

Article

Not peer-reviewed version

Exploring AI's Potential in Papilledema Diagnosis to Support Dermatological Treatment Decisions in Rural Healthcare

[Jonathan Shapiro](#)*, [Naomi Fridman](#), [Itay Cohen](#), [Ziad Khamaysi](#), Mahdi Awwad, [Tom Kozlovsky](#), [Idit Maharshak](#)

Posted Date: 4 July 2025

doi: 10.20944/preprints202507.0109.v2

Keywords: papilledema; teleophthalmology; artificial intelligence; dermatologic medication side effects
fundus photography; fundus photography



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Exploring AI's Potential in Papilledema Diagnosis to Support Dermatological Treatment Decisions in Rural Healthcare

Jonathan Shapiro ^{1,*}, Naomi Fridman ², Itay Cohen ^{3,4}, Ziad Khamaysi ^{3,4}, Mahdi Awwad ⁵, Tom Kozlovsky ^{6,7} and Idit Maharshak ^{6,7}

- ¹ Maccabi Healthcare Services, Tel Aviv, Israel
- ² NF Algorithms & AI, Israel
- ³ Department of Dermatology, Rambam Health Care Campus, Haifa, Israel
- ⁴ The Bruce and Ruth Rappaport Faculty of Medicine, Technion – Israel Institute of Technology, Haifa, Israel
- ⁵ Tzafon Medical Center, Ophthalmology Unit, Tiberias, Israel
- ⁶ Department of Ophthalmology, Edith Wolfson Medical Center, Holon, Israel
- ⁷ Gray School of Medicine, Tel-Aviv University, Tel-Aviv, Israel
- * Correspondence: jonmidi@gmail.com; Tel +972505823675

Abstract

Background: Papilledema, an ophthalmic finding associated with increased intracranial pressure, can be related to dermatological medications including corticosteroids, isotretinoin, and tetracyclines. Early detection is critical to prevent irreversible optic nerve damage, but access to ophthalmologic expertise is often limited in rural settings. Artificial intelligence (AI) may offer a potential for automated, accurate papilledema detection from fundus images, supporting timely diagnosis and management. **Objective:** Evaluating and comparing the diagnostic performance of two AI models, including a ResNet-based convolutional neural network (CNN) and GPT-4, against human ophthalmologists in identifying papilledema from fundus photographs, with a focus on applications relevant to dermatological care in resource-limited environments. **Methods:** A dataset of 1,389 fundus images (295 papilledema, 295 pseudo-papilledema, 799 normal) was preprocessed and partitioned into a test set of 198 images (99 papilledema, 99 normal). The ResNet model was fine-tuned using discriminative learning rates and a one-cycle learning rate policy. Both AI models and two human evaluators (a senior ophthalmologist and an ophthalmology resident) independently assessed the test images. Diagnostic metrics such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and Cohen's Kappa were calculated for each evaluator. **Results:** All human evaluators demonstrated high sensitivity for papilledema detection (96.91%–99.0%). The ResNet AI model achieved the highest specificity (100%), PPV (100%), NPV (98.99%), and overall accuracy (99.49%), outperforming both human evaluators (accuracy: 95.96%) and GPT-4 (accuracy: 85.86%, specificity: 78.86%, PPV: 73.74%). Cohen's Kappa indicated excellent agreement for the ResNet model (0.9899) and human experts (0.9192), with substantial agreement for GPT-4 (0.71). **Conclusions:** These results highlight the potential of AI-assisted detection of papilledema in underserved settings. However, they also underscore the need for validation on external datasets and real-world clinical environments before such tools can be broadly implemented.

Keywords: papilledema; artificial intelligence; fundus photography; deep learning; dermatology; rural healthcare

Introduction

Papilledema is a condition characterized by swelling of the optic discs due to increased intracranial pressure (ICP). Ophthalmologists consider papilledema an emergency clinical finding

and, thus, urgently refer such patients for further evaluation and imaging in the ER to rule out a space-occupying lesion in the brain and to prevent permanent damage to the compressed optic nerves [1]. Papilledema is traditionally associated with neurological or systemic disorders, but can be caused by medications used in dermatology. Withdrawal from prolonged systemic corticosteroid treatment, vitamin A derivatives, i.e., isotretinoin, and tetracyclines, are associated with intracranial hypertension, thereby causing optic disc edema [2–5]. This presents a diagnostic challenge, particularly in rural areas with limited access to specialized ophthalmic evaluation.

Corticosteroids have a useful role as anti-inflammatory and immunosuppressant agents for treating an array of dermatological conditions [6]. However, long-term therapy and abrupt discontinuation of corticosteroids are associated with various neurologic complications, including intracranial hypertension (IH). They may alter CSF dynamics and possibly increase ICP, ultimately creating papilledema [3,7]. **Vitamin A derivatives**, such as **isotretinoin**, are commonly prescribed for severe acne and other keratinization disorders. Although isotretinoin is highly effective for treating severe acne, it is associated with IH, particularly if used in high doses and combined with tetracyclines. It is postulated that isotretinoin may act on CSF production or reabsorption, causing a rise in ICP [8,9]. A physician reporting system identified 181 cases of intracranial hypertension linked to isotretinoin, with symptoms appearing on average 2.3 months after exposure. 24% of patients had taken tetracycline around the same time. Six patients experienced recurrent symptoms when re-challenged with isotretinoin after discontinuation [10].

Tetracyclines, including doxycycline and minocycline, are broad-spectrum anti-bacterial drugs generally employed for treating acne and rosacea due to their anti-bacterial and inflammatory properties. Yet, there is a risk that they may serve as a predisposing factor for IH, most common in younger women, in the age group that is usually treated with these medications. The mechanism is uncertain but may involve altered CSF dynamics or vascular effects [2,11–13].

