

Review

A Systematic Review of Diagnostic Accuracy and Clinical Validation Studies Using Artificial Intelligence for Detection of Non-Diabetic Retinal Diseases

Abdulaziz Salman Almadani^{1*}, Mohammed Naji Almutairi², Elyas Ali Kirat¹

¹ Department of Medicine, College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

² Emergency Department, Diriyah Hospital, Riyadh, Saudi Arabia

Correspondence should be addressed **Abdulaziz Salman Almadani**, College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia. Email: Abdulaziz200035@gmail.com

Copyright © 2026 **Abdulaziz Salman Almadani**. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 12 April 2026, Accepted: 27 May 2026, Published: 31 May 2026.

Abstract

Background: Retinal diseases are often associated with diabetes; however, non-diabetic retinal diseases can be associated with various ocular conditions. Retinal diseases are diagnosed using imaging, such as fundus photography and optical coherence tomography (OCT). The implementation of artificial intelligence (AI) in ophthalmology improves diagnostic accuracy, enabling early diagnosis, optimizing workflows, and enhancing the overall quality of patient care. However, AI-based detection of retinal diseases and challenges with validation across different patient populations remain. This systematic review aims to examine existing research on the diagnostic accuracy and clinical validation of AI-based detection of non-diabetic retinal diseases.

Methods: A systematic search of studies published in PubMed, Cochrane Library, and Science Direct was completed from inception through February 23, 2026, without geographic restriction. Major outcomes of interest included diagnostic accuracy and clinical validation outcomes of AI models and retinal imaging modalities outcomes. The target population was human participants of any age or sex who are diagnosed with non-diabetic retinal diseases. The QUADAS-2 assessment tool was used to evaluate the methodological quality and risk of bias in the included studies.

Results: 15 studies were included for the systematic review. The included studies showed that the used AI models had high sensitivity, specificity, and diagnostic accuracy for several non-diabetic retinal conditions, with the highest performance being associated with age-related macular degeneration (AMD) and the lowest sensitivity being associated with glaucoma. Additionally, AI models improved efficiency, reduced examination time, and working load. However, the majority of the included models lacked external validation and provided low sensitivity for rare condition detection.

Conclusion: AI-based models demonstrate high diagnostic accuracy and specificity for detecting non-diabetic retinal diseases and could serve as effective tools for screening and triage, particularly in resource limited areas. While promising, prospective studies and careful implementation strategies are essential to ensure efficiency, safety, and improved patient outcomes.

Keywords: *artificial intelligence, non-diabetic retinal disease, OCT, fundus examination, deep learning, external validation*

Introduction

Retinal diseases are any disorders that damage the retina or the supporting structures, resulting in vision impairment. They represent a persistent burden on individuals and healthcare systems, contributing to significant global blindness and vision impairment (1). Although retinal diseases are common among diabetics, they occur in 6 % to 13.6% of non-diabetic individuals (2). Non-diabetic retinal diseases can be associated with various ocular conditions, including retinal vein occlusions, retinal telangiectasia, and retinal macro-aneurysms. Other common systemic causes include hypertension, atherosclerosis, blood dyscrasias, systemic infections, and past radiotherapy (3). Non-diabetic retinopathy is considered an early indicator of hypertensive damage and other cardiovascular risk factors, such as increased internal carotid intima media thickness (4). Retinopathy is characterized by the presence of microaneurysms, retinal hemorrhages, cotton wool spots, hard exudates, intraretinal microvascular abnormalities, venous beading, and new vessels (3, 4). Moreover, non-diabetic retinopathy often progresses into diabetes (5).

Current approaches for the diagnosis of retinal diseases involve multimodal imaging along with retinal examination (6). Key retinal imaging modalities include fundus photography, optical coherence tomography (OCT), fluorescein angiography, and fundus autofluorescence (7). Fundus photography and OCT are the most commonly used imaging techniques in ophthalmology. Fundus photography is considered the primary fundus test for the diagnosis of several retinal diseases, as it can identify lesions in the retina, while OCT utilizes low-coherence light for scanning biological tissues in cross-section and converts the obtained information into numbers, providing quantitative diagnostic indicators (8). Despite advancements in retinal imaging techniques and diagnostic tools, the manual identification of retinal diseases remains time-consuming, labor-intensive, and dependent on the ophthalmologist's experience (7). Moreover, the

diagnosis of subtypes of retinal diseases requires a higher level of expertise than simple screening for abnormal cases, due to the significant phenotypic overlap in several retinal diseases (9, 10).

Recently, artificial intelligence (AI) has been widely used for the identification and grading of several diseases using medical image analysis, specifically retinal images (6). The implementation of AI in ophthalmology significantly improves diagnostic accuracy, enabling early diagnosis, optimizing workflows, and enhancing the overall quality of patient care (11). Most modern AI-based models are built upon the foundation of machine learning (ML), which enables algorithms to learn patterns from data, analyze information, and make predictions without relying on predefined, explicit rules. Deep Learning (DL) is a subfield of ML that has been remarkably efficient in the analysis of image and sequential data (11). Despite the progress made in evaluating the diagnostic performance and clinical applicability of AI-based detection of retinal diseases, challenges with validation across different patient populations remain, as these models are highly dependent on patients' demographic and clinical characteristics and require extensive training (9, 12).

This systematic review aims to critically examine and synthesize existing research on the diagnostic accuracy and clinical validation of AI-based detection of non-diabetic retinal diseases, and to assess the methodological quality and risk of bias of current studies. A comprehensive evaluation of the challenges of AI systems for the identification of non-diabetic retinal diseases is crucial for enhancing patient care, prompting future research and public health initiatives towards sustainable healthcare systems.

Methodology**Study Design**

This systematic review study followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, specifically the updated PRISMA 2020 (13).

Definition of Outcomes and Inclusion Criteria

Major outcomes of interest included AI-based models for the diagnosis of non-diabetic retinal diseases (diagnostic accuracy or clinical validation outcomes, such as sensitivity, specificity, area under the curve (AUC), and related performance measures) and retinal imaging modalities outcomes, including fundus photography, OCT, and angiography. The search strategy of this systematic review was developed based on the Population, Intervention, Comparison, and Outcomes (PICO) framework and the study designs. The target population was human participants of any age or sex who are diagnosed with non-diabetic retinal diseases. Eligible interventions comprised the utilization of AI-based models for the detection or diagnosis of non-diabetic retinal diseases. Comparator groups included different AI-based models, ML models that diagnose non-diabetic retinal diseases, and conventional methods of diagnosis of non-diabetic retinal diseases.

