

Artificial Intelligence in Neuro-Ophthalmology for Optic Disc Pathologies and Neurodegenerative Disease

Abhimanyu S Ahuja^{1,*}, Alfredo A Paredes III^{2,*}, Mallory LS Eisel^{3,*}, Cole Miller⁴, Nina Truong³, Julie Falardeau¹

¹Department of Ophthalmology, Casey Eye Institute, Oregon Health and Science University, Portland, OR, USA; ²Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL, USA; ³College of Medicine, Florida State University, Tallahassee, FL, USA; ⁴Leonard M. Miller School of Medicine, University of Miami, Miami, FL, USA

*These authors contributed equally to this work

Correspondence: Abhimanyu S Ahuja; Julie Falardeau, Department of Ophthalmology, Casey Eye Institute, Oregon Health and Science University, 515 SW Campus Drive, Portland, OR, 97239, USA, Email ahuja@ohsu.edu; falardea@ohsu.edu

Abstract: Artificial intelligence (AI) is rapidly reshaping neuro-ophthalmic care by extracting clinically significant information from imaging, biomarkers, and patient-level clinical data. We review recent advances across neurodegenerative disease detection using retinal biomarkers, automated recognition of optic disc swelling and its mimics, glaucoma screening and quantification, and classification of hereditary optic neuropathies. Using fundus photography and optical coherence tomography (OCT), contemporary machine learning (ML) systems, including deep learning as well as other supervised learning models, report strong discrimination for papilledema versus pseudopapilledema, non-arteritic anterior ischemic optic neuropathy (NAION) against similar presenting entities, and glaucomatous damage including indirect estimation of retinal nerve fiber layer (RNFL) thickness. Early work also suggests that retinal features can aid detection of mild cognitive impairment (MCI) and major neurocognitive disease. However, despite promising results, most studies remain retrospective and single-center, while focusing on imaging-only, limiting generalizability and clinical interpretability. Therefore a variety of challenges related to dataset heterogeneity, overfitting, limited external validation, and the gap between high diagnostic accuracy and practical clinical utility remain unresolved. Future prospective, multicenter evaluations focusing on integrating multimodal clinical data through explainable AI systems are necessary to improve diagnostic consistency, shorten time to care, and expand access for underserved populations.

Keywords: artificial intelligence, neuro-ophthalmology, deep learning, machine learning, support vector machine, extreme learning machine

Introduction

Neuro-ophthalmic disorders, such as papilledema, non-arteritic anterior ischemic optic neuropathy (NAION), and hereditary optic neuropathies carry substantial risks of irreversible visual morbidity if diagnosis is delayed.¹ Glaucoma, while not a neuro-ophthalmologic disorder, is a common optic neuropathy which impacts millions of people worldwide.² These conditions often hinge on subtle, image-based findings and nuanced clinical context, yet access to subspecialty care is constrained as neuro-ophthalmology faces a persistent workforce shortage and long wait lists for appointment scheduling, limiting timely evaluation.¹

Expert diagnosis is critical in the field of neuro-ophthalmology. Diagnostic error rates preceding neuro-ophthalmology consultation can be substantial, and patients can be subjected to unnecessary tests and treatments, highlighting opportunities for tools that improve consistency and expedite appropriate referrals.³ In parallel, ophthalmology has already demonstrated that autonomous or assistive artificial intelligence (AI) systems can function in front-line settings as Food and Drug Administration (FDA)-approved diagnostic models for diabetic retinopathy are increasing in



availability across primary care clinics.⁴ While the integration of AI in the medical sphere may show promise in improving diagnostic accuracy, it is important to interpret these results with caution and the perspective that AI may serve as a complement to, rather than substitute for expert clinical judgement.

Several AI modalities are described in this paper, such as machine learning (ML), deep learning (DL), support vector machine (SVM), and extreme learning machine (ELM). ML refers to a subset of AI composed of algorithms that learn patterns from data to perform predictive tasks.⁵ DL is a subset of ML that specializes in analyzing unstructured data, such as raw text or images, and is capable of automatically differentiating data into meaningful categories.⁶ As a result, DL is particularly well suited for analyzing large volumes of data that may contain subtle patterns or distinctions that are difficult for human observers to detect.⁶ SVM models are a type of supervised machine learning algorithm that classify data by identifying patterns that best separate predefined groups.^{7,8} In medical imaging, SVMs are commonly used when relevant image features can be extracted and labeled, allowing the model to distinguish between disease and normal states based on these features.^{7,8} ELMs are similar to SVM, but are designed for rapid training and are often applied when computational efficiency is prioritized, such as in exploratory clinical studies or smaller datasets.^{7,8}

Applications of AI in neuro-ophthalmology could help detect subtle structural and microvascular signatures across diverse conditions, while expanding access to care by enabling earlier, more consistent detection of disease in both specialty and primary-care settings. This manuscript reviews emerging evidence on AI for neuro-ophthalmic disease detection and differentiation, analyzes AI’s clinical promise, and outlines practical next steps to accelerate safe, equitable adoption (Figure 1).

