

RESEARCH

Open Access



Differentiation of optic disc edema and pseudopapilledema with deep learning on near-infrared reflectance images

Cumhur Ozbas^{1*}, Furkan Kirik², Ersin Akbulut², Gulsah Eldemir³, Betul Tugcu² and Mehmet H. Ozdemir²

Abstract

Background This study aimed to develop an artificial intelligence-based deep learning (DL) algorithm using near-infrared reflectance (NIR) images to differentiate between optic disc edema and pseudopapilledema, and to evaluate the diagnostic performance of the developed model.

Methods NIR images were divided into 2 groups for training and testing of the model. 85% (714 images) were used for training the model and 15% (126 images) for testing the trained model. Sensitivity, specificity and accuracy of the model were calculated in detecting optic disc edema, pseudopapilledema and normal optic discs. Receiver operating characteristic curve and area under the curve (AUC) values were also analyzed.

Results The developed model was tested with 24 optic disc edema, 52 pseudopapilledema, and 50 normal optic disc images not used in training. Sensitivities were 100%, 98%, and 96%; specificities were 99%, 97%, and 100%; and accuracy rates were 99%, 98%, and 98%, respectively. In addition, the AUC values of the groups were 0.995 (95% Confidence interval [CI]: 0.98-1); 0.983 (95% CI: 0.96-1); and 0.973 (95% CI: 0.94-1).

Conclusions The developed DL model demonstrated high diagnostic performance in distinguishing optic disc edema and pseudopapilledema using NIR images and may serve as a reliable clinical decision support tool.

Keywords Artificial intelligence, Deep learning, Optic disc edema, Pseudopapilledema, Near infrared reflectance

Introduction

Optic disc edema can develop as a secondary manifestation of numerous pathological processes, some benign and others with serious visual and neurological consequences. Making distinctions among various etiologies depends on a comprehensive history and a thorough

examination with attention to the optic disc. Common causes of optic disc edema include demyelinating optic neuritis, non-arteritic ischemic optic neuropathy (NAION), retinal vein occlusion (RVO), diabetic papillopathy, papilledema, and toxic neuropathies [1]. Papilledema signifies swelling of the optic disc due to increased intracranial pressure and can be differentiated from other causes of disc edema based on its clinical features. Pseudopapilledema, on the other hand, refers to an abnormal and elevated appearance of the optic nerve head, which is not caused by increased intracranial pressure or edema in the nerve fiber layer. In the differential diagnosis of optic disc edema, papilledema should be a primary

*Correspondence:

Cumhur Ozbas
cumhurozbas92@gmail.com

¹Department of Ophthalmology, Sanliurfa Training and Research Hospital, Yenice, Yenice Road No:1, Sanliurfa, Eyyubiye 63250, Turkey

²Department of Ophthalmology, Faculty of Medicine, Bezmialem Vakif University, Fatih, Istanbul, Turkey

³Istanbul, Turkey



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

consideration due to its potential for vision loss and significant health risks [2].

In recent years, internet or cloud-based artificial intelligence (AI) platforms, both paid and free, have become accessible to users. Clinicians can now develop AI models for tasks like image classification, lesion detection, and tissue segmentation. These AI models are increasingly preferred for diagnosing neuro-ophthalmic pathologies. While color fundus photographs are commonly used for AI evaluations of optic disc pathologies, studies utilizing other imaging methods remain limited. High-quality near infrared reflectance (NIR) imaging in optical coherence tomography (OCT) allows lesion visualization even on the undersurface.

In recent years, applications of AI in ophthalmology have expanded rapidly. These applications provide high diagnostic accuracy and decision support in various clinical domains, including anterior segment imaging, retinal and optic nerve analysis, glaucoma evaluation,

and refractive prediction. Table 1 summarizes well-established applications of AI in ophthalmology with selected recent studies, highlighting their diagnostic performance, clinical contributions, and current limitations.

In this study, NIR images were processed using deep learning (DL), which is the most effective learning model in AI. The study aimed to evaluate the effectiveness of the developed AI model in diagnosing optic disc edema and pseudopapilledema [8].

Methods

Study design

This retrospective, single-center study analysed spectral domain OCT (SD-OCT) (Heidelberg Engineering, Inc., Heidelberg, Germany) images from patients evaluated for suspected optic disc edema or pseudopapilledema at the Department of Ophthalmology, Bezmialem Vakıf University, between June 2019 and December 2022.

Table 1 Recent applications of AI in ophthalmology

Author	Pathology (Condition Studied)	Model used	Dataset Information	Results
Szanto et al., 2025 [3].	Optic nerve head (ONH) atrophy classification: glaucoma, NAION, optic neuritis, and healthy controls	Three ResNet-3D-18 models: whole volume, peripapillary region only, ONH only; 5-fold stratified cross-validation; unsegmented 3D OCT volumes	Multicenter; Device: Cirrus ONH OCT; 7,014 scans/1,382 eyes. Classes: glaucoma $n=113$, NAION $n=311$, optic neuritis $n=163$, healthy $n=715$. Mean age 54.2 ± 16.9 ; 65% male, 35% female	Whole-volume: Accuracy 88.9%; Macro AUC-ROC 0.977 (95% CI 0.974–0.979); F1 (glaucoma 0.94, NAION 0.87, optic neuritis 0.78, healthy 0.91). Peripapillary region: Accuracy 85.9%; AUC 0.970 (0.967–0.972). ONH: Accuracy 87.0%; AUC 0.972 (0.970–0.975).
Madadi et al., 2025 [4].	Neuro-ophthalmic diseases (varied chronic and acute cases; diagnosis from case reports)	LLM approach: ChatGPT (GPT-3.5 and GPT-4); each case entered as a prompt to obtain the most likely diagnosis	22 neuro-ophthalmic case reports from a public database; same information presented to 2 neuro-ophthalmologists (NO); responses compared	Accuracy: GPT-3.5 = 13/22 (59%), GPT-4 = 18/22 (82%), NOs = 19/22 (86%). Agreement: GPT-4 & NO-1 = 17/22 (77%); GPT-4 & NO-2 = 16/22 (73%); GPT-3.5 & GPT-4 = 13/22 (59%); NO-1 & NO-2 = 17/22 (77%).
Gu et al., 2025 [5].	Neuromyelitis optica (NMO): onset and stage prediction	Multimodal AI integrating ultrawide-field fundus (UWF) photographs with clinical examination data	UWF of 330 eyes from 285 NMO patients + 1,288 eyes from 770 non-NMO participants; clinical exam reports included	AUC: 0.9923; Max Youden index: 0.9389; Sensitivity: 97.0%; Specificity: 96.9% (test dataset for NMO prevalence prediction).
Gungor et al., 2025 [6].	Papilledema and other optic neuropathies (optic disc abnormalities)	CNN-based DL System: optic disc and peripapillary region segmentation; 3-class classification (papilledema/other abnormalities/normal)	International, multiethnic, multicenter: 20,533 retinal photos (10,647 patients) from 31 centers. Train + internal validation: 18,981 mydriatic photos (9,830 patients; 22 countries). External test: 1,552 prospective photos (817 patients) with a handheld nonmydriatic camera (Aurora, Optomed, Finland). Outcomes evaluated at eye and patient levels.	For differentiating papilledema from others and healthy: Accuracy 99.5% (95% CI 99.1–99.8), Sensitivity 81.0% (95% CI 64.1–97.7), Specificity 99.7% (95% CI 99.5–99.9), AUC 98.3% (95% CI 97.6–99.9).
Szanto et al., 2025 [7].	Papilledema vs. NAION vs. healthy	DL (ResNet-50 backbone); disc localization normalization; Grad-CAM explainability; 5-fold cross-validation with majority voting for external validation	Train/validation: 15,088 fundus photos from 5,866 eyes; External validation: 1,126 images from 928 eyes	Internal validation: Accuracy $\approx 96.2\%$, Macro AUC ≈ 0.995 ; External validation: Accuracy $\approx 93.6\%$, Macro AUC ≈ 0.980