Studies related to diabetes mellitus or studies that focused solely on diabetic retinopathy or other diabetes-related retinal changes, in addition to studies that do not provide sufficient data to assess diagnostic accuracy or clinical validation, were excluded. There were no restrictions on the study designs or geographical areas. Review papers, case reports, editorials, commentaries, conference abstracts without complete text, and studies involving non-human subjects, simulations, or purely technical algorithm development without clinical evaluation were eliminated. Only studies published in the English language were included.

Search Strategy

A comprehensive literature search was conducted among multiple electronic databases, including PubMed, Cochrane Library, and Science Direct, from inception to the present. The database search was conducted on February 23, 2026. Electronic searches were conducted using the following Boolean string keyword search strategy:

PubMed: (("artificial intelligence"[Mesh] OR "machine learning" OR "deep learning" OR "neural network*" OR "computer-aided diagnosis") AND ("retinal disease*" OR "macular degeneration" OR glaucoma OR "retinal vein occlusion" OR "retinal detachment" OR "retinitis pigmentosa" OR "optic neuropathy") AND ("diagnostic accuracy" OR sensitivity OR specificity OR validation OR "clinical validation" OR screening OR detection OR diagnosis OR ROC OR AUC) AND ("fundus photograph*" OR "retinal imaging" OR OCT OR "optical coherence tomography"))).

Cochrane Library: ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("retinal disease*" OR "macular degeneration" OR glaucoma OR "retinal disorder*") AND ("diagnostic accuracy" OR detection OR screening OR validation)

Science Direct: ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("retinal disease" OR "macular degeneration" OR glaucoma) AND ("diagnostic accuracy" OR "clinical validation")

Reference lists of included studies and relevant reviews were manually screened to identify additional eligible studies.

Screening and Extraction

All records identified through the database search were imported into reference management software (EndNote X8), and duplicates were removed. Two reviewers independently screened titles and abstracts for eligibility. Full-text articles were then retrieved and assessed independently by the same reviewers against the predefined inclusion and exclusion criteria. Disagreements at any stage of the screening process were resolved through discussion and, when necessary, consultation with a third reviewer. The study selection process was documented using a PRISMA flow diagram (**Figure 1**).

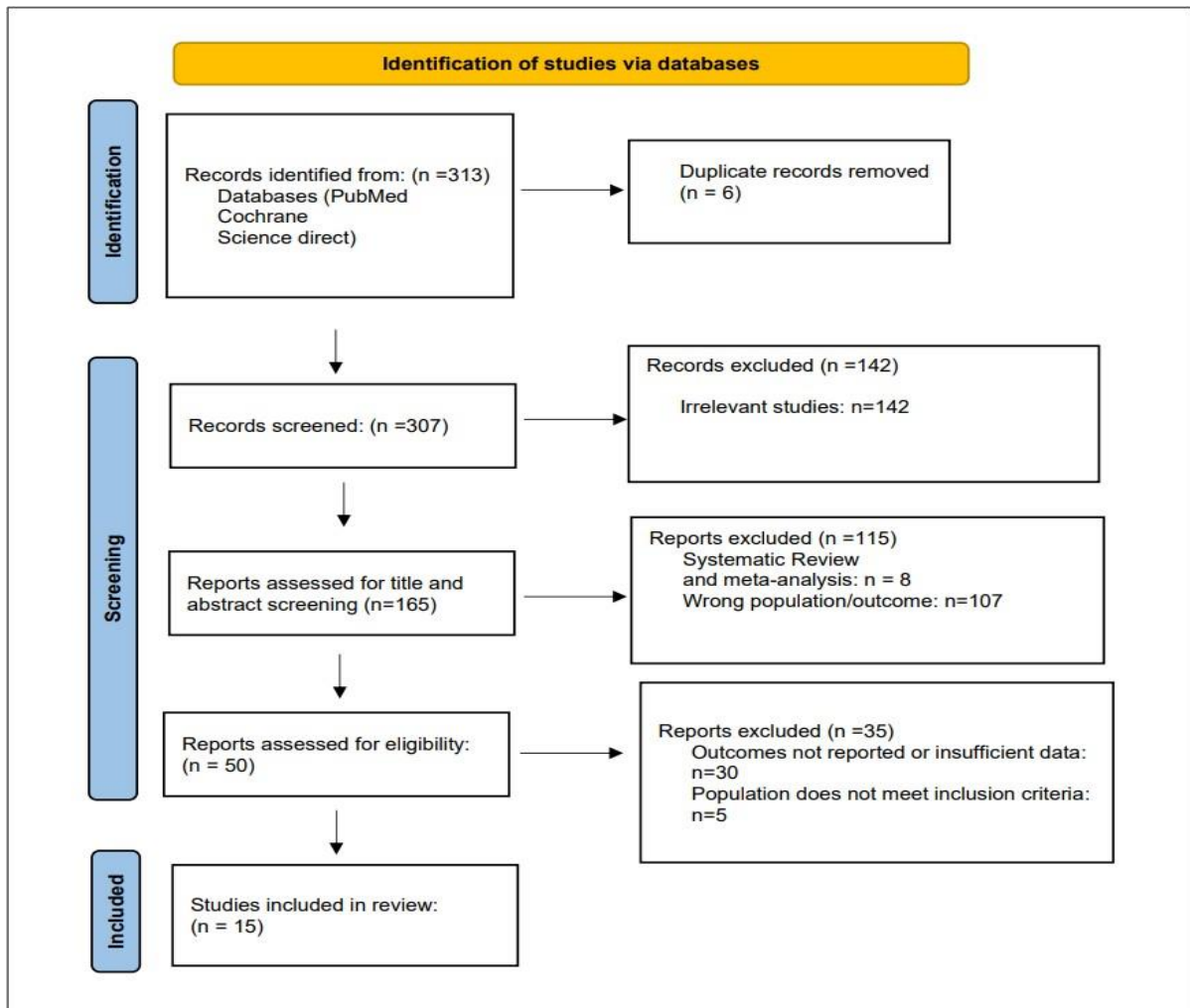


Figure 1: PRISMA flow diagram

Data was independently extracted by two reviewers using a standardized data extraction form. Extracted information included study characteristics and participant characteristics (author, year, country, study design, gender, and age of the study population) and key findings related to AI-based models and retinal imaging modalities. Any discrepancies in data extraction were resolved by consensus.

Quality Assessment

In our systematic review, we employed the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) tool as a critical method to assess the methodological quality and diagnostic accuracy of the included studies. QUADAS-2 assessment evaluated the risk of bias and applicability concerns across five domains: patient selection, index test, reference standard, flow and

timing, and data management (14). Each domain was judged as having low, high, or unclear risk of bias. Two reviewers independently performed the assessment, and any discrepancies were resolved through discussion or consultation with a third reviewer.

