Clinical Ophthalmology

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

PAIR: Evaluating the Limits of Agreement Among Non-Retinal Specialist Using PathFinder Artificial Intelligence Tool for Retinal Disease Referrals: A Prospective Observational Study

Kenneth CS Fong[1], Wilson J Wong[1], Amir Samsudin[1], Diymitra K Ganasan[1], Wan Ni Goh[1], Ju Yi Tai[2], Shir-Ning Fong[3], Crystal YJ Cheong[1], Ganeshwary Rajamanickam[1], Manoharan Shunmugam[1]

[1]Ophthalmology, OasisEye Specialists, Kuala Lumpur, Malaysia; [2]School of Clinical Medicine, University of Cambridge, Cambridge, UK; [3]School of Medicine, University of Nottingham, Nottingham, UK

Correspondence: Kenneth CS Fong, Email ken.fong@oasiseye.my

Purpose: To evaluate the diagnostic and referral agreement between non-retina specialists (NRS) using the PathFinder artificial intelligence (AI) assistant and fellowship-trained retina specialists *(RS* - gold standard) in interpreting macular optical coherence tomography (OCT) scans.

Methods: This cross-sectional study included 202 consecutive patients undergoing OCT on the CIRRUS platform with the PathFinder AI module. Three *RS* independently graded all scans without clinical data, while three NRS interpreted the same scans using PathFinder assistance and full clinical information. The gold standard for both diagnosis and referral was defined by agreement of at least two of the three RS. The NRS recorded diagnostic confidence and time to AI-assisted decision. Agreement was assessed with Cohen's and Fleiss' κ and sensitivity, and specificity were computed for four major pathologies.

Results: Among 202 eyes (mean age 62.7 ± 12.3 years), *RS* inter-agreement was moderate for diagnosis (overall κ = 0.59) and referral (κ = 0.47). NRS showed substantial (NRS1 κ = 0.78) to moderate (NRS2 κ = 0.64; NRS3 κ = 0.54) diagnostic agreement and high specificity (> 90% for all). Sensitivity varied across raters and diagnoses (0.57–0.89), with comparatively lower sensitivity observed for certain vision-threatening conditions such as age-related macular degeneration and macular hole. Referral agreement varied (κ = 0.68, 0.24, 0.34) amongst NRS. Visual acuity was the only significant predictor of referral discordance (OR 0.66, 95% CI 0.55–0.80, $p < 0.001$). Median NRS with PathFinder assistance processing time was < 20s. The most significant false-positive diagnoses made by PathFinder-assisted NRS in eyes deemed normal by *RS* was ERM (40.0%), followed by PED (20.0%) and AMD (13.3%), observed across a small number of eyes and does not affect the results.

Conclusion: PathFinder is a valuable real-time decision-support tool in resource-limited settings; however, disease-specific refinements and clinical oversight remain important, particularly for vision-threatening conditions.

Keywords: artificial intelligence, AI, optical coherence tomography, OCT, macular disease, diagnostic imaging, ophthalmology

Introduction

Artificial intelligence (AI) has emerged as a powerful tool in ophthalmic imaging, particularly for the automated interpretation of optical coherence tomography (OCT) scans.[1] By reducing reliance on subjective human interpretation, AI enables faster and more consistent detection of retinal pathologies such as choroidal neovascularization (CNV), diabetic macular edema (DME), and age-related macular degeneration (AMD). Traditional OCT analysis can be time-consuming and prone to interobserver variability, whereas deep learning (DL) based models have shown remarkable success in automating tasks such as image classification, layer segmentation, and fluid detection.[2,3]

https://doi.org/10.2147/OPTH.S584717
Received: 2 January 2026
Accepted: 28 February 2026
Published: 7 March 2026

Several studies have demonstrated that convolutional neural networks (CNNs) can accurately classify retinal diseases on OCT images and even outperform conventional methods. Nigjeh et al[4] reported enhanced accuracy for CNV and DME classification using VGG16 and ResNet18 models, while Umair et al developed OculusNet, an explainable AI framework capable of identifying complex pathological patterns.[5] Similarly, Ying et al showed that AI-based segmentation of hyperreflective foci improves diagnostic precision across multiple retinal diseases.[6] Beyond classification, AI systems also enhance image quality through noise reduction and super-resolution, improving interpretability and diagnostic workflow.[7]

Automated OCT interpretation is particularly valuable in settings with limited access to retina specialists. Studies have shown that machine-learning-based systems can reliably identify vision-threatening diseases and assist general ophthalmologists in triage and referral decisions.[8,9] These advances signal a paradigm shift toward integrated, clinician-assistive imaging technologies.

The CIRRUS PathFinder AI (Carl Zeiss Meditec AG, Germany) represents one such system, embedded within the OCT console to provide real-time, automated analysis. Trained on more than 75,000 OCT B-scans and validated by expert retina specialists, PathFinder demonstrates approximately 88% sensitivity and 93% specificity in detecting macular findings such as subretinal and intraretinal fluid, retinal pigment epithelium (RPE) changes, and photoreceptor disruption.[10] It also facilitates longitudinal monitoring of the fellow eye by automatically tracking anatomical changes over time.

Despite extensive research on AI-based OCT interpretation, most existing models are device-agnostic, and their accuracy varies across platforms due to differences in imaging technology. Unlike most device-agnostic OCT AI models evaluated retrospectively, PathFinder is an on-device, real-time decision-support system integrated directly into the CIRRUS OCT workflow, providing image quality assessment and pathology flagging at the point of acquisition and interpretation. However, its real-world diagnostic performance and clinical utility have not been independently validated. The present study aims to evaluate the agreement between non-retina specialists assisted by PathFinder AI and fellowship-trained retina specialists in diagnosing macular pathologies and determining referral decisions in a real-world setting. This work seeks to establish whether on-console and real-time AI assistance can enhance diagnostic consistency and streamline referral workflows in everyday clinical practice.

Methods

This was a cross-sectional diagnostic agreement study that included all consecutive patients undergoing OCT evaluation for suspected or confirmed retinal pathology during the April to July 2025 study period. The study adhered to the tenets of the Declaration of Helsinki and received approval from OasisEye Specialists Medical Advisory Board Research Ethics Committee (Approval number: OASIS-PF-025). The OasisEye Specialists Medical Advisory Board Research Ethics Committee determined that patient consent to review their medical record was not required as it was a non-invasive prospective study, and the patients' data maintains confidentiality.

Each patient's demographic and clinical information, including age, gender, presenting complaints, best-corrected visual acuity (BCVA), refraction, intraocular pressure, slit-lamp and fundus examination findings, and relevant systemic comorbidities, was recorded.