Artificial intelligence (AI) transforms ophthalmic practice by improving eye disease identification, recognition, and management. AI platforms, primarily through deep learning algorithms, are reaching a new level of screening and diagnosis of ocular diseases like diabetic retinopathy, age-related macular degeneration, and glaucoma, achieving sensitivities and specificities beyond 90% [14,15]. Such systems detect subtle pathological changes in retinal images and allow for early detection that can support large-scale screening activities, especially in underserved areas [16,17]. In the management of glaucoma, AI employs optical coherence tomography (OCT) data to predict disease progression and future functional losses, which enables timely interventions [18]. AI is reshaping ophthalmology with its skills to improve diagnostics, personalized treatment, and bridge gaps in care delivery. These advances are democratizing access to high-quality care and reducing disparities while improving patient outcomes worldwide [19,20].

Recent advances in AI and machine learning are reshaping clinicians' approach to diagnosing and managing complicated medical conditions such as papilledema, principally by utilizing tools such as ChatGPT to analyze and interpret clinical data, including fundoscopic images, for preliminary diagnoses [20,21]. By importing the appropriate high-definition fundoscopic images and clinical backgrounds into an AI model, a clinician using AI can detect critical findings such as swelling of the optic disc, hemorrhages, or obliteration of the optic cup, which are typical indicators of papilledema [22]. When provided with the relevant specific patient data, such as the medical history including utilization of medications, ChatGPT could generate a differential diagnosis while providing further investigation directions [23].

For example, a recent study [22] revealed that a specialized deep learning system (DLS) can reliably differentiate between optic disc drusen (ODD) and papilledema, even in cases of buried ODD and mild-to-moderate papilledema (Sensitivity: 78.4% (95% CI, 72.2%-84.7%), Specificity: 91.3% (95% CI, 87.0%-96.4%) [24]. Moreover, the U-Net deep learning model, the first automated system for clinical detection and grading of papilledema, achieved outstanding performance with 99.82% sensitivity, 98.65% specificity, and 99.89% accuracy [25]. Another large study validated a deep-learning system for detecting papilledema from 14,341 fundus images, achieving high accuracy (AUC

0.99), strong sensitivity (96.4%), and specificity (84.7%), highlighting its potential for automated diagnosis [26]. Relatively, ChatGPT-4o demonstrated superior diagnostic accuracy over Gemini Advanced, correctly identifying 52% vs. 30% of surgical retina cases and 78% vs. 43% of medical retina cases, while Gemini Advanced failed to recognize OCTA scans without structural images, mistaking them for artwork, highlighting ChatGPT-4o's advantage despite its limited diagnostic range [27].

Studies show the rising benefits of artificial intelligence in aiding ophthalmological and dermatological practice. In ophthalmology, for instance, the AI algorithms can detect papilledema and other optic neuropathies using fundoscopic images and show promising sensitivities and specificities in supporting early diagnosis and timely referrals [28,29]. Similarly, in dermatology, AI has been successfully applied to identifying skin lesions and determining systemic associations [30–32]. The platform's ability to obtain visual and dermatological data analysis further facilitates this multidisciplinary management. The combination of clinical expertise and AI-based diagnostic tools enables clinicians to improve diagnostic accuracy, streamline workflows, and effectively manage papilledema. This integration thus both improves patient outcomes and eases the load on healthcare systems, exposing the transformative potential of AI in modern medicine.

An important potential application of artificial intelligence (AI) in medicine is its ability to enhance diagnostic capabilities in resource-limited settings. One such scenario involves dermatological patients receiving isotretinoin treatment who may present with headache symptoms. In rural or underserved areas where access to ophthalmologists may be limited, AI tools capable of analyzing fundus images and detecting papilledema can provide a critical diagnostic resource. The availability of such AI-driven diagnostic tools could enable healthcare providers, such as dermatologists or general practitioners, to quickly identify signs of papilledema and make informed decisions about whether further specialized evaluation is needed. By improving early detection of papilledema, AI technology can enhance patient safety and optimize the management of patients treated with isotretinoin in remote settings. Our study compared two AI models, including ChatGPT, for identifying papilledema from fundus photos and evaluated their performance against human ophthalmologists.

Methods

The dataset we used was introduced by Ahn et al. [33], including full preprocessing details, consists of 1389 fundus photographs collected at Kim's Eye Hospital using a non-mydratic auto fundus camera (AFC-330, Nidek, Japan). It includes 295 images from patients with papilledema, 295 images from patients with pseudo papilledema, and 799 images from control patients. To standardize the images, preprocessing included resizing to a fixed width of 500 pixels while maintaining the aspect ratio. Additionally, local contrast normalization was applied using Gaussian filtering, pixel normalization to zero mean and unit variance, and cropping to 240×240 pixels centered on the optic nerve. To tackle the class imbalance, we defined a test (hold out) set, with 99 randomly chosen papilledema images, and 99 randomly chosen normal images. Validation data, used to fine-tune training, contained 88 patients, randomly chosen from the training set. For the classification task, we adopted transfer learning utilizing a powerful CNN (Convolutional Neural Network), ResNet (residual learning neural network) [34] model, already pretrained on the ImageNet [35] dataset for general image classification.

Given the limited size of our dataset, we chose to use discriminative fine-tuning [36], as it is particularly suitable for mitigating overfitting by assigning different learning rates to different pretrained model layers. Lower layers are updated conservatively to retain generalizable features, while higher layers are fine-tuned more aggressively to adapt to the task. This targeted adaptation helps prevent the model from fitting to noise while effectively leveraging pretrained knowledge. In addition, we also chose the one-cycle learning rate policy [37] to enhance training efficiency and generalization. This approach allows the model to explore a broader parameter space early in training by gradually increasing the learning rate, followed by a controlled decrease that stabilizes

convergence and helps prevent overfitting. The one-cycle policy has been shown to be particularly effective in transferring learning settings involving limited data.

