnosis								
Image-based AI technology								
Zoning								
		screening for ROP						
	Cole et al. (Nepal & Mongolia, 2022)[23]	Evaluate the performance of a DL algorithm for ROP screening	i-ROP, U-Net	Classification	Forus 3nethra neo, RetCam Portable	Not reported	Private	AUC-ROC = 1 in Nepal and 0.97 in Mongolia
	Morrison et al. (USA, 2022)[74]	Cost-effectiveness of AI-based ROP screening	Decision trees	-	-	-	Public	Sensitivity = 75–99%, Specificity = 60–99%
Development of AI models								
	Zhang et al. (China, 2019)[130]	Development of an automated screening system for ROP	AlexNet, VGG16, GoogLeNet	Classification	RetCam	420,365 (training: test = 10:1)	Private	Accuracy = 98.8%, Sensitivity = 94.1%, Specificity = 99.3%, Precision = 93.0%
	Ramachandran et al. (India, 2020)[90]	A deep learning framework for the detection of Plus disease	CNN	Classification	RetCam	289	Private	Sensitivity = 0.99, Specificity = 0.98
	Attallah et al. (Egypt, 2021)[4]	Automated tool for ROP	ResNet-50, Inception	Classification	RetCam	17,801	Public	AUC = 0.98, Accuracy = 93.2%
	Wallace et al. (USA, 2007)[114]	Quantify plus disease in ROP	ROP tool	Segmentation	RetCam	20	Public	Sensitivity = 95%, Specificity = 78%
	Gelman et al. (USA, 2007)[35]	Computer-based diagnosis plus disease	Integrated Curvature, Diameter, Tortuosity Index	Classification	RetCam	34	Private	AUC = 0.97
	Johnson et al. (USA, 2007)[57]	Quantify retinal vascular in ROP	VesselMap	Classification	NM200D	99	Private	Sensitivity = 71%, Specificity = 75%
	Hu et al. (China, 2019)[48]	Recognize ROP existence and severity	CNN	Classification	RetCam	2668 (training: validation: test = 20:3:3)	Private	Accuracy = 97%, Sensitivity = 96%, Specificity = 98%
	Redd et al. (USA, 2018)[92]	Detecting severe ROP	i-ROP	Classification	RetCam	4861	Private	AUC = 0.96, Sensitivity = 94%, Specificity = 79%
	Demirci et al. (USA, 2024)[26]	Segmentation for ROP diagnosis	SSL,Adversarial domain frameworks	Segmentation	RetCam	108	Public	mIoU = 29.6%, DSC = 45.2%, Accuracy = 97.1%
	Lu et al. (USA, 2022)[69]	Diagnose ROP using federated learning	i-ROP	Classification	RetCam	5245	Private	AUC = 0.93–0.96
Public dataset								
	Timkovic et al. (Czech Republic, 2023)[107]	Retinal image dataset of infants and ROP	CNN	Classification, Segmentation	RetCam,Phoenix	6004	Public	Not reported
	Yildiz et al. (USA, 2019)[126]	Diagnose plus disease in ROP	CNN	Classification	RetCam	5612	Public	AUC = 0.98 and 0.94
						6082 (training: validation: test = 6:1:3)	Public	Accuracy = 67%
	Akbari et al. (Iran, 2024)[3]	Detect ROP	Not specified	Classification	RetCam	1533	Public	Not reported
	Agrawal et al. (India, 2023)[2]	Segmentation of retinal structures for ROP diagnosis	AG U-Net	Segmentation	RetCam,Neo	600	Public	Accuracy = 86–99%
	Trivedi et al . (India, 2025)[111]	Diagnose ROP	CNN	Classification	3nethra Neo	1432	Public	Accuracy = 82–92%
	Zhao et al. (China, 2024)[132]	Develop ROP dataset for AI system	ResNet50, ResNet101, ConvNeXt, ViT-B	Classification	RetCam,SW-8000, Nautilus	1099 (training: validation: test = 8:1:1)	Public	AUC = 0.96–0.97, Accuracy = 79.7–87.6%, Sensitivity = 75.7–83.3%, Specificity = 95.1–97.0%

of an overall stage label, without qualification.[49,67,86] Some studies have begun segmenting the ridge, injecting richer information into AI staging models.[30,52,62,80,89] Jia and colleagues[55] employed the Sequence-encoding and Block-balancing Segmentation Network (SeBS-Net) to perform pixel-level ridge segmentation on 350 RetCam images, significantly outperforming the traditional U-Net. Adding the segmentation results as an extra input boosted overall staging accuracy, indicating that fine-grained ridge information can further improve AI staging robustness. Furthermore, Burt and coworkers[13] used a U-Net model to automatically segment and quantify anomalous neurovascular tissue volume (ANVTV) from UWF-OCT images, which includes the ridge structure. Their results indicated that ANVTV was significantly correlated with ROP stage and aligned closely with ROP progression. This study introduces the first volumetric biomarker for ROP and emphasizes the potential of OCT to complement or improve traditional staging methods based on fundus images. These studies demonstrate that AI can deliver consistent, scalable ROP screening across populations and devices when augmented with quantitative features.

Application in ROP plus disease. Plus disease is a key sign of ROP progression, and accurate diagnosis is critical for treatment decisions. Traditional subjective diagnosis based on retinal vessel dilation and tortuosity is inconsistent, while AI-based imaging technologies have provided new solutions.

Brown and coworkers[12] first developed a DL model using 5511 RetCam images to classify the severity of plus disease. The model achieved 91% accuracy on an independent set of 100 images, with an AUC of 0.98 for plus disease diagnosis, outperforming 6 out of 8 experts. This study demonstrated that the AI model learned a continuous representation of vascular severity but lacked explicit quantitative features, restricting clinical interpretability and utility. To address this limitation, Mao and coworkers[71] developed a 2-stage pipeline: a U-Net first segmented vessels and optic disc, after which a DenseNet classified images as normal, pre-plus, or plus. Using 5711 training and 450 test images, they reported 95.1% sensitivity and 97.8% specificity for classifying plus disease; however, the extra segmentation step doubled the computational overhead and required expert-quality ground-truth labels, limiting scalability in clinical practice. Within the same study, Mao's group[71] further quantified vascular morphology using segmentation outputs, demonstrating that derived indices declined dynamically after treatment, offering an interpretable basis for objective monitoring. Expanding on these segmentation-based approaches, Deng and coworkers[27] further transformed vessel-segmentation results into four continuous parameters, including vessel angle, density, fractal dimension, and tortuosity, to automatically distinguish plus disease. Collectively, plus disease segmentation not only improves diagnostic consistency, but also provides interpretable, quantitative evidence for dynamic treatment monitoring.

To evaluate real-world applicability, Bai and coworkers[7] conducted a multicenter validation of an independent algorithm, ROP.AI, using 8052 retinal images across multiple centers. For plus disease detection, it achieved an AUC of 0.75 and a sensitivity of 84%; however, its specificity was lower (43%) and performance declined compared to its development stage, indicating a need for further training to improve generalizability.

These studies demonstrate AI's promise in plus disease diagnosis through architectural innovations. Bai and coworkers' findings, however, underscore that balancing performance across diverse populations and real-world conditions remains a critical challenge for clinical adoption.[7]

2.2.2. Text-based AI technology in ROP diagnosis

Three studies further refined the diagnosis of text models by classifying ROP. Takeda and coworkers[103] applied multiple machine-learning algorithms on Japanese neonates to distinguish "Any-ROP" from "No ROP" using BW and GA, achieving AUC 0.93. Eckert and colleagues[31] developed the ROPScore, a linear-regression model applied to 474 Brazilian infants that predicted both any stage ROP (AUC 0.77) and severe ROP (AUC 0.88), providing a clear referral trigger. In addition, Ricard and coworkers[95] developed a screening tool for early postnatal prediction of ROP (STEP-ROP) logistic-regression score using on American neonates, classifying them as "no ROP", "mild ROP", or "severe ROP". Across these studies, sensitivity for detecting ROP remained high, yet specificity for discriminating severe disease was moderate. Eckert's model achieved 96% sensitivity, but only 56% specificity, for severe ROP, while Ricard's reached 95% sensitivity with 17% specificity, highlighting the remaining challenge of fine-grained classification when only clinical text data are available.

2.2.3. Smartphone video-based AI technology ROP diagnosis

Smartphone video data has emerged as a promising, low-cost alternative to specialized imaging devices for ROP screening, particularly in resource-limited settings. AI technologies leveraging smartphone-captured video frames have shown potential in enabling scalable, real-time ROP detection.

Young and coworkers[127] developed 2 smartphone-indirect ophthalmoscopy systems in India. They trained a ResNet18 model, which achieved 100% sensitivity and 58.6% specificity for TR-ROP, while expert-assigned vascular severity score (VSS) reached an AUC of 0.96, showing smartphones can spot severe ROP yet often over-refer. Building on this, Ortiz and colleagues[82] extended this idea in Mexico and Argentina. They filmed 524 videos. Frame-level sensitivity was 76.7% and patient-level sensitivity was 93.3%, both surpassing a 3-ophthalmologist consensus, but with lower specificity. Together, these studies demonstrate that smartphone video-based AI can deliver scalable ROP screening without costly wide-field cameras; refining specificity remains the common next step. Therefore, continued data collection and model development are still required in this area.

2.2.4. Multi-modal fusion in ROP diagnosis

To achieve more accurate ROP intelligent diagnosis, multi-modal fusion technology has become a promising approach. By combining multimodal retinal images, or clinical data, it provides a more comprehensive view than single-modal input, helping AI systems better approximate expert ophthalmologists, particularly in resource-limited settings.

Legocki and coworkers[64] developed multivariate models that combined handheld OCT-derived imaging biomarkers with demographic variables such as BW and GA for early referral-warranted ROP. Their results highlight that fusing posterior retinal features with clinical data can significantly outperform models using either modality alone, underscoring the translational potential of multi-modal integration in ROP screening.

Nevertheless, comprehensive ROP diagnostic models that integrate retinal images, clinical text, and genomic data remain scarce; large-scale, multicenter cohorts are urgently needed to explore and validate such cross-modal frameworks, thereby accelerating their clinical translation and scalable deployment in resource-limited settings worldwide.

2.3. AI technology for ROP prediction

AI-based ROP prediction studies have advanced along multiple approaches, with key distinctions in data inputs and model adaptability to diverse populations.

Among these approaches, one employs available neonatal characteristics such as BW and GA to enable early and practical screening. Chen and coworkers[21] developed a model showing that BW and GA alone achieve an AUC of 0.81 for ROP prediction, identifying 23% of high-risk infants while excluding 97.75% of low-risk cases. For predicting TR-ROP, adding gender, delivery mode, and pregnancies improved the AUC to 0.83. This model's reliance on non-imaging data makes it ideal for resource-limited settings, though its 23% high-risk detection rate raises concerns about potential missed cases. A larger

retrospective analysis further demonstrated that machine learning models based on demographic and clinical screening variables can effectively predict TR-ROP, showing high accuracy and sensitivity.[100] Building upon static clinical text, large scale machine learning approaches have further enhanced risk stratification using longitudinal clinical data. In a multicenter study, the machine learning models were developed at biweekly intervals to predict the need for ROP treatment, with a 100% sensitivity threshold, could potentially reduce unnecessary screening.[47] Furthermore, advancements have been achieved in forecasting visual outcomes for ROP patients following laser therapy or intravitreal injections. Huang and coworkers[50] employed a straightforward artificial neural network and an error backpropagation learning algorithm, to predict visual results. This study involved 60 infants with a history of ROP treatments, revealing that the algorithm provides a reasonably accurate prediction of future visual health for these patients.

