Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD30009

Deep Learning in Glaucoma with Fundus Photography

Yidan Zhang

Glasgow College, University of Electronic Science and Technology of China, Chengdu, China
2024190905011@std.uestc.edu.cn

Abstract. Deep learning (DL) is a key branch of artificial intelligence (AI). It has made remarkable progress in medical imaging, particularly in image classification and pattern recognition. In ophthalmology, the application of DLto fundus images for glaucoma assessment has become a rapidly developing research area. In recent years, this approach has represented promising results in terms of both analytical efficiency and accuracy, assisting in the differentiation of glaucoma patients from healthy eyes. This trend suggests that DL technology has the potential to improve existing diagnostic and treatment practices and optimize the clinical workflow for glaucoma diagnosis. However, challenges remain, such as a shortage of high-quality annotated data and limited model interpretability. This article reviews recent research progress on DL-based glaucoma assessment using fundus images, explores its potential clinical significance, and proposes future research directions. Through a systematic review of relevant literature, this study provides a comprehensive understanding of the application of DL in glaucoma diagnosis for both academia and clinical practice, and offers reference and guidance for future research.

Keywords: Artificial intelligence, Deep learning, Gaucoma

1. Introduction

As the leading cause of irreversible blindness globally, glaucoma is often associated with elevated intraocular pressure (IOP). It is marked by injury to the optic nerve head (ONH) and retinal nerve fiber layer (RNFL), resulting in peripheral and, at times, central vision loss [1, 2]. The cup-to-disc ratio in patients with glaucoma is often smaller than that in healthy individuals, and the RNFL is thinner [2].

Artificial intelligence (AI) has been widely applied in medical image examination these days. Through studies, it is shown that various AI methods can reach high specificity and sensitivity in glaucoma detection, whether applied to structural modalities like optical coherence tomography (OCT) and fundus photography (FP) or to functional modalities like visual field (VF) testing. AI algorithms fall into two distinct categories: conventional machine learning (ML) algorithms and deep learning (DL) [3-5]. The challenge with conventional ML lies in selecting which features are critical within each image. Given the substantial differences in the shape and size of pathological manifestations among individuals, extracting features proves to be a challenging task. Conventional ML also exhibits limited generalization capabilities. In contrast, DL employs an end-to-end learning process that takes labeled datasets as input and produces classifications as output. Multi-layer

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD30009

nonlinear information processing is employed by DL models for the purposes of feature extraction, transformation, pattern analysis, and classification. Consequently, these models possess the advantage of automatically identifying relevant patterns within images, rather than requiring domain experts to manually craft optimal features. Theoretically, through automatic feature learning and large-scale modeling capabilities, DL can achieve superior generalization performance when trained on diverse datasets.

2. Global image methods

This approach takes the entire fundus image as input to the model for training and inference. The model automatically learns discriminative features from the global image without relying on prior anatomical structure information. Its advantages include a simplified preprocessing pipeline and reduced deployment overhead; however, the features extracted by the model often suffer from limited clinical interpretability.

Raghavendra et al. developed an 18-layer convolutional model that demonstrated high classification ability while eliminating the need for handcrafted feature extraction or additional preprocessing, as it directly processed raw images. Using a dataset from Kasturba Medical College (Manipur, India), the model achieved a mean accuracy of 95.60%, with sensitivity, specificity, and positive predictive value of 95.50%, 95.10%, and 96.96%, respectively [6].

Existing DL models has the “black box” nature, a factor contributing to a decline in physician trust. Deperlioglu et al. introduced an approach that incorporates image processing and DL , enhanced with explainable artificial intelligence (XAI). This approach aims to guarantee the credibility of decisions made during AI-based glaucoma diagnosis. Image processing techniques, including Histogram Equalization (HE) and Contrast-Limited Adaptive Histogram Equalization (CLAHE), were applied to enhance color fundus image data. To facilitate diagnosis, the enhanced image data was entered into an explainable CNN. The achievement of XAI required the implementation of Category Activation Maps (CAM), which facilitates the interpretation of the CNN’s image analysis through heatmap-based methods. A comparative analysis was conducted on the Drishti-GS, ORIGA-Light, and HRF datasets. The results indicated that ORIGA-Light exhibited optimal performance, with an average accuracy of 93.5% and a sensitivity of 97.7%. A group of fifteen physicians evaluated the effectiveness of the XAI, and their findings indicated high rates of concordance between conventional and AI-based methods. This enhancement in physician confidence provides a reliable solution for automated glaucoma diagnosis [7].