OCT scans were acquired using the Cirrus OCT platform (Carl Zeiss Meditec AG, Germany) following a standardised 6×6 mm macular raster scan protocol. Only scans of high image quality were included for analysis, defined as those with a signal strength ≥80%. Images below this threshold or those affected by segmentation artefacts were excluded. Additionally, eyes with significant media opacity resulting in poor scan quality (eg., dense cataract or corneal scarring) were excluded, as these could compromise OCT signal strength and limit reliable AI-assisted interpretation. We also excluded eyes with primary pathologies outside the 6×6 mm macular raster scan protocol (eg., optic nerve head disorders or advanced glaucomatous damage), as these conditions fall beyond the anatomical scope of the PathFinder macular analysis.

The PathFinder AI module (Carl Zeiss Meditec AG, Germany), integrated within the Cirrus OCT platform, employs deep-learning algorithms to automatically assess OCT image quality and detect areas of potential pathology within the macular cube. It functions by acquiring dense 512×128 Smart Cube scans of the macula, automatically centered on the

fovea, and subsequently qualifying each scan for image quality using neural networks trained to distinguish good from poor-quality images. During analysis, the software detects the presence of structural abnormalities consistent with eight predefined retinal pathologies or conditions and visually flags them for the operator, yellow indicators for poor-quality scans and red markers for B-scans containing suspicious findings. The system is fully embedded within the Cirrus review software and can be activated or deactivated as needed on the reviewer console.

All OCT scans were independently reviewed by three fellowship-trained retina specialists (RS1, RS2, RS3) and three non-retina specialists (NRS1, NRS2, NRS3). Each observer provided a diagnosis and a referral opinion on whether the patient should be referred to a retina specialist. The retina specialists were masked to all clinical and demographic details, including age, gender, BCVA, refractive status, intraocular pressure, cataract status, presenting complaints, and fundus findings, and based their evaluations solely on the OCT images. They did not have access to the Pathfinder AI software either.

The non-retina specialists had access to the full clinical profile, including BCVA, and systemic information, and made their diagnoses using PathFinder assistance. They were asked to record their confidence level in each diagnosis using a five-point Likert scale (*Not confident, Slightly confident, Moderately confident, Very confident*, and *Completely confident*). The time taken by the non-retina specialists with PathFinder AI assistance to indicate the probable pathology was recorded and categorised into six predefined intervals (<5, 6–10, 11–15, 16–20, 21–25, 26–30, and >30 seconds).

The primary outcome measure of this study was the level of agreement between the three non-retina specialists and the retina specialists (taken as the gold standard) in two key domains: (1) the diagnosis assigned based on OCT interpretation, and (2) the referral decision, ie., whether the case warranted referral to a retina specialist.

Statistical Analysis

Descriptive statistics were used to summarise the data. Continuous variables (eg., age) were reported as means with standard deviations, while categorical variables (eg., sex, diagnostic categories, referral status) were summarised as frequencies and percentages (n, %).

When retina specialists differed in their assessments, the gold standard for both diagnosis and referral was defined as the category on which at least two of the three specialists agreed. If all three specialists disagreed, the diagnosis or referral status assigned by the most experienced specialist (KF) was taken as the reference standard. The majority rule was also applied to the NRS group to determine the final NRS diagnosis for each case, ensuring true diagnostic discordance rates between the three *RS* and three NRS graders.

Inter-rater agreement among the three retina specialists was assessed using Cohen's kappa (κ) with 95% confidence intervals (Wilson's Method) for pairwise comparisons and Fleiss' kappa for overall agreement. Agreement between each non-retina specialist and the gold standard was also quantified using Cohen's kappa.

For the four most common diagnostic categories, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for each non-retina specialist relative to the gold standard. Discordance rates in referral decisions were determined by comparing each non-retina specialist's decision with the gold standard. To explore determinants of discordant referral, logistic regression analysis was performed with discordance as the dependent variable and age, baseline visual acuity, diagnosis made by the non-retina specialist, their confidence in the diagnosis and time taken by the Pathfinder assistant to arrive at the diagnosis as predictors. Model performance was evaluated through goodness-of-fit (Hosmer–Lemeshow test), discrimination (area under the ROC curve), and calibration.

All analyses were performed using R (version 4.3.3, packages: *psych, irr, epiR, dplyr, tidyr, ggplot2*). Statistical significance was set at $p < 0.05$.

Results

Participant Demographics

A total of 202 participants were included, with a mean age of 62.7 ± 12.3 years (median 64, IQR 56–72, range 30–91 years). There were 96 males (48%) and 106 females (52%). Common systemic comorbidities included hypertension (n=81, 40%), hyperlipidemia (n=75, 37%) and diabetes mellitus (n=47, 23%). The mean visual acuity at time of

assessment was 0.44 ± 0.47 logMAR (6/15 Snellen's equivalent) (median = 0.3 logMAR (IQR=0.14 to 0.5logMAR, range= 0 to 2 logMAR). The commonest gold-standard diagnosis, as agreed by at least 2 retina specialists, was epiretinal membrane (n=86, 43%), followed by AMD (n=35, 17%), normal OCT (n=30, 15%), macular hole (n=10, 5%) and macular oedema (n=9, 4.5%).

Agreement Between Retina Specialists

For diagnosis (Table 1), pairwise agreement between the three retina specialists ranged from moderate to substantial, with an overall Fleiss' kappa of 0.59, indicating moderate agreement. When considering the ratings of the three retina specialists, complete agreement across all three was seen in 111 cases (55%), while partial agreement (two specialists agreeing) was observed in 74 cases (38%). In 14 cases (7%), all three specialists disagreed. Using the majority rule (at least two raters in agreement), the overall agreement was 93% (95% CI: 88–96%). Amongst the individual disease entities, the highest agreement was seen in ERM (k=0.86) and the lowest in AMD (k=0.56) (Figure 1).

For referral decisions, the agreement was consistently moderate, with an overall Fleiss' kappa of 0.47 (Table 1). In absolute terms, the three retina specialists showed total agreement in 157 cases (78%) and partial agreement (two specialists agreeing) in 45 cases (22%). There were no instances of complete disagreement across all three raters. When applying the majority rule (at least two specialists agreeing), this translated into 100% agreement on referral versus non-referral decisions.

