**Research Article**

# High Confidence Artificial Intelligence (AI) Predictions in Glaucoma Detection: A RIM ONE Database Study

**Fernando Ly Yang[1*] and Enrique Santos Bueso[1]**

## Abstract

**Introduction and Objectives:** Glaucoma is a progressive optic neuropathy that can lead to irreversible blindness. This study evaluates the use of neural networks in glaucoma prediction with high confidence.

**Patients or Materials and Methods:** The RIM One dataset was used, training an EfficientNetV2B0 model on fundus images. A 95% threshold was set for high-confidence predictions.

**Results:** Sensitivity was 91% and specificity was 99%. Applying the high-confidence threshold increased the AUC to 100%.

**Conclusions:** This study demonstrates the feasibility of using high-confidence AI predictions for glaucoma diagnosis, improving clinical relevance.

**Keywords:** Glaucoma; Artificial Intelligence; Deep learning; Diagnosis; Fundus imaging.

## Introduction

Glaucoma is a progressive optic neuropathy which develops without noticeable symptoms, leading to gradual and irreversible deterioration in visual function, eventually resulting in total visual field loss [1,2].

Globally, glaucoma is the second most prevalent cause of blindness, affecting approximately one in every two hundred individuals below the age of 50 years, and one in ten individuals above the age of 80 years. It is projected that by 2040, around 111.8 million people globally between the ages of 40 and 80 years will be afflicted with glaucoma [3,4]

The diagnosis of glaucoma is based on intraocular pressure measurement, visual field testing [5,6], optic disc examination and imaging and, increasingly [7,8], optical coherence tomography (OCT) [9,10] to examine features of the optic nerve head. The increasing prevalence of glaucoma will correspond to an increase in healthcare costs required for disease management.

Over the last decade, there has been research focus 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 some promise in ophthalmology [13], particularly in enhancing diagnostic capabilities

These AI technologies have the potential to reduce the burden on existing healthcare services by increasing the accuracy and efficiency of diagnosis, to intelligently target healthcare resources.

**Affiliation:**

[1]Hospital Clinico San Carlos

**\*Corresponding author:**
Fernando Ly Yang, Hospital Clinico San Carlos

## Material and Methods

This study utilized the publicly available RIM One dataset [14] of 485 optic disc images which comprises 313 normal control cases and 172 cases of glaucoma. Of the 485 total images, 248 of glaucoma and normal control images were used to train the EfficientNetV2B0 neural network model [15] while 63 were reserved for validation, and 174 for testing purposes.

Data augmentation was performed using the albumentations library, specifically rotation with a limit of 30 degrees. The image shape was set to 224x224 pixels with 3 channels. The last 165 layers corresponding to the model were fine-tuned using the Adam optimizer with a learning rate of $10^{-4}$ and binary cross-entropy loss function. The inverse frequency formula was employed, and accuracy and F1 score metrics were utilized for evaluation.

Measures of model performance; area under the curve (AUC) and confusion matrix were obtained using the 174 test images. Subsequently, in a novel extension of the model, all 174 images were processed by the trained neural network to identify images with a prediction probability exceeding 95%.

Delong's t-test was then used to compare the AUC obtained from the first test with 174 images and the second test with only images with a prediction probability above 95%. Sensitivity and specificity were also compared between the two tests.

## Results

The results from the analysis of 174 test images, with 118 labelled as normal and 56 labelled as glaucoma, revealed an AUC of 96% with the newly trained EfficientNetV2B0 model. The sensitivity was 91%, specificity 99%, positive predictive value 98% and negative predictive value was 95%. The confusion matrix illustrating these findings is presented in Figure 1.
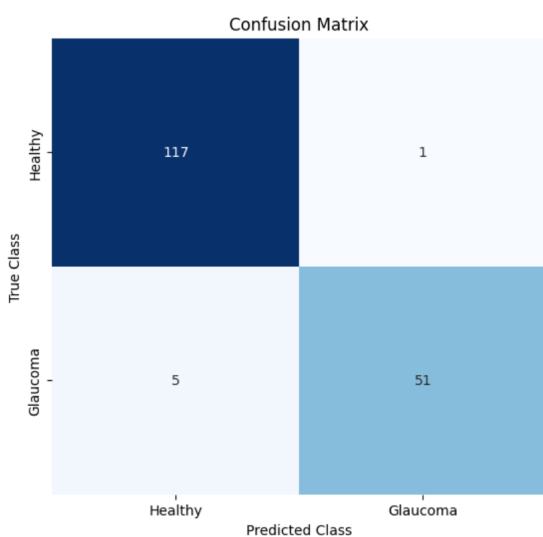


**Figure 1:** Confusion matrix test 174 images.

Of the 174 test images assessed, a protocol was established to exclude patients from further examination if the neural network's prediction probability did not meet the 95% threshold for certainty. Subsequently, 152 out of the 174 images surpassed this probability threshold and were included for further analysis as image results with high predictive confidence.