Abbreviations: OCT (Optical coherence tomography), ONH (Optic nerve head), NAION (Non-arteritic anterior ischemic optic neuropathy), LLM (Large language model), NMO (Neuromyelitis optica), UWF (Ultrawide-field), CNN (Convolutional neural network), AUC (Area under the curve), CI (Confidence interval), F1 (F1-score), NO (Neuro-ophthalmologist)

All patients included in the study underwent a comprehensive ophthalmological examination at the time of referral, which included best-corrected visual acuity, anterior segment examination, intraocular pressure measurement, color vision assessment using Ishihara charts, evaluation for afferent pupillary defect, and a detailed fundoscopic examination.

The study was approved by the Bezmialem University Ethics Committee (Decision No: 2022/406) and conducted in line with the Declaration of Helsinki.

In this study, patients from the optic disc edema, pseudopapilledema, and healthy control groups were selected based on specific inclusion criteria. The optic disc edema group included patients with fundoscopic and SD-OCT findings confirming optic disc swelling. This group consisted of various etiologies, including papilledema, NAION, RVO, demyelinating optic neuropathy, and diabetic papillopathy. Papilledema was characterized by optic disc swelling due to increased intracranial pressure, confirmed by fundoscopic examination and, when necessary, brain magnetic resonance imaging (MRI) [9]. NAION was diagnosed in patients presenting with sudden, painless vision loss, with segmental or diffuse optic disc edema observed on fundoscopy, along with systemic vascular risk factors such as hypertension or diabetes [10]. RVO was identified by dilated and tortuous retinal veins, retinal hemorrhages, and macular edema observed on OCT [11]. Demyelinating optic neuropathy was defined by acute, painful vision loss, delayed responses on visual evoked potentials (VEP), and demyelinating lesions detected on brain and optic nerve MRI

[12]. Diabetic papillopathy was characterized by minimal vision loss and mild-to-moderate optic disc swelling in diabetic patients [13].

The pseudopapilledema group included patients diagnosed based on specific imaging findings. This group comprised patients with optic disc drusen (ODD), confirmed through B-scan ultrasonography or autofluorescence imaging, showing hyperreflective nodular structures on SD-OCT with minimal or no visual field defects [14]. Additionally, only cases with peripapillary hyperreflective ovoid mass-like structures (PHOMS) showing clinical and imaging features consistent with pseudopapilledema were included. Inclusion criteria for this subgroup required blurred optic disc margins in the absence of elevated intracranial pressure or significant visual impairment [15]. Given that PHOMS may be associated with other conditions such as optic neuropathies and optic neuritis, patients with any clinical or radiological evidence suggestive of such etiologies were excluded following detailed ophthalmic examination, brain and orbital MRI, and, when available, VEP testing. This approach ensured accurate classification and reduced the risk of diagnostic overlap.

Healthy controls were aged 18–65 years, had best-corrected visual acuity of 20/20, and showed no ocular or systemic disease or optic disc abnormalities on OCT or fundoscopy.

Exclusion criteria included low-quality images, ocular surgery or trauma, incomplete data, and coexisting retinal or optic nerve disorders. These measures aimed to provide a consistent and reliable dataset for training the DL model. These criteria are summarized in Table 2 to provide a concise overview of the inclusion and exclusion parameters used in this study.

Although OCT is not considered the definitive gold standard for diagnosing all optic disc pathologies, it serves as a highly valuable tool for visualizing structural changes in the optic nerve head and the peripapillary region. In this study, high-resolution NIR images from SD-OCT were used for DL analysis. Diagnoses were confirmed through multimodal assessment including ophthalmic exam, fundoscopy, MRI, and ultrasonography when needed, ensuring reliable labelling for model training.

Data set preparation

The final dataset comprised 840 OCT images: 158 from eyes with optic disc edema (24 papilledema, 16 NAION, 12 RVO, 3 demyelinating optic neuropathy, and 5 diabetic papillopathy); 346 from 132 patients with ODD or pseudopapilledema secondary to PHOMS; and 336 from 168 healthy controls. This class distribution falls within the 20–40% range considered a mild imbalance according to the Imbalanced Data guideline [16]. To minimise any bias

Table 2 Inclusion and exclusion criteria

Group	Criteria
Inclusion – Optic disc edema	- Fundoscopic and SD-OCT findings confirming optic disc swelling - Etiologies: papilledema, NAION, RVO, demyelinating optic neuropathy, diabetic papillopathy
Inclusion – Pseudopapilledema	- ODD confirmed by B-scan ultrasonography or autofluorescence - PHOMS with blurred disc margins, consistent with pseudopapilledema, without elevated intracranial pressure or significant vision loss
Inclusion – Healthy controls	- Age 18–65 years - Best-corrected visual acuity 20/20 - No ocular/systemic disease or optic disc abnormality on OCT or fundoscopy
Exclusion criteria	- Low-quality images - History of ocular surgery or trauma - Incomplete clinical/imaging data - Coexisting retinal or optic nerve disorders - Clinical/radiological evidence of optic neuropathies or optic neuritis (for PHOMS subgroup)

arising from this skew, a single 85/15 train–test split was generated with eye-level stratified sampling so that the original class proportions were preserved in each split. Follow-up images of the same eyes, acquired when disc swelling persisted, were retained as independent samples because optic-disc morphology evolves over time; importantly, every image from a given eye was always assigned to the same split, preventing data leakage. These follow-up images were obtained at clinically distinct time points, with a minimum interval of one month between visits, to enhance temporal independence. A similar eye-level protocol was recommended and successfully applied by De Fauw et al. in their longitudinal OCT study [17], supporting the methodological soundness of our allocation strategy. The procedure, implemented via GroupShuffleSplit on unique eye identifiers, yielded 714 training images (85%) and 126 test images (15%), preserved class proportions, and provided the independent performance estimate for the held-out test set.

NIR images were cross-sectionally obtained from the OCT images taken for standard retinal nerve fiber layer assessment. Representative examples are provided in Fig. 1.

Development of the deep learning classification model

The project utilized Kaggle, a widely recognized and highly regarded platform for data science and DL projects, which has been successfully applied in both ophthalmology and other medical fields [18]. Python, a versatile and commonly used programming language in data analysis, AI, and scientific computing, was employed for this study.

A DL model was developed using the Convolutional Neural Networks (CNN) algorithm to classify optic disc images. CNNs were chosen due to their capability to process large and complex datasets [19]. The model was trained on labeled data and tested on the 15% eye-level test set described above to evaluate its performance. Various Python libraries, including TensorFlow and Keras, were utilized for building and training the CNN model.