Agreement of Non-Retina Specialists with the Gold-Standard Diagnosis

When compared with the gold-standard diagnosis, agreement varied across the three non-retina specialists. NRS1 demonstrated substantial agreement (κ = 0.78, Table 2), whereas NRS2 showed moderate agreement (κ = 0.64). NRS3 had the lowest performance, with moderate agreement at the lower end (κ = 0.54). Overall, non-retina specialists expressed high levels of confidence when using PathFinder for diagnosis. Both NRS1 and NRS3 reported being very confident in the vast majority of cases (98% and 96% respectively), while NRS2 showed lower confidence, with just over half of the cases (55%) rated as very confident (Table 2).

The time taken by non-retina specialists to arrive at a diagnosis with PathFinder was generally short. For all three specialists, the majority of cases were interpreted within 11–20 seconds, with smaller proportions requiring either shorter (<10 seconds) or longer (>20 seconds) durations. Very rapid interpretations (<5 seconds) were uncommon, while only a minority of cases extended beyond 30 seconds. Overall, this indicates that the software enabled relatively quick decision-making across users.

Diagnostic Accuracy of Non-Retina Specialists

The ability of non-retina specialists to detect specific retinal pathologies using PathFinder varied across diseases and raters (Table 2). Sensitivity was generally high for cystoid macular edema and epiretinal membrane, but comparatively lower for age-related macular degeneration and macular hole, particularly for some raters (Figure 2). Specificity was consistently high across all conditions, often exceeding 90%, indicating a strong ability to identify normal cases correctly. Positive predictive values and kappa statistics, however, showed more variability, with lower values for certain raters in conditions such as cystoid macular edema and macular hole, reflecting occasional false positives and inconsistent diagnostic agreement with the gold standard. Of note, rater-level analysis revealed marked variability in positive predictive value for macular hole. NRS3 demonstrated a substantially lower PPV (0.39) for macular hole compared with the other non-retina specialists. Closer examination showed that NRS3 overdiagnosed macular hole in 14 cases, which were classified as epiretinal membrane (n = 12) or vitreomacular

Table 1 Agreement Between the Three Retina Specialists for Diagnosis and Referral (Kappa, 95% CI)

	RS1 – RS2	RS1 – RS3	RS2 – RS3	Overall
Diagnosis	0.53 (0.46 to 0.61)	0.60 (0.52 to 0.67)	0.64 (0.55 to 0.71)	0.59 (0.55–0.63)
Referral	0.44 0(0.28–0.58)	0.51 (0.34–0.65)	0.47 (0.29–0.62)	0.47 (0.44–0.51)

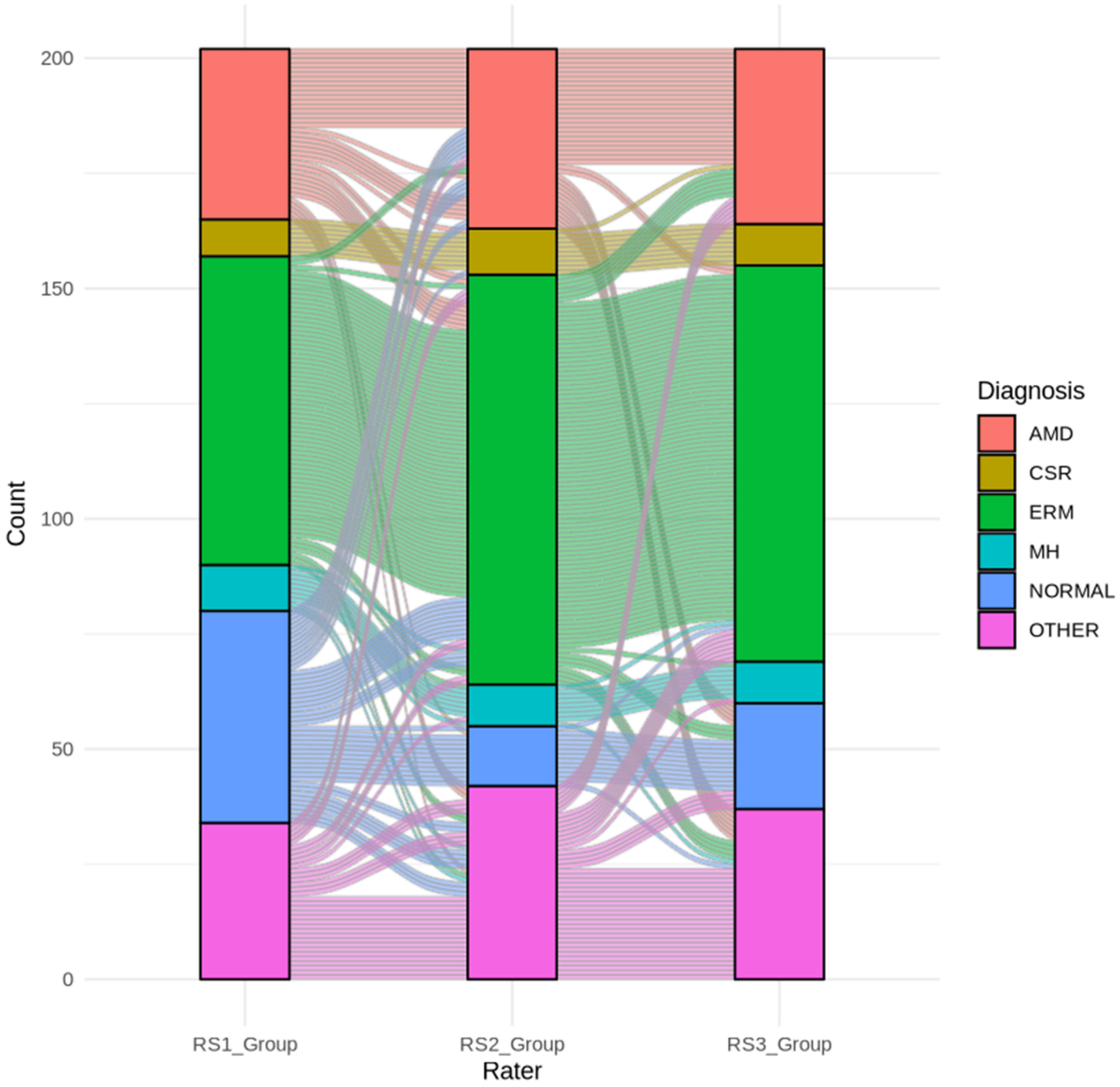

Figure 1 Alluvial diagram showing inter-rater agreement among the three retina specialists (RS1, RS2, RS3) for the top five primary diagnoses and "other" category.

traction (n = 2) by majority retina specialist consensus. Representative OCT images of these cases show ERM-associated pseudoholes in these cases (Figure 3).