Results

Search Results

We executed the search methodologies outlined previously, resulting in the identification of a total of 313 citations, subsequently reduced to 307 following the removal of duplicates. Upon screening titles and abstracts, only 165 citations met the eligibility criteria for further consideration. Through full-text screening, this number was further refined to 15 articles (15-29) aligning with our inclusion and exclusion criteria. **Figure 1**

provides an in-depth depiction of the search strategy and screening process.

Baseline Characteristics of the Included Studies

In this systematic review, a total of 15 studies published between 2018 and 2026 were included. The combined sample size of participants across these studies was 200,000, with a mean age of 40–70 years. Gender distribution was relatively balanced, with the majority of studies representing 13% to 56%. However, there was one study conducted by Woodward-Court et al. (24) in which participants were predominantly females. Geographically, the studies were conducted in diverse locations, including China, the United States, Taiwan, Turkey, Iran, and France. The study designs were varied, including seven retrospective cohort studies, three cross-sectional studies, two diagnostic accuracy studies, two AI development studies, and one prospective comparative study (**Table 1**).

Diseases Evaluated and Imaging Modalities

We evaluated a wide range of non-diabetic retinal diseases. The most evaluated condition was age-related macular degeneration (AMD), followed by inherited retinal diseases (retinitis pigmentosa (RP) and Stargardt disease (STGD)), retinal detachment (RD), vascular occlusions, such as central retinal artery occlusion (CRAO) and central retinal vein occlusion (CRVO), and glaucoma (**Table 1**).

The most used imaging modalities were color fundus photography (CFP) and OCT, followed by fundus autofluorescence (FAF) and ultra-widefield imaging (UWF). Whereas most studies utilized DL models, specifically models utilizing convolutional neural networks (CNNs), such as ResNet, EfficientNET, and Inception (**Table 1**).

Table 1: Baseline characteristics of the included studies

Study ID (Author, Year)	Country	Study Design	Setting (Clinic/Screening)	Sample Size (Images/Eyes/Patients)	Mean Age \pm SD	Gender (% Male)	Disease Type	Imaging Modality (Fundus/OCT)	AI Model Type
Dong Li et al., 2022 (15)	China	Multicenter, diagnostic study	Screening at 65 public medical centers & hospital	Development: 120,002 images / 63,400 patients; Prospective Validation: 208,758 images / 110,784 patients	Development: median 44 (IQR 32-55, range 9-85) Prospective: median 42 (range 8-87)	Development : 46.3% Prospective: 44.6%	10 retinal diseases: DR, AMD, Glaucoma, PM, RVO, MH, Epiretinal Macular Membrane, Hypertensive Retinopathy, Myelinated Fibers, RP	Fundus (color fundus photography, nonmydriatic 45°)	DL (Multitask CNN with YOLOv3 for macula & optic disc detection)
Durmaz Engin et al., 2025 (16)	Turkey	Comparative diagnostic study	Not specified (retrospective OCT image analysis)	1,500 OCT images (500 normal, 500 dry AMD, 500 wet AMD)	Not reported	Not reported	AMD (dry, wet)	OCT	Expert-designed CNN (EfficientNet V2), AutoML CNN (ResNet-50 V2 via LobeAI)

Gu et al., 2024 (17)	China	Real-world, multicentre, cross-sectional	Primary healthcare (6 clinics in Shanghai & Xinjiang)	4795 patients (1 image per eye, both eyes imaged)	Median 57.0 (IQR 39–66)	33.80%	Multiple retinal diseases (DR, AMD, GON, PM, RVO, RD, MH, ME, CSC, ERM, RP, retinal drusen, macular neovascularisation, geographic atrophy)	Fundus photography (45°–50°)	DL – Airdoc retinal artificial intelligence system (ARAS) (Yolo-V3 + EfficientNet-B3 multitask classification network)
Gungor et al., 2026 (18)	Multicountry (France, Germany, China, India, US, Denmark)	Retrospective, Cross-sectional	Clinic (9 neuro-ophthalmology centers)	1322 images / 771 patients	Training: 68.1 (16–97), Testing: 68.5 (16–100)	Training: 42.7%, Testing: 50.4%	CRAO, CRVO, NAION, Healthy Controls	Fundus Photography	CNN/DL
Katuru et al., 2025 (19)	USA	Retrospective cohort	Tertiary glaucoma center	16,936 image sets (DP + OCT), 1 eye per patient	63.73 ± 14.57	46.70%	Glaucoma	Fundus (Disc Photos), OCT (RNFLT maps)	DL (CNN-based)
Keenan et al., 2020 (20)	USA	Retrospective cohort	Multicenter retinal specialty clinics (AREDS2)	11,275 images / 4,724 eyes / 2,443 patients	72.9 ± ~5.5	42.5	AMD (intermediate to late)	FAF and CFP	DL (CNN, DenseNet-152 primary; also VGG16/19, InceptionV3, ResNet101)

Keenan et al., 2020 (21)	USA	Prospective comparison	Retinal specialty clinics	1127 eyes / 651 patients	80.0 ± 7.6	40.50%	AMD (neovascular)	SD-OCT (Cirrus & Spectralis)	ML-based (Notal OCT Analyzer)
Ou et al., 2023 (22)	China	Multicenter retrospective cohort	Clinic-based	2330 volumetric scans (1833 training, 497 external test); eyes: 1833 + 497; patients not specified	Center 1: 54.9 ± 12.3; Center 2: 52.2 ± 12.0	Center 1: 54.2; Center 2: 57.4 (female %) → male % derived: 45.8 & 42.6 → Male 54.2% & 57.4%)	AMD, DR, RVO, MH, CSC, EM, RS, Normal	OCT (B-scan + en face)	DL (Multiview Fusion Network – CNN backbone: ResNet-50; fusion via Random Forest)
Peng et al., 2018 (23)	USA	Retrospective analysis (AREDS cohort)	Multi-center clinic-based	59,302 images from 4,549 patients	Not explicitly reported	Not explicitly reported	AMD	CFP	DL (CNN, Inception-v3)
Woodward-Court et al., 2026 (24)	USA & UK (multi-center)	Retrospective, multi-institutional	Clinical setting	8251 B-scans / 286 eyes / 409 patients	61.6 ± 13.5 years	~13% male (87% female)	Hydroxychloroquine retinopathy	SD-OCT	DL (CNN; EfficientNet-b4)
Chen et al., 2021 (25)	Taiwan	Retrospective study	Clinical database (hospital-based)	1670 images (935 RP, 324 normal for training; 386 test)	Not reported	Not reported	RP	CFP	DL (Transfer learning: Xception, Inception V3, Inception ResNet V2)