Pixel values were rescaled to the [0, 1] range (rescale=1/255), and all NIR images were resized to 200×200 pixels prior to model training to optimize computational efficiency while preserving diagnostically relevant features. This resolution was selected based on preliminary experiments, which showed no meaningful improvement in classification performance at higher resolutions, while significantly reducing memory usage and training time. To reduce overfitting, data augmentation was applied during training. This included random rotations of ±20 degrees, horizontal flips, zooming between 0.9× and 1.1×, and brightness shifts of ±10%. A batch size of 32 was selected as it provided a good balance between GPU memory usage and convergence speed. Within each eye-level stratified 85% training subset, 10% of the images were further reserved for internal validation using the validation_split=0.1 parameter in Keras. This internal validation set also respected eye identifiers. As a result, approximately 76.5% of the total data contributed to weight updates, 8.5% was used for early stopping and monitoring, and 15% was reserved as an independent test set for final evaluation. In summary, of the 840 NIR images included in this study, 714 (85%) were allocated to the training set, of which 10% were further reserved for internal validation, and 126 (15%) were reserved as an independent test set. Eye-level stratified sampling ensured that all three categories (optic disc edema, pseudopapilledema, and healthy controls) were proportionally represented in each subset, while images from the same eye were kept together to prevent data leakage. This allocation strategy was designed to balance learning efficiency with rigorous evaluation, ensuring both robust training and fair assessment of model performance.

The model architecture comprised an input layer, a ResNet50 backbone with frozen layers, followed by global average pooling, a dense layer with 128 units and ReLU activation, and a softmax output layer for three-class classification (optic disc edema, pseudopapilledema, and normal). To preserve general low-level visual features, the first ~ 143 layers were frozen, while the remaining 34

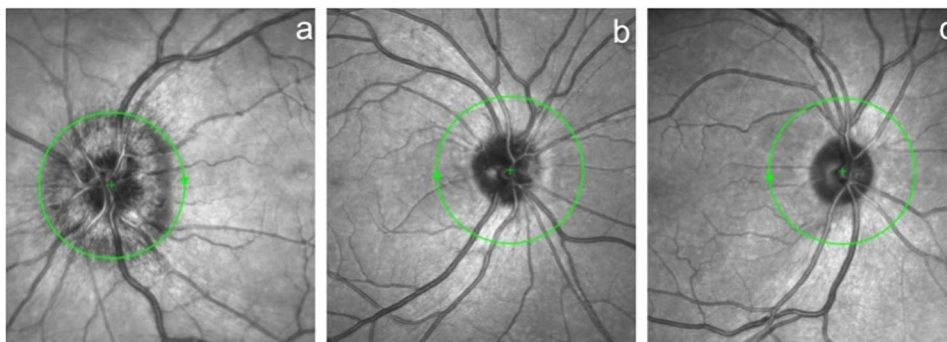


Fig. 1 Representative NIR images from the dataset. **a** Optic disc edema due to papilledema. **b** Pseudopapilledema secondary to ODD. **c** Normal optic disc. These images were used during training and testing of the DL model

layers and the dense layer were fine-tuned. Preliminary trials with deeper, fully trainable custom CNNs containing over one million parameters yielded no meaningful improvement in accuracy and tended to overfit earlier. Therefore, the ResNet50-based transfer learning pipeline was retained as the most effective trade-off between model depth, generalizability, and computational cost [20]. Training was initiated with an upper limit of 500 epochs using the Adam optimizer with an initial learning rate of 0.001. This upper limit was chosen based on findings from pilot experiments, where the validation loss continued to decrease until approximately epochs 80–100 in some random initializations, due to the small dataset and low learning rate. Lowering the cap to 100 or 200 epochs occasionally resulted in incomplete convergence and increased performance variability. To ensure adequate convergence across all training repetitions, we set the upper limit to 500 epochs. However, early stopping (patience = 5) effectively halted training earlier; in the final model, training stopped at epoch 64. Thus, the 500-epoch limit served solely as a safeguard and did not prolong the actual training duration. Early stopping monitored the validation loss and terminated training at epoch 64, when the loss had failed to improve for five consecutive epochs. Model performance was evaluated on a single 15% test subset generated using the GroupShuffleSplit method, which ensured that all images from a given eye were kept in the same subset and class proportions were preserved. For the test set, sensitivity, specificity, accuracy, and area under the curve (AUC) were calculated. Additionally, to enhance interpretability, we applied the SmoothGrad technique to representative test images. SmoothGrad generates refined saliency maps by averaging multiple noisy gradient calculations, producing more stable and less noisy visualizations. These maps highlighted the optic nerve head and peripapillary regions most influential for the model's predictions (Figs. 2 and 3).

Statistical analysis

The performance of the DL model on the test dataset was evaluated by calculating sensitivity, specificity, and accuracy rates. To further assess the model's effectiveness in image classification, Receiver Operating Characteristic (ROC) curve analysis was conducted, and the AUC was calculated. For the comparison of demographic variables between the study groups, the Shapiro-Wilk test was initially used to assess the normality of the data. Since the age data did not follow a normal distribution, the Kruskal-Wallis test was applied to compare age among the three groups, followed by post-hoc pairwise comparisons using Dunn's test with Bonferroni correction to identify significant differences. To compare gender distribution across the groups, the Chi-square test was employed. All

statistical analyses were conducted using SPSS (Version 22.0, SPSS Inc., IBM, Chicago, IL, USA), and a *p*-value of less than 0.05 was considered statistically significant.

Results

The mean age for the optic disc edema group was 57.08 ± 5.16 years, for the pseudopapilledema group 32.84 ± 20.20 years, and for the normal group 52.80 ± 25.99 years. The age distribution was significantly different in the pseudopapilledema group compared to the other groups ($p < 0.001$).

The gender distribution showed 40.0% males and 60.0% females in the optic disc edema group, 39.4% males and 60.6% females in the pseudopapilledema group, and 40.5% males and 59.5% females in the normal group. There was no significant difference in gender distribution across the groups ($p: 0.982$). Demographic data are presented in Table 3.

The study utilized a total of 840 images, obtained from 192 patients and 168 healthy individuals. A set of 126 OCT images (52 pseudopapilledema, 24 optic disc edema, 50 normal) was used to evaluate the performance of the developed DL model in classifying optic disc edema, pseudopapilledema, and normal images. The distribution of the test images and the comparison of the results with the reference classes are summarized in the confusion matrix (Fig. 4).

Among the 126 test images, the model correctly classified all 24 cases of optic disc edema, achieving 100% sensitivity for this clinically critical category. Only one pseudopapilledema image was mistakenly classified as optic disc edema, resulting in a specificity of 99%. Importantly, no optic disc edema image was misclassified, and confusion between optic disc edema and pseudopapilledema occurred in only one direction. Additionally, two normal optic disc images were incorrectly classified as pseudopapilledema.