To improve predictive performance, image-based systems have been developed that automatically analyze retinal features. For example, Wu and coworkers[122] designed OC-Net (for ROP prediction) and SE-Net (for severe ROP prediction), with internal validation AUCs of 0.90 and 0.87, respectively, rising to 0.94 and 0.88 in external validation, demonstrating robust generalizability. A large secondary analysis from the e-ROP Study further demonstrated that machine learning models integrating retinal image findings with demographic and clinical characteristics moderately predict TR-ROP, with deep neural networks outperforming other approaches, highlighting the benefit of multimodal data fusion.[43]

Focusing on Chinese cohorts, Lu and coworkers[70] validated models and found that Colorado ROP (CO-ROP) achieved 100% sensitivity for severe ROP (specificity 39.9%), outperforming Alexandria ROP (Alex--ROP) (56% sensitivity). They noted that higher GA and BW in eastern Chinese populations led to missed diagnoses in less-adapted models like Alex-ROP. Similarly, Sun and coworkers[102] compared ROPScore and Children's Hospital of Philadelphia (CHOP) ROP models. CHOP showed 100% sensitivity for type 1 ROP, but poor specificity (21.4%), while ROPScore had lower sensitivity (55.8%). These findings highlight the need for population-specific adaptations and locational fine tuning, as a single-time-point weight data may miss rapid progression, advocating for weight optimization using Chinese newborn growth curves.

Collectively, these studies demonstrate that AI-based ROP prediction combining clinical variables with imaging now reaches strong validation performance and can serve as a standardized screening while adapting to regional population characteristics.

2.4. AI technology for severity assessment and treatment suggestion

AI has demonstrated remarkable potential in zoning, staging, and prediction of ROP, expecting to become a comprehensive clinical decision-support system. Among these, severity assessment is particularly important for guiding clinical decision-making. Accordingly, researchers have developed AI models for evaluating ROP severity.

Coyner and coworkes[24] trained an AI model on 1579 fundus images, combining GA, BW, and an AI-derived VSS. The model achieved 100% sensitivity and specificity ranging from 48.9% to 80.8% for TR-ROP, enabling early identification approximately 1 month before clinical diagnosis and safely reducing unnecessary examinations in low-risk neonates. Wu and coworkers[122] utilized 7033 first-screening fundus images with 46 clinical variables to develop a dual-task DL system that predicts ROP, attaining an AUC of 0.87.

Yenice and coworkers[61] built a classification model on 634 images from 317 infants that distinguishes Type-2 from Type-1ROP with 98% accuracy and specificity, and Stage 2 from Stage 3 with 90% accuracy and 95% specificity, highlighting AI's potential to replace manual grading at critical decision points. Building upon these classification approaches, Wang and coworkers[116] introduced a more scalable framework, DeepROP, trained on fundus images. It employs 2 models: Id-Net, excelling in rapid screening with 96.6% sensitivity and 99.3% specificity, and Gr-Net achieved 88.5% sensitivity and 92.3% specificity, both reducing dependence on expert graders and enhancing its utility for telemedicine. Beyond accuracy and scalability, Yang and coworkers[125] advanced beyond "black-box" systems by developing an explainable AI on 6100 images. Their framework sequentially detects lesion type and location, stitches panoramic views for zone determination, and classifies plus disease, ultimately assigning overall severity according to international guidelines.[22] The system achieved an AUC of 0.95 and provides interactive visual feedback, significantly increasing clinician confidence.

Our team[68] further extended the scope by constructing a multi-task model on 24,495 images from 651 infants that simultaneously screens for ROP, identifies severe cases (AUC = 0.91), and recommends treatment modalities with an AUC of 0.94, outperforming 4 experienced ophthalmologists. In addition, our team also developed two AI systems for detecting regular ROP and A-ROP. Trained on fundus images, the system achieved classification accuracies exceeding 90% for regular ROP and A-ROP. These models combine high diagnostic accuracy with markedly reduced complexity, making them practical for real-world ROP severity classification.[124,131]

In addition to CFPs, AI is expanding to FA images for the assessment of ROP vascular activity. Lepore and coworkers[65] trained a CNN model on 835 FA images to evaluate the need for treatment. The model achieved an AUC of 0.91, with 88% accuracy, and demonstrated that FFA images contributed more power. FA-based AI can detect active-phase vascular changes that may be missed on standard photography, offering an adjunctive channel for risk stratification.

Collectively, AI now spans the entire ROP workflow for severity grading and treatment suggestion. Nevertheless, current studies remain limited by single-center cohorts, scarcity of advanced-stage cases, and moderate specificity. Future work should embrace multi-center, multi-ethnic prospective validation, integrate explainable architectures, and fuse multimodal data to enhance clinical generalizability and sustained adoption.

2.5. AI technology for vascular severity score quantification

VSS, which reflects the spectrum of plus disease, has been proven useful for ROP screening,[16] monitoring disease progression,[9] and evaluating treatment response.[32] Research on VSS quantification has advanced through successive stages, from foundational establishment to multi-scenario clinical validation.

Foundational research laid the groundwork for VSS as a reliable indicator of plus disease. Sharafi and coworkers[99] extracted 5 vascular features from 76 RetCam images to identify plus disease, achieving an accuracy of 0.86, outperforming one of three participating experts.

Clinical consensus validation strengthened VSS's credibility across expert evaluations. Campbell and coworkers[15] compared the continuous 1–9 VSS generated by the imaging and informatics in ROP (i-ROP) DL system with annotations from 34 ICROP3 experts using 30 images with plus disease and 28 images with ROP stages. The results revealed a high correlation between VSS and plus disease grading and significant differences across disease stages, establishing VSS as an objective benchmark for standardizing expert judgments; however, the reliance on data from a single North American device underscored the need for external validation.

Beyond plus disease identification, VSS has shown promise in treatment monitoring and risk prediction. In treatment monitoring, Eilts and coworkers[32] analyzed the Comparing Alternative Ranibizumab Dosages for Safety and Efficacy in ROP (CARE-ROP) trial, tracking VSS changes in 19 infants via 1046 follow-up images. After anti-vascular endothelial growth factor (anti-VEGF) therapy, VSS dropped sharply from 6.7 to 2.7 within 1 week, then rebounded to 6.5 when retreatment was initiated for ROP reactivation. Notably, higher baseline VSS correlated strongly with earlier recurrence, marking the first use of VSS for longitudinal tracking of treatment dynamics. In risk prediction, Coyner

and coworkers[24] integrated VSS with GA to develop a model for predicting TR-ROP. Using data from 1579 eyes, a single RetCam image at 32–33 weeks postmenstrual age enabled early prediction: the GA+VSS model achieved an area under the precision-recall curve (AUPR) of 0.35, maintaining 100% sensitivity and 48.9%-80.8% specificity across 2 external cohorts, and could identify TR-ROP approximately one month in advance. Furthermore, VSS has also demonstrated value as a support tool for clinical decision. A recent study showed that providing clinicians with an AI-generated VSS significantly improved interobserver agreement in plus disease diagnosis, suggesting that VSS may help reduce subjectivity and enhance diagnostic consistency in ROP screening.[25]

In summary, VSS has evolved from a foundational marker of plus disease to a versatile tool validated for expert consensus, treatment monitoring, and risk prediction. Future research should focus on validating VSS stability across multi-center settings and diverse imaging modalities to facilitate its integration into low-cost screening protocols.

2.6. AI technology for telemedicine and economic evaluation

The deployment of telemedicine in real-world settings further facilitates AI-based screening at the primary care level. In a retrospective analysis of 613 tele-screening sessions, Greenwald and coworkers[39] reported that the i-ROP DL model achieved an AUC of 0.99 in identifying ROP requiring referral. Cole and coworkers[23] further validated the algorithm in telemedicine projects in Nepal and Mongolia, showing that the AUC for detecting plus disease remained above 0.96 across images acquired from two different platforms. These studies suggest that, in settings with limited personnel or high referral costs, AI-based tele--screening can maintain clinical detection efficiency and provide reliable, quantifiable data for subsequent economic evaluations.

Economic evaluation is key to measuring the cost-effectiveness of AI screening technologies. In a 2024 study, Wu and coworkers[123] conducted a meta-modeling of 382 global economic studies on AI eye disease screening. The results showed that 73.3% of the studies believed AI had clear economic benefits, mainly reflected in reducing labor costs for repeated screenings, cutting referral transportation expenses, and shortening intensive care unit stay time. The remaining 26.7% of the results were affected by differences in screening protocols, local labor prices, and infrastructure. Specific to ROP, a model by Morrison and coworkers[74] analyzed data from 52,000 US preterm infants and confirmed that autonomous AI screening is more cost-effective than traditional remote reading when AI sensitivity reaches 90% and screening costs remain below $34 per patient. At 99% sensitivity, delayed treatment cases decreased dramatically from 265 to 40. Across all willingness-to-pay thresholds, autonomous AI consistently demonstrated superior economic value, highlighting the importance of maintaining low AI screening costs to maximize healthcare savings.

Together, these studies show AI-ROP screening cuts direct medical costs and long-term social costs from vision impairment in high-risk, labor-short, or high-transport-cost scenarios. In low-labor-cost or infrastructure-poor regions, local recalibration of cost thresholds is needed for viability.

2.7. Development of AI models

Beyond clinical deployment scenarios, methodological innovations such as multi-task learning, transfer learning, and federated learning have further driven the evolution of AI for ROP. In the field of intelligent diagnosis of ROP, AI models have advanced rapidly. Early efforts relied on simple binary classifiers that distinguished only between “ROP / no ROP”[4,90,130] or employed single-task “plus / no-plus”[35,57,114] networks. While these approaches required modest data, they lacked the granularity to capture the full complexity of ROP.

Subsequent work introduced multi-task and semi-supervised paradigms. Multi-task models simultaneously predict multiple dimensions of ROP such as ROP stage, plus disease severity, and progression risk,[48,49,65,92,109,116] offering a holistic view of the condition. Semi-supervised models, in contrast, leverage large volumes of fundus images alongside limited pixel-level annotations, dramatically reducing the need for exhaustive manual labeling. For instance, Demirci and colleagues proposed the adversarial vessel-unveiling semi-supervised segmentation framework that was trained on only 18 publicly available ROP images with full vascular labels plus 90 unlabeled in-house images, yet achieved precise retinal-vessel segmentation.

To further mitigate data scarcity, transfer learning has been widely adopted. Models pre-trained on large, general-purpose datasets such as ImageNet are fine-tuned on ROP-specific data, allowing the network to leverage knowledge acquired from millions of non-ROP images. Brown and coworkers[12] adopted this strategy for plus-disease detection: an Inception model pre-trained on ImageNet’s 1.2 million images was fine-tuned on 5511 ROP fundus images, yielding strong performance on the plus-disease classification task.