Jafarbeglou et al., 2025 (26)	Iran	Cross-sectional	Clinical (registry-based ophthalmic evaluation)	391 subjects (158 RP, 62 STGD, 171 healthy)	RP: 40±13; STGD: 30±12; Healthy: 38±12	RP: 47%; STGD: 44%; Healthy: 43%	Inherited retinal diseases (RP and STGD)	CFP and infrared imaging	DL (MobileNetV2 multi-input CNN; compared with ML and other DL models)
Karimi et al., 2025 (27)	Iran	Retrospective AI development study	Likely clinical registry-based (IRDReg®)	5844 unlabeled + 782 labeled images (316 RP, 124 STGD, 342 normal)	RP: 40 ± 13; STGD: 30 ± 12; Healthy: 38 ± 12	RP: 47%; STGD: 44%; Healthy: 43%	RP and STGD	CFP and Infrared (IR) (pretraining), CFP (fine-tuning)	Self-supervised DL (EfficientNet-B1 backbone)
Li et al., 2020 (28)	China	Retrospective, diagnostic	Clinic (Zhongshan Ophthalmic Centre)	11,087 UWF images from 7,966 patients	47.5 ± N/A	56.40%	RD (rhegmatogenous, tractional, exudative, recurrent post-surgery), Non-RD retinopathies	UWF	Cascaded DL (two models: 1st RD detection, 2nd macula-on/off classification)
Miere et al., 2020 (29)	France	Retrospective	Ophthalmology Clinic	503 FAF images (73 healthy, 125 STGD, 125 BD, 160 RP)	Not reported	Not reported	STGD, BD, RP, and healthy controls	FAF	DL CNN (ResNet-101, transfer learning)

AMD: Age-Related Macular Degeneration, AREDS: Age-Related Eye Disease Study, BD: Best Disease, CFP: Color Fundus Photography, CNN: Convolutional Neural Networks, CRAO: Central Retinal Artery Occlusion, CRVO: Central Retinal Vein Occlusion, CSC: Central Serous Chorioretinopathy, DL: Deep Learning, DR: Diabetic Retinopathy, ERM: Epiretinal Membrane, FAF: Fundus Autofluorescence, GON: Glaucomatous Optic Neuropathy, MH: Macular Hole, ME: Macular Edema, ML: Machine Learning, NAION: Non-Arteritic Anterior Ischemic Optic Neuropathy, PM: Pathological Myopia, RD: Retinal Detachment, RP: Retinitis Pigmentosa, RS: Retinoschisis, RVO: Retinal Vein Occlusion, SD-OCT: Spectral-Domain Optical Coherence Tomography, STGD: Stargardt Disease, UWF: Ultra-Widefield Fundus Imaging

Diagnostic Performance of AI Models

AI models depicted high diagnostic performance across most of the evaluated non-diabetic retinal diseases. For AMD, sensitivity ranged between 59% and 100%, specificity ranged between 91% to 100%, accuracy ranged between 67% to 99.7%, with AUC ranging between 0.88 and 0.97. For inherited retinal diseases, AI models showed sensitivity ranging between 95% and 100%,

specificity 97% and 99%, accuracy ranging between 91% and 99%, with AUC ranging between 0.995 and 0.999. For CRAO, AI-models detection sensitivity ranged between 92.6% and 100% with external validation. Whereas glaucoma detection showed the lowest sensitivity ranging between 53% to 70%, with high specificity, accuracy, and AUC values. Additionally, multidisease screening AI models demonstrated high diagnostic performance (**Table 2**).

Table 2: Summary of artificial intelligence and expert performance for detection of non-diabetic retinal diseases across multiple studies

Disease / Comparison	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC (ROC)	External Validation
Normal Fundus	57–78	89–99	76–100	–	Yes/No
AMD	59–100	91–100	67–99.7	0.88–0.97	No
Wet AMD vs Dry AMD	85–100	87–100	86–99.5	Not reported	No
RP / STGD	95–100	97–99	91–99	0.995–0.999	No / Internal
RD	93.8–96.1	90.9–99.6	91.7–98.9	0.975–0.989	Yes
CRAO	92.6–100	80–85	85–88.7	0.96–0.97	Yes (external)
Retinal Fluid (any / intraretinal / subretinal)	40.3–94	85.7–97.8	80.5–94.6	0.925	Yes / No
Glaucoma / RPD	53–70	90–97	80–90	0.832–0.939	No / Internal
Hydroxychloroquine Retinopathy	100	98.3	98.7	–	Yes

Multidisease Screening (AMD, RVO, MH, CSC, ERM, RS, CNV)	80–95	93–97	92.5–95.2	0.976–0.994	Yes
---	-------	-------	-----------	-------------	-----

AMD: Age-Related Macular Degeneration, AUC: Area Under the Curve, CNV: Choroidal Neovascularization, CRAO: Central Retinal Artery Occlusion, CSC: Central Serous Chorioretinopathy, ERM: Epiretinal Membrane, MH: Macular Hole, RD: Retinal Detachment, ROC: Receiver Operating Characteristic, RP: Retinitis Pigmentosa, RPD: Reticular Pseudo-Drusen, RS: Retinoschisis, RVO: Retinal Vein Occlusion, STGD: Stargardt Disease

Advantages and Disadvantages of Clinical Use of AI models

External validation was reported in some of the studies, with some AI models utilizing independent datasets and multicenter datasets. Multiple studies were conducted in a real-world setting or clinical settings, depicting the possibility of integrating AI models into daily clinical practice. These AI systems reported decreased examination time, operating with a few ophthalmologists, reducing regional disparities, supporting teleophthalmology and screening workflows, handling multiple diseases simultaneously, early prediction, in addition to the increased efficacy and accuracy (**Table 3**).

The majority of studies were retrospective and lacked external validation, which may have limited generalizability. Additionally, several AI models required high-quality images, in addition to depending on OCT, which is not available in non-specialized areas. The limited sample for rare conditions yielded a low sensitivity for detection. However, multiple models reported high readiness for clinical deployment (**Table 3**).