Referral Agreement of Non-Retina Specialists with the Gold Standard

Referral patterns showed greater variability across non-retina specialists compared to diagnostic agreement. NRS1 demonstrated substantial agreement with the gold standard ($\kappa = 0.68$), while both NRS2 ($\kappa = 0.24$) and NRS3 ($\kappa = 0.34$) achieved only poor agreement in referral decisions (Figure 4). The lower referral agreement for NRS3 was primarily driven by false-positive diagnoses of macular hole, which led to over-referral in cases that were classified as epiretinal membrane or vitreomacular traction by the retina specialist panel.

https://doi.org/10.2147/OPTH.S584717

Table 2 Agreement of Non-Retina Specialists to Gold-Standard Diagnosis and Referral

	NRS1	NRS2	NRS3
Combined (All Cases)			
Diagnosis (Kappa, 95% CI)	0.78 (0.71–0.85)	0.64 (0.56–0.72)	0.54 (0.46–0.62)
Referral (Kappa, 95% CI)	0.68 (0.56–0.81)	0.24 (0.14–0.34)	0.34 (0.24–0.45)
Confidence levels in diagnosis with PathFinder			
Not confident	–	11 (5%)	
Slightly confident	–	29 (14%)	7 (3.5%)
Moderately confident	–	49 (24%)	1 (0.5%)
Very Confident	197 (98%)	112 (55%)	194 (96%)
Completely confident	5 (2%)	1 (1%)	-
Age related macular degeneration			
Sensitivity	0.57	0.63	0.69
Specificity	0.92	0.98	0.97
False Positive Rate	0.19	0.00	0.03
Positive Predictive Value	0.61	0.92	0.86
Kappa	0.5	0.71	0.72
Epi Retinal Membrane			
Sensitivity	0.89	0.86	0.54
Specificity	0.85	0.75	0.92
False Positive Rate	0.23	0.14	0.03
Positive Predictive Value	0.82	0.63	0.78
Kappa	0.74	0.56	0.51
Cystoid Macular Edema			
Sensitivity	0.89	0.75	0.75
Specificity	0.96	0.97	0.97
False Positive Rate	0.03	0.03	0.03
Positive Predictive Value	0.57	0.54	0.54
Kappa	0.68	0.61	0.61
Macular Hole			
Sensitivity	0.6	0.6	0.9
Specificity	0.99	0.99	0.92
False Positive Rate	0.00	0.00	7.3%
Positive Predictive Value	0.86	0.99	0.39
Kappa	0.69	0.74	0.51

https://doi.org/10.2147/OPTH.S584717

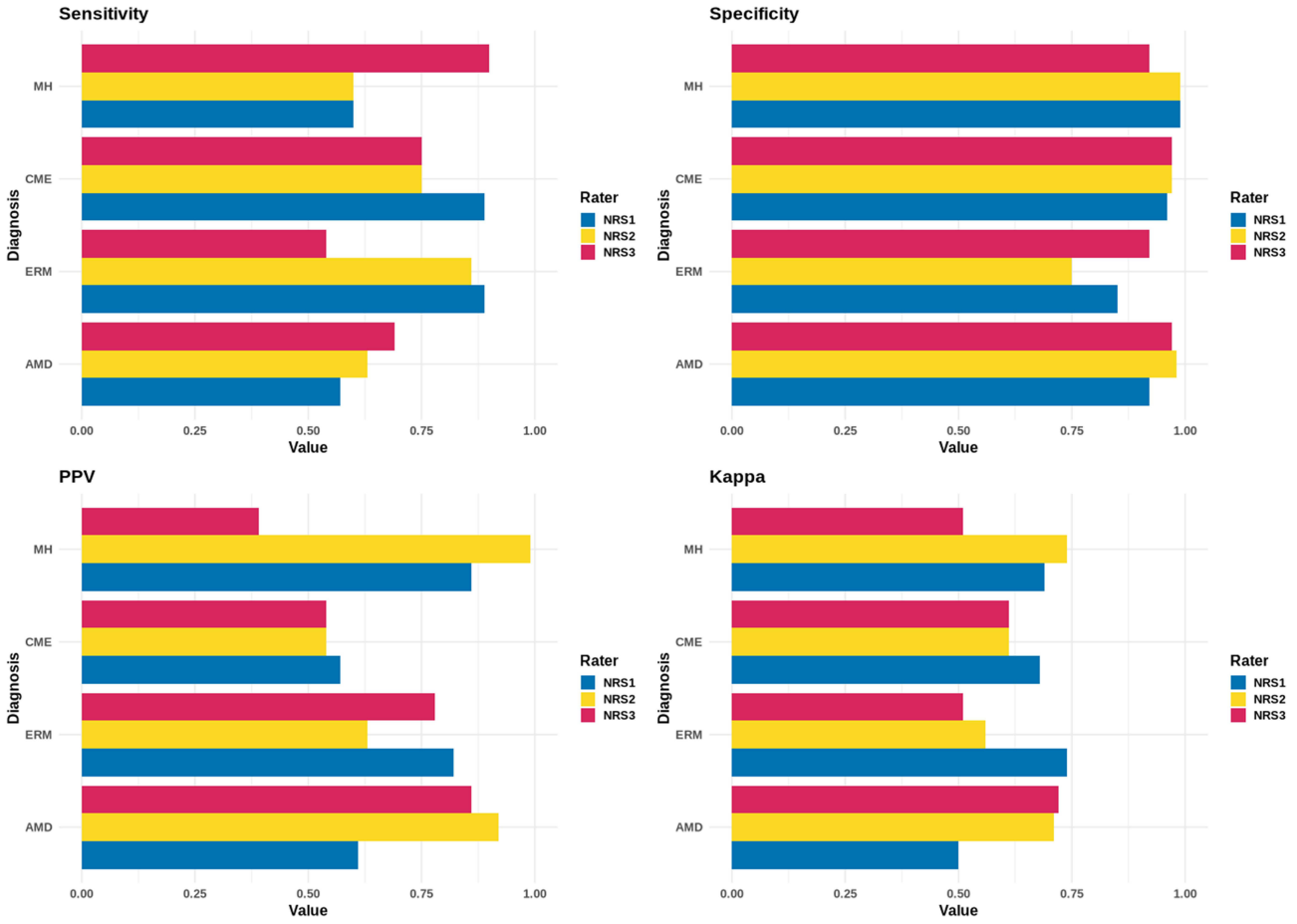

Figure 2 Diagnostic performance of non-retina specialists (NRS1–NRS3) compared with the gold-standard diagnosis established by retina specialists for the top four diagnostic categories.