Recently, federated learning (FL) has emerged as an attractive solution for multi-center collaboration. Lu and coworkers[69] conducted the first comparison between FL and centralized training in ophthalmology, using 5255 images. Employing ResNet-18, the FL achieved an AUC of 0.93 for plus-disease classification, statistically equivalent to the centralized model, while 57% of single-institution models performed significantly worse, highlighting dataset size and label consistency impacts. Based on the same FL framework, Hanif and coworkers[42] trained a model on bedside clinical labels to derive a ROP VSS. Cross-center analysis revealed significant differences in VSS for “no-plus” diagnoses and an inverse relationship between institutional mean VSS and GA. Yun and coworkers[128] embedded an AI-blockchain workflow into a multi-center randomized controlled trial data-management framework and validated it in a real-world cataract study. The system supported 500 concurrent users with model response times under 3 s. A 2-round Delphi survey concluded that the framework can address issues such as “inefficient data collection” and “vulnerability of data integrity to attacks.” Beyond laying a practical foundation for privacy-preserving ophthalmic computing, this work also offers a replicable model for collaborative AI development in ROP and other rare diseases. Collectively, these studies illustrate the potential for federated learning to yield generalizable ROP diagnostic models comparable to centralized training while preserving patient privacy, and they establish frameworks that facilitate objective cross-institutional comparisons of diagnostic patterns and disease epidemiology.

Beyond advances in model architectures and collaborative training strategies, emerging trends such as large language models (LLMs) and explainable AI (XAI) are beginning to reshape the landscape of ophthalmic AI research and warrant careful evaluation in the context of ROP.

LLMs, exemplified by transformer architectures trained on large-scale multimodal corpora, offer new opportunities to generate human language.[101] At present, LLMs have shown substantial potential in generating educational materials for ROP patients and are emerging as novel tools for patient education. Recent studies have demonstrated that LLMs, such as ChatGPT and Gemini, can effectively improve the readability of medical texts by reducing the reading level from high school standards to recommended levels, thereby enhancing comprehension among patients and their caregivers; [29,88] however, clear limitations remain. LLMs may not fully align with specific region clinical guidelines, [88] and the generated materials often lack personalized and actionable guidance tailored to individual patients. [77] In addition, concerns remain regarding insufficient domain expertise,[134] data quality and security.[94]and ethical considerations.[81] Despite these challenges, LLMs hold promise in the ROP field, with potential applications in screening analysis, patient education, and medical training.[51,106] Future efforts should focus on the use of high quality data, multidisciplinary collaboration, and rigorous clinical validation to establish LLMs as reliable tools, ultimately improving ROP management and clinical outcomes.

XAI has gradually become an important trend in AI research, driven by clinical demands for transparency and trustworthy. [96] A study has introduced explainability techniques, such as DeepSHAP and class activation mapping (CAM), into ROP diagnostic models, demonstrating that the regions highlighted by the algorithms are consistent with vascular features identified by clinical experts.[115] More recent studies have extended XAI to more advanced model architectures, including vision transformer–based models[129] and multimodal vision–language frameworks that provide joint explanations for both imaging and clinical information.[78]

Current XAI applications, however, still lack unified and standardized evaluation of explanation reliability and clinical utility.[96] Recent researh has noted that although XAI has the potential to facilitate clinical adoption, more rigorous domain specific validation studies and clinically oriented evaluations are required.[45] Overall, XAI represents a promising yet still evolving methodological direction in ROP related AI.

2.8. Challenges and clinical implementation considerations

DL trained on fundus images or FFA images now achieves diagnostic accuracy for ROP that is similar to experts.[63,84] Nevertheless, a clear gap persists between technical maturity and clinical deployment. On the one hand, non-representative training data readily introduces algorithmic bias[53]: datasets are often confined to single centers and exhibit skewed distributions of ethnicity, GA, and ROP stages, while variable image quality from different imaging devices causes performance to drop sharply when models are transferred to new settings.[58,60] On the other hand, the rapid progression of ROP leads to non-stationary data distributions and model drift, and real-world factors such as cloudy ocular media and imaging artifacts further increase false-negative rates.[93,112]

In addition, neonatal datasets contain highly sensitive personal health information. They remain vulnerable to re-identification, and the vague scope of informed consent continues to complicate privacy protection and regulatory compliance.[60] At the regulatory and clinical levels, ROP-specific approval pathways are still absent, and liability has not been clearly defined in law, leaving hospitals hesitant to adopt these systems.[38] "Black-box" models offer clinicians little insight into how critical decisions are reached, undermining both ophthalmologist and patient trust, while outdated hospital information technology infrastructures struggle to integrate AI seamlessly into existing workflows.[19] More fundamentally, without global resource allocation and ethical oversight, AI systems risk prioritizing infants in well-resourced regions, thereby exacerbating screening inequalities.[120] Only by enhancing model robustness and explainability, and by establishing dynamic regulation, privacy safeguards, shared accountability, and cross-disciplinary governance can ROP-AI truly move from bench to bedside.

Complementing these technical and clinical efforts, countries are formulating medical AI regulations tailored to local contexts, covering device approval, data governance, and safety. Such frameworks aim to ensure that AI systems advance public health effectively and ethically[83]. In short, overcoming issues in privacy, generalizability, accountability, and real-world integration will be key to translating ROP-AI from bench to bedside in a sustainable manner. These challenges also highlight the critical role of high-quality, well-annotated datasets, which are discussed in the next section.

2.9. Public datasets for ROP intelligent diagnosis

Amid the rapid growth of AI, public datasets have become essential. A comparative summary of representative ROP datasets is provided in Table 3. Timkovič and colleagues[107] released a validated ROP dataset of 6004 retinal images from 188 neonates, now serving as a key resource for diverse ROP research tasks. The U.S.-based iROP dataset has similarly supported early ROP image classification research, though access is still limited to qualified ophthalmologists.[126] More recently, Farabi and Ferdowsi University of Mashhad's ROP (FARFUM-ROP) dataset from Iran offered 1533 ROP fundus images from 68 patients, each annotated by five experts and accompanied by BW, GA, and gender, further enriching the data resources in the ROP field.[3] Complementing these classification-oriented collections, HVDROPDB was developed as the first public fundus segmentation dataset dedicated to ROP.[2] The dataset provides expert-reviewed images together with pixel-level masks of optic disc, vessels, and demarcation line/ridge, spanning normal retina and ROP stages 1–3. By bridging heterogeneous imaging devices and diverse clinical presentations, HVDROPDB narrows the gap between experimental single-center AI models and real-world clinical deployment. Similarly, Trivedi and coworkers[111] introduced the Macretina

Table 3
Comparison of publicly available datasets for AI research in ROP.

Dataset	Size	Annotation	Strengths	Limitations	Unique features
FARFUM-ROP[3]	1533 images	Grade:normal,plus,pre-plus;Stage;Diagnostic: no treatment,revisit, treatment	Multi-expert annotations with consensus; Rich clinical data; Multi-dimensional labels	Small sample size; Single-center data; Potential class imbalance; No longitudinal follow-up	Combines disease grading and clinical decision labels, enabling both diagnostic classification and treatment prediction tasks
Ostrava Infant Retinal Dataset[107]	6004 images	Grade:normal,plus,pre-plus;Diagnosis	Large sample size; Multi-device data; Longitudinal follow-up; Supports classification	Relatively few plus/pre-plus cases;Image quality variability	Provides longitudinal data with patient-level series organization, enabling analysis of disease changes over time
iROP dataset[126]	5561 images, later expanded not fully public	Grade:normal,plus,pre-plus	Based reference standard;Used for DL models; Clinically relevant	Restricted access;Dataset details not fully disclosed	Establishes a reference standard framework for plus disease grading, widely adopted to reduce expert variability in ROP research
HVDROPDB[2]	600 images with annotations	Pixel-level annotations: optic disc, blood vessels, and ridge	Provides pixel-level ground truth for segmentation; Includes multiple retinal structures;Multi-device data	Limited number of annotated images; Incomplete stage coverage; Image quality variability	One of the few datasets offering multi structure pixel-level annotations for specific ROP features
Macretina[111]	973 images	Ridge/ demarcation line, optic disc,Blood vessels	Multi-task dataset; Including A-ROP; Suitable for segmentation and classification tasks	Small segmentation dataset; Limited advanced-stage ROP;Lack of external validation	Specifically incorporates A-ROP cases, enabling targeted research on severe and rapidly progressing disease
Our dataset[132]	1099 images	Normal, stage 1–3 ROP, laser scars	Relatively large dataset for neonatal ROP; Multi-device acquisition; Includes multiple follow-ups for some infants; High annotation consistency	Single-center dataset; Potential annotation variability	Includes post-treatment (laser scar) images, enabling evaluation of treatment outcomes and post therapy retinal changes

AI: artificial intelligence, ROP: ROP: retinopathy of prematurity, DL: deep learning, A-ROP: aggressive ROP

dataset, a comprehensively annotated fundus images collection supporting ridge line detection, optic disc localization, and blood vessel segmentation.

And our team, working in a similar research direction, has also constructed a dataset tailored to specific needs. We assembled a ROP fundus dataset consisting of 1099 images from 483 premature infants, covering normal cases, stage 1–3 ROP, and laser-treated ROP, to support the development and validation of intelligent ROP systems. This dataset aims to bridge the gap between experimental single-center AI models and their real-world clinical application.[132]

2.10. Current gaps and future perspectives in AI for ROP

Despite the rapid development of AI in ROP research, several critical gaps remain. First, most existing studies rely on single-center datasets with limited ethnic and demographic diversity. As a result, the generalizability of AI models in different clinical settings is limited.[53] Second, current studies still show insufficient integration of heterogeneous data sources, including imaging and clinical data. In addition, standardized frameworks for multimodal data fusion are not yet well established.[85] Finally, another major limitation is the lack of unified evaluation criteria. This issue makes it difficult to directly compare model performance across studies. It also reduces reproducibility and limits the translation of research findings.[17]

Future research should focus on constructing large-scale, multicenter, and multi-ethnic datasets to improve model robustness and generalizability. Advancing multimodal fusion strategies and establishing standardized evaluation foundations will be essential for improving comparability across studies. Moreover, real world implementation studies are needed to validate the safety and effectiveness of AI systems. The development of XAI models and FL may further facilitate clinical adoption. Collectively, addressing these gaps will be critical for translating AI-based ROP systems from experimental settings into practical, equitable clinical applications.