The developed model was comprehensively evaluated for its diagnostic performance across the three categories using sensitivity, specificity, accuracy, ROC curves, AUC, precision, and F1 score. For optic disc edema, the model achieved 100% sensitivity, 99% specificity, 99% accuracy, and an AUC of 0.995 (95% confidence interval: 0.98–1.00). The precision for this class was 96%, and the F1 score was 98%. For pseudopapilledema detection, the model showed 98% sensitivity, 97% specificity, 98% accuracy, and an AUC of 0.983 (95% CI: 0.96–1.00). Precision and F1 score for this class were 96% and 97%, respectively. For normal optic discs, the model achieved 96% sensitivity, 100% specificity, 98% accuracy, and an AUC of 0.973 (95% CI: 0.94–1.00). In this class, precision was 100% and F1 score was 98%. These metrics are summarized in Table 4 and visualized in Fig. 5.

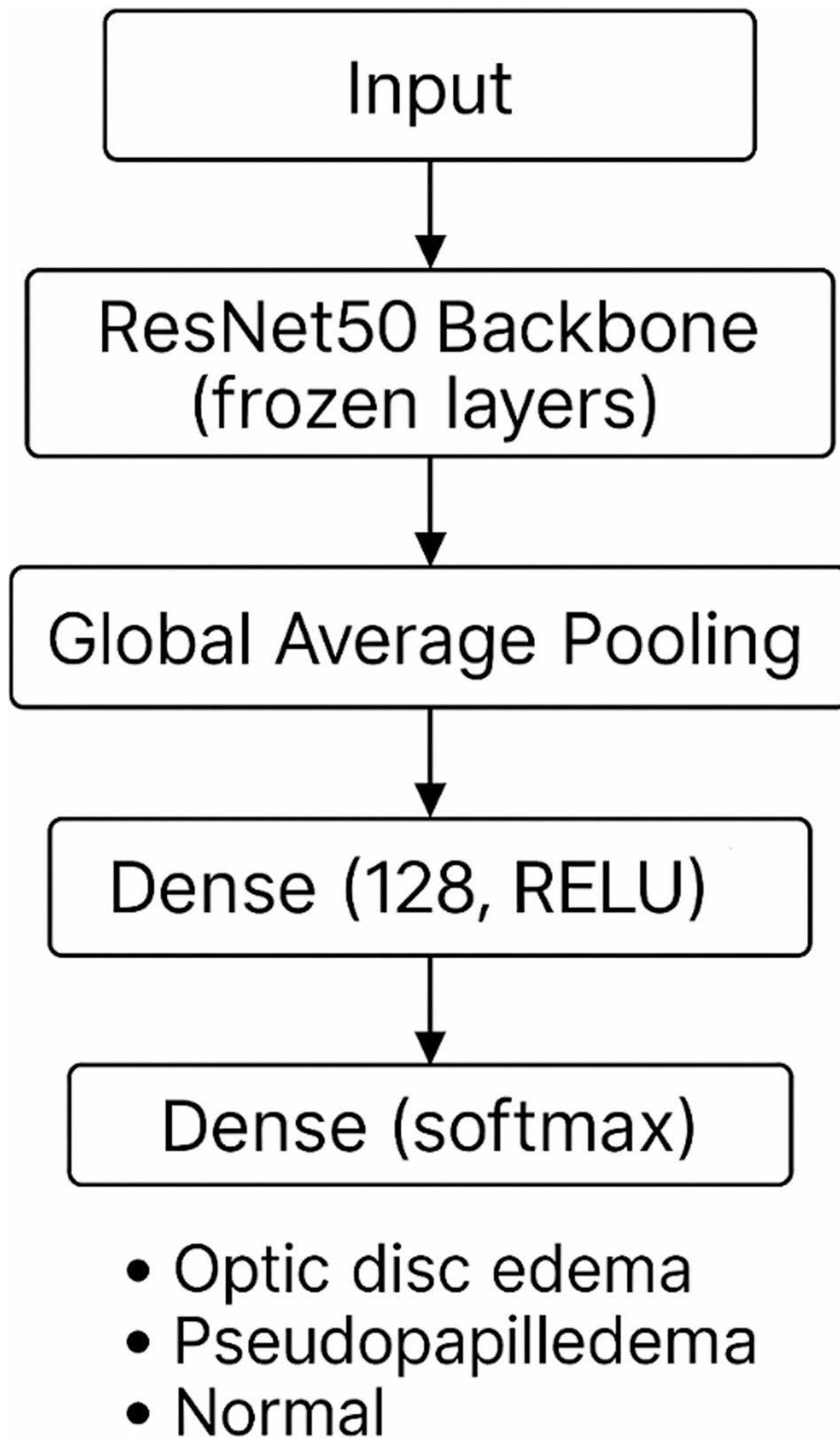


Fig. 2 Schematic diagram of the DL model architecture. The model comprises an input layer, a ResNet50 backbone with frozen layers, global average pooling, a dense layer with 128 units and ReLU activation, and a softmax output layer for classification into three categories: optic disc edema, pseudopapilledema, and normal

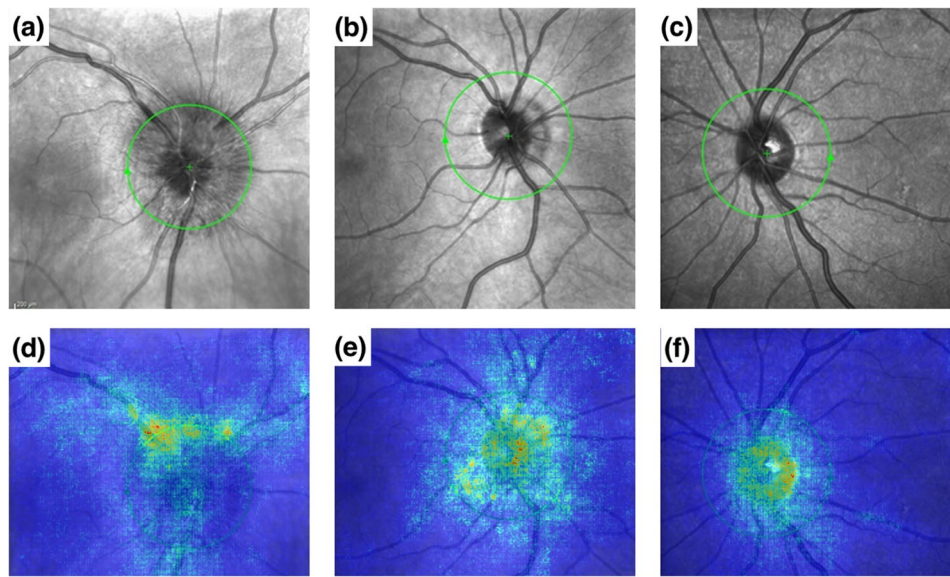


Fig. 3 Representative NIR images and their SmoothGrad saliency maps. **a** Optic disc edema, **(b)** Pseudopapilledema, **(c)** Normal optic disc, **(d)** Optic disc edema – SmoothGrad, **(e)** Pseudopapilledema – SmoothGrad, **(f)** Normal – SmoothGrad. SmoothGrad visualizations demonstrate the regions most influential for the model’s predictions. Warmer colors indicate higher contribution to the decision, showing that the model primarily focuses on the optic nerve head and peripapillary areas when classifying images

Table 3 Demographic characteristics of study groups

Group	Mean Age ± SD	Median Age (Min–Max)	Male (n, %)	Female (n, %)
Optic Disc Edema	57.08 ± 5.16	56.50 (46–70)	24 (40.0%)	36 (60.0%)
Pseudopapilledema	32.84 ± 20.20*	29.50 (7–77)	52 (39.4%)	80 (60.6%)
Normal	52.80 ± 25.99	55 (9–94)	68 (40.5%)	100 (59.5%)

**p* < 0.001, significant difference in age distribution between the pseudopapilledema group and both the optic disc edema and normal control groups

Table 4 Evaluation of classification metrics for optic disc Edema, Pseudopapilledema, and normal images

Group	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)	F1-score (%)
Optic disc edema	100.0	99.0	96.0	99.2	98.0
Pseudopapilledema	98.1	97.3	96.2	97.6	97.1
Normal optic disc	96.0	100.0	100.0	98.4	98.0

The obtained model, based on a ResNet50 architecture with transfer learning, successfully classified NIR images into optic disc edema, pseudopapilledema, and normal categories, achieving excellent performance on the internal test set; however, it has not yet been validated on external, unseen datasets.