These 198 test images were also evaluated using the GPT-4 model to assess its diagnostic accuracy in identifying papilledema. Additionally, to benchmark AI performance against human evaluators, a certified ophthalmologist and an ophthalmology resident independently analyzed the same test images, blinded to GPT-4’s predictions.

Statistical Analysis

Categorical variables were summarized using descriptive statistics. Interobserver agreement measures were represented using the Cohen’s Kappa coefficient. Sensitivity, specificity, positive predictive value, and diagnostic accuracy were calculated for each categorical variable and shown in Table 1. In addition, a two-sample proportion test was used to compare proportions. The data were analyzed using SPSS, version 26.0 for Windows (SPSS, Inc.). Confusion matrices were constructed for each evaluator (GPT-4, dermatologist, and ophthalmology resident), allowing for the calculation of key performance metrics, including accuracy, sensitivity, specificity, and precision.

Results

A total of 198 fundoscopic images were analyzed. All four evaluators demonstrated a high sensitivity for detecting papilledema, ranging from 96.91% (senior ophthalmologist) to 99.0% (ResNet AI model). Specificity varied more notably among evaluators, with the highest specificity achieved by the ResNet AI model at 100%, followed by the senior ophthalmologist and ophthalmology resident with specificities of 95.05% and 94.17%, respectively. GPT-4 exhibited the lowest specificity at 78.86%.

The positive predictive values (PPV) indicated that the ResNet AI model had a perfect PPV (100%), highlighting its reliability in confirming positive diagnoses. Human evaluators had slightly lower PPVs, 94.95% for the senior ophthalmologist and 93.94% for the resident, while GPT-4 showed the lowest PPV of 73.74%, indicating a higher rate of false positives. Negative predictive values (NPV) remained consistently high across all evaluators, with values ranging from 96.97% (senior ophthalmologist) to 98.99% (ResNet AI model), reflecting the high reliability in correctly identifying negative cases.

Overall diagnostic accuracy was highest for the ResNet AI model (99.49%), followed by both human evaluators at 95.96%, and lowest for GPT-4 at 85.86%. Cohen’s Kappa values, representing the agreement beyond chance, were very high for the ResNet AI model (0.9899) and similarly high for both human evaluators (0.9192), indicating excellent inter-rater reliability among human experts. GPT-4 had a lower kappa value (0.71), signifying moderate agreement. Confusion matrices are detailed in Table 1.

Table 1a-1d: Confusion Matrices Comparing Evaluator Performance Against Ground Truth Labels for Papilledema Detection. *P=Papilledema, **N=Normal, Kappa Agreement Scale [47], • 0.01 – 0.20: Slight agreement. • 0.21 – 0.40: Fair agreement. • 0.41 – 0.60: Moderate agreement. • 0.61 – 0.80: Substantial agreement. • 0.81 – 1.00: Almost perfect (or strong) agreement.

Table 1a. Confusion Matrix for ResNet AI Model in Detecting Papilledema Compared to Ground Truth.

	Resnet model P*	Resnet model N**	Total
Labeled P	99	0	99
Labeled N	1	98	99

Total	100	98	198
Accuracy (%)	99.5		
Cohen’s Kappa	0.99		

Table 1b. Confusion Matrix for Senior Ophthalmologists in Detecting Papilledema Compared to Ground Truth.

	Senior Ophthalmologist P*	Senior Ophthalmologist N**	Total
Labeled P	94	5	99
Labeled N	3	96	99
Total	97	101	198
Accuracy (%)	95.96		
Cohen’s Kappa	0.9192		

Table 1c. Confusion Matrix for Ophthalmology Resident in Detecting Papilledema Compared to Ground Truth.

	Ophthalmology Resident P*	Ophthalmology Resident N**	Total
Labeled P	93	6	99
Labeled N	2	97	99
Total	95	103	198
Accuracy (%)	95.96		
Cohen’s Kappa	0.92		

Table 1d: Confusion Matrix for GPT-4 Model in Detecting Papilledema Compared to Ground Truth.

	GPT-4o model P*	GPT-4o model N**	Total
Labeled P	73	26	99
Labeled N	2	97	99

Total	75	123	198
Accuracy (%)	85.86		
Cohen’s Kappa	0.72		

Table 2 summarizes the diagnostic performance metrics for each evaluator. The ResNet AI model demonstrated the highest overall performance, with perfect specificity (100.00%), PPV (100.00%), and an accuracy of 99.49%. It also achieved the highest inter-rater agreement (Cohen’s Kappa = 0.99), indicating near-perfect alignment with ground truth labels. In comparison, both human evaluators - the senior ophthalmologist and the ophthalmology resident—showed similar performance, each with an overall accuracy of 95.96% and Kappa values of 0.9192 and 0.92, respectively. GPT-4o exhibited substantially lower specificity (78.86%) and PPV (73.74%), resulting in the lowest accuracy (85.86%) and Kappa (0.72), despite a high NPV (97.98%). These findings suggest that while all evaluators performed well in identifying true negative cases, the ResNet model was the most precise in minimizing false positives and achieving consistent agreement with reference standards.

Table 2. Comparative Diagnostic Performance of Evaluators.

