# **Supplementary**

**Title: Diagnostic Accuracy of Artificial Intelligence in Glaucoma Screening and Clinical Practice**

Abadh K Chaurasia1\*, Connor J Greatbatch1,Alex W Hewitt1,2

# **Affiliations:**

1. Menzies Institute for Medical Research, School of Medicine, University of Tasmania, Australia.
2. Centre for Eye Research Australia, University of Melbourne, Australia.

# **\*Correspondence:**

Abadh K Chaurasia

Menzies Institute for Medical Research,

University of Tasmania, Australia.

Email: awadhaiims@gmail.com

# **Funding**

This research was supported by National Health and Medical Research Council Practitioner Fellowships (DAM, AWH), a Research Training Program Scholarship from the University of Tasmania (AKC), and by grants from the National Health and Medical Research Council (GNT1150144) and the Royal Hobart Hospital Research Foundation.

# **Rationale and Terminology**

Glaucoma assessment is mainly based on optic nerve head (ONH) imaging, visual field testing, gonioscopy, and intraocular pressure (IOP) monitoring to detect and identify structural and functional damage.[1](https://paperpile.com/c/EA8xj3/DIHWE),[2](https://paperpile.com/c/EA8xj3/Nc0fB),[3](https://paperpile.com/c/EA8xj3/KRJRd) An accurate diagnosis requires a glaucoma specialist with clinical tests such as optical coherence tomography (OCT) and Humphrey visual field (HVF). Furthermore, serial testing is required as a single test is insufficient to detect early-stage glaucoma.[4](https://paperpile.com/c/EA8xj3/JhamM) ONH evaluation has shown to be more effective in diagnosing glaucoma but is expensive, thus hindering high-throughput glaucoma screening.[5](https://paperpile.com/c/EA8xj3/UOwKm),[6](https://paperpile.com/c/EA8xj3/WdttG) However, imaging technology and data storage capacity have improved exponentially in recent years, allowing glaucoma specialists to assess patients with high-resolution imaging over many years to grade patients as healthy or glaucomatous. The accumulation of high-resolution imaging data (fundus photography and OCT images) allows for Artificial Intelligence (AI)-based models to be used to detect glaucomatous changes. These two imaging options provide a pivotal opportunity to develop an AI model for glaucoma detection. Such models may potentially classify glaucoma with a single eye scan that could be a cost-effective solution for population-based screening. Therefore, we extensively investigated and determined AI models’ diagnostic performance and factors affecting implementation into clinical practice.

## **Machine learning**

Machine Learning (ML) is the technology of computer algorithms that learn to extract patterns and features from datasets. Using this experience, ML models can then make predictions and generalizations on new data. The extraction of relevant features and pattern recognition from clinical data makes it the foundation of AI decision-making. The prominent use cases for ML application in medicines are image recognition and computer vision systems.

Clinical implementation of ML computer vision systems through pattern recognition and segmentation techniques has been substantially researched in the medical field. The two major ML approaches; supervised and unsupervised learning, have both been previously implemented in glaucoma detection. Supervised learning models include support vector machines (SVM), K-nearest neighbour (K-NN), random forest (RF), decision trees (DT), and artificial neural networks (ANN). Unsupervised learning techniques include principal component analysis (PCA), fuzzy C-means systems (F-CS), hierarchical clustering (HC), K-means clustering (K-MC). The most promising of these AI models for image recognition, and glaucoma classification employs ANNs in the practice of deep learning. This is distinguished from traditional ML by extracting high-level features automatically using neural networks.

## **Deep Learning (DL)**

Several studies have demonstrated that DL application in medicine using diverse imaging modalities have high diagnostic performance in terms of the area under the receiver operator characteristics curve (AUC), sensitivity, and specificity.[7](https://paperpile.com/c/EA8xj3/DuV4G),[8](https://paperpile.com/c/EA8xj3/3kNAK),[9](https://paperpile.com/c/EA8xj3/nfnIn),[10](https://paperpile.com/c/EA8xj3/ZTF9U) Recently, the Food and Drug Administration (FDA) approved multiple AI-based medical devices, including IDx-DR (de novo pathway clearance), for diagnosing diabetic retinopathy through retinal imaging.[11](https://paperpile.com/c/EA8xj3/NDJW6),[12](https://paperpile.com/c/EA8xj3/i8ipG) The diagnostic power of AI is only expected to grow with increased computational power and the expansion of datasets. Consequently, the emerging power of AI could revolutionize glaucoma management with early diagnosis and cost-effective population-based screening.

