Research Article

Leveraging Eye Fundus Images and Metadata Fusion for Glaucoma Detection with Artificial Intelligence

Fernando Ly Yang[1*], Chou M[1], Lauren Van Lancker[1], Chris Panos[1]

Abstract

Background/Aims: Accurate diagnosis of glaucoma relies on precise imaging techniques and comprehensive clinical data. While deep learning methods hold potential for enhancing diagnostic accuracy, the incorporation of clinical data into these models remains relatively unexplored.

Methods: In this study, we utilized the PAPILA dataset to investigate the integration of clinical data into machine learning models for glaucoma diagnosis. Two neural network architectures were compared: one trained solely on retinal fundus images and another incorporating both images and clinical data. The performance of these models was evaluated using standard metrics, including the DeLong test for statistical significance.

Results: Our findings reveal that the inclusion of clinical data resulted in a modest improvement in classification performance. However, the difference in performance between models using only images and those incorporating clinical data was not statistically significant according to the DeLong test.

Conclusion: Integrating clinical data into machine learning models for glaucoma diagnosis holds promise for enhancing diagnostic accuracy. While our study demonstrates a positive trend in classification performance with the inclusion of clinical data, further research is warranted to fully understand its impact and explore additional avenues for improvement.

Introduction

Glaucoma is a progressive condition that affects the optic nerve due to elevated intraocular pressure (IOP) caused by poor drainage of fluid within the eye. This ocular disorder often develops without noticeable symptoms, leading to a gradual and irreversible deterioration in visual function, eventually resulting in total vision loss [1,2].

Glaucoma stands as the second most prevalent cause of blindness globally, impacting approximately one in every two hundred individuals below the age of fifty and one in ten individuals above the age of eighty. Projections suggest that by 2040, around 111.8 million people aged 40–80 years will be afflicted with glaucoma [3,4].

The primary diagnostic procedures for glaucoma include tonometry for evaluating intraocular pressure, campimetry for assessing the visual field [5,6], and retinal fundus imaging [7,8] and optical coherence tomography (OCT) [9,10] for evaluating characteristics of the optic nerve head.

Affiliation:

[1]Epsom and St Helier University Hospitals NHS Trust, Glaucoma, Carshalton, United Kingdom

*Corresponding author:
Fernando Ly Yang, Epsom and St Helier University Hospitals NHS Trust, Glaucoma, Carshalton, United Kingdom.

Over the past ten years, research has predominantly focused on approaches rooted in deep learning techniques [11,12], which have demonstrated significant effectiveness in tasks such as image classification and segmentation. These methods have shown promising outcomes in the realm of ophthalmology [13], particularly in enhancing diagnostic capabilities.

Methods

In this study, the public databases G1020 [14], ORIGA [15], and PAPILA [16] were employed. G1020 and ORIGA, containing a combined total of 1670 images, were utilized for training a Segformer model (nvidia/mit-b0) from HuggingFace [17], aimed at segmenting the optic disc. Sixty percent of the 1670 images were designated for training, with 20% allocated for validation and another 20% for testing. Validation results revealed an overall accuracy of 99%, with a loss of 1%, while during training, the overall accuracy was 98%, with a loss of 3%. Following this, the transformer model was applied to segment the PAPILA [16] dataset and to crop the optic nerve at 1.5 diameters of the optic disc, covering the upward, downward, leftward, and rightward directions.

The analysis incorporated both clinical data and cropped PAPILA images. The dataset, known as PAPILA [16], consisted of three diagnostic categories: normal, glaucoma, and uncertain. Images with uncertain diagnoses were excluded, resulting in a total of 331 images—298 normal and 33 glaucoma images. Clinical data, encompassing parameters such as central corneal thickness, age, gender, axial length, and refractive defect, were combined with these images to classify glaucoma or healthy status. Notably, the clinical data were normalized using the StandardScaler from the scikit-learn library. For model training, 70% of the images were allocated for training, while 15% each were designated for validation and testing.