3. Overview of existing studies

In the field of intelligent diagnosis and treatment of ROP, research

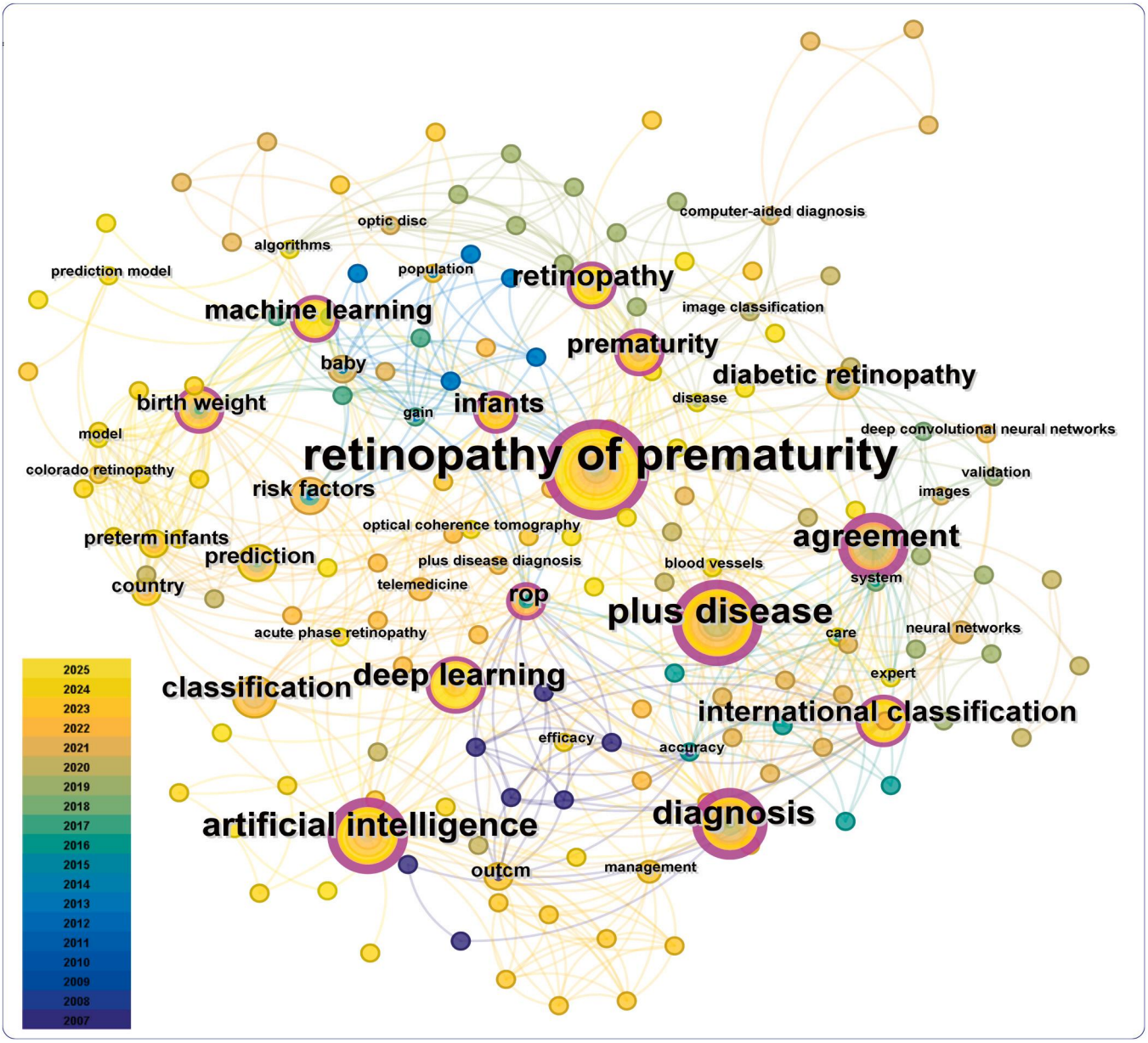


Fig. 2. Co-occurrence visualization of keywords.

hotspots generally refer to topics that generate a substantial body of studies within a given period and exhibit intrinsic interconnections. A recent bibliometric study has mapped the application of AI in ROP studies published between 2010 and 2023, providing an overview of publication patterns, major contributors, and emerge topics.[34] Building upon the prior effort, the present analysis extends the temporal scope to 2025 and integrates keyword centrality, clustering, and collaboration networks to further characterize recent research dynamics. Keywords in the literature can effectively capture the core content of an article, and analyzing their frequency and centrality helps to identify research hotspots in this domain. Using CiteSpace to analyze the keyword network (Figure 2), the most frequent keywords were ROP (30), plus disease (14), AI (12), diagnosis (11), DL (9) and agreement (9),. The highest centrality keywords included ROP (0.33), agreement (0.27), plus disease (0.22), AI (0.22), diagnosis (0.22), DL (0.20) and infants (0.20).

As shown in Figure 3, cluster analysis of the keywords indicates that research in the field of intelligent diagnosis and treatment of ROP can be divided into nine major directions, represented by preterm infants, fluorescein angiography, deep convolutional neural networks, deep learning, computer-aided diagnosis, validation, logistic regression, care and treatment modalities.

As shown in Figure 4, research in intelligent diagnosis and treatment of ROP has evolved through three stages: 2007–2018, 2019–2022, and 2023–2025. The first stage (2007–2018) was an exploratory phase with few keywords, mainly diagnosis, international classification, ROP, agreement, infants, and birth weight. The second stage (2019–2022) saw a sharp increase in keywords, focusing on plus disease, retinopathy, prematurity, AI, classification, deep learning, machine learning, country, neural networks, and OCT. In the third stage (2023–2025), the number of keywords declined, and hotspots shifted to management, random survival forest, and neonatal risk factor.

As shown in Figure 5, the cooperation network of countries and institutions in ROP intelligent diagnosis and treatment research includes 24 nodes. The United States had the largest number of publications, followed by China, with the U.S. showing closer collaboration with other countries.

The author collaboration network is depicted in Figure 6. The network comprises 215 nodes and 674 links, with a density of 0.0293, indicating relatively weak collaboration among scholars in this field. Nevertheless, several cohesive research teams can be identified. Team 1, consisting of Campbell, John Peter, Chiang, Michael F.; Kalpathy-Cramer, Jayashree; Ostmo, Susan; and others, forms a complex network with extensive collaborations. This team has focused on intelligent ROP diagnosis since 2007. Team 2, represented by Zhang, Guoming; Wu, Zhenquan; and Zhao, Xinyu; is characterized by a high publication output and close internal collaboration. In terms of academic influence, the most impactful author group includes Campbell, John Peter; Chiang. Michael F.; Ostmo, Susan; Kalpathy-Cramer. Jayashree; and Chan, R. V. Paul. The second most influential group consists of Zhang, Guoming; Wu, Zhenquan,;and Zhao, Xinyu.

4. Limitations

Current AI technologies have demonstrated clinical value in the tasks of ROP diagnosis, risk prediction, and treatment suggestion, achieving expert-level performance. Nevertheless, large-scale and standardized real-world deployment remains constrained. First, annotation consistency across training data has yet to be unified, and differences in resolution, illumination, and viewing field of imaging devices are not fully resolved. Moreover, clinical text models built primarily on European and North American populations have not been adequately validated worldwide. Second, most existing studies are single-center, and datasets are biased in GA, BW, and disease distribution limit model generalizability across institutions and ethnicities. Third, clinically validated interpretable frameworks and well-defined accountability are still lacking, leaving clinicians hesitant to adopt AI systems.

5. Conclusions

Current research on intelligent ROP diagnosis and treatment has focused on image quality assessment, zoning, staging, plus disease classification and scoring, severity prediction, and treatment suggestion,

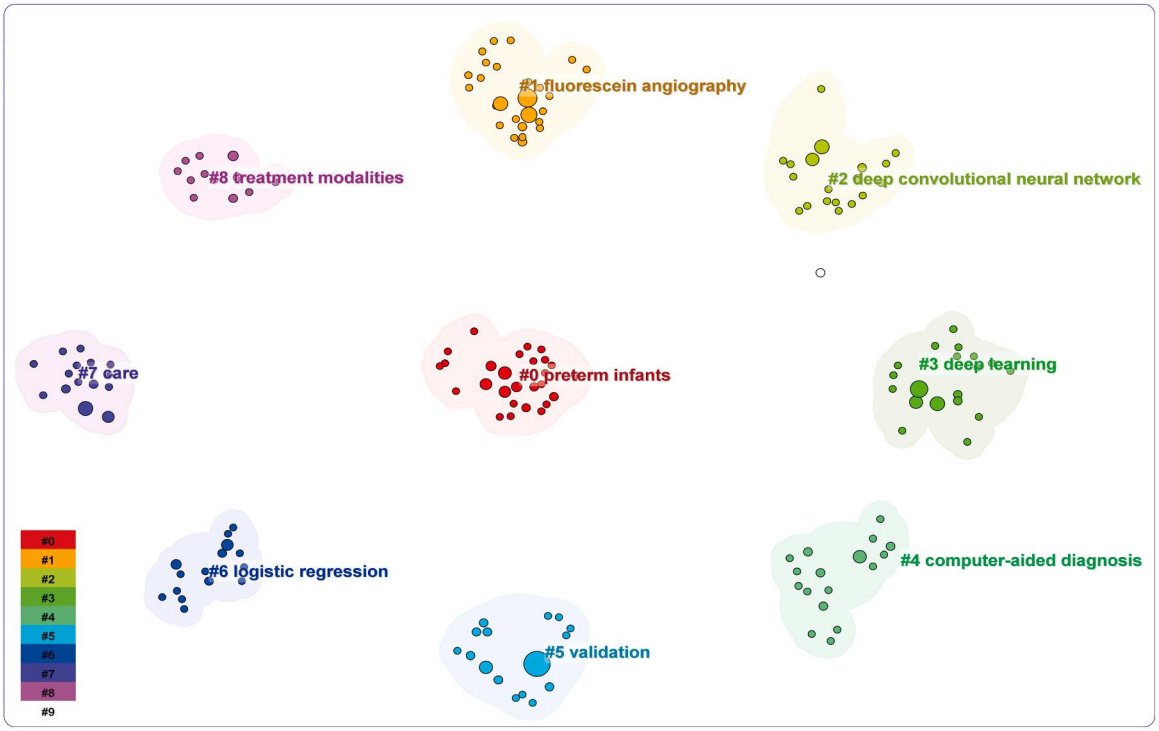


Fig. 3. Keyword cluster visualization.

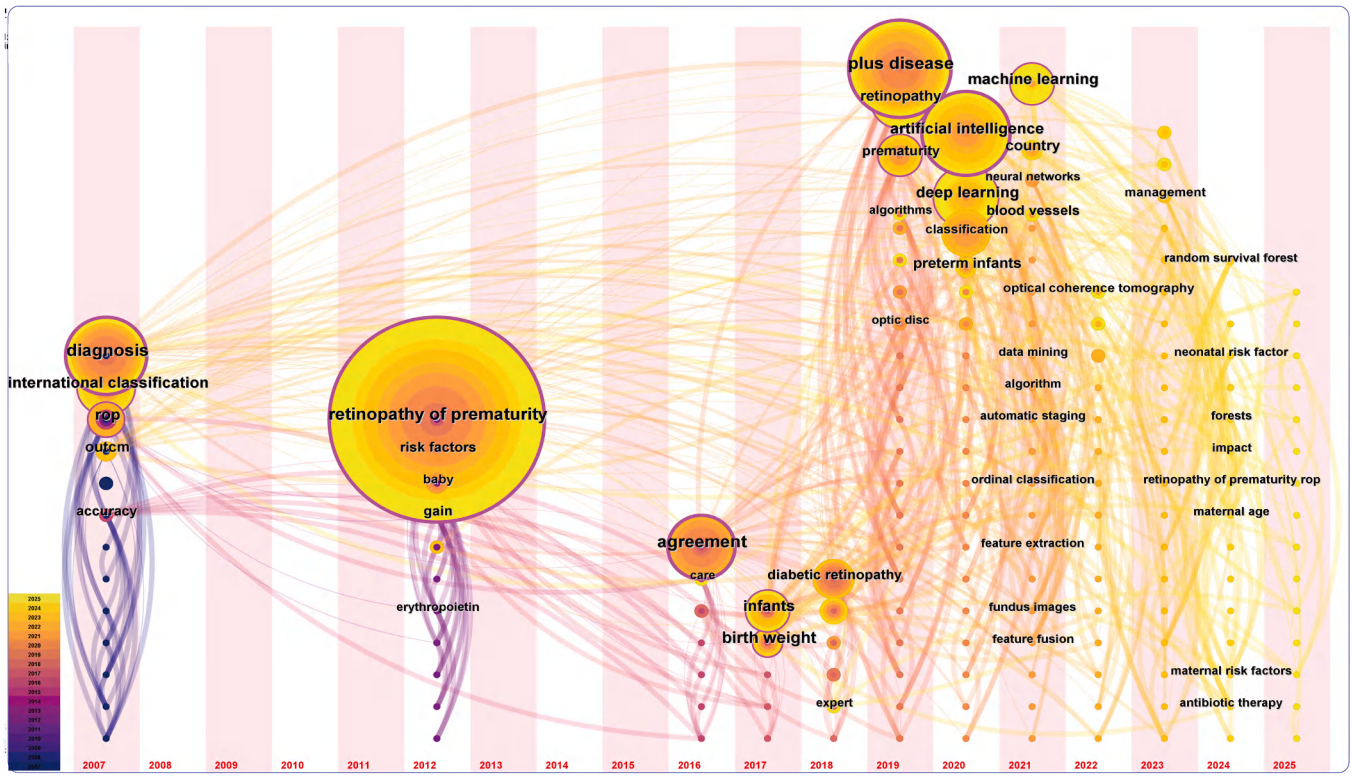


Fig. 4. Keyword time-zone map.