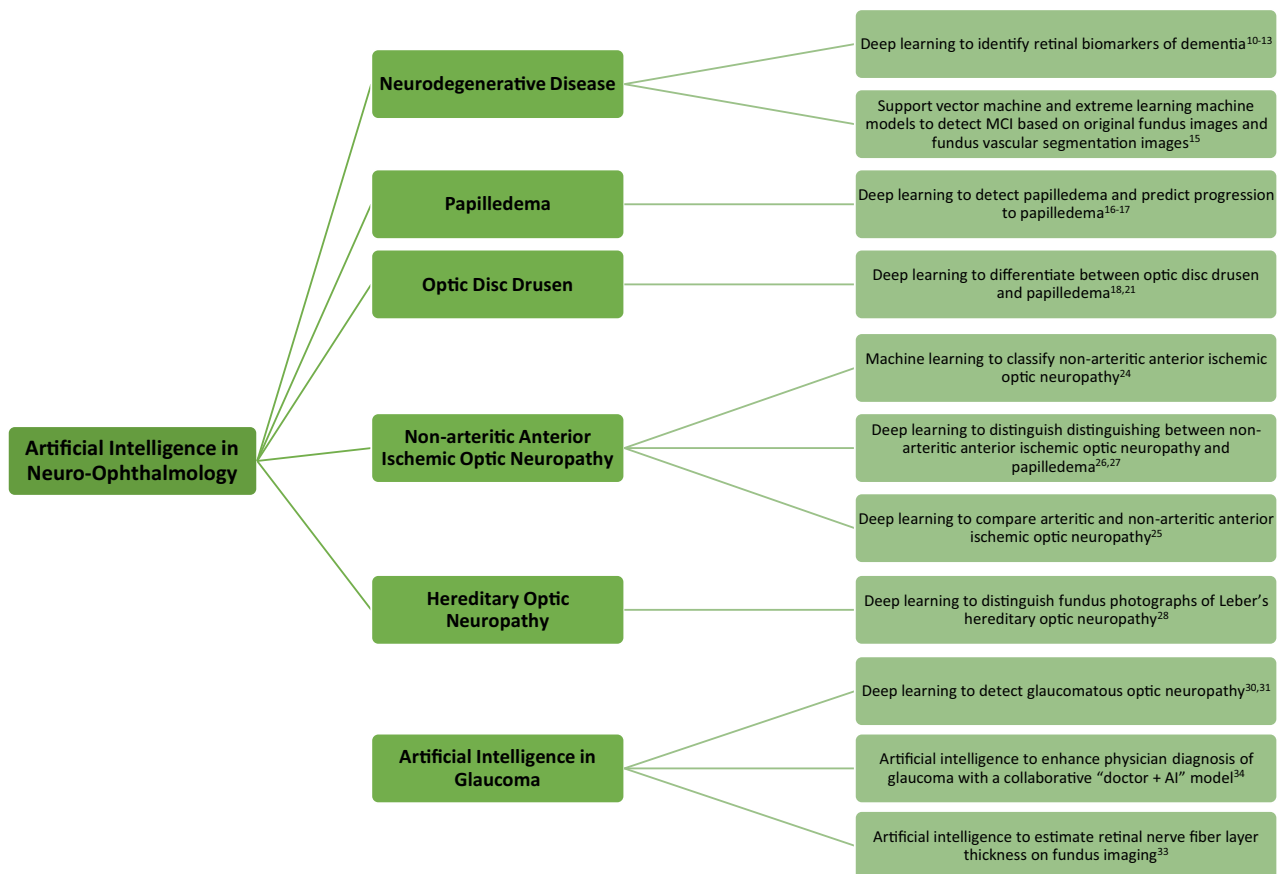


Figure 1 Artificial Intelligence in Neuro-Ophthalmology.

Methods

We utilized the scholarly literature database PubMed to identify relevant literature to compose our review. We conducted the literature search by accessing PubMed virtually from July 31, 2025 to September 29, 2025 using the keywords and phrases: “artificial intelligence in neuro-ophthalmology”, “machine learning”, “deep learning”, “optic disc”, “neurodegenerative disease”, “papilledema”, and “glaucoma” (Figure 2). Out of the identified primary data and review articles, we identified 32 articles to include in our review. We included articles published after 2014, thus articles published prior to 2014 were excluded from this review of the literature. The inclusion criteria we applied were relevance to the discussion of recent applications of AI in the field of neuro-ophthalmology and analysis of the utility and diagnostic accuracy of AI in disorders of the eye and nervous system. This selection process yielded articles related to various AI modalities including machine learning (ML), deep learning (DL), support vector machine (SVM), and extreme learning machine (ELM), as well as imaging modalities including optical coherence tomography (OCT), OCT angiography (OCTA), and color fundus photography (CFP).

Neurodegenerative Disease Detection

ML and DL, a subset of ML, methods have been applied in the context of retinal imaging-related biomarkers for neurodegenerative disease, including Alzheimer’s disease (AD), dementia, and mild cognitive impairment (MCI). Retinal microvascular changes have been observed in AD patients post-mortem,⁹ as well as in patients with AD and MCI.¹⁰ DL has been applied to identify retinal biomarkers of dementia using modalities including optical coherence tomography (OCT), OCT angiography (OCTA), and color fundus photography (CFP).¹¹ Eye-AD, described by Hao et al, is a DL model which uses OCTA images to detect early-onset AD and MCI.¹¹ The multicenter study utilized 1,671 patients (5,751 OCTA images), and Eye-AD exhibited clinical decision-making patterns and successful identification of patients with early-onset AD and MCI compared to healthy controls.¹¹ For early-onset AD detection, Eye-AD had an area under the curve (AUC) of 0.9007, accuracy of 0.8176, and precision of 0.8429.¹¹ For MCI detection, Eye-AD had an AUC of 0.8630, accuracy of 0.8487, and precision of 0.8506.¹¹ Cheung et al (2022) performed a retrospective, multicenter case-control study utilizing 12,949 retinal images from 648 patients with AD and 3,240 patients without AD to develop a DL model to detect AD.¹² The DL model utilized a EfficientNet-b2 backbone then further analyzed optic nerve head-centered and macula-centered retinal images from each eye using an adaptive feature fusion technique prior to determining a result of AD-dementia or no dementia for each patient.¹² The DL model exhibited 83.6% accuracy,