Evaluator	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	Cohen’s Kappa
ResNet AI	99.00	100.00	100.00	98.99	99.49	0.99
Senior Ophthalmologist	94.95	95.05	94.95	96.97	95.96	0.9192
Ophthalmology Resident	93.94	94.17	93.94	94.17	95.96	0.92
GPT-4o	73.74	78.86	73.74	97.98	85.86	0.72

Discussion

Large language models (LLMs) like ChatGPT are increasingly utilized in medical contexts due to their accessibility and public trust. However, their outputs must be interpreted cautiously, especially in high-stakes diagnostic scenarios. [38]

Ahn et al. (2019) [33] developed a DLS that distinguished between normal optic discs, papilledema, and other optic disc abnormalities with high accuracy (AUC 0.96). This laid the foundation for the BONSAI consortium, which later trained models for grading papilledema severity. [39] These systems have shown performance comparable to that of experienced neuro-

ophthalmologists and maintained robustness across imaging devices, including smartphones. [40] From a clinical perspective, these findings have particular relevance in rural and underserved areas, where timely ophthalmologic consultation is often unavailable. Dermatological treatments such as isotretinoin and tetracyclines are well-documented causes of intracranial hypertension, [4,10] and new-onset headache in patients on these medications should prompt urgent evaluation for papilledema. Fundus photography, especially non-mydriatic or smartphone-based systems, can facilitate remote diagnosis. [40]

Incorporating AI analysis at the point of care could streamline this process. Several studies support using digital fundus photography in emergency settings as a viable alternative to direct ophthalmoscopy [26] and AI interpretation could reduce reliance on delayed remote assessments. Such integration could prove lifesaving by accelerating diagnosis and referral for patients at risk of permanent damage to their optic nerves.

In our study, ChatGPT-4o achieved substantial agreement (Cohen's Kappa = 0.72) with expert evaluations in the detection of papilledema from fundus photographs. However, the model's sensitivity and specificity (78.86% and 73.74%, respectively) indicate a need for improvement, particularly in minimizing false positives. These findings reflect known challenges large language models face when interpreting complex or ambiguous visual input [23,38], underscoring the need for further refinement before clinical application. [23,38]

Our ResNet-based model significantly outperformed ChatGPT-4o and human ophthalmologists, achieving 99.49% accuracy, perfect specificity (100%), and near-perfect agreement with ground truth (Kappa = 0.99). These results align with recent studies showing that ChatGPT-4o underperforms compared to dedicated systems in ophthalmic diagnostics. [27]

The high performance of our ResNet model is consistent with previous studies using deep learning systems (DLSs) trained for papilledema detection. For instance, Milea et al. (2020) [26] achieved an AUC of 0.99 on a dataset of over 14,000 images, while Saba et al. (2021) [25] reported a sensitivity of 99.82% and specificity of 98.65% using a U-Net architecture.

The disparity in performance between ResNet and ChatGPT-4o likely stems from their respective architectures and training paradigms. Our ResNet model was specifically fine-tuned on papilledema detection using a domain-specific dataset, benefiting from discriminative learning rates and a one-cycle policy—techniques known to improve generalization on limited data. [36,37] ChatGPT, in contrast, is a generalist model with limited visual training for ophthalmic images, making it less suitable for nuanced fundoscopic interpretation. ChatGPT was presumably not trained on this specific dataset; therefore, its performance can be considered independent, similar to that of human physicians.

The COVID-19 pandemic has accelerated the adoption of AI-powered telehealth, and continued technological innovation is likely to support its integration into clinical practice further [26]. Cost, however, remains a practical consideration. While AI has proven to be cost-effective in screening for diabetic retinopathy [41], its economic impact in the diagnosis of papilledema remains underexplored. Specialized AI models currently demonstrate superior diagnostic performance compared to ChatGPT, but their implementation may incur higher costs [42,43]. Although ChatGPT currently falls short of these specialized tools in accuracy, its accessibility, zero cost, and ongoing development suggest it could become a valuable tool for papilledema screening in the future. Continued evaluation of newer ChatGPT versions is necessary to assess their evolving diagnostic capabilities and limitations. Ethical challenges must also be addressed, including concerns around data privacy, algorithmic bias, and clinical accountability [20,44].

Direct, on-site AI-assisted analysis of fundus photographs captured in emergency settings could serve as a rapid and effective screening method—particularly in rural areas where ophthalmologic expertise is limited. Given that papilledema may signal a medical emergency, early detection via AI could be lifesaving. Hand-held fundus cameras (e.g., Optomed Aurora) have shown high sensitivity and specificity in detecting optic disc abnormalities, making them especially useful in emergency

settings [45]. Additionally, patient-generated “selfie” fundus images offer a novel option for remote AI-assisted screening [46].

Limitations

Our study has several important limitations. First, the ResNet model was trained and validated using a single-institution dataset, which may limit its generalizability. External validation with larger, multi-center datasets is needed to confirm the robustness and applicability of the results. Second, our classification task was limited to distinguishing between papilledema and normal fundus images, excluding clinically relevant conditions such as pseudo-papilledema or borderline cases. This limits the clinical realism of our evaluation. Third, we did not incorporate important patient-level clinical metadata such as medication history, symptoms, or neurological findings, which typically play a significant role in real-world diagnostic decision-making [22,23]. Additionally, the exact training datasets and methods used for GPT-4 are unknown, creating uncertainty regarding the ophthalmologic content it was exposed to during training and limiting our ability to interpret its diagnostic accuracy fully. Lastly, the standardized preprocessing applied to images may not fully reflect the real-world variability in imaging quality and modalities encountered in routine clinical practice, which could potentially affect diagnostic performance in practical settings.

Conclusions

This study demonstrates the potential of artificial intelligence to assist in detecting papilledema from fundus images, particularly in resource-limited settings. Effective implementation of AI models requires training on diverse datasets to ensure robust and clinically meaningful performance. While specialized AI models currently show higher accuracy, general-purpose models like GPT-4, though promising for future development, are not yet reliable enough for clinical use. Continued advancement and validation of these AI technologies may ultimately enhance diagnostic capabilities, improve patient outcomes, and expand access to essential ophthalmologic care.