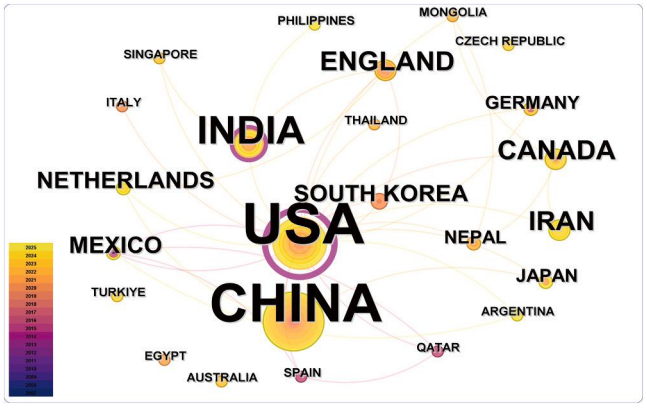


Fig. 5. Country co-occurrence map.

using multimodal data such as textual records, fundus photographs, and OCT images; however, publications and international collaboration have slightly declined in recent years, and few AI models have been applied clinically. Future work should prioritize multicenter, multiethnic prospective studies, expand health economic evaluations in telemedicine, and strengthen international cooperation, aiming to develop clinically applicable, accurate, affordable, and equitable intelligent ROP management solutions.

Methods

This study was designed as a comprehensive narrative review to summarize recent advances, emerging trends, and clinical applications of AI in ROP. A broad literature search was conducted across major electronic databases, including PubMed, Web of Science, Google Scholar, EMBASE, Scopus, and IEEE Xplore, covering publications from database inception to 31 December 2025. The search strategy combined disease-related terms ("retinopathy of prematurity," "prematurity retinopathy," and "retrolental fibroplasia") with AI-related terms ("artificial intelligence," "AI," "machine learning," "deep learning," "transfer learning," "neural network," "supervised learning," "unsupervised learning," "computer vision," "computational intelligence," "ensemble learning," "reinforcement learning," "large language model," and "LLM"). Reference lists of relevant articles were additionally screened to identify further eligible studies. The review was conducted in accordance with the SANRA checklist to ensure methodological rigor and transparency,[5] and the results assessed by two reviewers are presented in Figure S1.

Studies were considered for inclusion if they (1) investigated the application of AI in ROP, (2) involved multimodal data or clinically relevant imaging modalities, (3) were original research articles, and (4) provided sufficient methodological detail to allow meaningful interpretation. Editorials, letters, conference abstracts without full text, case reports, non–peer-reviewed publications, studies unrelated to ROP, and duplicate or overlapping datasets were excluded.

Retrieved articles were screened at the title and abstract level by multiple authors to assess relevance to the scope of this review, with discrepancies resolved through discussion. Given the objective of providing a comprehensive overview rather than a formal systematic synthesis, no quantitative meta-analysis or standardized quality scoring was performed. Instead, a qualitative synthesis was undertaken to summarize key methodological developments, clinical applications, limitations, and future research directions in AI-assisted ROP management.

Following the literature search, a set of key references was selected based on their foundational contributions, methodological relevance, or representative findings in the field of multimodal AI applied to ROP. These references are highlighted below and marked in bold in the reference list.

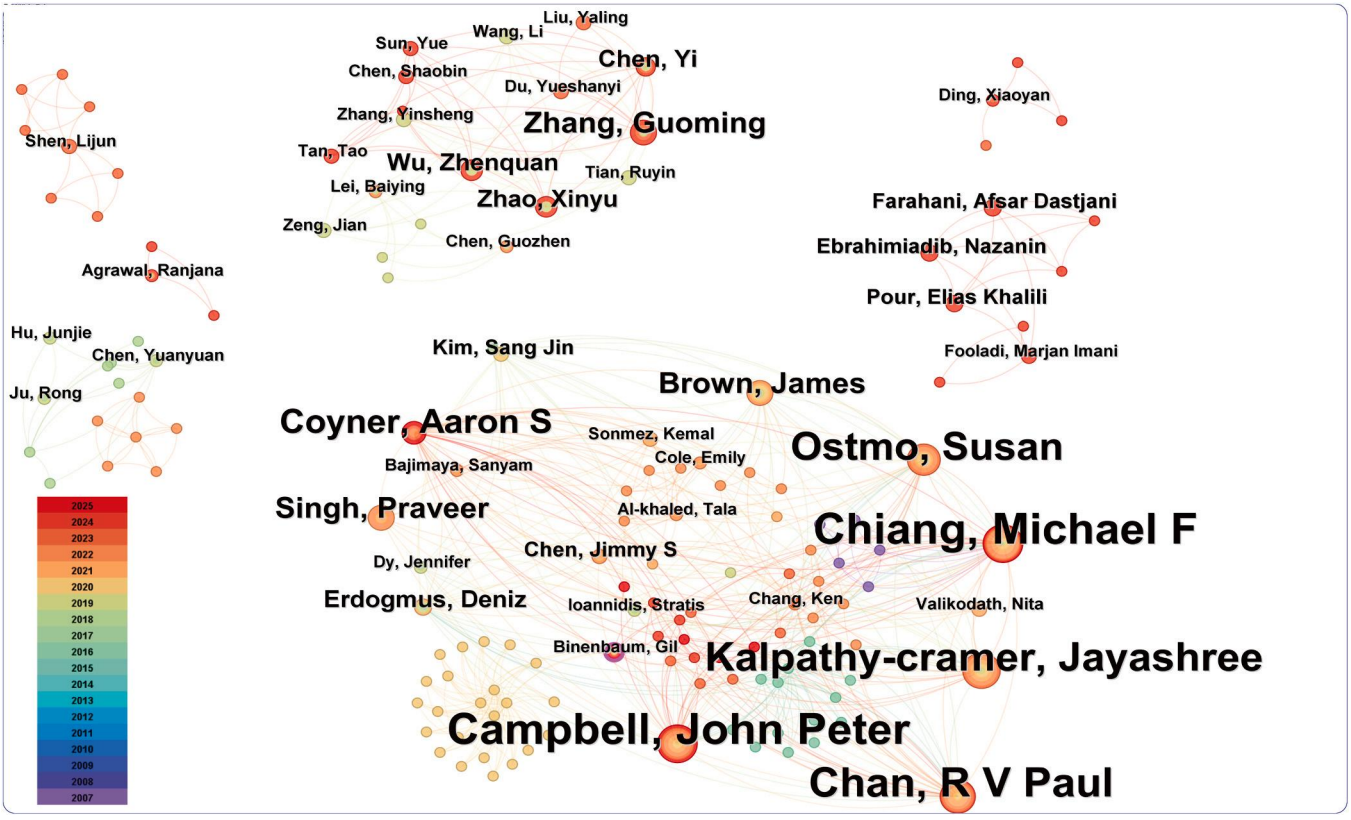


Fig. 6. Author co-occurrence map.

1. Zhao, X. et al. Sci Data 11, 543 (2024). This study provides a large, publicly available ROP fundus image dataset, essential for training and validating clinically applicable AI models.
2. Jia, X. et al. Med Image Anal 105, 103723 (2025). It introduces a novel segmentation network (SeBSNet) that significantly improves the precision of detecting subtle ROP lesions through sequence coding and block balancing.
3. Xie, H. et al. Med Image Anal 84, 102725 (2023). This work presents an adversarial knowledge distillation framework to create efficient, lightweight models for aggressive ROP detection, enhancing practical deployment.

Author contributions

Guoming Zhang, Wei Chi, Wenbin Wei, Xinyu Zhao, Zhenquan Wu, and Shirou Wu contributed to all aspects of the study, including conceptualization, study design, data analysis, manuscript drafting and revision. Kaixuan Cui, Wai Ching Lam, Wei-Chi Wu, Weihua Yang, Baiying Lei and Shuqiang Huang contributed to manuscript drafting and review. All authors reviewed and approved the final version of the manuscript.

CRediT authorship contribution statement

Wei Chi: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Wenbin Wei:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Zhenquan Wu:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Guoming Zhang:** Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Xinyu Zhao:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Weihua Yang:** Writing – review & editing, Writing – original draft. **Wei-Chi Wu:** Writing – review & editing, Writing – original draft. **Shuqiang Huang:** Writing – review & editing, Writing – original draft. **Baiying Lei:** Writing – review & editing, Writing – original draft. **Kaixuan Cui:** Writing – review & editing, Writing – original draft. **Shirou Wu:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Wai-Ching Lam:** Writing – review & editing, Writing – original draft.

Consent to participate

Not applicable.

Consent for publication

This article has been approved for publication by the authors.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, we used ChatGPT to improve language and readability. After using this tool, we reviewed and edited the content as needed and take full responsibility for the content of this publication.

Funding

This work was supported by National Natural Science Foundation of China (82271103, 82401315, 82301269, 82401272), Shenzhen Medical Research Fund (C2501034, C2301005, A2403020), Shenzhen Science and Technology Program (JCYJ20240813152703005, JCYJ20250604184009011), Sanming Project of Medicine in Shenzhen (SZSM202311018), Guangdong Basic and Applied Basic Research

Foundation (2026A1515012558).

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.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.survophthal.2026.05.006.