Predictors of Discordance for Referral

Given that NRS2 had the lowest agreement with the gold standard for referrals, a logistic regression analysis was performed to identify predictors of discordance (Table 3) with NRS2. The only significant factor was visual acuity, where each 1-line worsening in vision was associated with a 34% lower odds of discordance in referral (OR 0.66, 95% CI 0.55–0.80, $p < 0.001$), indicating that poorer vision increased the likelihood of concordance with the gold-standard referral. Other factors, including age, gender, underlying retinal diagnosis (AMD, CME, ERM), self-reported confidence level, and interpretation time with PathFinder, were not significantly associated with referral discordance. The logistic regression model showed good fit and discrimination (Hosmer-Lemeshow p = 0.84, AUC = 0.80, McFadden's $R^2 = 0.24$, Somers' D = 0.61), with good calibration.

Discordance of Gold Standard Graded Normal and Non-Retina Specialists Graded Pathology

The most significant finding of discordance in diagnosis for *RS* graded normal eyes is the high frequency of epiretinal membrane, accounting for 40.0% of all false-positive diagnoses (Table 4). This is followed by PED at 20.0% and AMD at 13.3%. These results are yielded across thirteen eyes with thirty false-positive diagnoses made by the three PathFinder assisted NRS, that were truly normal according to the retina specialists (Figure 5).

https://doi.org/10.2147/OPTH.S584717

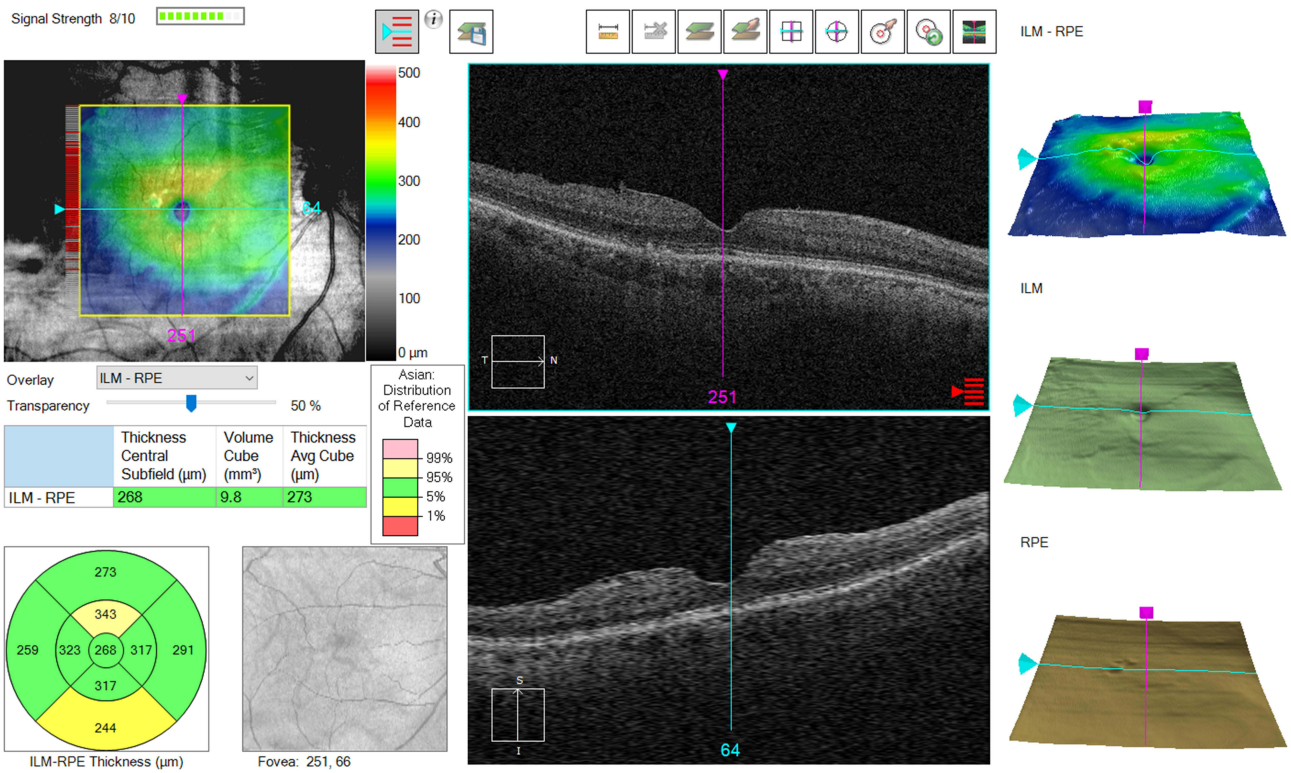

Figure 3 An OCT image with PathFinder evaluation of an ERM with pseudohole diagnosed as macular hole by NRS3 and referred but it was not in agreement to RS.