3. Structure-based methods

The core concept of this approach is to focus on anatomical structures about glaucoma and subsequently make diagnoses based on these structures. Common strategies covers two main aspects (1) segmenting the optic disc and optic cup to calculate a structural metric, namely the cup-to-disc ratio (2) estimating the retinal nerve fiber layer thickness (RNFT) from fundus images These methods more directly reflect the pathological characteristics of the disease and generally yield more interpretable results. However, algorithms that estimate RNFLT typically require a substantial amount of Spectral-Domain Optical Coherence Tomography (SD-OCT) data during the training stage.

DOI: 10.54254/2755-2721/2025.LD30009

3.1. Model based on cup-to-disc ratio

The cup-to-disc ratio is a significant metric that can be readily assessed from FPs. Nowadays, numerous DL algorithms have been developed using regions of interest (ROIs) centered on the optic disc, mainly for capturing this crucial structure, the cup-to-disc ratio.

3.1.1. Coarse-crop-based model

Coarse crop–based models crop a region centered on the optic disc that coarsely encompasses the target structure. Chen et al. have proposed a six-layer convolutional DL model. Utilizing the ORIGA and SCES datasets, they incorporated data-augmented ROI images as input to the proposed deep CNN, thereby achieving a substantial reduction in processing time compared to segmentation discs and cups. The dropout method was implemented during the training process, and the final softmax classifier yielded area under the curve (AUC) values of 0.831 and 0.887 on two test datasets [8].

Transfer learning has shown advantages in diagnosing diabetic eye disease. As part of their study, Mark Christopher and his fellow researchers examined the effectiveness of various DL architectures and transfer learning for detecting glaucomatous optic neuropathy (GON) and FPs. The evaluation process involved the analysis of three classical CNN architectures: VGG16, Inception v3, and ResNet50. Each architecture was assessed in both “native training” and “transfer learning” versions. The inputs comprised “cropped and normalized ONH region images.” The results demonstrated that transfer learning versions exhibited a substantial enhancement in performance when compared to native versions across all architectures [9].

Naoto Shibata et al. employed CFPs captured at KOWANonmyd WX Matsue Red Cross Hospital as the dataset, with test data sourced from the medical faculties of the University of Tokyo and Kitasato University. The Hofstede transformation was employed for cropping, and data augmentation was applied to the optic disc region as input to train ResNet 18. This approach resulted in an AUC of 0.965 across all test data [10].

Using medical images and deep learning to diagnose diseases often faces challenges of insufficient samples or inadequate labeling. However, semi-supervised and weakly supervised learning can uncover patterns within limited data. Zhao et al. proposed a weakly supervised multi-task learning (WSMT) framework that can perform three tasks simultaneously: glaucoma diagnosis based on ODH line patterns, evidence identification, and optic disc segmentation. This framework uses only binary classification labels. This framework comprises four components: (1) a novel CNN with skip connections and dense blocks that automatically captures multi-scale feature representations, (2) a pyramid ensemble architecture that learns high-resolution evidence maps solely from diagnostic labels through multi-layer global pooling and activation pyramid mapping, which enables evidence identification and segmentation, (3) a deep neural network named the Constrained Clustering Branch (CCB) that segments the optic disc, and (4) a fully connected discriminator that automates glaucoma diagnosis. The proposed framework is a tree-like network architecture that uses three branches to perform the tasks of evidence identification, glaucoma diagnosis, and optic disc segmentation while sharing feature representations constructed by the CNN backbone. This architecture achieves higher-level diagnostic tasks guiding lower-level localization/segmentation tasks [11].

DOI: 10.54254/2755-2721/2025.LD30009

3.1.2. Model based on precise segmentation

Models based on precise segmentation segment the optic disc and optic cup regions, and then use the resulting masks either to extract structural features or to train subsequent diagnostic models [12].