Table 3: Advantages and Disadvantages of Clinical Use of AI models

Study ID	Real-world Setting	External Validation	Pros for Clinical Use	Limitations	Deployment Readiness (Low/Moderate/High)
Dong Li et al 2022 (15)	65 screening centers & hospitals across 19 Chinese provinces	Yes, Beijing Eye Study & Kailuan Eye Study	<ul style="list-style-type: none"> - High diagnostic accuracy for 10 retinal diseases - Saves 95% examination time - Heatmap visualization for explainability - Can operate in areas with few ophthalmologists - Can be combined with ophthalmologists for efficiency 	<ul style="list-style-type: none"> - Slightly lower performance for PM - Image quality control by ophthalmologists may limit real-world application - Small number of RP images 	High
Durmaz Engin et al., 2025 (16)	Retrospective analysis of publicly available OCT images	No	<ul style="list-style-type: none"> - AutoML allows physicians to generate models without coding - expert models achieve high accuracy and near-perfect F1 scores 	<ul style="list-style-type: none"> - Dataset may not fully represent real-world variability - AutoML performance lower than expert models - geographic atrophy excluded 	High (Expert model), Moderate (AutoML)
Gu et al., 2024 (17)	Primary healthcare clinics in Shanghai (high-income) and Xinjiang (low-income)	No	<ul style="list-style-type: none"> - High accuracy for multiple retinal abnormalities - can reduce regional disparities - multitask detection of macula and optic disc lesions 	<ul style="list-style-type: none"> - Sensitivity varies across diseases - low detection for rare conditions - external validation pending 	Moderate
Gungor et al., 2026 (18)	Multi-center neuro-ophthalmology clinics; potential application in stroke/emergency settings	Yes	<ul style="list-style-type: none"> - Early hyperacute CRAO detection - aids fibrinolysis eligibility - robust multiclass classification - outperforming stroke neurologists 	<ul style="list-style-type: none"> - Limited hyperacute CRAO images - requires high-quality fundus photos - needs further prospective validation 	Moderate–High

Katuru et al., 2025 (19)	Tertiary glaucoma center	No	<ul style="list-style-type: none"> - High accuracy using OCT - quantitative measurement of retinal nerve fiber layer thickness - improved detection across diverse demographics - visual field (VF)-based ground truth reduces subjective bias 	<ul style="list-style-type: none"> - DP-based models less accurate - demographic disparities - OCT access may be limited in non-specialist settings - sensitivity/specificity not reported 	Moderate-High (OCT-based DL feasible, but requires OCT devices and integration)
Keenan et al., 2020 (20)	Multicenter retinal specialty clinics (AREDS2)	No	<ul style="list-style-type: none"> - FAF-based DL outperforms ophthalmologists for RPD detection - CFP-based DL still surpasses human graders - Can assist early identification of high-risk AMD eyes - Could be integrated into teleophthalmology or screening workflows 	<ul style="list-style-type: none"> - CFP-based DL lower accuracy than FAF - FAF less widely available than CFP - No prospective or external validation - Requires computational resources 	Moderate
Keenan et al., 2020 (21)	Multicenter, 19 USA retinal specialty clinics	Yes (Reading center as ground truth)	<ul style="list-style-type: none"> - High accuracy and sensitivity for AI - assists in detecting intraretinal and subretinal fluid - rapid and consistent analysis - could reduce missed fluid in clinical practice - prospective design 	<ul style="list-style-type: none"> - AI may overcall low-volume fluid - requires good image quality - some scans excluded due to poor quality or artifacts 	Ready for clinical decision support; not autonomous diagnosis; can aid human specialists in routine OCT evaluation
Ou et al., 2023 (22)	Multicenter clinic-based (2 centers)	Yes (Center 2 external set; OCTA-500 public dataset)	<ul style="list-style-type: none"> - High diagnostic performance (AUC 0.994) - handles multiple diseases simultaneously - interpretable via activation heatmaps - robust across different backbones and fusion algorithms 	<ul style="list-style-type: none"> - Limited sample size for rare diseases - preprocessing dependent on OCT device - exact patient-level labels not reported - lack of prospective real-world testing 	Medium-High – strong technical performance but needs prospective clinical trials and integration with hospital systems
Peng et al., 2018 (23)	Retrospective AREDS cohort (multi-center clinics)	No	<ul style="list-style-type: none"> - Patient-level scoring - mimics human grading - interpretable via sub-networks 	<ul style="list-style-type: none"> - Limited external validation - some performance lower for late AMD detection 	Early-stage readiness; requires prospective/real-world validation before deployment

- high accuracy for drusen/pigment detection
- public availability

- retrospective design

- High accuracy
- early prediction (years before diagnosis)
- single modality (OCT)
- automated (no manual input)
- works across devices
- reduces need for multimodal testing

- Retrospective design
- limited reporting of AUC/F1
- small number of pericentral (Asian phenotype) cases
- potential dataset imbalance
- requires OCT availability

High (multi-center validation, device generalizability, clinically relevant outputs)

Woodward-Court et al., 2026 (24)

Yes (routine clinical SD-OCT data from multiple centers)

Yes (3 external datasets across USA & UK)

Chen et al., 2021 (25)

Potential use in rural/low-resource settings; telemedicine applicability

No true external validation

- High diagnostic accuracy
- comparable to specialists
- uses widely available fundus imaging
- interpretable via Grad-CAM
- supports early detection

- Retrospective design
- no external validation
- limited demographic reporting
- false positives in high myopia
- false negatives with media opacity (e.g., cataract)

Moderate – promising but requires external validation and real-world testing

Jafarbeglou et al., 2025 (26)	Simulated real-world dataset (registry-based, prevalence-balanced)	No external validation dataset	<ul style="list-style-type: none"> - High diagnostic accuracy - multi-modal imaging improves detection - lightweight MobileNetV2 suitable for low-resource settings - Grad-CAM enhances interpretability - potential for telemedicine screening 	<ul style="list-style-type: none"> - No external validation - limited to two inherited retinal diseases - relatively small STGD sample - lack of threshold reporting - cross-sectional design 	Moderate – promising but requires external validation and prospective testing
Karimi et al., 2025 (27)	No (experimental dataset; registry-based)	No	<ul style="list-style-type: none"> - Very high accuracy (98.15%) and AUC (99.68%) - robust with limited labeled data - interpretable via Grad-CAM - allows for scalability and biologically informed augmentation (left-right eye pairing) 	<ul style="list-style-type: none"> - No external validation; relatively small, labeled dataset - limited generalizability - no prospective or real-world testing - unclear threshold tuning - potential class imbalance - Limited generalizability to non-UWF imaging devices 	Moderate (requires external validation and clinical trials before deployment)
Li et al., 2020 (28)	Clinical setting (large ophthalmology center)	Yes, dataset from another institution	<ul style="list-style-type: none"> - High sensitivity suitable for screening - macula-on/off detection aids surgical timing - automated guidance for preoperative posturing - interpretable heatmaps 	<ul style="list-style-type: none"> - misclassifications in shallow or distorted retinal diseases - some misalignment in patient gaze during capture 	High readiness for deployment in clinics with UWF imaging; can assist ophthalmologists and aid screening in under-resourced areas
Miere et al., 2020 (29)	Clinic FAF images	No	<ul style="list-style-type: none"> - High accuracy for classification of inherited retinal diseases - interpretable via integrated gradients - automated detection 	<ul style="list-style-type: none"> - Retrospective, single-center - limited sample size - no external validation 	Preliminary; requires larger multicenter validation before clinical deployment

AMD: Age-Related Macular Degeneration, AREDS: Age-Related Eye Disease Study, AUC: Area Under the Curve, AutoML: Automated Machine Learning, CFP: Color Fundus Photography, CRAO: Central Retinal Artery Occlusion, FAF: Fundus Autofluorescence, Grad-CAM: Gradient-weighted Class Activation Mapping, PM: Pathological Myopia, RP: Retinitis Pigmentosa, SD-OCT: Spectral-Domain Optical Coherence Tomography, STGD: Stargardt Disease, UWF: Ultra-Widefield Fundus Imaging.