Future Directions and Broader Implications

Beyond the diagnostic accuracy results presented, this study opens several critical avenues for further development and implementation of AI in clinical ophthalmology. One promising direction lies in the creation of human-in-the-loop systems, where AI functions as a screening or triage tool but final decisions remain with trained clinicians. Such hybrid workflows could be particularly effective in rural or overstretched healthcare environments, enabling non-specialists to identify cases warranting urgent ophthalmologic referral while maintaining clinical oversight for ambiguous or borderline presentations [48,49].

Another area warranting exploration is AI threshold calibration. Depending on the use case—screening versus diagnosis—the AI model’s sensitivity and specificity can be optimized by adjusting its decision thresholds. For instance, in teler dermatology platforms monitoring patients on isotretinoin, prioritizing high sensitivity may help reduce the risk of missed papilledema cases, even at the cost of more false positives. Incorporating ROC and decision-curve analyses can help tailor model performance to specific clinical workflows [50].

From a global health equity perspective, AI tools such as the one evaluated here can significantly enhance access to timely diagnosis in resource-limited settings. The use of smartphone-based fundus photography, as illustrated in Figure 2, offers a practical and scalable solution for capturing retinal images in underserved areas [51]. When paired with AI algorithms and embedded into electronic medical record (EMR) systems, these tools can facilitate real-time risk stratification and referral.

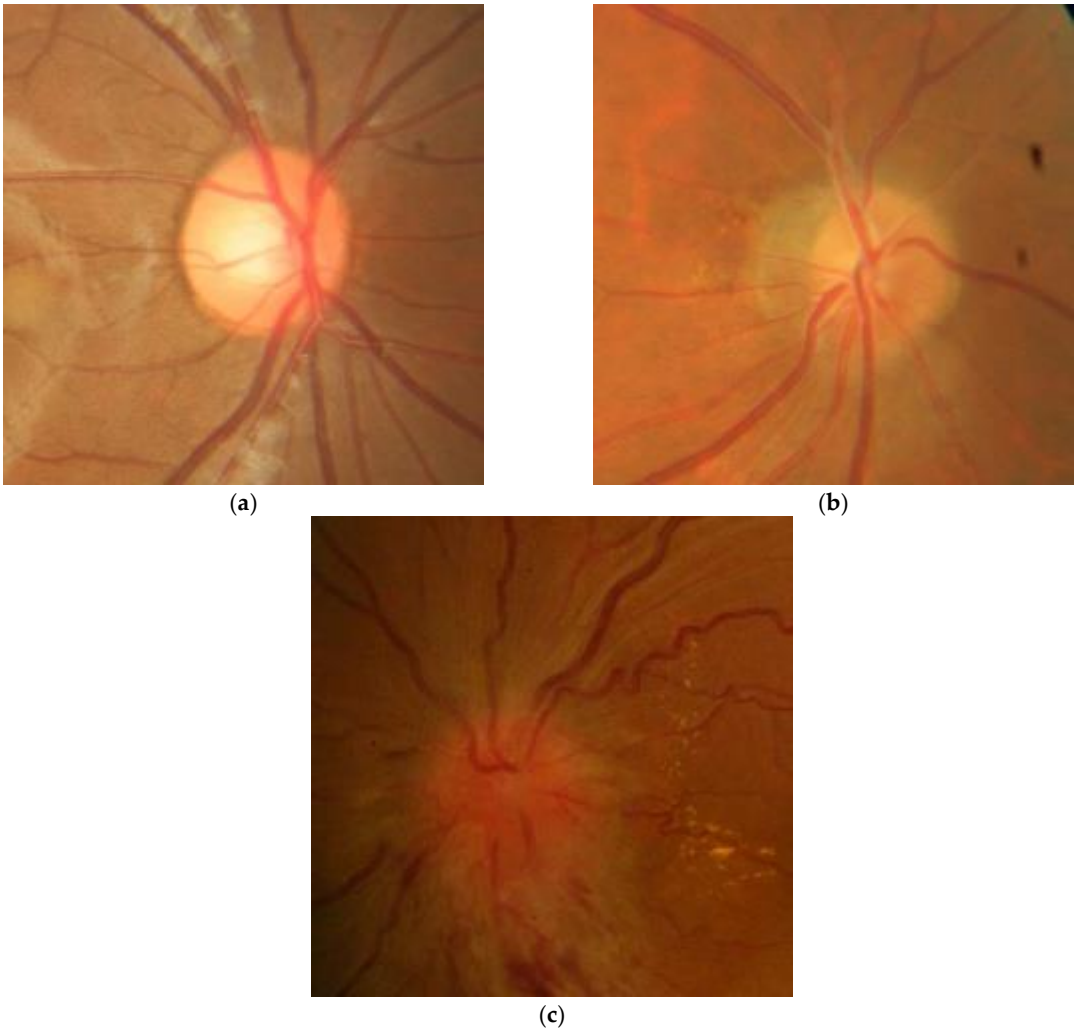


Figure 1. (a) Normal optic disc (b) Pseudo-Papilledema (c) Papilledema.