Fine-tuning of EfficientNetV2B0 [18] involved utilizing the last 200 layers with images, employing the AdamW optimizer with a learning rate of 1e-2 and weight decay of 1e-4. Binary crossentropy was used as the loss function, and weights were adjusted using the inverse frequency formula. Subsequently, image features were concatenated with metadata features to facilitate the classification of healthy and glaucoma.

The analysis comprised two stages of fine-tuning. Initially, fine-tuning was conducted exclusively with cropped PAPILA images, distinguishing between healthy and glaucoma using EfficientNetV2B0. In a subsequent phase, clinical data were integrated into the neural network architecture to enhance the prediction of healthy and glaucoma statuses.

Following these fine-tuning stages, a comparative analysis was performed using the DeLong test to evaluate the Area Under the Curve (AUC) of both versions of the EfficientNet on the test set. One version utilized only images, while the other incorporated both images and clinical data.

Results

Using the Segformer neural network, the entire PAPILA dataset underwent cropping (Figure 1). Instances of uncertain diagnoses within the PAPILA dataset were omitted. Within the dataset, 90% of images were categorized as normal, while 10% were identified as depicting glaucoma. In terms of gender distribution, men comprised 34% and women 66% of the dataset. The average age in the dataset was 59 years, with a standard deviation of 12. The mean refractive error stood at 0.85, with a standard deviation of 2.11. The average central corneal thickness measured 534, with a standard deviation of 41. Additionally, the mean axial length was 23.46, with a standard deviation of 1.10, and the mean intraocular pressure (IOP) was 16.01, with a standard deviation of 3.35.

A total of 331 images were utilized in the analysis. These were divided into 70% for training, 15% for validation, and 15% for testing. Data augmentation techniques such as horizontal and vertical flipping, as well as rotation up to 15 degrees, were employed to augment the dataset. The fine-tuning of the EfficientNetV2B0 model on the 15% test subset, consisting of 62 images, yielded an area under the curve (AUC) of 81%. The resulting confusion matrix is depicted in Figure 2.

The same procedure was repeated, incorporating clinical data alongside the images. Again, the dataset was divided into 70% for training, 15% for validation, and 15% for testing. The identical data augmentation techniques and neural network architecture EfficientNetV2B0 were employed. This combined approach resulted in an improved AUC of 84%, accompanied by the confusion matrix illustrated in Figure 3.