Figure 6 illustrates the learning curves. The training and validation losses declined in parallel until approximately epoch 60, after which the validation loss plateaued and began to rise slightly. Early stopping was therefore

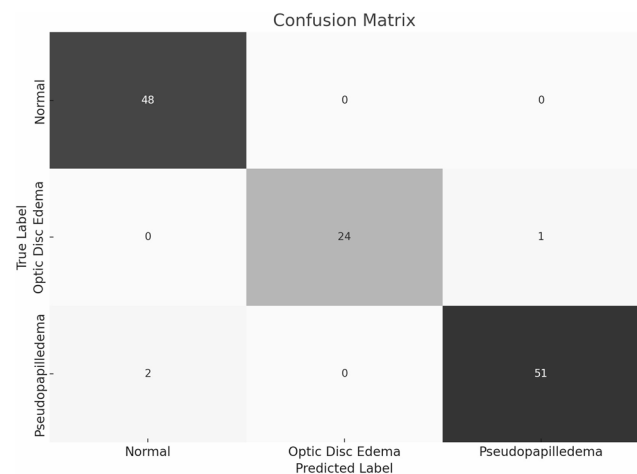


Fig. 4 Confusion matrix summarizing the classification performance of the DL model. The matrix displays the number of correctly and incorrectly classified instances for optic disc edema, pseudopapilledema, and normal optic discs. Diagonal cells represent correct classifications, while off-diagonal cells indicate misclassifications

triggered at epoch 64, successfully preventing over-fitting. The accuracy curves stabilised around 0.97–0.98, with only minimal divergence between training and validation sets.

Discussion

The advancement of AI research has significantly increased its role in healthcare, especially in ophthalmology. AI’s ease of use, cost-effectiveness, and accuracy have made it integral. Techniques like color fundus photography, OCT, and computerized visual field testing

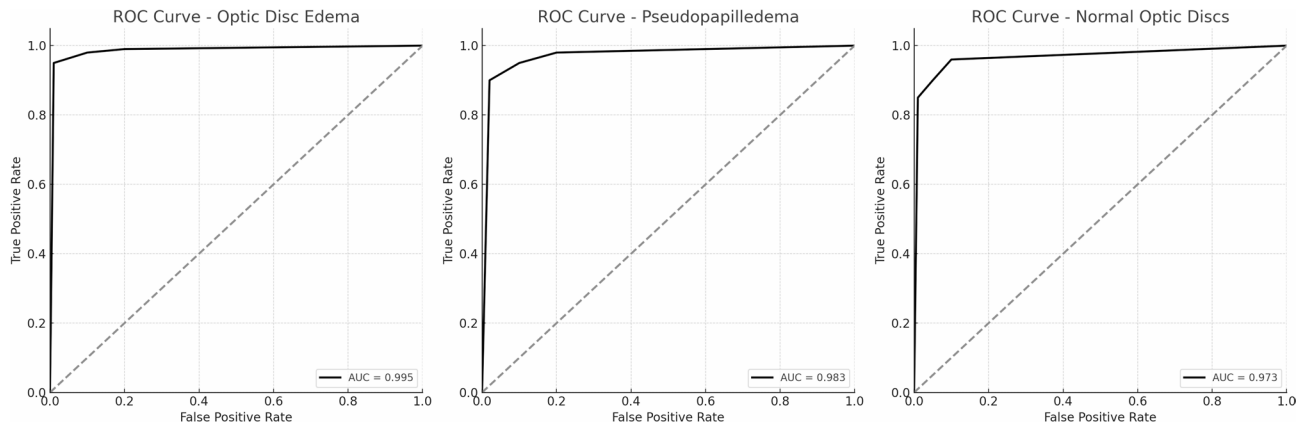


Fig. 5 ROC curves illustrating the diagnostic performance of the DL model in detecting optic disc edema, pseudopapilledema, and normal optic discs. AUC values indicate the model's discriminative ability for each class, with values closer to 1.0 reflecting higher accuracy

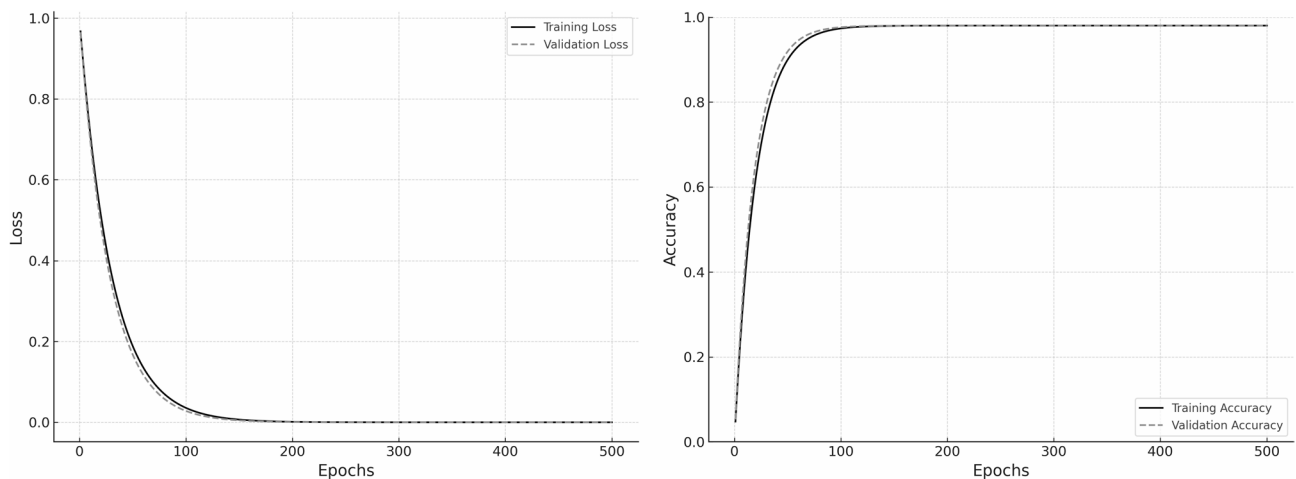


Fig. 6 Training and validation accuracy and loss curves plotted over 500 epochs. The dotted vertical line indicates the epoch (64) at which early stopping was triggered. Panel (a) shows accuracy; panel (b) shows categorical cross-entropy loss

have made ophthalmology a leading AI user in medicine. This success fuels AI studies across healthcare, with DL models making significant progress in addressing vision impairments linked to longer life expectancy. DL models, which have been widely adopted in ophthalmology, continue to demonstrate strong potential in real-world clinical applications [21–24]. Initial studies focus on common causes of vision loss like diabetic retinopathy, age-related macular degeneration, glaucoma, and retinopathy of prematurity. Numerous studies in these areas have achieved high sensitivity and specificity values, some approaching nearly %100 [25–28].