For this subset of 152 images identified by the newly trained model with high predictive confidence, the AUC improved to 100%, with corresponding 100% sensitivity, specificity, positive predictive value, and negative predictive value. The confusion matrix illustrating these findings is presented in Figure 2.
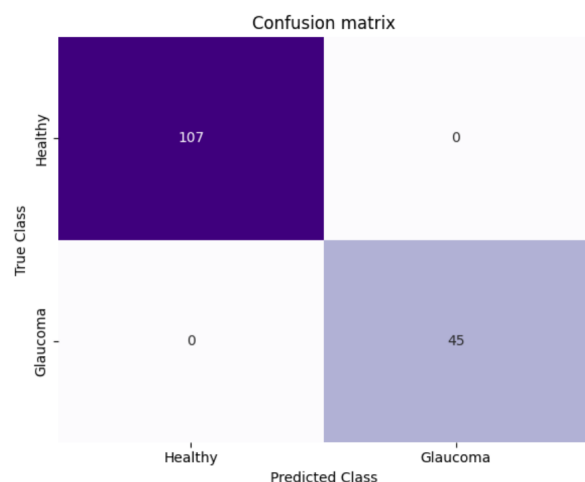


**Figure 2:** Confusion matrix test 152 images.

The DeLong t-test revealed a statistically significant difference in the AUC between the two groups, with a p-value of 0.007. Additionally, significant differences were observed in sensitivity (p = 0.0003) and negative predictive value (p = 0.002). However, the differences in specificity (p = 0.18) and positive predictive value (p = 0.055) were not statistically significant.

## Discussion

Several studies have demonstrated the proficiency of neural networks in accurately distinguishing between glaucomatous and healthy using retinal fundus images, typically yielding AUC values of around 99% across a number of publicly available databases [16-19]. Deep learning techniques have even proven adept at detecting glaucoma using retinal images excluding the optic nerve, achieving an AUC of 88% [20].

Fumero et al. achieved a 99% AUC with the RIM ONE14 database, while Phasuk et al [21]. attained 94%. The 96% AUC obtained in this study is in line with published literature.

To date, no published study according to our knowledge, has considered the use of prediction probability obtained from neural networks and its potential in clinical practice.

This study focuses not only on the prediction of glaucomatous versus healthy discs of a trained neural network on the RIM-ONE test data but also on the certainty probability of each prediction. 95% was selected as a confidence level as it is a standard threshold in scientific research. If the probability exceeds 95%, we can consider the prediction of the artificial intelligence as accurate. This approach to the use of AI could be adopted in clinical practice to identify those patients who require further investigation for possible glaucoma. This mirrors current clinical practice of initial disc examination by a clinician and further investigation only if there is clinical suspicion of glaucoma.

While other neural networks may be mathematically superior to this model, this novel extension of the model to calculate certainty probability gives this model clinical relevance previously lacking in other models. In this study, in the high confidence prediction test set the AUC is 100%, whereas with the original test set, which included results with predictive probabilities under 95%, an AUC of 96% was obtained. The comparison with the De Long test yielded statistically significant results ($p < 0.05$).

In our analysis, we found that not only was the increase of Area Under the Curve (AUC) statistically significant, indicating an improved overall performance of the model, the sensitivity and negative predictive value also showed statistically significant improvements.

Specificity and positive predictive did not improve to a statistically significant degree, despite reaching 100%, because the baseline values for specificity and positive predictive value of the full test of 172 images were already high, at 98% and 95% respectively.

All previous studies have primarily focused on achieving a 100% AUC without considering individual prediction probabilities. This approach of assigning predictive probabilities to each outcome from the AI model takes into consideration the clinical relevance of neural networks in diagnosing glaucoma.

This study is, to our knowledge, the first to demonstrate the potential clinical relevance of incorporating high-confidence AI predictions into artificial intelligence models to assess glaucoma from fundus images.

## Conclusion

Our study demonstrates the effectiveness of neural networks in diagnosing glaucoma from retinal fundus images. By calculating high confidence AI predictions with a probability of certainty exceeding 95%, we highlight a potential clinical application. This approach bridges the gap between AI research and clinical practice, offering a promising tool for efficient and accurate screening for glaucoma in community and hospital settings.

## 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. 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 (2014): 107-147.

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

6. 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.

7. 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.

8. 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.

9. Michelessi M, Lucenteforte E, Oddone F, Brazzelli M, Parravano M, Franchi S, Ng SM, Virgili G. Optic nerve head and fibre layer imaging for diagnosing glaucoma. Cochrane Database Syst Rev 2015 (2015): CD008803.

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

11. 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.

12. 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.

13. Fumero F, Alayon S, Sanchez JL, et al. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation," 2011 24th International Symposium on Computer-Based Medical Systems (CBMS), Bristol, UK (2011): 1-6.

14. Keras (nd). EfficientNetV2B0 (2022).

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

16. 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.

17. 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.

18. 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.

19. 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.

20. Phasuk S, Tantibundhit C, Poopresert P, et al. Automated Glaucoma Screening from Retinal Fundus Image Using Deep Learning. Annu Int Conf IEEE Eng Med Biol Soc 2019 (2019): 904-907.