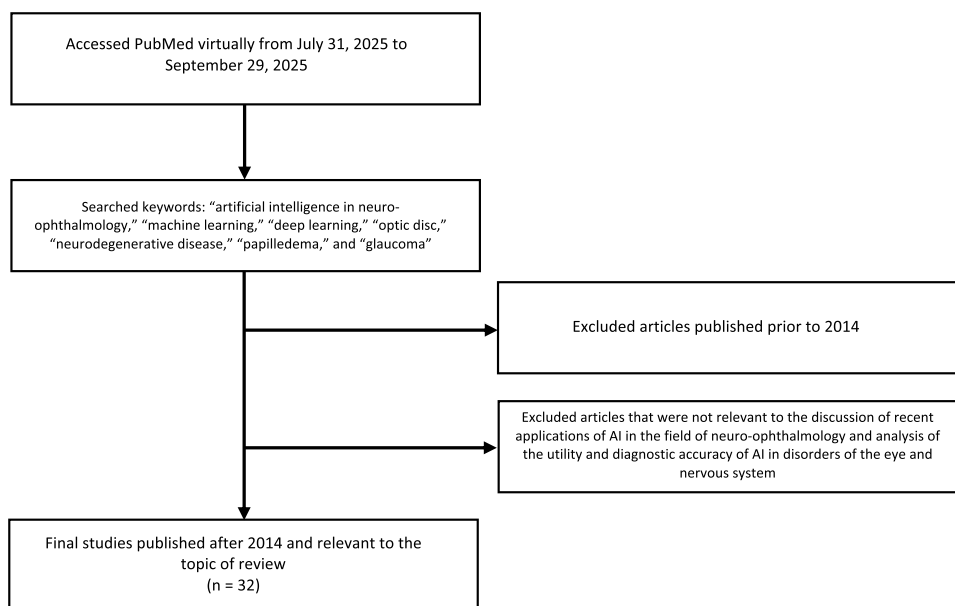


Figure 2 Article Selection Criteria.

93.2% sensitivity, 82% specificity, and area under the receiver operating characteristic curve (AUROC) of 0.93.¹² Furthermore, DL models have used segmented OCTA images to analyze retinal microvasculature at multiple layers of the retina and foveal avascular zone in patients with AD, MCI, and no dementia.¹⁰ Xie et al (2024) used DL and logistic regression in retinal imaging of 158 patients (55 with AD, 41 with MCI, and 62 healthy controls) to assess retinal microvascular parameters in each group of patients.¹⁰ The authors found that vessel area, length densities, and number of vascular bifurcations in inner vascular complexes were significantly decreased in AD patients compared to healthy controls.¹⁰ Decreased vascular area, length density, vascular fractal dimension, number of vascular bifurcations in the superficial and inner vascular complexes, larger vascular tortuosity in the inner vascular complexes, and increased foveal avascular zone roundness were observed in patients with MCI compared to healthy controls.¹⁰ Furthermore, the AlzEye retinal imaging dataset includes 353,157 patients and 6,261,931 retinal images gathered from Moorfields Eye Hospital NHS Foundation Trust and aims to characterize the link between retinal biomarkers and neurodegenerative and neurovascular diseases.¹³ As the data cohort is still under development, the authors highlight the goal for AlzEye to be utilized for the development of DL models.¹³ It is important to note that the use of AUC as a measure of diagnostic accuracy is limited by a potential decreased generalizability to novel patient populations in a real-world clinical setting due to reported AUC values reflecting overfitting to retrospective training datasets.¹⁴

SVM and ELM models represent supervised ML models because they utilize known labeled data to train and make decisions on future data.¹⁵ SVM and ELM models were tested in a clinical trial [ChiCTR1900027404] on their ability to detect MCI based on original fundus images (using SVM) and fundus vascular segmentation images (using ELM) from 86 patients between ages 18 and 80 (mean 46.37 ± 1.79 years).¹⁵ The patients were divided into normal ($n = 38$), MCI ($n = 26$), and dementia ($n = 22$) groups based on cognitive assessment and ICD-11 and DSM-5 MCI diagnostic criteria.¹⁵ The authors report that the SVM model trained on original fundus images showed significantly greater predictive efficacy ($p = 0.0012$) than the ELM model.¹⁵ While not necessarily validated as a standalone diagnostic tool, the use of AI in neuro-ophthalmology presents the opportunity to screen elderly patients for early identification of neurodegenerative disease biomarkers through retinal imaging.

Detecting Optic Disc Abnormalities

Papilledema

Papilledema is defined as optic disc edema, typically bilateral, secondary to intracranial hypertension.¹⁶ Failure to detect papilledema may result in permanent vision loss and delay the recognition of serious underlying neurologic conditions, including intracranial mass lesions, hydrocephalus, and venous sinus thrombosis.¹⁶ Thus, researchers have developed AI models to aid in the diagnosis of papilledema. One group conducted a retrospective study using a DL model to analyze digital color ocular fundus photographs collected by the Brain and Optic Nerve with Artificial Intelligence (BONSAI) consortium.¹⁶ The model was trained and validated using 14,341 fundus photographs and was externally tested on 1,505 photographs.¹⁶ Across testing sets, AUC ranged from 0.93 to 0.98 in the ability to discriminate optic discs with papilledema from all other optic discs.¹⁶ The DL model was found to have a sensitivity of 96.4% and specificity of 84.7% for the detection of papilledema, suggesting that clinicians may utilize AI in combination with fundus photography to diagnose papilledema and differentiate it from other ophthalmologic conditions.¹⁶

In addition to diagnosis of papilledema at clinical presentation, AI models have been tested in their ability to detect early optic nerve changes on OCT that might predict the future development of papilledema.¹⁷ The ability to do so would allow for early detection and better risk stratification for future intervention.¹⁷ One such study utilized the publicly available AI model Visual Geometry Group-19, testing its ability to predict progression to papilledema from early OCT images.¹⁷ Ninety-three subjects with both an official diagnosis of papilledema and a normal OCT prior to the diagnosis were included in the experimental group, and 254 healthy subjects were included in the control group for the AI model to identify.¹⁷ The model was pretrained to identify abnormalities in retinal nerve fiber layer (RNFL) thickness maps, ganglion cell thickness maps, ILM-RPE thickness maps, and an extracted vertical tomogram.¹⁷ When trained to analyze the RNFL thickness map, the model achieved an area under the precision-recall curve (AUPRC) of 0.826 ± 0.033 with the initial study population.¹⁷ A subsequent trial was performed with a removal of 35 participants from the experimental

group to equalize age and gender populations between the control and experimental groups and an AUPRC of 0.713 ± 0.040 was achieved.¹⁷ This reduction in performance following demographic matching highlights the sensitivity of AI models to cohort composition and emphasizes the importance of controlling for confounding variables during model development and validation.¹⁷ This reduction displays how the model may perform differently in a real world setting.¹⁷ Nonetheless, these findings support the potential role of AI-assisted OCT analysis as a supplementary tool for early detection, provided its limitations are explicitly acknowledged.¹⁷