Quality Assessment

The methodological quality of the included studies was assessed using the QUADAS-2 tool. Overall, five studies demonstrated low risk of bias across all the assessed domains, however, a high risk of bias in patient selection was observed in two studies. Several studies had unclear risk of bias, especially in the patient selection and data management domains. Unclear risk of bias in patient selection resulted from limited information about

recruitment methods and eligibility criteria. Similarly, unclear risk of bias in data management was due to insufficient reporting of data handling and analysis procedures (**Table 4**).

QUADAS-2 assessment results for the included studies, indicating low risk (green), high risk (red), or unclear risk (orange) of bias across the following domains: Patient Selection, Index Test, Reference Standard, Flow & Timing, and Data Management.

Table 4: QUADAS-2 assessment results for included studies

Study	Risk of Bias				Concerns regarding Applicability		
	Patient Selection	Index Test	Reference Standard	Flow & Timing	Patient Selection	Index Test	Reference Standard
Dong Li et al., 2022 (15)	Low	Low	Low	Low	Low	Low	Low
Durmaz Engin et al., 2025 (16)	High	Low	Low	Unclear	Unclear	Low	Low
Gu et al., 2024 (17)	High	Low	Low	Unclear	Unclear	Low	Low
Gungor et al., 2026 (18)	Low	Low	Low	Low	Low	Low	Low
Katuru et al., 2025 (19)	Unclear	Unclear	Low	Low	Unclear	Unclear	Low
Keenan et al., 2020 (20)	Low	Low	Unclear	Low	Low	Low	Unclear
Keenan et al., 2021 (21)	Low	Low	Low	Low	Low	Low	Low
Ou et al., 2023 (22)	Low	Low	Low	Unclear	Low	Low	Low
Peng et al., 2018 (23)	Low	Unclear	Low	Low	Low	Unclear	Low
Woodward-Court et al., 2026 (24)	Low	Low	Low	Low	Low	Low	Low
Chen et al., 2021 (25)	Low	Unclear	Low	Unclear	Low	Unclear	Low
Jafarbeglou et al., 2025 (26)	Low	Unclear	Low	Unclear	Low	Unclear	Low
Karimi et al., 2025 (27)	Low	Unclear	Unclear	Low	Unclear	Unclear	Unclear
Li et al., 2020 (28)	Low	Low	Low	Low	Low	Low	Low
Miere et al., 2020 (29)	Unclear	Low	Unclear	Low	Low	Low	Low

Figure 2 depicts the results for QUADAS-2 assessment of included studies. Most of the studies were rated as having low risk of bias in the flow and timing, index test, and reference standard domains. In the patient selection and data

management domains, the majority of studies were judged as having unclear risk of bias.

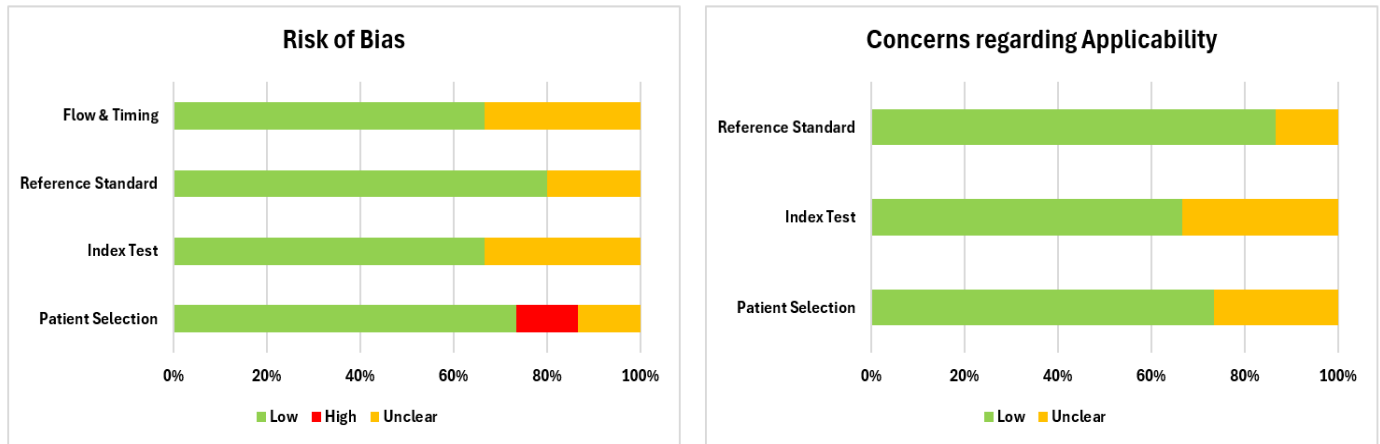


Figure 2: QUADAS-2 assessment results for included studies

Discussion

This systematic review aimed to examine the diagnostic accuracy and clinical validation of AI-based models for the detection of non-diabetic retinal diseases, and to evaluate their potential clinical applicability. 200,000 participants were included in this study, with a mean age of 40–70 years. Our results show that AI models depicted high diagnostic performance across most of the evaluated non-diabetic retinal diseases. Results also show the possibility of integrating AI models in daily clinical practice demonstrating several advantages including decreased examination time, operating with a few ophthalmologists, reducing regional disparities, supporting teleophthalmology and screening workflows, and handling multiple diseases simultaneously. Additionally, AI models have been shown to assist in early prediction of disease.

Our results revealed high diagnostic accuracy for several non-diabetic retinal diseases. Included studies showed high specificity with various sensitivity rates. Similar findings have been reported by Ashrafi et al. (30) in which AI models showed high diagnostic accuracy for inherited retinal diseases, particularly for RP and STGD. Similarly,

Cen et al. (31) used AI-trained models for the detection and classification of fundus diseases. These findings highlight the potential role of AI-based models in assisting ophthalmologists in clinical diagnosis of retinal diseases, especially in remote areas.

Moreover, our findings highlight the superiority of DL models in identifying non-diabetic retinal diseases. These findings are consistent with Parrey et al. (32) where AI-based systems, particularly DL models, showed high accuracy and precision in the identification of retinal diseases, such as glaucoma, AMD, and DR, demonstrating strong potential for clinical application. DL is particularly advantageous when these models are trained on large, diverse datasets, leading to better diagnostic accuracy and generalizability across different patient populations. This continuous learning and improvement from vast amounts of retinal imaging data render AI valuable as a diagnostic tool in ophthalmic practice.