Data Availability

Further information and requests for resources should be directed to the corresponding author (zhangguoming@sz-eyes.com)

References

1. Agrawal R, Kulkarni S, Walambe R, Kotecha K. Assistive framework for automatic detection of all the zones in retinopathy of prematurity using deep learning. *J Digit Imaging*. 2021 Aug;34(4):932–947. https://doi.org/10.1007/s10278-021-00477-8.
2. Agrawal R, Walambe R, Kotecha K, et al. HVDROPDB datasets for research in retinopathy of prematurity. *Data Brief*. 2023 Nov 25;52, 109839. https://doi.org/10.1016/j.dib.2023.109839.
3. Akbari M, Pourreza HR, Khalili Pour E, et al. FARFUM-RoP, a dataset for computer-aided detection of retinopathy of prematurity. *Sci Data*. 2024 Oct 30;11:1176. https://doi.org/10.1038/s41597-024-03897-7.
4. Attallah O. DIAROP: automated deep learning-based diagnostic tool for retinopathy of prematurity. *Diagnostics*. 2021 Nov 3;11(11):2034. https://doi.org/10.3390/diagnostics11112034.
5. Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev*. 2019 Dec;4(1):5–11. https://doi.org/10.1186/s41073-019-0064-8.
6. Bai A, Carty C, Dai S. Performance of deep-learning artificial intelligence algorithms in detecting retinopathy of prematurity: a systematic review. *Saudi J Ophthalmol*. 2022;36(3):296–307. https://doi.org/10.4103/sjopt.sjopt_219_21.
7. Bai A, Dai S, Hung J, et al. Multicenter validation of deep learning algorithm ROP. AI for the automated diagnosis of plus disease in ROP. *Transl Vis Sci Technol*. 2023 Aug 14;12(8):13. https://doi.org/10.1167/tvst.12.8.13.
8. Bao Y, Ming WK, Mou ZW, et al. Current application of digital diagnosing systems for retinopathy of prematurity. *Comput Methods Programs Biomed*. 2021 Mar;200:105871. https://doi.org/10.1016/j.cmpb.2020.105871.
9. Bellsmith KN, Brown J, Kim SJ, et al. Aggressive posterior retinopathy of prematurity: clinical and quantitative imaging features in a large north american cohort. *Ophthalmology*. 2020 Aug;127(8):1105–1112. https://doi.org/10.1016/j.ophtha.2020.01.052.
10. Binenbaum G, Tomlinson LA, De Alba Campomanes AG, et al. Validation of the postnatal growth and retinopathy of prematurity screening criteria. *JAMA Ophthalmol*. 2020 Jan 1;138(1):31. https://doi.org/10.1001/jamaophthalmol.2019.4517.
11. Bowl W, Lorenz B, Stieger K, et al. Correlation of central visual function and ROP risk factors in prematures with and without acute ROP at the age of 6-13years: the giessen long-term ROP study. *Br J Ophthalmol*. 2016 Sept;100(9):1238–1244. https://doi.org/10.1136/bjophthalmol-2015-307855.
12. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018 July 1;136(7):803. https://doi.org/10.1001/jamaophthalmol.2018.1934.
13. Burt SS, Coyner AS, Roti EV, et al. Automated quantification of retinopathy of prematurity stage via ultrawidefield OCT. *Ophthalmol Sci*. 2024 Nov 28;5(2), 100663. https://doi.org/10.1016/j.xops.2024.100663.
14. Campbell JP, Ataer-Cansizoglu E, Bolon-Canedo V, et al. Expert diagnosis of plus disease in retinopathy of prematurity from computer-based image analysis. *JAMA Ophthalmol*. 2016 June 1;134(6):651–657. https://doi.org/10.1001/jamaophthalmol.2016.0611.
15. Campbell JP, Chiang MF, Chen JS, et al. Artificial intelligence for retinopathy of prematurity: validation of a vascular severity scale against international expert diagnosis. *Ophthalmology*. 2022 July;129(7):e69–e76. https://doi.org/10.1016/j.ophtha.2022.02.008.
16. Campbell JP, Singh P, Redd TK, et al. Applications of artificial intelligence for retinopathy of prematurity screening. *Pediatrics*. 2021 Mar;147(3), e2020016618. https://doi.org/10.1542/peds.2020-016618.
17. Carter RE, Attia ZI, Lopez-Jimenez F, Friedman PA. Pragmatic considerations for fostering reproducible research in artificial intelligence. *npj Digit Med*. 2019 May 22;2(1):42. https://doi.org/10.1038/s41746-019-0120-2.
18. Chan-Ling T, Gole GA, Quinn GE, Adamson SJ, Darlow BA. Pathophysiology, screening and treatment of ROP: a multi-disciplinary perspective. *Prog Retin Eye Res*. 2018 Jan;62:77–119. https://doi.org/10.1016/j.preteyeres.2017.09.002.
19. Chau M, Rahman MG, Debnath T. From black box to clarity: strategies for effective AI informed consent in healthcare. *Artif Intell Med*. 2025 Sept;167, 103169. https://doi.org/10.1016/j.artmed.2025.103169.
20. Chen JS, Coyner AS, Ostmo S, et al. Deep learning for the diagnosis of stage in retinopathy of prematurity: accuracy and generalizability across populations and cameras. *Ophthalmol Retin*. 2021 Oct;5(10):1027–1035. https://doi.org/10.1016/j.oret.2020.12.013.
21. Chen S, Zhao X, Wu Z, et al. Multi-risk factors joint prediction model for risk prediction of retinopathy of prematurity. *EPMA J*. 2024 June;15(2):261–274. https://doi.org/10.1007/s13167-024-00363-7.
22. Chiang MF, Quinn GE, Fielder AR, et al. International classification of retinopathy of prematurity, third edition. *Ophthalmology*. 2021 Oct;128(10):e51–e68. https://doi.org/10.1016/j.ophtha.2021.05.031.
23. Cole E, Valikodath NG, Al-Khaled T, et al. Evaluation of an Artificial Intelligence System for Retinopathy of Prematurity Screening in Nepal and Mongolia. *Ophthalmol Sci*. 2022 Apr 25;2(4), 100165. https://doi.org/10.1016/j.xops.2022.100165.
24. Coyner AS, Chen JS, Singh P, et al. Single-exam risk prediction of severe retinopathy of prematurity. *Pediatrics*. 2021 Dec 1;148(6), e2021051772. https://doi.org/10.1542/peds.2021-051772.
25. Coyner PhDAS, Young BK, Ostmo SR, et al. Use of an artificial intelligence-generated vascular severity score improved plus disease diagnosis in retinopathy of prematurity. *Ophthalmology*. 2024 Nov;131(11):1290–1296. https://doi.org/10.1016/j.ophtha.2024.06.006.
26. Demirci GM, Yao J, Ho MC, et al. Adversarial vessel-unveiling semi-supervised segmentation for retinopathy of prematurity diagnosis [Internet] [cited 2025 Aug 7] *arXiv*. 2024. https://doi.org/10.48550/arXiv.2411.09140.
27. Deng X, Chen K, Chen Y, et al. Vessels characteristics in familial exudative vitreoretinopathy and retinopathy of prematurity based on deep convolutional neural networks. *Front Pedia*. 2023 Aug 24;11, 1252875. https://doi.org/10.3389/fped.2023.1252875.
28. Derks LA, Tekin YS, Loudon SE, et al. External validation of an artificial intelligence–based model for retinopathy of prematurity screening using phoenix ICON retinal images. *J Am Assoc Pedia Ophthalmol Strabismus*. 2025 Dec;29(6), 104696. https://doi.org/10.1016/j.jaapos.2025.104696.
29. Dihan QA, Brown AD, Zaldivar AT, et al. Implementing generative AI to enhance patient education on retinopathy of prematurity. *J Pedia Ophthalmol Strabismus*. 2025 Nov;62(6):443–452. https://doi.org/10.3928/01913913-20250515-01.
30. Ding A, Chen Q, Cao Y, Liu B. Retinopathy of prematurity stage diagnosis using object segmentation and convolutional neural networks [cited 2025 Aug 9] *2020 Int Jt Conf Neural Netw (IJCNN) [Internet]*. 2020:1–6. https://doi.org/10.1109/IJCNN48605.2020.9207288.
31. Eckert GU, Fortes Filho JB, Maia M, Procianoy RS. A predictive score for retinopathy of prematurity in very low birth weight preterm infants. *Eye*. 2012 Mar;26(3):400–406. https://doi.org/10.1038/eye.2011.334.
32. Eilts SK, Pfeil JM, Poschkamp B, et al. Assessment of retinopathy of prematurity regression and reactivation using an artificial intelligence–based vascular severity score. *JAMA Netw Open*. 2023 Jan 19;6(1), e2251512. https://doi.org/10.1001/jamanetworkopen.2022.51512.
33. Fung THM, Kuet ML, Patel CK, et al. Retinal imaging in infants. *Surv Ophthalmol*. 2021 Nov;66(6):933–950. https://doi.org/10.1016/j.survophthal.2021.01.011.
34. Gao J, Fang N, Xu Y. Application of artificial intelligence in retinopathy of prematurity from 2010 to 2023: a bibliometric analysis. *Health Sci Rep*. 2025 Apr 18;8(4), e70718. https://doi.org/10.1002/hsr2.70718.
35. Gelman R, Jiang L, Du YE, et al. Plus disease in retinopathy of prematurity: pilot study of computer-based and expert diagnosis. *J AAPOS Publ Am Assoc Pedia Ophthalmol Strabismus / Am Assoc Pedia Ophthalmol Strabismus*. 2007 Dec;11(6):532–540. https://doi.org/10.1016/j.jaapos.2007.09.005.
36. Gilbert C, Fielder A, Gordillo L, et al. Characteristics of infants with severe retinopathy of prematurity in countries with low, moderate, and high levels of development: implications for screening programs. *Pediatrics*. 2005 May 1;115(5):e518–e525. https://doi.org/10.1542/peds.2004-1180.
37. Gonçalves MB, Nakayama LF, Ferraz D, et al. Image quality assessment of retinal fundus photographs for diabetic retinopathy in the machine learning era: a review. *Eye*. 2024 Feb;38(3):426–433. https://doi.org/10.1038/s41433-023-02717-3.
38. González-Gonzalo C, Thee EF, Klaver CCW, et al. Trustworthy AI: closing the gap between development and integration of AI systems in ophthalmic practice. *Prog Retin Eye Res*. 2022 Sept;90, 101034. https://doi.org/10.1016/j.preteyeres.2021.101034.
39. Greenwald MF, Danford ID, Shahrawat M, et al. Evaluation of artificial intelligence-based telemedicine screening for retinopathy of prematurity. *J AAPOS Publ Am Assoc Pedia Ophthalmol Strabismus*. 2020 June;24(3):160–162. https://doi.org/10.1016/j.jaapos.2020.01.014.
40. Gschließer A, Stifter E, Neumayer T, et al. Inter-expert and intra-expert agreement on the diagnosis and treatment of retinopathy of prematurity. *Am J Ophthalmol*. 2015 Sept;160(3):553–560.e3. https://doi.org/10.1016/j.ajo.2015.05.016.
41. Hanif AM, Jian Y, Young BK, Campbell JP. Implementation of OCT in retinopathy of prematurity screening. *Curr Opin Ophthalmol*. 2024 May 1;35(3):252–259. https://doi.org/10.1097/ICU.0000000000001030.
42. Hanif A, Lu C, Chang K, et al. Federated learning for multi-center collaboration in ophthalmology: implications for clinical diagnosis and disease epidemiology. *Ophthalmol Retin*. 2022 Aug;6(8):650–656. https://doi.org/10.1016/j.oret.2022.03.005.
43. He D, Luo X, Ying B, et al. Machine Learning Models for Predicting Treatment-Requiring Retinopathy of Prematurity in the e-ROP Study. *Transl Vis Sci Technol*. 2025 Aug 11;14(8):14. https://doi.org/10.1167/tvst.14.8.14.