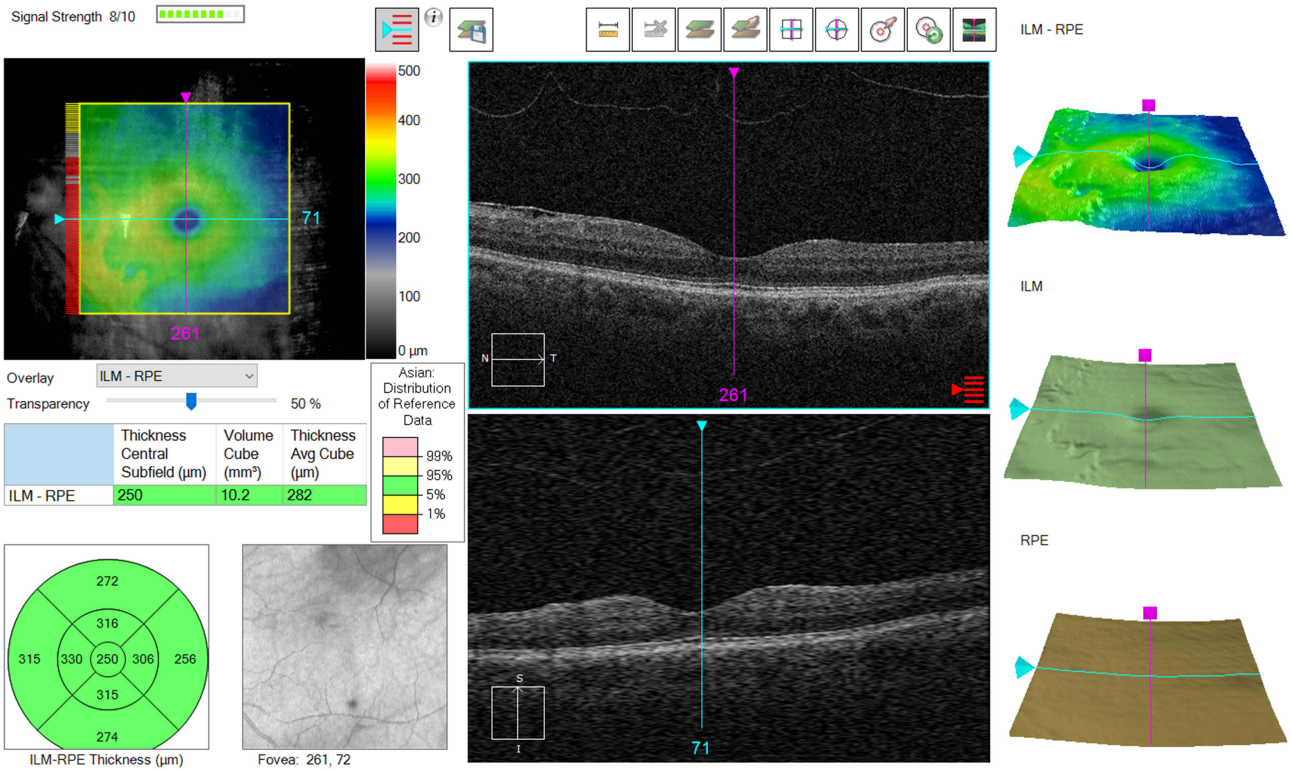

Figure 4 An OCT image of an Epiretinal Membrane with 20/20 vision with PathFinder evaluation that NRS diagnosed in agreement to *RS* but did not refer.

https://doi.org/10.2147/OPTH.S584717

Table 3 Logistic Regression for Predictors of Discordance Between the Referral Preference by NRS2 and the Majority Referral Patterns Expected by Retina Specialists

Term	Interval	OR	95% CI	p.value
Vision	0.1 logMAR increment	0.66	0.55–0.8	<0.001
Age	5 year increment	0.88	0.74–1.05	0.16
Men	Vs. Women	0.85	0.39–1.85	0.68
AMD	Vs. normal	3.49	0.8–15.33	0.1
CME	Vs. normal	0.32	0.05–2.26	0.25
ERM	Vs. Normal	1.02	0.34–3.06	0.97
Confidence	Vs. Very Confident	0.72	0.33–1.55	0.4
Time taken by NRS (PathFinder assisted)	Vs ≤10 sec	2.79	0.48–16.13	0.25

Table 4 Discordance of Gold Standard Graded Normal and Non-Retina Specialists Graded Pathology

NRS PathFinder-Assisted Pathology	Frequency	Percentage
ERM (Epi Retinal Membrane)	12	40.0%
PED (Pigment Epithelial Detachment)	6	20.0%
AMD (Age-Related Macular Degeneration)	4	13.3%
CME (Cystoid Macular Edema)	2	6.7%
Unknown	2	6.7%
PVD (Posterior Vitreous Detachment)	1	3.3%
Maculoschisis	1	3.3%
Lamellar Hole	1	3.3%
Myopic Maculopathy	1	3.3%

Discussion

In this study, we evaluated the diagnostic and referral performance of non-retina specialists using the PathFinder AI assistant, benchmarked against gold-standard diagnoses and referrals derived from retina specialists. Overall, retina specialists demonstrated moderate agreement with each other for both diagnosis and referral, though majority-rule analysis where at least two agreed, yielded very high agreement. Non-retina specialists, when aided by PathFinder, achieved substantial to moderate agreement with the gold-standard diagnosis, with performance varying across raters and disease categories. Importantly, specificity remained consistently high across conditions, while sensitivity and predictive values showed greater variability. Confidence in diagnosis was generally high and interpretation times were short, underscoring the usability of the software in real-world settings. However, referral agreement was notably weaker, particularly for two non-retina specialists. Further analysis suggested that NRS2, and likely NRS3, relied more on visual acuity than OCT characteristics when deciding on referrals. Since retina specialists in our study were masked to visual status, this difference in approach likely contributed to higher discordance rates. A key concern is that patients with preserved vision but with pathologies such as ERM or AMD could be under-referred, despite retina specialists considering referral appropriate. It is also noted that subtle and early signs of pathologies such as ERM, PED, and