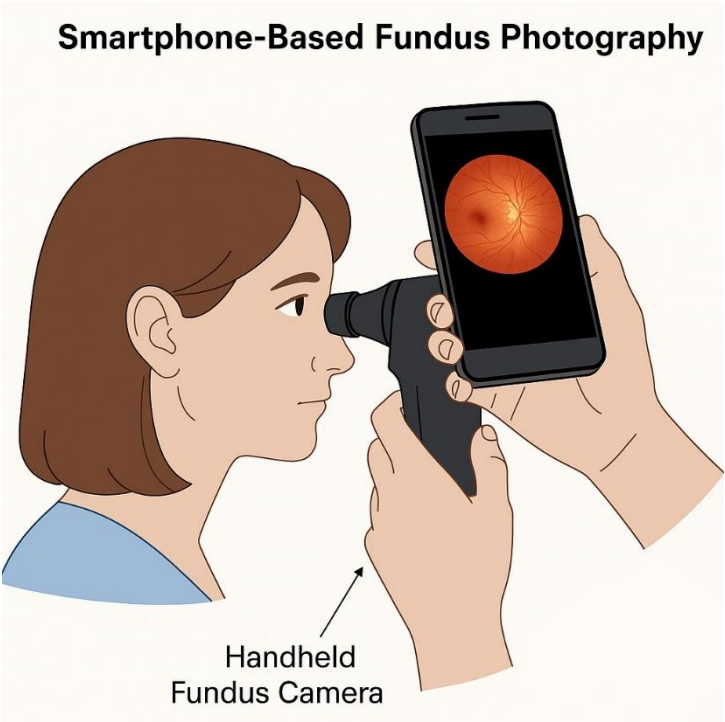


Figure 2. Smartphone-Based Fundus Photography.

However, ethical considerations remain paramount. AI systems may underperform when applied to populations underrepresented in training datasets, thereby increasing the risk of biased or inaccurate results. Transparent reporting, external validation across diverse settings, and continual performance monitoring are necessary to mitigate these risks. In parallel, issues of clinical accountability, explainability, and data privacy must be addressed—particularly in the context of black-box models or cloud-based diagnostic platforms [53].

Finally, while the clinical promise of AI is evident, its economic viability and integration into reimbursement frameworks remain unresolved. Health economic modeling and real-world implementation studies are needed to assess whether AI-assisted fundus screening cost-effectively improves outcomes, particularly for high-risk patients on medications such as isotretinoin [54].

Multimodal Models and Long-Term Adaptation

While our study focused on analyzing fundus photographs alone, future diagnostic AI systems will likely benefit from a multimodal architecture, one that incorporates not only imaging data but also structured clinical metadata. In real-world practice, the interpretation of optic disc swelling rarely occurs in isolation. Clinicians consider a range of contextual clues, including patient age, sex, medication history, symptoms such as headaches or blurred vision, and comorbid conditions. These factors can help differentiate true papilledema from other disc anomalies or non-neurogenic visual disturbances. By integrating such inputs, future AI models may better approximate clinical reasoning and improve diagnostic accuracy, particularly in cases with atypical imaging features [20,21].

Emerging research in medical AI already supports the potential of models that fuse visual and non-visual inputs. For example, in diabetic retinopathy, combining retinal images with laboratory values, such as hemoglobin A1c, improves model specificity and triage outcomes [16]. Applying a similar approach to papilledema detection—where fundus images are evaluated alongside medication usage (e.g., isotretinoin, corticosteroids), time since symptom onset, and patient risk factors—could reduce both false positives and false negatives. This is particularly relevant in dermatology and primary care settings, where clinical presentation is often multifactorial, and access to subspecialty review may be delayed [4,40].

In parallel, it is important to recognize that AI systems, including ours, are not static. Diagnostic performance can fluctuate over time as imaging technologies evolve, population demographics shift, and clinical practices change. For example, if a new type of fundus camera becomes standard in rural clinics or if image quality varies due to different lighting conditions or patient positioning, the original AI model may perform suboptimally unless updated [26]. Additionally, the introduction of new medications with papilledema risk profiles could alter the clinical context in which AI operates.

To ensure sustained accuracy and relevance, AI systems must be designed with mechanisms for ongoing calibration and retraining. This could involve periodic revalidation using newly acquired data, incorporation of real-world feedback loops from clinician users, or implementation of federated learning strategies that allow models to learn from decentralized datasets without compromising patient privacy [14,38]. Such adaptability is critical not only for maintaining diagnostic precision but also for preserving clinician trust in AI recommendations.

Moreover, health systems deploying these tools will need governance structures to oversee updates, version control, and the ethical use of continuously learning models. Without proper oversight, models may drift from their original performance metrics or inadvertently propagate biases from new data sources. Ultimately, as AI in ophthalmology transitions from research to deployment, its long-term success will depend on building systems that are not only accurate at baseline but also resilient, context-aware, and capable of evolving alongside clinical care.

Statement of funding sources: None

Conflicts of interest: The authors have no conflict of interest to declare.

Data availability statement: The data supporting the findings of this study are available from the corresponding author upon request, owing to ethical restrictions.

Declaration of generative AI and AI-assisted technologies in the writing process: During the preparation of this work, the author used ChatGPT-4o to improve language and readability. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

IRB approval status: Reviewed and approved by the Edith Wolfson Medical Center Ethics Committee, approval number 0115-24-WOMC.