Optic Disc Drusen

Pseudopapilledema describes a group of optic disc abnormalities that mimic papilledema.¹⁸ Distinguishing pseudopapilledema from papilledema can be challenging, resulting in potentially avoidable diagnostic investigations for increased intracranial pressure, including costly neuroimaging studies and invasive procedures such as lumbar puncture.¹⁹ Optic disc drusen (ODD) is a common cause of pseudopapilledema, and it is present in 1.0–2.0% of the general population.¹⁸ ODD is the formation of acellular calcium, mucopolysaccharide, and amino and nucleic acid deposits in the prelaminar optic nerve head.¹⁹ It leads to visual field deficits in up to 87% of cases.¹⁹ The use of AI technology in concurrence with optical imaging has improved clinicians' ability to differentiate ODD from papilledema while using less invasive and costly techniques, and in doing so, has made the diagnosis of ODD more efficient.¹⁹

One retrospective study assessed the ability of a DL system to differentiate between ODD and papilledema using standard color ocular fundus photographs.¹⁸ In this study, the DL system was trained on over 4,087 fundus photographs from 1,959 patients from the international BONSAI consortium and externally validated on 421 independent images from 221 patients.¹⁸ Diagnoses were confirmed with OCT, ultrasound, autofluorescence, or elevated intracranial pressure, ensuring strong reliability.¹⁸ The model achieved an AUC of 0.97, with 90.5% overall accuracy in distinguishing ODD from papilledema.¹⁸ Performance was strongest when comparing visible ODD with severe papilledema (AUC 0.99, 96.3% accuracy) and remained high in clinically challenging cases of buried ODD versus mild papilledema (AUC 0.93, 84.2% accuracy).¹⁸ Despite these strong results, differentiating buried ODD from early papilledema is a clinically nuanced task that frequently relies on multimodal imaging, longitudinal follow-up, and integration of clinical context beyond fundus appearance alone.²⁰ Image-based DL classifiers, while effective in constrained classification settings, may inadequately capture this broader neuro-ophthalmic reasoning process, and near-ceiling AUC values should therefore be interpreted cautiously.¹⁸

Another study developed a DL algorithm that identifies structures on OCT scans and uses this information to distinguish between ODD, papilledema, and healthy controls.²¹ This retrospective study included 241 total patients and classification of the 256 images taken showed AUC of 0.99 ± 0.001 for ODD detection, 0.98 ± 0.01 for the detection of healthy controls, and 0.99 ± 0.005 for the detection of papilledema.²¹ These results highlight the promise of AI applied to both standard fundus photographs and OCT scans as an accessible, non-invasive, and cost-effective tool to support clinicians in differentiating ODD from papilledema.^{18,21} However, their clinical adoption requires interpretability and transparency, particularly in understanding which structural features drive model predictions.²²

Non-Arteritic Anterior Ischemic Optic Neuropathy

Non-arteritic anterior ischemic optic neuropathy (NAION) is a common cause of sudden unilateral vision loss in individuals typically over the age of 50.²³ Current literature demonstrates the potential of AI to improve diagnostic precision and the distinguishment of NAION from similarly presenting conditions, such as optic neuritis (ON), papilledema, and arteritic anterior ischemic optic neuropathy (AAION).^{24–26}

In a retrospective observational study, Jalili et al (2024) evaluated ML on vessel density features of peripapillary OCTA scans to classify healthy, NAION, and ON eyes.²⁴ Utilizing SVM, random forest, and Gaussian Naive Bayes models, the classifiers achieved AUC of 1.0 and accuracy of 100% in distinguishing between the three groups.²⁴ Although notable, perfect performance in small retrospective datasets may reflect limited generalizability to real-world clinical applications where imaging quality, comorbid disease, and atypical presentations are more variable. Nonetheless, these findings demonstrate how utilizing AI through automated vessel density-based ML may assist in distinguishing between NAION and ON.²⁴

Szanto et al (2025) similarly studied the use of DL in distinguishing NAION and papilledema.^{26,27} In a development and validation study using fundus photographs, their DL model achieved over 96% accuracy in internal testing and 93% accuracy in external validation, with AUC values of 0.98–0.99 and F1 scores above 0.90.²⁷ Szanto et al (2025) also conducted a retrospective review evaluating DL models using unsegmented 3D OCT volumes.²⁶ These DL models demonstrated similar performance, with nearly 95% accuracy internally and 90% accuracy on external validation, supported by high AUC (~0.98) and F1 scores (~0.89–0.95).²⁶ Together, these results demonstrate that DL models applied to fundus photos and OCT scans may assist clinicians in differentiating between NAION and papilledema.^{26,27}

Gungor et al (2024) applied the use of DL to the distinction between AAION and NAION in a multicenter, international cohort study.²⁵ This study included 961 color fundus images from 802 patients and demonstrated strong external validity.²⁵ The DL model achieved an accuracy of 92.6% while neuro-ophthalmologists achieved accuracies ranging from 74–82%.²⁵ These findings emphasize the potential value of DL as a tool to aid neuro-ophthalmologists in rapidly and accurately distinguishing between AAION and NAION, which is critical considering the urgent need for corticosteroid treatment in giant cell arteritis-related AAION to prevent bilateral vision loss.²⁵

Taken together, these studies highlight the clinical utility of AI in neuro-ophthalmology, demonstrating robust performance in differentiating NAION from other causes of optic disc swelling including ON, papilledema, and AAION in a timely manner. However, future research should investigate the incorporation of clinical variables (eg., cardiovascular risk factors, laboratory markers) and modeling approaches that also simulate disease evolution and treatment response over time to better reflect real-world neuro-ophthalmic decision making in NAION management.