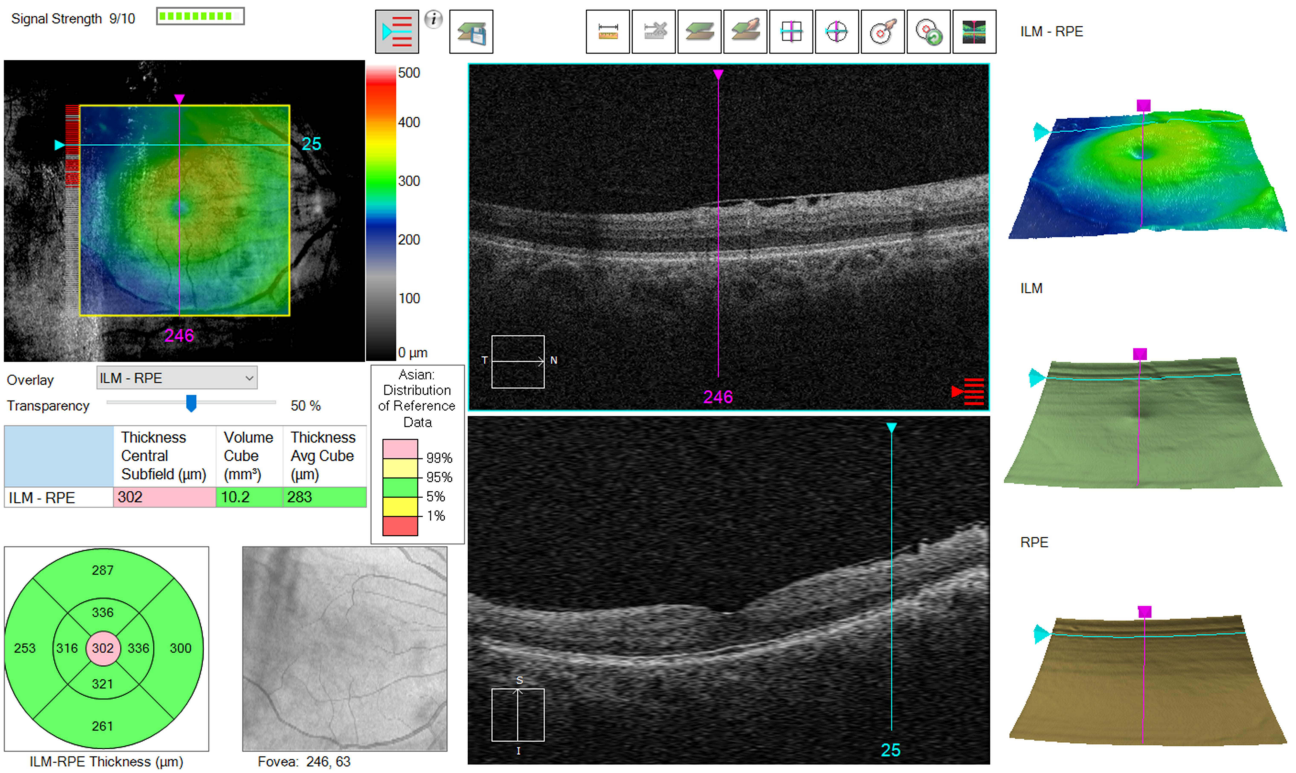

Figure 5 An OCT image with PathFinder evaluation of a case ERM that one NRS got right and two got wrong.

AMD, can be picked up by PathFinder. This is seen in the diagnosis discordance of non-retina specialists to those graded normal by gold standard. This is due to retina specialists who would consider these findings clinically insignificant.

Artificial-intelligence applications in ophthalmic imaging have evolved rapidly, particularly in OCT interpretation, where large volumes of data demand efficient and reproducible analysis. Early work by De Fauw et al[11] demonstrated that a deep-learning system could identify retinal diseases and generate referral recommendations directly from OCT volumes, matching the diagnostic performance of expert ophthalmologists. Similarly, Schlegl et al[12] developed a fully automated pipeline for fluid detection and quantification on OCT, capable of distinguishing intraretinal and subretinal fluid in neovascular AMD. Later studies refined these approaches, such as Cao et al,[13] who used deep-learning segmentation to monitor treatment response in DME, and Yim et al,[14] who built a model predicting future conversion from dry to neovascular AMD using baseline OCT features. Collectively, these investigations established that AI can interpret OCT images with diagnostic accuracy comparable to human graders while also enabling longitudinal disease tracking.

Most of these systems, however, were device-agnostic research models requiring external data processing and dedicated computational resources. They typically functioned offline or via cloud-based platforms and were not integrated into the OCT acquisition workflow. In contrast, the PathFinder AI is embedded directly within the Cirrus OCT console, performing real-time image-quality assessment, retinal-layer segmentation, and pathology detection during scan acquisition. This design eliminates the need for data export or external analysis, enabling immediate feedback to technicians and clinicians. The present study is, to our knowledge, among the first to evaluate such a commercially available, on-device OCT-AI system in a real-world clinical setting.

Our results show that non-retina specialists, when assisted by PathFinder, could diagnose common macular pathologies such as epiretinal membrane and cystoid macular edema with high accuracy and confidence. Specificity across observers remained consistently high, suggesting that the AI tool effectively reduces false positives and helps rule out normal scans. Sensitivity was lower for subtle or early manifestations of age-related macular degeneration and macular holes, indicating that AI assistance complements but does not replace clinical expertise. Importantly, the substantial

interobserver confidence observed in this study suggests that AI integration not only enhances diagnostic accuracy but also reinforces clinician assurance, an essential factor for widespread adoption. However, isolated deviations in diagnostic performance for certain conditions and users (for example, a higher false-positive rate for macular hole in one rater) warrant caution. Closer examination revealed that most of these cases represented pseudoholes secondary to ERM rather than true full-thickness macular holes, reflecting classification challenges within the vitreomacular interface spectrum rather than systematic AI error. This suggests that optimal use of AI assistance may benefit from additional training in OCT interpretation and effective use of AI-based decision-support tools.

The implications are particularly relevant for resource-constrained or high-volume clinical environments, where retina specialists may not be readily available. Integrating AI-enabled OCT devices at the point of care could allow comprehensive ophthalmologists, optometrists, or even trained technicians to detect pathology earlier and make more accurate referral decisions. Such an approach could transform routine imaging centers into front-line retinal screening hubs, improving accessibility to specialized care.

Previous studies of OCT-based AI systems have reported very high diagnostic performance. For example, in a community screening platform by Liu et al,[15] the OCT-AI model achieved a sensitivity of 96.6% (95% CI 91.8–98.7%) and specificity of 98.8% (95% CI 98.0–99.3%) for detecting urgent retinal cases, and 98.5% sensitivity (95% CI 96.5–99.4%) with 96.2% specificity (95% CI 94.6–97.3%) when both urgent and routine cases were included. Similarly, Bai et al[16] evaluated AI-assisted OCT for detection of 15 retinal disorders and found sensitivities ranging from 87.65% to 100% and specificities from 80.12% to 99.41%. In comparison, our study using the PathFinder AI assisted by non-retina specialists demonstrated similarly high specificity across conditions, but somewhat lower sensitivity for certain pathologies (for example, our sensitivity for AMD and macular hole was below those ranges). Several factors likely contribute to this difference in sensitivity: in our real-world clinical workflow, non-retina specialists were making diagnostic and referral decisions rather than the AI operating in isolation; the pathology spectrum included subtle and mixed presentations rather than screening-only cases; the PathFinder is device-embedded and optimized for the CIRRUS OCT environment rather than a retrospective algorithm trained on multiple devices; and non-retina specialists may rely differently on clinical information in conjunction with AI output, potentially affecting detection thresholds, particularly for referral decisions that are influenced by visual acuity, symptom burden, anticipated treatment benefit, local management capability, and cost considerations. This is also substantiated by the fact that even the three retina specialists agreed only moderately on the final diagnosis and referral decisions. These differences suggest that while PathFinder performs well in a practical setting with high specificity, careful attention is required to sensitivity limitations when deploying in non-specialist hands. Additionally, known methodological challenges such as the lack of large-image training datasets,[17] inconsistent reporting metrics, non-standardised imaging or post-processing protocols, and limited graphics processing unit capabilities for exploiting 3-dimensional features can be addressed to improve the performance of future iterations of the PathFinder AI.