References

1. Reier L, Fowler JB, Arshad M, et al. Optic Disc Edema and Elevated Intracranial Pressure (ICP): A Comprehensive Review of Papilledema. *Cureus* 2022;14(5):e24915. DOI: 10.7759/cureus.24915.
2. Orylska-Ratynska M, Placek W, Owczarczyk-Saczonek A. Tetracyclines-An Important Therapeutic Tool for Dermatologists. *Int J Environ Res Public Health* 2022;19(12). DOI: 10.3390/ijerph19127246.
3. Yasir M, Goyal A, Sonthalia S. Corticosteroid Adverse Effects. *StatPearls*. Treasure Island (FL) ineligible companies. Disclosure: Amandeep Goyal declares no relevant financial relationships with ineligible companies. Disclosure: Sidharth Sonthalia declares no relevant financial relationships with ineligible companies.2025.
4. Tan MG, Worley B, Kim WB, Ten Hove M, Beecker J. Drug-Induced Intracranial Hypertension: A Systematic Review and Critical Assessment of Drug-Induced Causes. *Am J Clin Dermatol* 2020;21(2):163-172. DOI: 10.1007/s40257-019-00485-z.
5. Rigi M, Almarzouqi SJ, Morgan ML, Lee AG. Papilledema: epidemiology, etiology, and clinical management. *Eye Brain* 2015;7:47-57. DOI: 10.2147/EB.S69174.
6. Gabros S, Nessel TA, Zito PM. Topical Corticosteroids. *StatPearls*. Treasure Island (FL) ineligible companies. Disclosure: Trevor Nessel declares no relevant financial relationships with ineligible companies. Disclosure: Patrick Zito declares no relevant financial relationships with ineligible companies.2025.
7. Trayer J, O'Rourke D, Cassidy L, Elnazir B. Benign intracranial hypertension associated with inhaled corticosteroids in a child with asthma. *BMJ Case Rep* 2021;14(5). DOI: 10.1136/bcr-2021-242455.
8. Reifenrath J, Rupprecht C, Gmeiner V, Haslinger B. Intracranial hypertension after rosacea treatment with isotretinoin. *Neurol Sci* 2023;44(12):4553-4556. DOI: 10.1007/s10072-023-07039-6.
9. Pile HD, Patel P, Sadiq NM. Isotretinoin. *StatPearls*. Treasure Island (FL) ineligible companies. Disclosure: Preeti Patel declares no relevant financial relationships with ineligible companies. Disclosure: Nazia Sadiq declares no relevant financial relationships with ineligible companies.2025.
10. Friedman DI. Medication-induced intracranial hypertension in dermatology. *Am J Clin Dermatol* 2005;6(1):29-37. DOI: 10.2165/00128071-200506010-00004.
11. Gardner K, Cox T, Digre KB. Idiopathic intracranial hypertension associated with tetracycline use in fraternal twins: case reports and review. *Neurology* 1995;45(1):6-10. DOI: 10.1212/wnl.45.1.6.
12. Shutter MC, Akhondi H. Tetracycline. *StatPearls*. Treasure Island (FL) ineligible companies. Disclosure: Hossein Akhondi declares no relevant financial relationships with ineligible companies.2025.
13. Del Rosso JQ, Webster G, Weiss JS, Bhatia ND, Gold LS, Kircik L. Nonantibiotic Properties of Tetracyclines in Rosacea and Their Clinical Implications. *J Clin Aesthet Dermatol* 2021;14(8):14-21. (<http://www.ncbi.nlm.nih.gov/pubmed/34840653>).
14. Hashemian H, Peto T, Ambrosio R, Jr., et al. Application of Artificial Intelligence in Ophthalmology: An Updated Comprehensive Review. *J Ophthalmic Vis Res* 2024;19(3):354-367. DOI: 10.18502/jovr.v19i3.15893.
15. Balyen L, Peto T. Promising Artificial Intelligence-Machine Learning-Deep Learning Algorithms in Ophthalmology. *Asia Pac J Ophthalmol (Phila)* 2019;8(3):264-272. DOI: 10.22608/APO.2018479.
16. Joseph S, Selvaraj J, Mani I, et al. Diagnostic Accuracy of Artificial Intelligence-Based Automated Diabetic Retinopathy Screening in Real-World Settings: A Systematic Review and Meta-Analysis. *American journal of ophthalmology* 2024;263:214-230. DOI: 10.1016/j.ajo.2024.02.012.