Hereditary Optic Neuropathy

Leber's hereditary optic neuropathy (LHON) is a rare mitochondrial genetic disorder characterized by subacute, bilateral sequential or simultaneous vision loss.²⁸ Accurate diagnosis can be challenging as its characteristic findings including hyperemic, pseudo-edematous optic disc and peripapillary telangiectasias may be subtle or mimic other conditions (eg., optic neuritis) that require different management.²⁸

In a recent retrospective study, Lee et al (2024) developed a DL model to distinguish fundus photographs of LHON, ON, and normal eyes.²⁸ The dataset included 30 genetically confirmed LHON eyes, 30 ON eyes, and 120 normal eyes.²⁸ The DL model distinguished between LHON, ON, and normal eyes with high AUROC values of 0.988, 0.990, and 1, respectively, and an overall accuracy of 0.93.²⁸ It is important to consider that these performance estimates were derived from a small dataset, raising concerns for potential overfitting and limited generalizability, especially given the rarity and heterogeneity of LHON in clinical practice.

Reported precision, recall, and F1 scores were 0.8 when distinguishing LHON from other conditions, ON from other conditions, and between LHON and ON, which highlights the potential of DL systems as rapid, cost-effective, and non-invasive diagnostic adjuncts.²⁸ This would be especially valuable in rare hereditary optic neuropathies where traditional tests such as genetic analysis, are often expensive and less accessible. As these metrics are derived from image-based classification and do not incorporate longitudinal data, future research incorporating LHON evolution and treatment responses over time would be insightful.

Other Optic Nerve Pathology

While glaucoma is not traditionally categorized as a neuro-ophthalmic disorder, this chronic optic neuropathy carries a substantial and growing global burden.² As the leading cause of irreversible blindness worldwide, it underscores the critical need for scalable strategies to facilitate earlier detection and triage.^{2,29} The disease is marked by characteristic damage to the optic nerve and RNFL, which results in gradual and often asymptomatic vision loss until advanced stages.²⁹ While elevated intraocular pressure is the only modifiable risk factor identified to date, glaucoma can also occur at statistically normal pressures, reflecting the multifactorial nature of its pathogenesis.²⁹ Despite advances in screening and therapy, glaucoma is frequently underdiagnosed, and many patients suffer significant, irreversible vision loss before detection.²⁹ Researchers have evaluated the ability of AI models to diagnose glaucoma using both color fundus photographs and OCT images.^{30,31} Li et al (2018) trained a DL model to detect glaucomatous optic neuropathy on over 31,745 colored fundus photographs and later tested it on a remaining 8000 images selected randomly.³⁰ The model

demonstrated an AUC of 0.986 for referable glaucomatous optic neuropathy, sensitivity of 95.6%, and specificity of 92.0%, surpassing the accuracy of many human graders.³⁰ These findings suggest that AI-based assessment of monoscopic fundus photographs could provide an effective, scalable, and affordable approach to glaucoma screening, particularly in underserved populations.³⁰ Asaoka et al (2019) built a transfer-learning DL model using OCT macular RNFL and ganglion cell complex grids, pre-trained on RS-3000 data and fine-tuned on Topcon OCT-1000/2000 data, to detect early glaucoma.³¹ In an independent test set, it achieved an AUROC of 93.7% with 82.5% sensitivity and 93.9% specificity.³¹ This finding displays that AI models may also be effective in diagnosing glaucoma using OCT imaging.³¹ Although these studies consistently report high diagnostic accuracy, glaucoma exemplifies the limitations of image-centric AI approaches, as disease progression and management depend on longitudinal trends, intraocular pressure dynamics, functional testing, and patient-specific risk factors that are not fully captured by static image classifiers.³²

Another group of researchers tested the ability of AI to estimate RNFL thickness on fundus imaging in order to quantify neural damage in glaucoma.³³ They then compared AI RNFL thickness estimates to RNFL thickness measured directly by spectral-domain optical coherence tomography (SDOCT).³³ Successful ability to determine RNFL thickness using fundus imaging would allow for assessment of neural damage using fundal photographs, and thus faster and cheaper evaluation than SDOCT.³³ The AI model was trained using 26,528 pairs of disc photos and SDOCT scans from 1,849 eyes of 958 subjects and subsequently tested on 6,292 pairs of disc photos and SDOCTs from 463 eyes of 240 subjects.³³ Researchers found that AI predictions of RNFL obtained from optic disc photographs were highly correlated to actual RNFL thickness measurements using SDOCT.³³ Correlation between the predicted and the observed RNFL thickness values was 0.832.³³ The predictions performed well to discriminate eyes with glaucomatous visual field loss from healthy eyes, as AI was able to accurately quantify ocular nerve damage through RNFL estimations.³³ This work represents a shift toward more quantitative and potentially interpretable AI outputs that align imaging biomarkers with clinically meaningful parameters.³³

Gong et al (2022) further studied the ability of AI to enhance physician diagnosis of glaucoma with a collaborative “doctor + AI” model.³⁴ Using 1,000 fundus images, the authors demonstrated that doctors working with AI showed significantly improved diagnostic accuracy compared to either working independently of each other.³⁴ From the first round to the second round, doctor A improved in diagnostic accuracy from 86% without AI assistance to 92.5% with AI assistance, doctor B improved from 83.5% to 93.5%, doctor C from 93% to 95.5%, and doctor D from 84% to 95.5%.³⁴ The findings suggest that this collaborative model enhances diagnostic precision, supports clinical decision-making, and may be particularly valuable in glaucoma screening settings.³⁴

DL systems have demonstrated diagnostic accuracy comparable to that of neuro-ophthalmologists, highlighting their potential as valuable adjuncts in the detection of optic disc abnormalities. Brousse et al (2020) evaluated the performance of a DL system, trained on 14,341 fundus photographs from 19 international centers to that of two neuro-ophthalmologists (referred to as Expert 1 and 2).³⁵ Testing was conducted using 400 fundus photographs of normal optic discs, 201 depicting papilledema, and 199 showing other optic disc abnormalities.³⁵ The DL system showed an overall accuracy of 84.7%, which was comparable to Expert 1 (84.4%) and higher than Expert 2 (80.1%).³⁵ The DL system also demonstrated strong discrimination with AUC values of 0.97 for normal discs, 0.96 for papilledema, and 0.89 for other optic disc abnormalities.³⁵ The DL system’s accuracy, sensitivity, and specificity were similar to or exceeded those of the neuro-ophthalmologists.³⁵ However, a key limitation in the DL system was the absence of visual field data, OCT, and clinical history, all of which are essential in comprehensive neuro-ophthalmic decision making.³⁵ Furthermore, future studies should compare DL systems with a larger, more diverse group of neuro-ophthalmologists to better contextualize comparative diagnostic performance. Despite these limitations, these findings suggest that DL systems can enhance diagnostic accuracy and streamline clinical workflows, such as in triaging and identifying optic disc abnormalities for further neuro-ophthalmological assessment.