Concerns regarding AI-induced over-diagnosis are increasingly relevant in clinical imaging. In the present study, however, false positive rates were low and PPV values were high across most disease categories, indicating that AI-assisted interpretations rarely labeled normal eyes incorrectly. These findings suggest that PathFinder-assisted OCT interpretation favours specificity over sensitivity, with missed pathology rather than over-calling representing the dominant response, and highlighting the need for human oversight. Importantly, this pattern underscores the role of human oversight, as non-retina specialists integrated clinical context, particularly visual acuity, when making referral decisions, rather than relying on AI outputs in isolation. Taken together, this pattern highlights both the strengths and the current boundaries of AI-assisted OCT interpretation in real-world clinical workflows.

This study has certain limitations. It was conducted at a single tertiary center with a limited sample size, which may restrict generalizability. The diagnostic distribution was heavily weighted toward ERM, reflecting real-world OCT practice with consecutive case inclusion; however, the relatively small number of other pathologies, particularly CME and macular hole, limits the precision with which PathFinder's diagnostic performance for these less frequent conditions can be assessed. The AI model evaluated was proprietary to the Cirrus OCT system and cannot be extrapolated to other devices or image formats. Furthermore, the retina specialists were

Furthermore, retina specialists were intentionally masked to clinical data to establish an OCT-based reference standard, whereas non-retina specialists had full access to these details, a design choice that may introduce contextual differences but reflects real-world AI-assisted decision-making. Finally, PathFinder currently identifies predefined macular pathologies and does not yet quantify lesion burden or disease activity, features that would further enhance clinical utility.

Future work should include multicenter studies with larger and more diverse populations, cross-platform validation to test device independence, and longitudinal designs evaluating the impact of AI-assisted triage on patient outcomes. Integrating predictive modules, such as AI-based disease progression forecasting, could further expand its utility from diagnostic support to comprehensive retinal management.

Conclusion

In summary, this study provides early clinical evidence that PathFinder AI-assisted OCT interpretation can enhance the diagnostic and referral capability of non-retina specialists, in selected clinical contexts. By embedding deep-learning algorithms directly within the OCT console, PathFinder bridges the gap between imaging acquisition and interpretation, offering a practical and scalable solution for extending retina-level clinical decision support to general ophthalmic practice. With further validation and improving iterations, such device-integrated AI systems may play a pivotal role in improving access to retinal care, especially in resource-limited settings.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Kapoor R, Whigham BT, Al-Aswad LA. Artificial intelligence and optical coherence tomography imaging. *Asia Pac J Ophthalmol*. 2019;8(2):187–194.
2. Lakshmi P, Veerapandu G, Gamini S, Singh MK. CNN classification of multi-scale ensemble OCT for macular image analysis. *Int. J. Electr. Electron. Res*. 2022;10(4):858–861. doi:10.37391/ijeer.100417
3. Abdi AS, Abdulazeez AM. A comprehensive review of deep learning in OCT image segmentation and classification. *Med. Nov. Technol. Devices*. 2025;28:100396. doi:10.1016/j.medntd.2025.100396
4. Nigjeh MK, Nigjeh MK. Umbaugh SE. Automated retinal disorders classification: leveraging digital image enhancement techniques and deep learning on OCT images", Proc. SPIE 13137, Applications of Digital Image Processing XLVII, 131370K. doi:10.1117/12.3028346. Accessed March 4, 2026.
5. Umair M, Ahmad J, Saidani O, Alshehri MS, Al Mazroa A, Hanif M. OculusNet: detection of retinal diseases using a tailored web-deployed neural network and saliency maps for explainable AI. *Front Med*. 2025;12:1596726. doi:10.3389/fmed.2025.1596726
6. Ying JN, Li H, Zhang YY, Li WD, Yi QY. Application and progress of artificial intelligence technology in the segmentation of hyperreflective foci in OCT images for ophthalmic disease research. *Int J Ophthalmol*. 2024;17(6):1138–1143. doi:10.18240/ijo.2024.06.20
7. Tao Z, Zhang W, Yao M, Zhong Y, Sun Y, Li XM. A joint model for macular edema analysis in optical coherence tomography images based on image enhancement and segmentation. *Biomed Res Int*. 2021;2021:6679556. doi:10.1155/2021/6679556
8. Elsharkawy M, Elrazzaz M, Ghazal M, Alhalabi M, Soliman A, Mahmoud A. Role of optical coherence tomography imaging in predicting progression of age-related macular disease: a survey. *Diagnostics*. 2021;11(12):2313. doi:10.3390/diagnostics11122313
9. Müller PL, Liefers B, Treis T, Rodrigues FG, Olvera-Barrios A, Paul B. Reliability of retinal pathology quantification in age-related macular degeneration: implications for clinical trials and machine learning applications. *Transl Vis Sci Technol*. 2021;10(3):4. doi:10.1167/tvst.10.3.4
10. Talcott KE, Valentim CCS, Perkins SW, et al. Automated Detection of Abnormal Optical Coherence Tomography B-Scans Using a Deep Learning Artificial Intelligence Neural Network Platform. *Int Ophthalmol Clin* 64;1.115–127.
11. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342–1350. doi:10.1038/s41591-018-0107-6
12. Schlegl T, Waldstein SM, Bogunovic H, Endstraßer F, Sadeghipour A, Philip AM. Fully automated detection and quantification of macular fluid in oct using deep learning. *Ophthalmology*. 2018;125(4):549–558. doi:10.1016/j.ophtha.2017.10.031
13. Cao J, You K, Jin K, Lou L, Wang Y, Chen M. Prediction of response to anti-vascular endothelial growth factor treatment in diabetic macular oedema using an optical coherence tomography-based machine learning method. *Acta Ophthalmologica*. 2021;99(1):e19–27. doi:10.1111/aos.14514
14. Yim J, Chopra R, Spitz T, Winkens J, Obika A, Kelly C. Predicting conversion to wet age-related macular degeneration using deep learning. *Nat Med*. 2020;26(6):892–899. doi:10.1038/s41591-020-0867-7
15. Liu X, Zhao C, Wang L, Wang G, Lv B, Lv C. Evaluation of an OCT-AI–based telemedicine platform for retinal disease screening and referral in a primary care setting. *Transl Vis Sci Technol*. 2022;11(3):4. doi:10.1167/tvst.11.3.4
16. Bai J, Wan Z, Li P, et al. Accuracy and feasibility with AI-assisted OCT in retinal disorder community screening. *Front Cell Dev Biol*. 2022;10:1053483.
17. Yanagihara RT, Lee CS, Ting DSW, Lee AY. Methodological challenges of deep learning in optical coherence tomography for retinal diseases: a review. *Transl Vis Sci Technol*. 2020;9(2):11. doi:10.1167/tvst.9.2.11