44. Hellström A, Smith LEH, Dammann O. Retinopathy of prematurity. *Lancet (Lond Engl)*. 2013 Oct 26;382(9902):1445–1457. https://doi.org/10.1016/S0140-6736(13)60178-6.
45. Hettikankanamage N, Shafiabady N, Chatteur F, et al. eXplainable artificial intelligence (XAI): a systematic review for unveiling the black box models and their relevance to biomedical imaging and sensing. *Sens (Basel Switz)*. 2025 Oct 30;25(21):6649. https://doi.org/10.3390/s25216649.
46. Hoyek S, Cruz NFSD, Patel NA, et al. Identification of novel biomarkers for retinopathy of prematurity in preterm infants by use of innovative technologies and artificial intelligence. *Prog Retin Eye Res*. 2023 Nov;97:101208. https://doi.org/10.1016/j.preteyeres.2023.101208.
47. Hp F., Yj O., Ko Z. Machine learning risk prediction for treated retinopathy of prematurity in infants.:0–15.
48. Hu J, Chen Y, Zhong J, Ju R, Yi Z. Automated analysis for retinopathy of prematurity by deep neural networks. *IEEE Trans Med Imag*. 2019 Jan;38(1):269–279. https://doi.org/10.1109/TMI.2018.2863562.
49. Huang YP, Basanta H, Yu-Chuan Kang E, et al. Automated detection of early-stage ROP using a deep convolutional neural network. *Br J Ophthalmol*. 2021 Aug;105(8):1099–1103. https://doi.org/10.1136/bjophthalmol-2020-316526.
50. Huang CY, Kuo RJ, Li CH, et al. Prediction of visual outcomes by an artificial neural network following intravitreal injection and laser therapy for retinopathy of prematurity. *Br J Ophthalmol*. 2020 Sept;104(9):1277–1282. https://doi.org/10.1136/bjophthalmol-2019-314860.
51. Iqbal U, Tanweer A, Rahmanti AR, et al. Impact of large language model (ChatGPT) in healthcare: an umbrella review and evidence synthesis. *J Biomed Sci*. 2025 May 7;32:45. https://doi.org/10.1186/s12929-025-01131-z.
52. J K, Bin Khalid S, Vittal S, et al. Automated diagnosis of early-stage retinopathy of prematurity. *Eye*. 2026 Jan;40(1):134–138. https://doi.org/10.1038/s41433-025-04122-4.
53. Jafarizadeh A, Maleki SF, Pouya P, et al. Current and future roles of artificial intelligence in retinopathy of prematurity. *Artif Intell Rev*. 2025 Mar 29;58(6):188–243. https://doi.org/10.1007/s10462-025-11153-6.
54. Jayanna S, Agarwal K, Doshi V, et al. A retrospective analysis of ultra-widefield photograph (optos) documentation of retinopathy of prematurity at a tertiary eye care outpatient setup: the indian twin cities ROP study, report number 11. *J Am Assoc Pedia Ophthalmol Strabismus*. 2022 Apr;26(2):68.e1–68.e6. https://doi.org/10.1016/j.jaapos.2021.12.007.
55. Jia X, Qiu J, Nie D, Liu T. ROP lesion segmentation via sequence coding and block balancing. *Med Image Anal*. 2025 Oct;105, 103723. https://doi.org/10.1016/j.media.2025.103723.
56. Jin K, Gao Z, Jiang X, et al. MSHF: a multi-source heterogeneous fundus (MSHF) dataset for image quality assessment. *Sci Data*. 2023 May 17;10:286. https://doi.org/10.1038/s41597-023-02188-x.
57. Johnson KS, Mills MD, Karp KA, Grunwald JE. Semiautomated analysis of retinal vessel diameter in retinopathy of prematurity patients with and without plus disease. *Am J Ophthalmol*. 2007 Apr;143(4):723–725. https://doi.org/10.1016/j.ajo.2006.11.024.
58. Kazemzadeh K. Artificial intelligence in ophthalmology: opportunities, challenges, and ethical considerations. *Med Hypothesis Discov Innov Ophthalmol*. 2025 May 10;14(1):255–272. https://doi.org/10.51329/mehdiophthal1517.
59. Kearns H, Hull S. Perspectives on screening in retinopathy of prematurity: new algorithms and AI tools. *Clin Exp Optom.* 2025;26:1–13. https://doi.org/10.1080/08164622.2025.2589393.
60. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019 Oct 29;17:195. https://doi.org/10.1186/s12916-019-1426-2.
61. Kıran Yenice E, Kara C, Erdaş ÇB. Automated detection of type 1 ROP, type 2 ROP and a-ROP based on deep learning. *Eye (Lond Engl)*. 2024 Sept;38(13):2644–2648. https://doi.org/10.1038/s41433-024-03184-0.
62. Kurup GG, Kavya JS, Sivakumar R. Computer aided detection of demarcation line/ridges in the retinal fundus images of preterm infants [cited 2025 Aug 13] *2018 Int CET Conf Control Commun Comput (IC4) [Internet]*. 2018:171–175. https://doi.org/10.1109/CETIC4.2018.8530942.
63. Labib K.M., Ghumman H., Jain S., Jarstad J.S. A Review of the Utility and Limitations of Artificial Intelligence in Retinal Disorders and Pediatric Ophthalmology. Cureus. 16(10):e71063. doi:10.7759/cureus.71063.
64. Legocki AT, Lee AY, Ding L, et al. Multivariate models to diagnose early referral-warranted retinopathy of prematurity with handheld optical coherence tomography. *Transl Vis Sci Technol*. 2023 May 24;12(5):26. https://doi.org/10.1167/tvst.12.5.26.
65. Lepore D, Ji MH, Pagliara MM, et al. Convolutional Neural Network Based on Fluorescein Angiography Images for Retinopathy of Prematurity Management. *Transl Vis Sci Technol*. 2020 July 7;9(2):37. https://doi.org/10.1167/tvst.9.2.37.
66. Li L, Lin D, Lin Z, et al. DeepQuality improves infant retinopathy screening. *npj Digit Med*. 2023 Oct 16;6:192. https://doi.org/10.1038/s41746-023-00943-3.
67. Li P, Liu J. Early diagnosis and quantitative analysis of stages in retinopathy of prematurity based on deep convolutional neural networks, 17 *Transl Vis Sci Technol*. 2022 May;11(5):17. https://doi.org/10.1167/tvst.11.5.17.
68. Liu Y, Du Y, Wang X, et al. An artificial intelligence system for screening and recommending the treatment modalities for retinopathy of prematurity. *Asia-Pac J Ophthalmol*. 2023 Sept;12(5):468–476. https://doi.org/10.1097/APO.0000000000000638.
69. Lu C, Hanif A, Singh P, et al. Federated learning for multi-center collaboration in ophthalmology: improving classification performance in retinopathy of prematurity. *Ophthalmol Retin*. 2022 Aug;6(8):657–663. https://doi.org/10.1016/j.oret.2022.02.015.
70. Lu Y, Lv Z, Cen J, et al. Retrospective validation of G-ROP, CO-ROP, alex-ROP, and ROPscore predictive algorithms in two chinese medical centers. *Front Pedia*. 2023;11, 1079290. https://doi.org/10.3389/fped.2023.1079290.
71. Mao J, Luo Y, Liu L, et al. Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks. cited 2025 Aug 5]; *Acta Ophthalmol (Copenh) [Internet]*. 2020 May;98(3). https://doi.org/10.1111/aos.14264.
72. Mao J, Shao Y, Lao J, et al. Ultra-wide-field imaging and intravenous fundus fluorescein angiography in infants with retinopathy of prematurity. *Retin (Philos Pa)*. 2020 Dec;40(12):2357–2365. https://doi.org/10.1097/IAE.0000000000002761.
73. Marra KV, Chen JS, Nudleman E, Robbins SL. Review of retinopathy of prematurity management in the anti-VEGF era: evolving global paradigms, persistent challenges and our AI-assisted future. *Clin Experiment Ophthalmol*. 2025 Dec;53(9):1202–1217. https://doi.org/10.1111/ceo.14598.
74. Morrison SL, Dukhovny D, Chan RVP, Chiang MF, Campbell JP. Cost-effectiveness of artificial intelligence–based retinopathy of prematurity screening. *JAMA Ophthalmol*. 2022 Apr 1;140(4):401. https://doi.org/10.1001/jamaophthalmol.2022.0223.
75. Mulay S, Ram K, Sivaprakasam M, Vinekar A. Early detection of retinopathy of prematurity stage using deep learning approach. cited 2025 Aug 14]. *Med Imaging 2019 Comput-Aided Diagn [Internet]*. 2019:107. https://doi.org/10.1117/12.2512719
76. Nakayama LF, Mitchell WG, Ribeiro LZ, et al. Fairness and generalisability in deep learning of retinopathy of prematurity screening algorithms: a literature review. *BMJ Open Ophthalmol.* 2023;8(1), e001216. https://doi.org/10.1136/bmjophth-2022-001216.
77. Nastasi AJ, Courtright KR, Halpern SD, Weissman GE. Does ChatGPT provide appropriate and equitable medical advice?: a vignette-based, clinical evaluation across care contexts [Internet] [cited 2026 Jan 30]. *Health Inform*. 2023. https://doi.org/10.1101/2023.02.25.23286451
78. Neima H.Z. Trustworthy retinopathy of prematurity diagnosis using explainable vision-language model.:0–17.
79. Nguyen TTP, Ni S, Khan S, et al. Advantages of widefield optical coherence tomography in the diagnosis of retinopathy of prematurity. *Front Pedia*. 2021;9, 797684. https://doi.org/10.3389/fped.2021.797684.
80. Nisha MA, G. S KL, et al. An efficient ridge detection method for retinopathy of prematurity severity analysis [cited 2025 Aug 13] *2020 4th Int Conf Intell Comput Control Syst (ICICCS) [Internet]*. 2020:1210–1215. https://doi.org/10.1109/ICICCS48265.2020.9121101.
81. Ong JCL, Chang SYH, William W, et al. Ethical and regulatory challenges of large language models in medicine. *Lancet Digit Health*. 2024 June;6(6):e428–e432. https://doi.org/10.1016/S2589-7500(24)00061-X.
82. Ortiz A, Patiño S, Torres J, et al. AI-enabled screening for retinopathy of prematurity in low-resource settings. *JAMA Netw Open*. 2025 Apr 29;8(4), e257831. https://doi.org/10.1001/jamanetworkopen.2025.7831.
83. Palaniappan K, Lin EYT, Vogel S. Global regulatory frameworks for the use of artificial intelligence (AI) in the healthcare services sector. *Health (basel Switz)*. 2024 Feb 28;12(5):562. https://doi.org/10.3390/healthcare12050562.
84. Parmar UPS, Surico PL, Singh RB, et al. Artificial intelligence (AI) for early diagnosis of retinal diseases. *Med (Mex)*. 2024 Mar 23;60(4):527. https://doi.org/10.3390/medicina60040527.
85. Pawłowski M, Wróblewska A, Sysko-Romańczuk S. Effective techniques for multimodal data fusion: a comparative analysis. *Sens (Basel Switz)*. 2023 Feb 21;23(5):2381. https://doi.org/10.3390/s23052381.
86. Peng Y, Zhu W, Chen Z, et al. Automatic Staging for Retinopathy of Prematurity With Deep Feature Fusion and Ordinal Classification Strategy. *IEEE Trans Med Imag*. 2021 July;40(7):1750–1762. https://doi.org/10.1109/TMI.2021.3065753.
87. Poppe JA, Fitzgibbon SP, Taal HR, et al. Early prediction of severe retinopathy of prematurity requiring laser treatment using physiological data. *Pedia Res*. 2023;94(2):699–706. https://doi.org/10.1038/s41390-023-02504-6.
88. Postacı S.A., Dal A. The ability of large language models to generate patient information materials for retinopathy of prematurity: evaluation of readability, accuracy, and comprehensiveness. Turk J Ophthalmol. 54(6):330–336. doi:10.4274/tjo.galenos.2024.58295.
89. Raja Sankari VM, Umapathy S. Computer-aided diagnosis of early-stage retinopathy of prematurity in neonatal fundus images using artificial intelligence. *Biomed Phys Eng Express*. 2025 Jan 31;11(1):15015. https://doi.org/10.1088/2057-1976/ad91ba.
90. Ramachandran S, Niyas P, Vinekar A, John R. A deep learning framework for the detection of plus disease in retinal fundus images of preterm infants. *Biocybern Biomed Eng*. 2021 Apr;41(2):362–375. https://doi.org/10.1016/j.bbe.2021.02.005.
91. Ramanathan A, Athikarisamy SE, Lam GC. Artificial intelligence for the diagnosis of retinopathy of prematurity: a systematic review of current algorithms. *Eye*. 2023 Aug;37(12):2518–2526. https://doi.org/10.1038/s41433-022-02366-y.
92. Redd TK, Campbell JP, Brown JM, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. bjophthalmol-2018-313156 *Br J Ophthalmol*. 2018 Nov 23. https://doi.org/10.1136/bjophthalmol-2018-313156.
93. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. *J Am Med Inf Assoc JAMIA*. 2019 Nov 4;27(3):491–497. https://doi.org/10.1093/jamia/ocz192.
94. Ren Y, Luo X, Wang Y, et al. Large language models in traditional chinese medicine: a scoping review. *J Evid-Based Med*. 2025 Mar;18(1), e12658. https://doi.org/10.1111/jebm.12658.