Discussion

AI has shown substantial promise in neuro-ophthalmology, achieving diagnostic accuracy comparable to, or exceeding that of human experts across a range of conditions. From identifying retinal biomarkers of AD to detecting optic nerve pathologies such as papilledema, glaucoma, and NAION, AI models have demonstrated consistent reliability. These

applications underscore the potential of AI as a powerful adjunct to clinical practice, with the potential to improve early detection, reduce diagnostic uncertainty, limit unnecessary and costly investigations, and ultimately enhance patient outcomes. However, when interpreting comparisons between AI models and human diagnostic accuracy, it is important to consider the context in which these evaluations were performed. In many studies, AI systems and human graders were assessed using the same restricted input modalities, such as fundus photographs or OCT images, without access to clinical history, visual field testing, or ancillary investigations. Under these controlled conditions, AI performance was often compared to that of ophthalmologists or neuro-ophthalmologists performing image-based classification rather than comprehensive clinical diagnosis. Accordingly, reported performance metrics reflect the ability of AI to interpret specific imaging features and should not be construed as direct replacements for full clinical decision-making.

Importantly, AI has demonstrated utility in differentiating between conditions with overlapping presentations, such as papilledema and optic disc drusen,^{18,21} or NAION and optic neuritis.^{24–27} In glaucoma, DL systems not only detect disease earlier than many human graders but can also estimate RNFL thickness from fundus photographs, potentially reducing dependence on costly OCT imaging.^{30,31,33} Similarly, in rare conditions such as LHON, AI tools have achieved high classification accuracy, offering a rapid, cost-effective alternative when genetic testing is less accessible.²⁸

Beyond diagnostic accuracy, AI has the potential to expand access to care. Its ability to analyze standard fundus photographs and OCT scans makes it theoretically feasible for deployment in primary care, community clinics, and telemedicine platforms, a process already initiated for certain retinal diseases.³⁶ This could significantly improve screening in underserved populations where access to neuro-ophthalmologists and advanced imaging technologies is limited. In addition, the use of computerized tools in conjunction with AI could provide an objective layer of evaluation that complements traditional clinical assessments, which is particularly valuable in diseases such as AD where early diagnosis often relies on more subjective cognitive testing.³⁷ These advances highlight the broader purpose of AI in neuro-ophthalmology, to enhance diagnostic precision while democratizing access to high-quality diagnostic care.

Despite the promising results across these studies, several limitations warrant consideration. Most investigations to date are retrospective, and while some incorporate external validation datasets, few have undergone prospective clinical trials. Dataset heterogeneity also poses a challenge as many models are trained on images from single institutions, raising concerns about generalizability across diverse populations and imaging modalities.^{28,33} Additionally, AI models often rely solely on imaging data without incorporating multimodal clinical information such as systemic history, visual field testing, or laboratory findings, which remain integral to neuro-ophthalmic decision-making.³⁵ More broadly, issues of patient privacy and the ethical implementation of AI systems at a community or population level must be carefully addressed before widespread clinical deployment.³⁸ Finally, a substantial proportion of studies reported diagnostic performance primarily using AUC, a metric with recognized limitations.¹⁴ In some cases, very high AUC values approaching 1 may reflect a model that is highly optimized to the training dataset rather than one that generalizes reliably across diverse patient populations.¹⁴ As a result, models reporting strong AUC performance may still demonstrate reduced effectiveness when evaluated on external datasets with differing demographic or disease characteristics.¹⁴

Future Directions

A major challenge in the adoption of AI for neuro-ophthalmology involves DL models achieving high diagnostic accuracy but offering little transparency into how predictions are made, sometimes referred to as the “black box problem”.³⁹ This opacity can undermine clinician trust, limit accountability, and create barriers to patient acceptance, particularly when treatment decisions hinge on understanding the rationale behind a diagnosis.³⁹ The importance of a model’s ability to explain its results is amplified in ophthalmology, where subtle imaging features, such as early optic disc swelling or nerve fiber layer changes, must be clearly identified for clinical decision-making.⁴⁰ To address this, researchers have begun developing explainable AI (XAI) approaches that make model predictions more interpretable without sacrificing performance.³⁹ Examples include generating heatmaps, which highlight regions of fundus or OCT images most influential to a model’s decision.³⁹ By creating models capable of providing interpretable reasoning, legal and ethical barriers to the implementation of AI could be mitigated, facilitating safer integration into clinical practice.