Chen et al. proposed a neural network convolutional model named C-CNN and trained it on the ORIGA and SCES datasets. The researchers employed peripapillary atrophy (PPA) removal and elliptical fitting segmentation to obtain a clean optic disc (OD) region. The model is composed of six layers, including five multi-layer perceptron (MLP) layers, convolutional layers and one fully connected layer. It incorporates response normalization and overlapping pooling layers to mitigate the risk of overfitting. The innovation in the MLPconv layers involves the replacement of traditional CNN linear filters with an MLP, in conjunction with the employment of ReLU activation to achieve complex nonlinear transformations. This enhancement in data capture facilitates the detection of hidden pathological patterns in the fundus. C-CNN employs a context-based training strategy, diverging from conventional independent CNN training. For instance, the 5 C-CNN model consists of five CNNs concatenated in series, where the convolutional layer output from the preceding CNN serves as contextual input to the fully connected layer of the subsequent CNN. The discriminative power of features is enhanced by the final glaucoma prediction generated through the softmax classifier in the last CNN layer. The findings of the study indicate that the 5-C-CNN model demonstrates the optimal performance across the datasets, with an average AUC of 0.833 on ORIGA and 0.890 on SCES [13].

Chai et al. developed a fully automated multi-branch neural network (MB-NN). The dataset has fundus images from different machines at Beijing Tongren Hospital. The network extracts global and ROI - specific local features simultaneously. The first branch uses a CNN to extract features from the whole image. Faster - RCNN was used to outline the optic disc region for the second branch's input. Subsequently, CNNs were employed to extract locally significant features. For the third branch, a fully convolutional network (FCN) model was employed to perform segmentation of the optic disc, cup, and PPA regions, followed by metric calculation [14].

3.2. Model based on RNFLT

RNFLT is another important biomarker for glaucoma diagnosis in ML studies, showing high AUC values with low standard error [15]. It can be measured using SD-OCT. However, although OCT is broadly viewed as the benchmark in ophthalmology, its high cost restricts its availability mainly to large ophthalmic centers and research laboratories. Some DL algorithms focus on how to measure it using FPs. Medeiros et al developed and validated a DL algorithm trained on SD-OCT data to quantitatively assess optic nerve damage from FPs. The dataset comprised 32,820 paired optic disc photographs and SD-OCT RNFL scans from 1,198 subjects (2,312 eyes), with 80% allocated for training and validation and 20% reserved for testing. Using SD-OCT–derived mean RNFLT as the reference standard, the model was constructed based on the ResNet-34 architecture. Training incorporated image preprocessing and data augmentation, and heatmaps were employed to identify salient regions. The results exhibited a strong correlation between model-predicted RNFLT and actual measurements, with a mean absolute error (MAE) of 7.39 μm. During differentiating glaucoma from healthy eyes, the model attained an AUC of 0.944—on par with SD-OCT—boasting a classification accuracy of 83.7%. This approach mitigates the limitations of manual annotation and provides a low-cost solution for glaucoma screening, although approximately 30% of the measurement variance remained unexplained [16].

DOI: 10.54254/2755-2721/2025.LD30009

4. Conclusions

Although DL algorithms show considerable promise for diagnosing glaucoma through fundus photography, several limitations remain in the current literature. First, data-related issues remain, such as dataset heterogeneity, class imbalance, and limited labeling. These weaken generalizability. Second, most models lack sufficient explainability. DL algorithms often operate as "black boxes". Although with the aid of explainable AI frameworks, their clinical interpretability remains limited. Third, clinical adaptability requires further improvement. Models trained on retrospective data perform poorly when facing variations in image quality, patient populations, and comorbidities. Finally, cross-modality prediction of RNFLT remains a challenge. Thickness estimation from fundus images continues to exhibit high variance, reflecting the difficulty of capturing complex pathological mechanisms.

Looking ahead, future research should prioritize the development of large, diverse, and well-annotated multi-center datasets to enhance generalizability. Integrating multimodal data from OCT, fundus photography, and visual field testing, could improve diagnostic accuracy by capturing both structural and functional aspects of glaucoma. Advances in semi-supervised, weakly supervised, and federated learning hold promise for overcoming limitations of data scarcity and privacy concerns. Additionally, explainable AI frameworks and physician-in-the-loop systems will be crucial for fostering clinical trust and facilitating real-world translation. Ultimately, the future direction lies in building robust, interpretable, and clinically deployable AI systems that can support early detection, monitoring, and personalized management of glaucoma on a global scale.