Findings indicate the effectiveness of AI-based models in early prediction and identification of non-diabetic retinal diseases, especially AMD. This is consistent with the findings of Saha et al. (33) who trained deep convolution neural networks (CNN) for

automated detection and classification of early AMD biomarkers from OCT images with an overall accuracy of 87%. Early diagnosis is particularly beneficial for asymptomatic patients, who might otherwise go undiagnosed until advanced stages, such as outer retinal atrophy or development of exudative neovascular membranes, that could lead to irreversible loss of vision. These findings underscore the importance of early disease identification, which allows for accurate stage identification and timely intervention to slow disease progression.

Strengths and Limitations

This systematic review has several strengths, such as employing a multi-database search strategy and following the PRISMA 2020 guidelines, which strengthen the reproducibility and transparency of the study findings. The review included a broad range of studies with geographically diverse populations, which strengthens the generalizability of the study findings. The study also reviewed major outcomes of interest, including diagnostic accuracy and clinical validation outcomes of AI models and retinal imaging modalities outcomes, which provide a comprehensive assessment of the use of AI-based models for detection of non-diabetic retinal diseases. Additionally, the use of QUADAS-2 tool for assessing methodological quality of the included studies led to bias reduction. None of the included studies showed high risk of bias, which strengthens the study findings, although some of the included studies were rated as having unclear risk of bias, specifically in the patient selection and data management domains due to insufficient methodological reporting.

The majority of studies were retrospective and lacked external validation, which may have limited generalizability and robustness of the study findings, particularly across diverse populations and healthcare settings. This also raises concerns regarding the reproducibility of AI models in detection of non-diabetic retinal diseases. Moreover, the use of some datasets such as the AREDS (Age-Related Eye Disease Study) fundus photographs. Additionally, integration into clinical practice requires appropriate infrastructure, regulatory oversight, and ophthalmologist training.

poses a major limitation due to overrepresentation of white participants, which may lead to algorithmic bias, which have been addressed by inclusion of multicenter studies from countries with non-white populations. This highlights the importance of multi-modal imaging for diagnosis along with external validation. Another major limitation is the substantial heterogeneity across studies in terms of diseases, imaging modalities, and AI models; therefore, a narrative synthesis was considered the most appropriate approach as it enabled a comprehensive interpretation of diagnostic performance, clinical applicability, methodological limitations, and sources of bias related to the use of AI-based systems for detection of non-diabetic retinal diseases. Furthermore, several AI models required high-quality images, in addition to depending on OCT in several studies, which is not available in non-specialized areas. Additionally, the limited sample for rare conditions yielded a low sensitivity for detection.

Implications and Recommendations

This systematic review highlights the high accuracy and specificity of AI-based systems along with their potential as effective tools for screening and identification of non-diabetic retinal diseases. Moreover, AI-based models are considered promising in triage settings. The use of models with high sensitivity could lead to reduced false negatives, ensuring early disease identification. While models with higher specificity might decrease the workload on retinal specialists through reducing false positive referrals, thus improving efficiency in clinical workflows. This is particularly important in low-to-middle income countries and resource limited areas, where the integration of AI could lead to eye care standardization and improved access.

However, most studies relied on retrospective and potentially non-diverse datasets, raising concerns about generalizability. Future research should focus on prospective, multicenter validation, inclusion of diverse populations, and standardization of evaluation metrics, to ensure generalizability.

Conclusion

AI-based models demonstrate strong diagnostic performance for detecting non-diabetic retinal diseases, particularly in terms of specificity. While promising for early disease identification, their clinical adoption should be guided by rigorous validation and careful implementation to ensure reliability, safety, and improved patient outcomes.

Disclosure

Conflict of interest

There is no conflict of interest.

Funding

All authors have declared that no financial support was received from any organization for the submitted work.

Ethical consideration

Not applicable.

Data availability

All data is available within the manuscript.

Author contribution

All authors contributed to conceptualizing, data drafting, collection and final writing of the manuscript.

References

1. Hariprasad SM, Holz FG, Asche CV, Issa A, Mora O, Keady S, et al. Clinical and Socioeconomic Burden of Retinal Diseases: Can Biosimilars Add Value? A Narrative Review. *Ophthalmology and therapy*. 2025;14(4):621-41.
2. Lei C, Jiang F, Zhang L, Ran Q, Zhang Y, Zhang M. The Association Between Vitamin D Deficiency and Nondiabetic Retinopathy in the American Population: National Health and Nutrition Examination Survey 2005-2008. *BioMed research international*. 2025;2025:2828949.
3. Venkatramani J, Mitchell P. Ocular and systemic causes of retinopathy in patients without diabetes mellitus. *BMJ (Clinical research ed)*. 2004;328(7440):625-9.
4. Ojaimi E, Nguyen TT, Klein R, Islam FM, Cotch MF, Klein BE, et al. Retinopathy signs in people without diabetes: the multi-ethnic study of atherosclerosis. *Ophthalmology*. 2011;118(4):656-62.
5. Klein R, Klein BE, Moss SE, Wong TY. The relationship of retinopathy in persons without diabetes to the 15-year incidence of diabetes and hypertension: Beaver Dam Eye Study. *Transactions of the American Ophthalmological Society*. 2006;104:98-107.
6. Jafarbeglou F, Ahmadi H, Soleimani F, Karimi A, Daftarian N, Fekri S, et al. A deep learning model for diagnosis of inherited retinal diseases. *Scientific reports*. 2025;15(1):22523.
7. Bilal H, Keles A, Bendeche M. Advances in disease detection through retinal imaging: A systematic review. *Computers in Biology and Medicine*. 2025;194:110412.
8. Leng X, Shi R, Wu Y, Zhu S, Cai X, Lu X, et al. Deep learning for detection of age-related macular degeneration: A systematic review and meta-analysis of diagnostic test accuracy studies. *PloS one*. 2023;18(4):e0284060.
9. Yoon J, Han J, Ko J, Choi S, Park JI, Hwang JS, et al. Developing and Evaluating an AI-Based Computer-Aided Diagnosis System for Retinal Disease: Diagnostic Study for Central Serous Chorioretinopathy. *Journal of medical Internet research*. 2023;25:e48142.
10. Markakis D, Britten-Jones AC, Guymer RH, Edwards TL, Hall AJ, Kerr NM, et al. Retrospective audit reviewing accuracy of clinical diagnosis of geographic atrophy in a single centre private tertiary retinal practice in Australia. *Scientific reports*. 2025;15(1):8528.
11. Savastano MC, Rizzo C, Fossataro C, Bacherini D, Giansanti F, Savastano A, et al. Artificial intelligence in ophthalmology: Progress, challenges, and ethical implications. *Progress in Retinal and Eye Research*. 2025;107:101374.
12. Lee CS, Brandt JD, Lee AY. Entering the Exciting Era of Artificial Intelligence and Big Data