17. Olawade DB, Weerasinghe K, Mathugamage M, et al. Enhancing Ophthalmic Diagnosis and Treatment with Artificial Intelligence. *Medicina (Kaunas)* 2025;61(3). DOI: 10.3390/medicina61030433.
18. Tonti E, Tonti S, Mancini F, et al. Artificial Intelligence and Advanced Technology in Glaucoma: A Review. *J Pers Med* 2024;14(10). DOI: 10.3390/jpm14101062.
19. Li Z, Wang L, Wu X, et al. Artificial intelligence in ophthalmology: The path to the real-world clinic. *Cell Rep Med* 2023;4(7):101095. DOI: 10.1016/j.xcrm.2023.101095.
20. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J* 2021;8(2):e188-e194. DOI: 10.7861/fhj.2021-0095.
21. Chen X, Wang X, Zhang K, et al. Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal* 2022;79:102444. DOI: 10.1016/j.media.2022.102444.
22. Li Z, Chen W. Solving data quality issues of fundus images in real-world settings by ophthalmic AI. *Cell Rep Med* 2023;4(2):100951. DOI: 10.1016/j.xcrm.2023.100951.
23. Chen J, Liu L, Ruan S, Li M, Yin C. Are Different Versions of ChatGPT's Ability Comparable to the Clinical Diagnosis Presented in Case Reports? A Descriptive Study. *J Multidiscip Healthc* 2023;16:3825-3831. DOI: 10.2147/JMDH.S441790.
24. 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.
25. Saba T, Akbar S, Kolivand H, Ali Bahaj S. Automatic detection of papilledema through fundus retinal images using deep learning. *Microsc Res Tech* 2021;84(12):3066-3077. DOI: 10.1002/jemt.23865.
26. Milea D, Najjar RP, Zhubo J, et al. Artificial Intelligence to Detect Papilledema from Ocular Fundus Photographs. *The New England journal of medicine* 2020;382(18):1687-1695. DOI: 10.1056/NEJMoa1917130.
27. Carla MM, Crincoli E, Rizzo S. RETINAL IMAGING ANALYSIS PERFORMED BY CHATGPT-4o AND GEMINI ADVANCED: The Turning Point of the Revolution? *Retina (Philadelphia, Pa)* 2025;45(4):694-702. DOI: 10.1097/IAE.0000000000004351.
28. Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. *Adv Ophthalmol Pract Res* 2022;2(3):100078. DOI: 10.1016/j.aopr.2022.100078.
29. Moraru AD, Costin D, Moraru RL, Branisteanu DC. Artificial intelligence and deep learning in ophthalmology - present and future (Review). *Exp Ther Med* 2020;20(4):3469-3473. DOI: 10.3892/etm.2020.9118.
30. Goktas P, Grzybowski A. Assessing the Impact of ChatGPT in Dermatology: A Comprehensive Rapid Review. *J Clin Med* 2024;13(19). DOI: 10.3390/jcm13195909.
31. Cuellar-Barboza A, Brussolo-Marroquin E, Cordero-Martinez FC, Aguilar-Calderon PE, Vazquez-Martinez O, Ocampo-Candiani J. An evaluation of ChatGPT compared with dermatological surgeons' choices of reconstruction for surgical defects after Mohs surgery. *Clin Exp Dermatol* 2024;49(11):1367-1371. DOI: 10.1093/ced/llae184.
32. Elias ML, Burshtein J, Sharon VR. OpenAI's GPT-4 performs to a high degree on board-style dermatology questions. *Int J Dermatol* 2024;63(1):73-78. DOI: 10.1111/ijd.16913.
33. Ahn JM, Kim S, Ahn KS, Cho SH, Kim US. Accuracy of machine learning for differentiation between optic neuropathies and pseudopapilledema. *BMC Ophthalmol* 2019;19(1):178. DOI: 10.1186/s12886-019-1184-0.
34. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016:770-778.
35. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition: IEEE; 2009: 248-255.
36. Howard J, Ruder S. Universal Language Model Fine-tuning for Text Classification. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics; 2018: 328-339.
37. Smith LN. A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay. 2018.

38. AlRyalat SA, Musleh AM, Kahook MY. Evaluating the strengths and limitations of multimodal ChatGPT-4 in detecting glaucoma using fundus images. *Front Ophthalmol (Lausanne)* 2024;4:1387190. DOI: 10.3389/fopht.2024.1387190.
39. Leong YY, Vasseneix C, Finkelstein MT, Milea D, Najjar RP. Artificial Intelligence Meets Neuro-Ophthalmology. *Asia Pac J Ophthalmol (Phila)* 2022;11(2):111-125. DOI: 10.1097/APO.0000000000000512.
40. Biousse V, Najjar RP, Tang Z, et al. Application of a Deep Learning System to Detect Papilledema on Nonmydriatic Ocular Fundus Photographs in an Emergency Department. *American journal of ophthalmology* 2024;261:199-207. DOI: 10.1016/j.ajo.2023.10.025.
41. Ruamviboonsuk P, Ruamviboonsuk V, Tiwari R. Recent evidence of economic evaluation of artificial intelligence in ophthalmology. *Curr Opin Ophthalmol* 2023;34(5):449-458. DOI: 10.1097/ICU.0000000000000987.
42. Rau A, Rau S, Zoeller D, et al. A Context-based Chatbot Surpasses Trained Radiologists and Generic ChatGPT in Following the ACR Appropriateness Guidelines. *Radiology* 2023;308(1):e230970. DOI: 10.1148/radiol.230970.
43. Savelka J, Ashley KD. The unreasonable effectiveness of large language models in zero-shot semantic annotation of legal texts. *Front Artif Intell* 2023;6:1279794. DOI: 10.3389/frai.2023.1279794.
44. Tan W, Wei Q, Xing Z, et al. Fairer AI in ophthalmology via implicit fairness learning for mitigating sexism and ageism. *Nature communications* 2024;15(1):4750. DOI: 10.1038/s41467-024-48972-0.
45. EyeWiki. Artificial Intelligence in Neuro-Ophthalmology. EyeWiki. December 20, 2023 (https://eyewiki.org/Artificial_Intelligence_in_Neuro-Ophthalmology).
46. Kumari S, Venkatesh P, Tandon N, Chawla R, Takkar B, Kumar A. Selfie fundus imaging for diabetic retinopathy screening. *Eye (Lond)* 2022;36(10):1988-1993. DOI: 10.1038/s41433-021-01804-7.
47. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37(5):360-3. (<https://www.ncbi.nlm.nih.gov/pubmed/15883903>).
48. Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., & Vardoulakis, L. M. (2020). A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. *CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3313831.3376718>
49. Park, S. H., Han, K., & Lee, J. H. (2023). Human-in-the-loop approach to medical AI: Framework and applications. *Journal of Digital Imaging*, 36(2), 413–423.
50. Vickers, A. J., Van Calster, B., & Steyerberg, E. W. (2019). A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic and Prognostic Research*, 3, 18. <https://doi.org/10.1186/s41512-019-0064-7>
51. Mackenzie, P. J., Kalevar, A., & Liu, I. (2019). Use of smartphone-based fundus photography in clinical practice. *Canadian Journal of Ophthalmology*, 54(1), 16–21.
52. He, J., Baxter, S. L., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 25(1), 30–36. <https://doi.org/10.1038/s41591-018-0307-0>
53. Char, D. S., Abràmoff, M. D., & Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. *The American Journal of Bioethics*, 20(11), 7–17. <https://doi.org/10.1080/15265161.2020.1819469>
54. Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., ... & Kittler, H. (2020). Human-computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8), 1229–1234. <https://doi.org/10.1038/s41591-020-0942-0>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.