A De Long test was conducted to compare the AUCs obtained from the model trained solely on images with those trained on both images and clinical data. The test results indicated no statistically significant difference between the two AUC values.
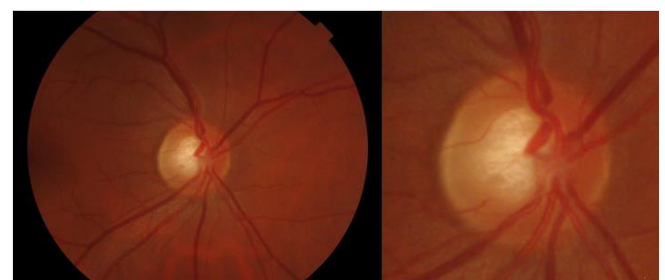
Figure 1: Shows a comparison of normal original PAPILA dataset image and the cropped one
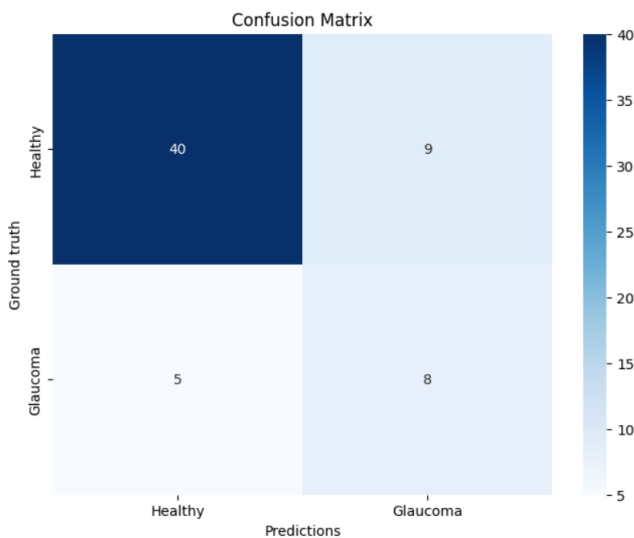
Figure 2: Displays the confusion matrix of EfficientNetV2B0 soly trained using eye fundus images.
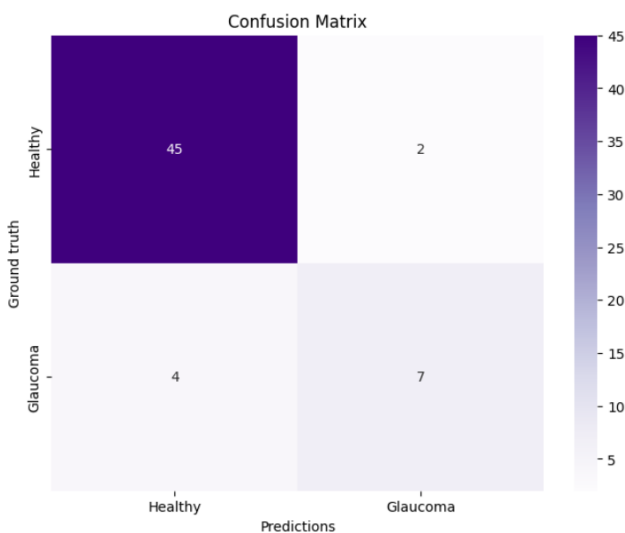
Figure 3: Displays the confusion matrix of EfficientNetV2B0 trained using eye fundus images and clinical data.

Discussion

Several studies have demonstrated the ability of neural networks to accurately classify patients with glaucoma and those without it using retinal fundus images, with AUC values typically around 99% [19-22] across various public databases. Deep learning techniques have proven useful in detecting glaucoma even without directly examining the optic nerve, achieving an AUC of 88% [23].

To date, the only study utilizing neural networks with the PAPILA dataset is by Kovalyk O et al [16], in which they reported AUC values ranging from 75% to 84% depending on the architecture of the neural network employed. However, the distribution of images for training, validation, and testing

is not specified in their study, and confusion matrix results are not provided. Consequently, it is unclear what percentage of AUC is achieved over the total number of images in the test set.

In comparison, our study using EfficientNetV2B0, trained solely on images with 15% of the total test set of 62 images, achieved an AUC of 81%, a result comparable to that of Kovalyk O et al [16]. However, when incorporating clinical data alongside images, our model achieved an AUC of 84%, surpassing all models utilized by Kovalyk O et al [16] except for one, which reported a slightly higher AUC. Nevertheless, according to the De Long test, this increase is not statistically significant.

These findings raise significant questions about the utility of including metadata in the disease diagnosis process. While it was initially expected that metadata would provide additional and complementary information to images, its effect on improving the model's performance seems to be limited in this particular dataset.

One possible explanation for the lack of significant improvement could lie in the nature of the clinical data used. It is possible that the available clinical variables are not capturing relevant clinical aspects or additional information that could enhance the model's discrimination ability. Typically, in ophthalmology, variables such as age, gender, central cornea thickness, refractive error, and axial length are not determinants for diagnosing glaucoma. Therefore, their inclusion may not contribute substantially to the predictive power of the model in this context.

It is important to acknowledge the limitations of our study, including the small size of the dataset and the limited representation of glaucoma cases compared to healthy ones, as well as the selection of clinical data used. In future research, other sources of clinical data such as variables from visual field or OCT parameters, in addition to incorporating eye fundus images, could potentially enhance glaucoma diagnosis and offer a more significant improvement in the model's performance.

Conclusions

Despite the lack of statistically significant improvement in the model's performance with the inclusion of metadata, our findings contribute to understanding the role of metadata in the disease diagnosis process. This research highlights the importance of continuing to explore new ways to integrate complementary clinical information into machine learning models to improve the accuracy and effectiveness of medical diagnosis.