95. Ricard CA, Dammann CEL, Dammann O. Screening tool for early postnatal prediction of retinopathy of prematurity in preterm newborns (STEP-ROP). *Neonatology*. 2017;112(2):130–136. https://doi.org/10.1159/000464459.
96. Saarela M, Podgorelec V. Recent applications of explainable AI (XAI): a systematic literature review. *Appl Sci*. 2024 Oct 2;14(19):8884. https://doi.org/10.3390/app14198884.
97. Sabri K, Ells AL, Lee EY, Dutta S, Vinekar A. Retinopathy of Prematurity: A Global Perspective and Recent Developments. *Pediatrics*. 2022 Sept 1;150(3), e2021053924. https://doi.org/10.1542/peds.2021-053924.
98. Shah S, Slaney E, VerHage E, et al. Application of artificial intelligence in the early detection of retinopathy of prematurity: review of the literature. *Neonatology*. 2023;120(5):558–565. https://doi.org/10.1159/000531441.
99. Sharafi SM, Ebrahimiadib N, Roohipourmoallai R, et al. Automated diagnosis of plus disease in retinopathy of prematurity using quantification of vessels characteristics. *Sci Rep*. 2024 Mar 16;14:6375. https://doi.org/10.1038/s41598-024-57072-4.
100. Shoeibi N, Abrishami M, Hosseini SM, et al. Development and validation of machine learning classifiers for predicting treatment-needed retinopathy of prematurity. *BMC Med Inf Decis Mak*. 2025 July;25(1):221. https://doi.org/10.1186/s12911-025-03057-w.
101. Shool S, Adimi S, Saboori Amleshi R, et al. A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC Med Inf Decis Mak*. 2025 Mar 7; 25:117. https://doi.org/10.1186/s12911-025-02954-4.
102. Sun H, Dong Y, Liu Y, et al. Using ROPScore and CHOP ROP for early prediction of retinopathy of prematurity in a chinese population. *Ital J Pedia*. 2021 Feb 18;47:39. https://doi.org/10.1186/s13052-021-00991-z.
103. Takeda Y, Kaneko Y, Sugimoto M, et al. Prediction models for retinopathy of prematurity using nonimaging machine learning approaches: a regional multicenter study. *Ophthalmol Sci*. 2025 July 1;5(4), 100715. https://doi.org/10.1016/j.xops.2025.100715.
104. Tan Z, Isaacs M, Zhu Z, et al. Retinopathy of prematurity screening: a narrative review of current programs, teleophthalmology, and diagnostic support systems. *Saudi J Ophthalmol*. 2022;14(36):283–295. https://doi.org/10.4103/sjopt.sjopt_220_21.
105. Tapak L, Farahani LN, Taleghani NT, et al. Risk factors for the time to development of retinopathy of prematurity in premature infants in Iran: a machine learning approach. *BMC Ophthalmol*. 2024 Aug 23;24:364. https://doi.org/10.1186/s12886-024-03637-w.
106. Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine. *Nat Med*. 2023 Aug;29(8):1930–1940. https://doi.org/10.1038/s41591-023-02448-8.
107. Timkovič J, Nowaková J, Kubíček J, et al. Retinal image dataset of infants and retinopathy of prematurity. *Sci Data*. 2024 July 23;11:814. https://doi.org/10.1038/s41597-024-03409-7.
108. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019 Feb;103(2):167–175. https://doi.org/10.1136/bjophthalmol-2018-313173.
109. Tong Y, Lu W, Deng Q qin, Chen C, Shen Y. Automated identification of retinopathy of prematurity by image-based deep learning. *Eye Vis (Lond)*. 2020 Aug;7(1):40. https://doi.org/10.1186/s40662-020-00206-2.
110. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019 Jan;25(1):44–56. https://doi.org/10.1038/s41591-018-0300-7.
111. Trivedi U, Srivastava A, Mahajan P. Macretina: a dataset, to support deep learning assisted retinopathy of prematurity diagnosis. *Sci Rep*. 2025 Dec 24;16:1993. https://doi.org/10.1038/s41598-025-31624-8.
112. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med*. 2018 Nov 6;15(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689.
113. Vermeer KA, Mo J, Weda JJA, Lemij HG, de Boer JF. Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography. *Biomed Opt Express*. 2013 Dec 23;5(1):322–337. https://doi.org/10.1364/BOE.5.000322.
114. Wallace DK, Zhao Z, Freedman SF. A pilot study using "ROPtool" to quantify plus disease in retinopathy of prematurity. *J AAPOS Publ Am Assoc Pedia Ophthalmol Strabismus*. 2007 Aug;11(4):381–387. https://doi.org/10.1016/j.jaapos.2007.04.008.
115. Wang J, Ji J, Zhang M, et al. Automated Explainable Multidimensional Deep Learning Platform of Retinal Images for Retinopathy of Prematurity Screening. *JAMA Netw Open*. 2021 May 5;4(5), e218758. https://doi.org/10.1001/jamanetworkopen.2021.8758.
116. Wang J, Ju R, Chen Y, et al. Automated retinopathy of prematurity screening using deep neural networks. *Ebiomedicine*. 2018 Aug 27;35:361–368. https://doi.org/10.1016/j.ebiom.2018.08.033.
117. Wang H, Li L, Wang W, et al. Development and validation of a deep learning image quality feedback system for infant fundus photography. *Sci Rep*. 2025 July 23;15, 26852. https://doi.org/10.1038/s41598-025-10859-5.
118. Wang S, Liu J, Zhang X, et al. Global, regional and national burden of retinopathy of prematurity among childhood and adolescent: a spatiotemporal analysis based on the global burden of disease study 2019. *BMJ Paediatr Open*. 2024 Jan 6;8(1), e002267. https://doi.org/10.1136/bmjpo-2023-002267.
119. Wang S, Shen W, Gao Z, et al. Enhancing the ophthalmic AI assessment with a fundus image quality classifier using local and global attention mechanisms. *Front Med*. 2024 Aug 7;11, 1418048. https://doi.org/10.3389/fmed.2024.1418048.
120. Whittlestone JNRune, Alexandrova Anna. *Ethical and societal implications of data and artificial intelligence: a roadmap for research*. London: Nuffield Foundation; 2019.
121. Witmer MT, Kiss S. Wide-field imaging of the retina. *Surv Ophthalmol*. 2013;58(2): 143–154. https://doi.org/10.1016/j.survophthal.2012.07.003.
122. Wu Q, Hu Y, Mo Z, et al. Development and validation of a deep learning model to predict the occurrence and severity of retinopathy of prematurity. *JAMA Netw Open*. 2022 June 16;5(6), e2217447. https://doi.org/10.1001/jamanetworkopen.2022.17447.
123. Wu H, Jin K, Yip CC, Koh V, Ye J. A systematic review of economic evaluation of artificial intelligence-based screening for eye diseases: From possibility to reality. *Surv Ophthalmol*. 2024;69(4):499–507. https://doi.org/10.1016/j.survophthal.2024.03.008.
124. Xie H, Liu Y, Lei H, et al. Adversarial learning-based multi-level dense-transmission knowledge distillation for AP-ROP detection. *Med Image Anal*. 2023 Feb;84, 102725. https://doi.org/10.1016/j.media.2022.102725.
125. Yang W, Zhou H, Zhang Y, et al. An interpretable system for screening the severity level of retinopathy in premature infants using deep learning. *Bioeng (basel Switz)*. 2024 Aug 5;11(8):792. https://doi.org/10.3390/bioengineering11080792.
126. Yildiz VM, Tian P, Yildiz I, et al. Plus disease in retinopathy of prematurity: convolutional neural network performance using a combined neural network and feature extraction approach. *Transl Vis Sci Technol*. 2020 Feb 14;9(2):10. https://doi.org/10.1167/tvst.9.2.10.
127. Young BK, Cole ED, Shah PK, et al. Efficacy of smartphone-based telescreening for retinopathy of prematurity with and without artificial intelligence in India. *JAMA Ophthalmol*. 2023 June 1;141(6):582. https://doi.org/10.1001/jamaophthalmol.2023.1466.
128. Yun D, Wu X, Chen X, et al. An artificial intelligence and blockchain technology-based data management framework for multicenter randomized controlled trials. *Sci Bull*. 2025 Mar;70(6):856–860. https://doi.org/10.1016/j.scib.2025.01.041.
129. Yurdakul M, Uyar K, Taşdemír Ş, Atabaş Í. ROPGCViT: a novel explainable vision transformer for retinopathy of prematurity diagnosis. *IEEE Access*. 2025;13: 77064–77079. https://doi.org/10.1109/ACCESS.2025.3564213.
130. Zhang Y, Wang L, Wu Z, et al. Development of an automated screening system for retinopathy of prematurity using a deep neural network for wide-angle retinal images. *IEEE Access*. 2019;7:10232–10241. https://doi.org/10.1109/ACCESS.2018.2881042.
131. Zhang R, Zhao J, Xie H, et al. Automatic diagnosis for aggressive posterior retinopathy of prematurity via deep attentive convolutional neural network. *Expert Syst Appl*. 2022 Jan;187, 115843. https://doi.org/10.1016/j.eswa.2021.115843.
132. Zhao X, Chen S, Zhang S, et al. A fundus image dataset for intelligent retinopathy of prematurity system. *Sci Data*. 2024 May 27;11:543. https://doi.org/10.1038/s41597-024-03362-5.
133. Zhao J, Lei B, Wu Z, et al. A deep learning framework for identifying zone I in RetCam images. *IEEE Access*. 2019;7:103530–103537. https://doi.org/10.1109/access.2019.2930120.
134. Zhao W., Yang X., Lyu Z., Xu C., Guan Z. Road of large language model: source, challenge, and future perspectives. Research. 8:655. doi:10.34133/research.0655.