# Battle of the Algorithms: A Head-to-Head Comparison of Four Leading AI Models in Retinal Disease Detection

Christopher Stewart[1], Jaskaran Bhangu Singh[1], Rhadika Rewal[1] , Gwyn Samuel Williams[2], Safa El-Hassan[2]

(1) Faculty of Medicine, Health and Life Science, Institute of Life Science 2, Swansea University, Swansea, UK

(2) Department of Ophthalmology, Singleton Hospital, Swansea Bay University Health Board, Swansea, UK.

## Introduction

The global burden of retinal diseases represents a significant healthcare challenge, with conditions such as diabetic retinopathy, glaucoma, and age-related macular degeneration remaining the leading causes of preventable blindness worldwide.[1-3] Early detection and intervention are crucial for preventing vision loss, yet traditional screening methods face substantial limitations in terms of accessibility, consistency, and scalability.[4] This has created an urgent need for more efficient and accurate diagnostic tools.

## Aim

To evaluate and compare the diagnostic accuracy and reliability of four artificial intelligence (AI) language models in diagnosing various retinal conditions through fundus image analysis across different pathologies and disease severities.

## Methodology

We conducted a systematic evaluation of the four AI models' performance across multiple retinal conditions, including diabetic retinopathy (DR), retinal vein occlusion (RVO), age-related macular degeneration (AMD), and glaucoma. Diagnostic metrics including sensitivity, specificity, and accuracy were calculated for each condition and model. Rigorous statistical analysis employed chi-square tests for comparing categorical outcomes, whilst t-tests were used for continuous variables. McNemar's test was applied for paired nominal data to assess the significance of differences between models.

## Results

For diabetic retinopathy:

- Model A (Google Deepmind) achieved a sensitivity of (99.0%), Model B (95.5%), Model C (94.8%), and Model D (93.2%)
- All models showed significantly lower performance for mild diabetic retinopathy (0/18 cases).

For AMD detection:

- Model A maintained its superior performance with 99.0% sensitivity and 97.8% specificity.

For glaucoma assessment:

- Model A demonstrated the highest accuracy in identifying advanced cases (28/30 cases, 93.3%), compared to Model B (85.6%), Model C (84.2%), and Model D (82.1%).
- Early-stage glaucoma detection remained challenging for all models, with Model A achieving the highest accuracy (22/28 cases, 78.6%).

One-way ANOVA F-statistic of 12.37 (p-value <0.001) and Post-hoc Tukey's HSD test showed significant performance differences between models ($p<0.05$), with Model A consistently demonstrating superior accuracy across all conditions.

| Condition | Google DeepMind | ChatGPT-4 | Microsoft AI | Bing AI |
|---|---|---|---|---|
| Diabetic Retinopathy | 95.3% | 93.5% | 94.2% | 91.5% |
| AMD | 94.2% | 92.1% | 93.0% | 90.0% |
| Glaucoma | 94.0% | 91.8% | 92.5% | 89.7% |

**Table 1 (above):** demonstrates the accuracy across all four models compared head-to-head for the 3 of the top 5 leading causes of blindness worldwide. Model A is Google Deepmind, Model B is ChatGPT-4, Model C is Microsoft AI, and Model D is Bing AI.

## Conclusion

This comparative analysis demonstrates Model A, DeepMind, had consistently higher accuracy across multiple conditions. Nevertheless, the universal challenge across all four AI. This is critical for AI use as a screening tool whereby identification of subtle changes in early disease can lead to prompt intervention and thus improved patient outcomes. These findings provide valuable insights for the future development and implementation of AI in ophthalmological diagnostics.

## References

1. Teo ZL, Tham YC, Yu M, Chee ML, Rim TH, Cheung N, et al. Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis. Ophthalmology [Internet]. 2021;128(11):1580–91. Available from: https://www.sciencedirect.com/science/article/pii/S0161642021003213

2. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global Prevalence of Glaucoma and Projections of Glaucoma Burden through 2040: A Systematic Review and Meta-Analysis. Ophthalmology [Internet]. 2014;121(11):2081–90. Available from: https://www.sciencedirect.com/science/article/pii/S0161642014004333

3. Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng CY, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. Lancet Glob Health [Internet]. 2014 Feb 1;2(2):e106–16. Available from: https://doi.org/10.1016/S2214-109X(13)70145-1

4. Vujosevic S, Aldington SJ, Silva P, Hernández C, Scanlon P, Peto T, et al. Screening for diabetic retinopathy: new perspectives and challenges. Lancet Diabetes Endocrinol [Internet]. 2020 Apr 1;8(4):337–47. Available from: https://doi.org/10.1016/S2213-8587(19)30411-5