Meeting presentation: Under consideration in European Glaucoma Congress 2024 and World Ophthalmology Congress 2024

Conflict of interest: No conflict of interest exists for any author

References

1. Casson RJ, Chidlow G, Wood JP, et al. Definition of glaucoma: clinical and experimental concepts. Clin Exp Ophthalmol 40 (2012): 341-349.

2. Voelker R. What Is Glaucoma? JAMA 330 (2023): 1594.

3. Tham YC, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology 121 (2014): 2081-2090.

4. Tham YC, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology 121 (11): 2081-2090.

5. De Moraes CG, Liebmann JM, Levin LA. Detection and measurement of clinically meaningful visual field progression in clinical trials for glaucoma. Prog Retin Eye Res 56 (2017): 107-147.

6. Nouri-Mahdavi K. Selecting visual field tests and assessing visual field deterioration in glaucoma. Can J Ophthalmol 49 (2014): 497-505.

7. Maupin E, Baudin F, Arnould L, et al. Accuracy of the ISNT rule and its variants for differentiating glaucomatous from normal eyes in a population-based study. Br J Ophthalmol 104 (2020): 1412-1417.

8. Law SK, Kornmann HL, Nilforushan N, et al. Evaluation of the "IS" Rule to Differentiate Glaucomatous Eyes From Normal. J Glaucoma 25 (2016): 27-32.

9. Moradi Y, Moradkhani A, Pourazizi M, et al. Diagnostic Accuracy of Imaging Devices in Glaucoma: An Updated Meta-Analysis. Med J Islam Repub Iran 37 (2023): 38.

10. Michelessi M, Lucenteforte E, Oddone F, et al. Optic nerve head and fibre layer imaging for diagnosing glaucoma. Cochrane Database Syst Rev 2015 (2015): CD008803.

11. Chan HP, Samala RK, Hadjiiski LM, et al. Deep Learning in Medical Image Analysis. Adv Exp Med Biol 1213 (2020): 3-21.

12. Chen X, Wang X, Zhang K, et al. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal 79 (2022): 102444.

13. Orlando JI, Fu H, Barbosa Breda J, et al. REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Med Image Anal 59 (2020): 101570.

14. Bajwa MN, Singh GAP, Neumeier W, et al. G1020: A Benchmark Retinal Fundus Image Dataset for Computer-Aided Glaucoma Detection. IJCNN 2020 (2020): 1-7.

15. Zhang Z, Yin FS, Liu J, et al. ORIGA(-light): an online retinal fundus image database for glaucoma analysis and research. Annu Int Conf IEEE Eng Med Biol Soc 210 (2010): 3065-3068.

16. Kovalyk O, Morales-Sánchez J, Verdú-Monedero R, et al. PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. Sci Data 9 (2022): 291.

17. Hugging Face. (n.d.). Hugging Face Model Hub (2020).

18. Keras (nd). EfficientNetV2B0 (2020).

19. Velpula VK, Sharma LD. Multi-stage glaucoma classification using pre-trained convolutional neural networks and voting-based classifier fusion. Front Physiol 14 (2023): 1175881.

20. Ganesh SS, Kannayeram G, Karthick A, et al. A Novel Context Aware Joint Segmentation and Classification Framework for Glaucoma Detection. Comput Math Methods Med 2021 (2021): 2921737.

21. Rehman AU, Taj IA, Sajid M, et al. An ensemble framework based on Deep CNNs architecture for glaucoma classification using fundus photography. Math Biosci Eng 18 (2021): 5321-5346.

22. Hemelings R, Elen B, Schuster AK, et al. A generalizable deep learning regression model for automated glaucoma screening from fundus images. NPJ Digit Med 6 (2023): 112.

23. Hemelings R, Elen B, Barbosa-Breda J, et al. Deep learning on fundus images detects glaucoma beyond the optic disc. Sci Rep 11 (2021): 20313.