Neuro-ophthalmology has fewer AI studies compared to other subfields of ophthalmology [29]. The main reasons for this are the lower prevalence and heterogeneous nature of neuro-ophthalmological conditions, leading to a lack of sufficient data. Unlike retinal diseases, where color fundus photography and macular OCT imaging are often sufficient for diagnosis, neuro-ophthalmological

conditions frequently require more complex and varied diagnostic tools.

The necessity of AI in neuro-ophthalmology stems from several factors. The shortage of specialists makes accessing expert diagnosis challenging. Additionally, the diagnostic process is complex and time-consuming, often requiring integration of various clinical and imaging data. AI has the potential to significantly enhance diagnostic accuracy and efficiency in this field. Failure to detect optic disc edema can lead to severe consequences, while misidentification can result in unnecessary investigations. The challenge lies in distinguishing mild disc swelling, often requiring multiple clinical tests. DL algorithms aim to objectively interpret optic disc appearance by reducing the variability of individual interpretations.

Initial AI studies in neuro-ophthalmology predominantly utilized classical machine learning (ML) methods, focusing on differentiating normal optic discs from those with papilledema and staging papilledema. For instance, Echegaray et al. analyzed fundus photographs

Table 5 Comparative performance of our model and three prior neuro-ophthalmic AI studies

Study (year)	Imaging modality	Clinical task/classest†	Test images (n)	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)
Our model	Near Infrared reflectance	ODE, PPE, normal	126	98.4	0.995 (ODE); 0.983 (PPE) 0.973 (N)	100 (ODE); 98 (PPE); 96 (N)	99 (ODE); 97 (PPE); 100 (N)
Milea et al., 2020 [21]	Dilated color fundus photographs	Papilledema vs. normal+other	1505	—	0.96	96.4	84.7
Ahn et al., 2019 [33]	Fundus photographs	Optic-neuropathy, PPE, normal	219	98.6 (ResNet)	0.999 (ResNet)	NR	NR
Lin et al., 2024 [35]	Mydriatic color fundus photographs	Papilledema vs. normal+other (pediatric)	898	89.6	0.99	98.0	94.1

ODE Optic Disc Edema, PPE Pseudopapilledema, N Normal, NR Not Reported, AUC Area Under the Curve

†Class labels shown as reported; “normal+other” indicates the one-vs-rest evaluation used by the authors

of papilledematous eyes and found a significant concordance between their model’s staging success and expert neuro-ophthalmologists’ assessments [30]. Similarly, Akbar et al. developed a model that achieved high accuracy in detecting and staging papilledema using fundus images [31]. However, unlike DL methods that enable self-learning, these models relied on classical ML techniques where specific features were manually defined and taught to the algorithm. Fatima et al. created a model for detecting papilledema from color fundus photographs, achieving a sensitivity of 84.1% and specificity of 90.6% [32]. Their approach required extensive manual preprocessing, including optic disc localization, image cropping, and vessel segmentation, to prepare the data for model training. In contrast, our study demonstrates that by utilizing NIR images, the developed DL model surpasses these performance metrics, benefiting from the higher contrast and structural details provided by NIR imaging. Our model achieved a sensitivity of 100% and specificity of 99% in detecting optic disc edema, with comparable success in identifying pseudopapilledema. One of the early AI studies aiming to differentiate pseudopapilledema from optic neuropathies was conducted by Ahn et al. in 2019 [33]. Their model achieved an impressive accuracy ranging from 95.89% to 98.63%, utilizing fundus photographs that were manually resized to a fixed width while maintaining aspect ratio. This manual preprocessing step, while effective, introduces variability and increases the complexity of data preparation. Similarly, Milea et al. developed a CNN that classified normal optic discs, papilledema, and other disc abnormalities using a large, racially diverse dataset of 15 846 fundus images [22]. Their dataset was markedly imbalanced (normal \gg papilledema), yet the model still attained 96.4% sensitivity and 84.7% specificity for papilledema, underscoring that class-imbalance is a recurring methodological challenge in neuro-ophthalmic AI studies. By comparison, our CNN achieved equal or better performance with far fewer images: sensitivities of 100% for optic-disc edema,

98% for pseudopapilledema, and 96% for normal discs; specificities between 97% and 100%; and an overall accuracy of 98–99%. These results, obtained after stratified splitting and minority-focused augmentation to counter our dataset’s mild imbalance, place our model at least on par with the current benchmarks. Unlike the study by Ahn et al., which required manual resizing, our model incorporates automated preprocessing, resizing all images to 200 \times 200 pixels within the model pipeline, thereby streamlining the data preparation process and minimizing user-dependent variability. While the study by Milea et al. utilized an extensive dataset, our model achieved similarly high performance using significantly fewer images (840 images in total, with 126 images in the test set), showcasing the efficiency and robustness of our DL approach. The transferability of our trained model allows for its seamless application in different clinical settings, enabling other researchers and clinicians to adopt the model without the need for extensive retraining. Furthermore, recent research has demonstrated that DL systems can outperform clinicians in identifying optic disc abnormalities, underscoring the potential of AI as a reliable and objective diagnostic tool in neuro-ophthalmology [34]. These findings further underscore the clinical utility of our model and highlight the transformative role of AI in enhancing diagnostic accuracy and reducing interobserver variability.

Our findings are consistent with previous studies that have employed DL techniques to differentiate between optic disc abnormalities, such as the BONSAI group’s approach for pediatric papilledema and the work by Chang et al. which focused on differentiating pediatric pseudopapilledema from true papilledema [35, 36]. To contextualize our results, Table 5 provides a direct comparison between the present model and three previously published AI systems developed for neuro-ophthalmic classification tasks. Despite using a relatively small test set, our model achieved comparable or superior accuracy and AUC values, highlighting the diagnostic potential

of near-infrared OCT imaging and task-specific model optimization.

While DL generally requires a large amount of data, neuro-ophthalmology typically has limited image datasets. We overcame this problem by using TL [20]. Additionally, during the training process, we leveraged well-established frameworks such as TensorFlow, which have demonstrated superior performance in previous studies [37]. This strategic choice further underscores the robustness and reliability of our model.