References

[1] Jayaram, H., Kolko, M., Friedman, D. S., & Gazzard, G. (2023). Glaucoma: Now and beyond. The Lancet, 402(10414), 1788–1801. https: //doi.org/10.1016/S0140-6736(23)01523-7

[2] Schuster, A. K., Erb, C., Hoffmann, E. M., Dietlein, T., & Pfeiffer, N. (2020). The diagnosis and treatment of glaucoma. Deutsches Ärzteblatt International, 117(13), 225–234. https: //doi.org/10.3238/arztebl.2020.0225

[3] Mursch-Edlmayr, A. S., Ng, W. S., Diniz-Filho, A., et al. (2020). Artificial intelligence algorithms to diagnose glaucoma and detect glaucoma progression: Translation to clinical practice. Translational Vision Science & Technology, 9(2), 55. https: //doi.org/10.1167/tvst.9.2.55

[4] Ashtari-Majlan, M., Dehshibi, M. M., & Masip, D. (2023). Deep learning and computer vision for glaucoma detection: A review. arXiv preprint arXiv: 2307.16528. https: //doi.org/10.48550/arXiv.2307.16528

[5] Panwar, N., Huang, P., Lee, J., Keane, P. A., Chuan, T. S., Richhariya, A., ... & Agrawal, R. (2016). Fundus photography in the 21st century—A review of recent technological advances and their implications for worldwide healthcare. Telemedicine and e-Health, 22(3), 198–208. https: //doi.org/10.1089/tmj.2015.0068

[6] Raghavendra, U., Fujita, H., Bhandary, S. V., Gudigar, A., Tan, J. H., & Acharya, U. R. (2018). Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Information Sciences, 441, 41–49. https: //doi.org/10.1016/j.ins.2018.02.074

[7] Deperlioglu, O., Kose, U., Gupta, D., Khanna, A., Giampaolo, F., & Fortino, G. (2022). Explainable framework for glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation. Future Generation Computer Systems, 129, 152–169. https: //doi.org/10.1016/j.future.2021.11.011

[8] Chen, X., Xu, Y., Wong, D. W. K., Wong, T. Y., & Liu, J. (2015, August). Glaucoma detection based on deep convolutional neural network. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 715–718). IEEE. https: //doi.org/10.1109/EMBC.2015.7318392

[9] Christopher, M., Belghith, A., Bowd, C., Proudfoot, J. A., Goldbaum, M. H., Weinreb, R. N., ... & Zangwill, L. M. (2018). Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Scientific Reports, 8(1), 16685. https: //doi.org/10.1038/s41598-018-35044-9

[10] Shibata, N., Tanito, M., Mitsuhashi, K., Fujino, Y., Matsuura, M., Murata, H., & Asaoka, R. (2018). Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Scientific Reports, 8(1), 14665. https: //doi.org/10.1038/s41598-018-32861-5

DOI: 10.54254/2755-2721/2025.LD30009

[11] Zhao, R., Liao, W., Zou, B., Chen, Z., & Li, S. (2019, July). Weakly-supervised simultaneous evidence identification and segmentation for automated glaucoma diagnosis. In Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 809–816. https: //doi.org/10.1609/aaai.v33i01.3301809

[12] Diaz-Pinto, A., Colomer, A., Naranjo, V., Morales, S., Xu, Y., & Frangi, A. F. (2019). Retinal image synthesis and semi-supervised learning for glaucoma assessment. IEEE Transactions on Medical Imaging, 38(9), 2211–2218. https: //doi.org/10.1109/TMI.2019.2902391

[13] Navab, N., Hornegger, J., Wells, W. M., & Frangi, A. (Eds.). (2015). Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III (Vol. 9351). Springer. https: //doi.org/10.1007/978-3-319-24574-4

[14] Chai, Y., Liu, H., & Xu, J. (2018). Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. Knowledge-Based Systems, 161, 147–156.

[15] Akter, N., Fletcher, J., Perry, S., Simunovic, M. P., Briggs, N., & Roy, M. (2022). Glaucoma diagnosis using multi-feature analysis and a deep learning technique. Scientific Reports, 12(1), 8064. https: //doi.org/10.1038/s41598-022-12122-9

[16] Medeiros, F. A., Jammal, A. A., & Thompson, A. C. (2019). From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. Ophthalmology, 126(4), 513-521.