Despite the dominance of image-based DL classifiers in neuro-ophthalmic AI research, alternative computational paradigms warrant greater attention.⁴¹ Dynamic modeling approaches such as emerging Digital Twin frameworks aim to

simulate disease evolution and treatment response over time rather than generating static diagnostic labels.⁴¹ These models hold promise for conditions like papilledema and glaucoma, where temporal dynamics and individualized trajectories are central to management, but remain largely unexplored in current neuro-ophthalmic AI literature.⁴² Medical digital twin frameworks offer a way to extend neuro-ophthalmic AI beyond diagnosis toward simulation of disease evolution and treatment response.⁴² Because digital twins are continuously updated as new data is gathered they may ultimately enable individualized forecasting.⁴² Existing studies using DL algorithms to analyze ophthalmic diagnostic imaging are promising; however, external validity may be limited and synchronized patient-specific modeling such as Digital Twins frameworks provide a coherent roadmap for moving neuro-ophthalmic AI from snapshot classification toward personalized simulation that more closely matches clinical decision-making.^{41,42}

Future research in neuro-ophthalmic AI should prioritize large-scale, prospective, multicenter clinical trials, such as the AlzEye project,¹³ to address the limitations of retrospective, single-institution studies and to establish real-world generalizability. Another key direction is the integration of multimodal data, moving beyond reliance on fundus photographs or OCT alone to incorporate OCTA, visual fields, systemic risk factors, and genetic information, an approach already highlighted in glaucoma research.³³ Finally, expanding deployment in primary care and telemedicine could extend the reach of neuro-ophthalmic diagnostics, as demonstrated by AI assisted OCT platforms piloted for retinal diseases.³⁶ Applying similar models to neuro-ophthalmology could enable earlier detection and triage in underserved settings.

Conclusion

In closing, AI is emerging as a transformative tool for the diagnosis of various neuro-ophthalmologic diseases. By leveraging retinal and optic nerve imaging, AI-based models have shown the potential to detect pathology and differentiate between clinically overlapping presentations. AI may enhance diagnostic consistency through standardization of diagnostic criteria, while serving as an adjunct to clinical judgement. Beyond accuracy, the incorporation of AI may broaden access to care by enabling screening in primary care and telemedicine settings where specialist resources are limited.

Despite this progress, it is important to note that many studies comparing AI with neuro-ophthalmologists report the AUC as a primary outcome, a metric that may reflect strong discrimination within the training dataset but does not necessarily indicate reliable diagnostic performance across diverse patient populations. In addition, advancing clinical utility will require a shift from single-timepoint image analysis toward approaches that account for patient-level variability and longitudinal change, better aligning AI outputs with the way true neuro-ophthalmic disease evolves over time. Further, the path toward clinical adoption requires continued validation and transparency. Prospective multicenter trials, multimodal data integration, and XAI frameworks will be essential to ensure generalizability, interpretability, and clinician trust. As these systems evolve, AI has the potential not only to augment neuro-ophthalmic decision-making but also to redefine diagnostic efficiency and accessibility in the field.

Disclosure

Abhimanyu S Ahuja, Alfredo A Paredes III, and Mallory LS Eisel are co-first authors for this study. The authors report no conflicts of interest in this work.

References

1. Liu YA, Ko MW, Moss HE. Telemedicine for neuro-ophthalmology: challenges and opportunities. *Curr Opin Neurol*. 2021;34(1):61–66. doi:10.1097/WCO.0000000000000880
2. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081–2090. doi:10.1016/j.ophtha.2014.05.013
3. Stunkel L, Newman NJ, Biousse V. Diagnostic error and neuro-ophthalmology. *Curr Opin Neurol*. 2019;32(1):62–67. doi:10.1097/WCO.0000000000000635
4. Rajesh AE, Davidson OQ, Lee CS, Lee AY. Artificial intelligence and diabetic retinopathy: AI framework, prospective studies, head-to-head validation, and cost-effectiveness. *Diabetes Care*. 2023;46(10):1728–1739. doi:10.2337/dci23-0032
5. Barragán-Montero A, Javaid U, Valdés G, et al. Artificial intelligence and machine learning for medical imaging: a technology review. *Phys Med*. 2021;83:242–256. doi:10.1016/j.ejmp.2021.04.016