This study's limitations should be considered. All 840 clear NIR images were acquired on a single Spectralis SD-OCT device at our institution. Scanner settings, acquisition protocols, and local patient demographics may therefore have introduced centre-specific biases, limiting external generalisability. Ground-truth labels were assigned by a single fellowship-trained neuro-ophthalmologist but were anchored to multimodal clinical assessments (fundoscopy, OCT B-scans, visual fields, and, when indicated, neuro-imaging). Although this approach ensured internal consistency, it may still suffer from observer bias; future studies should rely on consensus grading by a panel of experts or at least dual-masked adjudication to enhance labelling reliability and reduce inter-observer variability. Although the class distribution met the "mild imbalance" criterion and we applied stratified splitting with minority-focused augmentation, residual imbalance-related bias cannot be excluded. To enlarge the dataset we included longitudinal images from the same eyes; while images from a given eye and time point were never split across training and test sets, this strategy may still reduce sample independence. Finally, diagnostic performance was estimated on a single 15% internal test split ($n = 126$), which is relatively small; the reported 100% sensitivity and 99% specificity for optic disc edema must therefore be interpreted with caution. To address these limitations we are negotiating data-sharing agreements with at least three tertiary centres that use different OCT platforms (CIRRUS HD-OCT 5000, Topcon DRI Triton, Spectralis OCT2). The resulting prospective cohort (~2,000 images) will allow device-heterogeneous, multi-centre external validation and robust multi-split cross-validation. We also aim to deposit an anonymised subset in an open-access repository to facilitate independent benchmarking. Until such evidence becomes available, the present algorithm should be regarded as proof-of-concept and used only as a decision-support aid rather than an autonomous diagnostic system. Additionally, the optic disc edema category in our dataset included multiple underlying etiologies (e.g., papilledema, NAION, RVO), which may exhibit distinct morphological patterns. Due to the limited sample size for each etiology, class-wise performance analysis within the optic disc edema group was not feasible. Future

studies with larger and more balanced etiology-specific datasets are needed to evaluate intra-group variability and enhance clinical specificity.

In this study, a DL model based on NIR images was developed, demonstrating its effectiveness in distinguishing optic disc edema from pseudopapilledema. However, future research should incorporate several key advancements to further enhance the model's accuracy and generalizability. Firstly, integrating different imaging modalities (e.g., fundus autofluorescence, superimposed OCT, or ultra-widefield fundus imaging) could improve the model's versatility. The incorporation of various imaging techniques and biomarkers relevant to optic nerve pathologies into AI systems could further enhance diagnostic accuracy. Additionally, the use of multimodal AI systems could strengthen the decision-making process by combining patient-specific clinical data with imaging findings. Integrating supplementary information such as biomarkers, genetic data, or patient history into the AI model could facilitate a more personalized and reliable diagnostic approach. Finally, prospective, multi-center studies are essential to validate the model's effectiveness across diverse populations and clinical settings. Such studies are necessary to assess the model's robustness and facilitate its transition into clinical practice. In addition to its diagnostic accuracy, the proposed model has potential clinical utility as a decision-support tool. By integrating the model into OCT viewing software, clinicians could receive real-time alerts for suspected optic disc edema, prompting further evaluation. Given its relatively compact architecture and fast inference speed, the proposed model could be feasibly embedded into existing OCT systems for real-time clinical decision support. Most commercial OCT platforms already support programmable application interfaces that allow integration of custom algorithms. By incorporating the model into the OCT viewing software, the system could automatically generate a classification output immediately after image acquisition. This real-time feedback could assist clinicians in triaging patients, initiating further tests when needed, and reducing diagnostic delays. Moreover, the system could highlight regions of interest or confidence scores, further aiding interpretation in non-specialist settings. Future collaborations with OCT manufacturers would be essential to validate and deploy such integrated solutions in clinical environments. With further validation, the model could assist non-specialists in triaging referrals or flagging urgent cases, thereby improving workflow efficiency and patient care.

In summary, while AI shows promising results in ophthalmology, its current use is limited by dataset dependence, device variability, and the lack of large-scale external validation. Future studies should therefore prioritize multi-center collaborations, multimodal integration,

and prospective validation to enable the safe and effective translation of AI into routine neuro-ophthalmic practice.

Conclusion

In conclusion, our study demonstrates that the developed DL model can identify optic disc pathologies with high accuracy, even with a small sample size. This study may contribute to improving diagnostic efficiency, optimizing clinical decision-making, and ultimately reducing health-care costs. Integrating the DL model into clinical workflows can prevent medical errors and serve as a clinical decision support system, enhancing patient safety and reducing clinician workload and burnout. More broadly, AI has shown significant potential across many areas of ophthalmology, and particularly in neuro-ophthalmology, where diagnostic processes are complex and specialist access is limited. AI-based systems can enhance diagnostic accuracy, reduce interobserver variability, and facilitate earlier interventions, making them a valuable decision-support tool.

Abbreviations

AI	Artificial Intelligence
AUC	Area Under the Curve
CI	Confidence Interval
CNN	Convolutional Neural Networks
DL	Deep Learning
LLM	Large Language Model
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NAION	Non Arteritic Ischemic Optic Neuropathy
NIR	Near-Infrared Reflectance
NMO	Neuromyelitis Optica
NO	Neuro-ophthalmologist
OCT	Optical Coherence Tomography
ODD	Optic Disc Drusen
ONH	Optic Nerve Head
PHOMS	Peripapillary Hyperreflective Ovoid Mass-like Structures
ROC	Receiver Operating Characteristic
RVO	Retinal Vein Occlusion
SD-OCT	Spectral Domain Optical Coherence Tomography
TL	Transfer Learning
UWF	Ultrawide-field
VEP	Visual Evoked Potential

Acknowledgements

Not applicable.

Authors' contributions

C.O. and F.K. conceptualized the study. C.O., E.A., and G.E. developed the methodology. Formal analysis and investigation were carried out by C.O. and F.K. The original draft was written by C.O. and E.A. Review and editing were performed by B.T. and M.H.O. All authors reviewed and approved the final manuscript.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Bezmi Alem Vakif University (Decision No: 2022/406). Informed consent was obtained from all participants in accordance with ethical guidelines prior to their inclusion in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 May 2025 / Accepted: 22 September 2025