- in *Ophthalmology*. *Ophthalmol Sci*. 2024;4(2):100469.
13. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed)*. 2021;372:n71.
 14. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine*. 2011;155(8):529-36.
 15. Dong L, He W, Zhang R, Ge Z, Wang YX, Zhou J, et al. Artificial Intelligence for Screening of Multiple Retinal and Optic Nerve Diseases. *JAMA network open*. 2022;5(5):e229960.
 16. Durmaz Engin CB, U.; Özizmirli, D.; Selver, M. A. Comparative Analysis of Automated vs. Expert-Designed Machine Learning Models in Age-Related Macular Degeneration Detection and Classification. *Turkish journal of ophthalmology*. 2025;55(3):120-6.
 17. Gu CW, Y.; Jiang, Y.; Xu, F.; Wang, S.; Liu, R.; Yuan, W.; Abudureyimu, N.; Wang, Y.; Lu, Y.; Li, X.; Wu, T.; Dong, L.; Chen, Y.; Wang, B.; Zhang, Y.; Wei, W. B.; Qiu, Q.; Zheng, Z.; Liu, D.; Chen, J. Application of artificial intelligence system for screening multiple fundus diseases in Chinese primary healthcare settings: a real-world, multicentre and cross-sectional study of 4795 cases. *The British journal of ophthalmology*. 2024;108(3):424-31.
 18. Gungor AS, I.; Gilbert, A. L.; Hamann, S.; Lebranchu, P.; Hobeau, C.; Gohier, P.; Vignal-Clermont, C.; Dumitrascu, O. M.; Cohen, S. Y.; Lagrèze, W. A.; Feltgen, N.; van der Heide, F.; Lamirel, C.; Jonas, J. B.; Obadia, M.; Racoceanu, D.; Milea, D. Artificial Intelligence-Based Detection of Central Retinal Artery Occlusion Within 4.5 Hours on Standard Fundus Photographs. *Journal of the American Heart Association*. 2025;14(13):e041441.
 19. Katuru AC, In Young; Majid, Iyad; Shen, Lucy Q.; Wang, Mengyu. Deep Learning with Disc Photos or OCT Scans in Glaucoma Detection. *Ophthalmology Science*. 2025;5(6):100877.
 20. Keenan TDLC, Q.; Peng, Y.; Domalpally, A.; Agrón, E.; Hwang, C. K.; Thavikulwat, A. T.; Lee, D. H.; Li, D.; Wong, W. T.; Lu, Z.; Chew, E. Y. Deep Learning Automated Detection of Reticular Pseudodrusen from Fundus Autofluorescence Images or Color Fundus Photographs in AREDS2. *Ophthalmology*. 2020;127(12):1674-87.
 21. Keenan TDLC, T. E.; Domalpally, A.; Elman, M. J.; Havalio, M.; Agrón, E.; Benyamini, G.; Chew, E. Y. Retinal Specialist versus Artificial Intelligence Detection of Retinal Fluid from OCT: Age-Related Eye Disease Study 2: 10-Year Follow-On Study. *Ophthalmology*. 2021;128(1):100-9.
 22. Ou CW, X.; An, L.; Qin, J.; Zhu, M.; Jin, M.; Kong, X. A Deep Learning Network for Accurate Retinal Multidisease Diagnosis Using Multiview Fusion of En Face and B-Scan Images: A Multicenter Study. *Translational vision science & technology*. 2024;13(12):31.
 23. Peng Y, Dharssi S, Chen Q, Keenan TD, Agrón E, Wong WT, et al. DeepSeeNet: A Deep Learning Model for Automated Classification of Patient-based Age-related Macular Degeneration Severity from Color Fundus Photographs. *Ophthalmology*. 2019;126(4):565-75.
 24. Woodward-Court PH, J.; Lee, T.; Taribagil, P.; Zhao, C. S.; Otti, V.; Tucker, W. R.; Allingham, M.; Alekseev, O.; Wagner, S. K.; Myung, D.; Leung, L. S.; Lad, E. M.; Hasan, H.; Talks, J.; Alexander, D. C.; Keane, P. A.; Dow, E. R. Deep Learning Algorithm for the Diagnosis and Prediction of Hydroxychloroquine Retinopathy: An International, Multi-institutional Study. *Ophthalmology Retina*. 2025;9(12):1140-8.
 25. Chen TC, Lim WS, Wang VY, Ko ML, Chiu SI, Huang YS, et al. Artificial Intelligence-Assisted Early Detection of Retinitis Pigmentosa - the Most Common Inherited Retinal Degeneration. *Journal of digital imaging*. 2021;34(4):948-58.
 26. Jafarbeglou F, Ahmadi H, Soleimani F, Karimi A, Daftarian N, Fekri S, et al. A deep

learning model for diagnosis of inherited retinal diseases. *Scientific Reports*. 2025;15(1):22523.

27. Karimi A, Ahmadieh H, Sadeghi Tabrizi M, Daftarian N, Farshian Abbasi A, Fekri S, et al. A self-supervised learning method for detection of retinitis pigmentosa and Stargardt disease. *Scientific Reports*. 2025;15(1):43306.

28. Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen C, et al. Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. *Communications biology*. 2020;3(1):15.

29. Miere A, Le Meur T, Bitton K, Pallone C, Semoun O, Capuano V, et al. Deep Learning-Based Classification of Inherited Retinal Diseases Using Fundus Autofluorescence. *J Clin Med*. 2020;9(10).

30. Ashrafi S, Ahmadieh H, Bayat K, Baghban AA, Azadi A, Shahriari MH, et al. Diagnostic Accuracy of AI Models in Detecting Different Inherited Retinal Diseases: A Systematic Review and Meta-Analysis. *Translational vision science & technology*. 2025;14(7):22.

31. Cen LP, Ji J, Lin JW, Ju ST, Lin HJ, Li TP, et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nature communications*. 2021;12(1):4828.

32. Parrey MUR, Bhatti M, Abdul-Latif M, Rehman S, Ismail M, Hamid O. Meta-analysis of AI algorithm performance in detecting retinal diseases. *Anaesthesia, Pain & Intensive Care*. 2025;29:400-6.

33. Saha S, Nassisi M, Wang M, Lindenberg S, kanagasingam Y, Sadda S, et al. Automated detection and classification of early AMD biomarkers using deep learning. *Scientific reports*. 2019;9(1):10990.