6. Iyortsuun NK, Kim SH, Jhon M, Yang HJ, Pant S. A review of machine learning and deep learning approaches on mental health diagnosis. *Healthcare*. 2023;11(3). doi:10.3390/healthcare11030285
7. Huang GB, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern B Cybern*. 2012;42(2):513–529. doi:10.1109/TSMCB.2011.2168604
8. Liu X, Gao C, Li P. A comparative analysis of support vector machines and extreme learning machines. *Neural Netw*. 2012;33:58–66. doi:10.1016/j.neunet.2012.04.002
9. Koronyo Y, Rentsendorj A, Mirzaei N, et al. Retinal pathological features and proteome signatures of Alzheimer's disease. *Acta Neuropathol*. 2023;145(4):409–438. doi:10.1007/s00401-023-02548-2
10. Xie J, Yi Q, Wu Y, et al. Deep segmentation of OCTA for evaluation and association of changes of retinal microvasculature with Alzheimer's disease and mild cognitive impairment. *Br J Ophthalmol*. 2024;108(3):432–439. doi:10.1136/bjo-2022-321399
11. Hao J, Kwapong WR, Shen T, et al. Early detection of dementia through retinal imaging and trustworthy AI. *NPJ Digit Med*. 2024;7(1):294. doi:10.1038/s41746-024-01292-5
12. Cheung CY, Ran AR, Wang S, et al. A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study. *Lancet Digit Health*. 2022;4(11):e806–e815. doi:10.1016/S2589-7500(22)00169-8
13. Wagner SK, Hughes F, Cortina-Borja M, et al. AlzEye: longitudinal record-level linkage of ophthalmic imaging and hospital admissions of 353 157 patients in London, UK. *BMJ Open*. 2022;12(3):e058552. doi:10.1136/bmjopen-2021-058552
14. Kleppe A. Area under the curve may hide poor generalisation to external datasets. *ESMO Open*. 2022;7(2):100429. doi:10.1016/j.esmoop.2022.100429
15. Zhang Q, Li J, Bian M, et al. retinal imaging techniques based on machine learning models in recognition and prediction of mild cognitive impairment. *Neuropsychiatr Dis Treat*. 2021;17:3267–3281. doi:10.2147/NDT.S333833
16. Milea D, Najjar RP, Zubo J, et al. Artificial intelligence to detect papilledema from ocular fundus photographs. *N Engl J Med*. 2020;382(18):1687–1695. doi:10.1056/NEJMoal917130
17. Li A, Tandon AK, Sun G, Dinkin MJ, Oliveira C. Early detection of optic nerve changes on optical coherence tomography using deep learning for risk-stratification of papilledema and glaucoma. *J Neuroophthalmol*. 2024;44(1):47–52. doi:10.1097/WNO.0000000000001945
18. Sathianvichitr K, Najjar RP, Zhiqun T, et al. A deep learning approach for accurate discrimination between optic disc drusen and papilledema on fundus photographs. *J Neuroophthalmol*. 2024;44(4):454–461. doi:10.1097/WNO.0000000000002223
19. Allegrini D, Pagano L, Ferrara M, et al. Optic disc drusen: a systematic review: up-to-date and future perspective. *Int Ophthalmol*. 2020;40(8):2119–2127. doi:10.1007/s10792-020-01365-w
20. Aumiller MS. Optic disc drusen: complications and management. *Optometry*. 2007;78(1):10–16. doi:10.1016/j.optm.2006.07.009
21. Girard MJA, Panda S, Tun TA, et al. Discriminating between papilledema and optic disc drusen using 3D structural analysis of the optic nerve head. *Neurology*. 2023;100(2):e192–e202. doi:10.1212/WNL.000000000000201350
22. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56. doi:10.1038/s41591-018-0300-7
23. Kaur K, Margolin E. Nonarteritic anterior ischemic optic neuropathy. *StatPearls*; 2025. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK559045/>. Accessed October 17, 2025.
24. Jalili J, Nadimi M, Jafari B, et al. Vessel density features of optical coherence tomography angiography for classification of optic neuropathies using machine learning. *J Neuroophthalmol*. 2024;44(1):41–46. doi:10.1097/WNO.0000000000001925
25. Gungor A, Najjar RP, Hamann S, et al. Deep learning to discriminate arteritic from nonarteritic ischemic optic neuropathy on color images. *JAMA Ophthalmol*. 2024;142(11):1073–1079. doi:10.1001/jamaophthalmol.2024.4269
26. Szanto D, Wang JK, Woods B, et al. Deep learning differentiates papilledema, NAION, and healthy eyes with unsegmented 3D OCT volumes. *Am J Ophthalmol*. 2025;277:249–259. doi:10.1016/j.ajo.2025.05.036
27. Szanto D, Erekat A, Woods B, et al. Deep learning approach readily differentiates papilledema, non-arteritic anterior ischemic optic neuropathy, and healthy eyes. *Am J Ophthalmol*. 2025;276:99–108. doi:10.1016/j.ajo.2025.04.006
28. Lee DK, Choi YJ, Lee SJ, Kang HG, Park YR. Development of a deep learning model to distinguish the cause of optic disc atrophy using retinal fundus photography. *Sci Rep*. 2024;14(1):5079. doi:10.1038/s41598-024-55054-0
29. Stein JD, Khawaja AP, Weizer JS. Glaucoma in adults—screening, diagnosis, and management: a review. *JAMA*. 2021;325(2):164–174. doi:10.1001/jama.2020.21899
30. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018;125(8):1199–1206. doi:10.1016/j.ophtha.2018.01.023
31. Asaoka R, Murata H, Hirasawa K, et al. Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. *Am J Ophthalmol*. 2019;198:136–145. doi:10.1016/j.ajo.2018.10.007
32. Majid I, Mishra Z, Wang ZC, Chopra V, Heuer D, Hu ZJ. Automated detection and biomarker identification associated with the structural and functional progression of glaucoma on longitudinal color fundus images. *Appl Sci*. 2025;15(3):1627. doi:10.3390/app15031627
33. Medeiros FA, Jammal AA, Thompson AC. From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. *Ophthalmology*. 2019;126(4):513–521. doi:10.1016/j.ophtha.2018.12.033
34. Gong D, Hu M, Yin Y, et al. Practical application of artificial intelligence technology in glaucoma diagnosis. *J Ophthalmol*. 2022;2022:5212128. doi:10.1155/2022/5212128
35. Biousse V, Newman NJ, Najjar RP, et al. Optic disc classification by deep learning versus expert neuro-ophthalmologists. *Ann Neurol*. 2020;88(4):785–795. doi:10.1002/ana.25839
36. Liu X, Zhao C, Wang L, et al. 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
37. Henkel C, Seibert S, Nichols Widmann C. Current advances in computerized cognitive assessment for mild cognitive impairment and dementia in older adults: a systematic review. *Dement Geriatr Cognit Disord*. 2025;54(2):109–119. doi:10.1159/000541627
38. Nasir M, Siddiqui K, Ahmed S. Ethical-legal implications of AI-powered healthcare in critical perspective. *Front Artif Intell*. 2025;8:1619463. doi:10.3389/frai.2025.1619463

39. Muhammad D, Bendeche M. Unveiling the black box: a systematic review of explainable artificial intelligence in medical image analysis. *Comput Struct Biotechnol J.* 2024;24:542–560. doi:10.1016/j.csbj.2024.08.005
40. Singh A, Jothi Balaji J, Rasheed MA, Jayakumar V, Raman R, Lakshminarayanan V. Evaluation of explainable deep learning methods for ophthalmic diagnosis. *Clin Ophthalmol.* 2021;15:2573–2581. doi:10.2147/OPTH.S312236
41. Sadée C, Testa S, Barba T, et al. Medical digital twins: enabling precision medicine and medical artificial intelligence. *Lancet Digit Health.* 2025;7(7):100864. doi:10.1016/j.landig.2025.02.004
42. Kim JH, Kim CH, Jeon H, Jung HC, Lee S. Personalized treatment approaches in neurocritical care. *Acute Crit Care.* 2025. doi:10.4266/acc.003050

Eye and Brain

Dovepress
Taylor & Francis Group

Publish your work in this journal

Eye and Brain is an international, peer-reviewed, open access journal focusing on clinical and experimental research in the field of neuro-ophthalmology. All aspects of patient care are addressed within the journal as well as basic research. Papers covering original research, basic science, clinical and epidemiological studies, reviews and evaluations, guidelines, expert opinion and commentary, case reports and extended reports are welcome. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/eye-and-brain-journal>