Published online: 22 October 2025

References

- Urfalioglu S, Ozdemir G, Guler M, Duman GG. The evaluation of patients with optic disc edema: A retrospective study. *North Clin Istanbul*. 2021;8(3):280–5. <https://doi.org/10.14744/nci.2020.25483>.
- Van Stavern GP. Optic disc edema. *Semin Neurol*. 2007;27(3):233–43. <https://doi.org/10.1055/s-2007-979684>.
- Szanto D, Wang J, Woods B, et al. Optic nerve atrophy conditions associated with 3D unsegmented optical coherence tomography volumes using deep learning. *JAMA Ophthalmol*. 2025. <https://doi.org/10.1001/jamaophthalmol.2025.2766>. Published online August 21, 2025.
- Madadi Y, Delsoz M, Lao PA, Fong JW, Hollingsworth TJ, Kahook MY, Yousefi S. ChatGPT assisting diagnosis of neuro-ophthalmology diseases based on case reports. *J Neuroophthalmol*. 2025;45(3):301–6. <https://doi.org/10.1097/WNO.0000000000002274>.
- Gu S, Bao T, Wang T, Yuan Q, Yu W, Lin J, Zhu H, Cui S, Sun Y, Jia X, Huang L, Ling S. Multimodal AI diagnostic system for neuromyelitis Optica based on ultrawide-field fundus photography. *Front Med (Lausanne)*. 2025;12:1555380. <https://doi.org/10.3389/fmed.2025.1555380>.
- Gungor A, Tang Z, Loo JL, Choon STL, Singhal S, Ming RFC, Tadayoni L, Sarbout I, Newman NJ, Biousse V, Najjar RP, Milea D, BONSAI Group. Deep learning-based detection of papilledema on retinal photographs from handheld cameras: a prospective study. *J Neuroophthalmol*. 2025. <https://doi.org/10.1097/WNO.0000000000002394>. Epub ahead of print Aug 28.
- Szanto D, Erekat A, Woods B, Wang JK, Garvin M, Johnson BA, Kardon R, Linton E, Kupersmith MJ. Deep learning approach readily differentiates papilledema, non-arteritic anterior ischemic optic neuropathy, and healthy eyes. *Am J Ophthalmol*. 2025;276:99–108. <https://doi.org/10.1016/j.ajo.2025.04.006>.
- Faes L, Wagner SK, Fu DJ, Liu X, Korot E, Ledsam JR, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *Lancet Digit Health*. 2019;1(5):e232–42. [https://doi.org/10.1016/S2589-7500\(19\)30108-6](https://doi.org/10.1016/S2589-7500(19)30108-6).
- Friedman DI, Liu GT, Digre KB. Revised diagnostic criteria for the pseudotumor cerebri syndrome in adults and children. *Neurology*. 2013;81(13):1159–65. <https://doi.org/10.1212/WNL.0b013e3182a55f17>.
- Hayreh SS. Ischemic optic neuropathy. *Prog Retin Eye Res*. 2009;28(1):34–62. <https://doi.org/10.1016/j.preteyeres.2008.11.002>.
- Hayreh SS. Prevalent misconceptions about acute retinal vascular occlusive disorders. *Prog Retin Eye Res*. 2005;24(4):493–519. <https://doi.org/10.1016/j.preteyeres.2005.03.002>.
- Toosy AT, Mason DF, Miller DH. Optic neuritis. *Lancet Neurol*. 2014;13(1):83–99. [https://doi.org/10.1016/S1474-4422\(13\)70259-X](https://doi.org/10.1016/S1474-4422(13)70259-X).
- Sinclair SH, Schwartz SS, Watsky MA. Diabetic papillopathy: a clinical entity. *Arch Ophthalmol*. 1981;99(4):506–12. <https://doi.org/10.1001/archophth.1981.03930010508012>.
- Malmqvist L, Bursztyn L, Costello F, et al. Optic disc drusen: Understanding an old problem from a new perspective. *J Neuroophthalmol*. 2018;38(4):498–508. <https://doi.org/10.1097/WNO.0000000000000698>.
- Lee S, Kim JH, Hwang JM. Peripapillary hyperreflective ovoid mass-like structures (PHOMS) in various optic nerve head anomalies and diseases. *Ophthalmology*. 2020;127(8):1108–19. <https://doi.org/10.1016/j.ophtha.2020.02.013>.

16. Google Developers. Imbalanced Data [Internet]. Mountain View (CA): Google. 2023. Available from: <https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data>. Accessed 28 Jun 2025.
17. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342–50. <https://doi.org/10.1038/s41591-018-0107-6>.
18. Vayrynen J, Wuori E, Solares JRA, Xiao C, Shetty S, Shah N. Kaggle in health-care research: a systematic review. *Appl Sci*. 2024;14(10):1234. <https://doi.org/10.3390/app14101234>.
19. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in health-care: past, present and future. *Stroke Vasc Neurol*. 2017;2:230–43. <https://doi.org/10.1136/svn-2017-000101>.
20. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2009;22(10):1345–59. <https://doi.org/10.1109/TKDE.2009.191>.
21. Milea D, Najjar RP, Zhuho J, et al. BONSAI Group. Artificial intelligence to detect papilledema from ocular fundus photographs. *N Engl J Med*. 2020;382(18):1687–95. <https://doi.org/10.1056/NEJMoa1917130>.
22. Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, Vardoulakis LM. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020:1–12. <https://doi.org/10.1145/3313831.33767>
23. Liu H, Li L, Wormstone IM, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. *JAMA Ophthalmol*. 2019;137(12):1353–60. <https://doi.org/10.1001/jamaophthalmol.2019.3501>.
24. Arcadu F, Benmansour F, Maunz A, et al. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit Med*. 2019;2:92. <https://doi.org/10.1038/s41746-019-0172-3>.
25. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318(22):2211–23. <https://doi.org/10.1001/jama.2017.18152>.
26. Grassmann F, Mengelkamp J, Brandl C, et al. A deep learning algorithm for prediction of Age-Related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology*. 2018;125(9):1410–20. <https://doi.org/10.1016/j.ophtha.2018.02.037>.
27. 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):1199–206. <https://doi.org/10.1016/j.ophtha.2018.01.023>.
28. Redd TK, Campbell JP, Brown JM, Kim SJ, Ostmo S, Chan RVP, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J Ophthalmol*. 2019;103(5):580–4. <https://doi.org/10.1136/bjophthalmol-2018-313156>.
29. Leong YY, Vasseneix C, Finkelstein MT, Milea D, Najjar RP. Artificial intelligence Meets neuro-ophthalmology. *Asia Pac J Ophthalmol (Phila)*. 2022;11(2):111–25. <https://doi.org/10.1097/APO.0000000000000512>.
30. Echegaray S, Zamora G, Yu H, Luo W, Soliz P, Kardou R. Automated analysis of optic nerve images for detection and staging of papilledema. *Invest Ophthalmol Vis Sci*. 2011;52(10):7470–8. <https://doi.org/10.1167/iov.11-7484>.
31. Akbar S, Akram MU, Sharif M, Tariq A, Yasin UU. Decision support system for detection of papilledema through fundus retinal images. *J Med Syst*. 2017;41(4):66. <https://doi.org/10.1007/s10916-017-0712-9>.
32. Fatima KN, Hassan T, Akram MU, Akhtar M, Butt WH. Fully automated diagnosis of papilledema through robust extraction of vascular patterns and ocular pathology from fundus photographs. *Biomed Opt Express*. 2017;8(2):1005–24. <https://doi.org/10.1364/BOE.8.001005>.
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. <https://doi.org/10.1186/s12886-019-1184-0>.
34. Vasseneix C, Nusinovi S, Xu X, BONSAI Brain and Optic Nerve Study With Artificial Intelligence Group, et al. Deep learning system outperforms clinicians in identifying optic disc abnormalities. *J Neuroophthalmol*. 2023;43(2):159–67. <https://doi.org/10.1097/WNO.0000000000001800>.
35. Lin MY, Najjar RP, Tang Z, BONSAI Brain and Optic Nerve Study with Artificial Intelligence group, et al. The BONSAI brain and optic nerve study with artificial intelligence deep learning system can accurately identify pediatric papilledema on standard ocular fundus photographs. *J AAPOS*. 2024;28(1):103803. <https://doi.org/10.1016/j.jaapos.2023.10.005>.
36. Chang MY, Heidary G, Beres S, et al. Artificial intelligence to differentiate pediatric pseudopapilledema and true papilledema on fundus photographs. *Ophthalmol Sci*. 2024;4(4):100496. <https://doi.org/10.1016/j.xops.2024.100496>.
37. Liu J, Dutta J, Li N, Kurup U, Shah M. Usability study of distributed deep learning frameworks for convolutional neural networks. In: Deep Learning Day at SIGKDD Conference on Knowledge Discovery and Data Mining; 2018.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.