## **Convolutional Neural Network (CNN) Architectures**

The application of convolutional neural networks has shown to be the most robust approach for large amounts of image data. They are designed to analyze image data through convolution and pooling algorithms to extract features from image data. CNNs are designed to perform image classification and identification with high accuracy. The applicability of CNNs in glaucoma prediction is exponentially growing due to large amounts of image data being produced through fundus photography and OCT imaging. Furthermore, with the rise of graphical processing units (GPUs), the computational power available to researchers has skyrocketed.

## **Transfer learning**

Transfer learning is an ML technique whereby a model is first trained on a large dataset (e.g., ImageNet Database) and then fine-tuned on a smaller dataset pertaining to the task of interest. This somewhat alleviates the need for large datasets, which can be challenging to acquire in the medical field. For instance, tumor classification with a limited dataset.[13](https://paperpile.com/c/EA8xj3/cXho2)

## **Machine Learning Classifier**

The application of AI to glaucoma diagnosis through imaging began in 2010[14](https://paperpile.com/c/EA8xj3/UNNyl), using traditional ML techniques such as SVM and K-NN before the exponential increase in DL implementation over the last two years with CNNs. The most commonly used traditional ML classifier and DL classifier were SVMs and CNNs, respectively. Most authors trained the traditional ML algorithms using K-fold cross-validation techniques to reduce the risk of overfitting and overcome data insufficiency issues. In this technique, the dataset is randomly split into k subsets of equal size and an individual model is trained on k-1 subsets and tested on the final subset. This results in k-replicates of the model in which the performance is aggregated to establish the most generalized performance.

# **Dataset’s descriptions**

## **Fundus photographs**

It is an image of the retina, including ONH and retinal blood vessels. The fundus photograph was taken with either monoscopic or stereoscopic view; it depends on ophthalmic fundus cameras.

### **Monoscopic fundus images**

The AI models were analyzed on monoscopic images for glaucoma detection.

### **Stereoscopic fundus images**

The AI models were evaluated on the stereoscopic fundus images for glaucoma detection.

### **Mixed fundus images:**

### The AI models were determined on the mixed fundus images, monoscopic and stereoscopic fundus images for glaucoma detection.

## **Optical Coherence Tomography (OCT) images**

It is a non-invasive imaging technique that uses low-coherence light to capture a high resolution of two or three-dimensional images of the ONH. In this review, only ONH volume or images were used for analysis.

## **Retinal Nerve Fiber Layer (RNFL)**

It is unmyelinated axons of the retinal ganglion cells that form the optic nerve. The loss of ganglion cells and their axons results in thinning of the RNFL in glaucoma, which can be assessed more accurately with an OCT. The volume of RNFL thickness was analyzed to detect glaucoma through AI algorithms.

## **Optic Nerve Head (ONH)**

It is an oval, posterior depressed area where all retinal nerve fibres pass posteriorly through the lamina cribrosa. Deterioration of RNFL is usually followed by the change in ONH morphologically. Therefore, it is essential for glaucoma diagnosis or monitoring through fundus photography or OCT scanning.

**Optic Cup (OC) and Optic Disc (OD)**

OD is a demarcated part of an ONH where the central retinal vessels originate. The central part of the OD, which is light/pale, makes its OC. The ratio of vertical OC diameter to OD is an important diagnostic parameter for glaucoma through the fundus imaging modality.

# **Applied techniques**

## **Segmentation-based**

### It is a technique in which different regions (OD, OD and OC, RNFL or ONH) are detected and segmented in a fundus or OCT image. The classification of glaucoma was undertaken based on segmented images.

## **Non-segmentation-based**

Features were automatically extracted from an entire image without a particular anatomical region being utilized.

# **Region-of-Interest (ROI) Selection**

The final diagnosis of glaucoma was classified based on a region-of-interest selection from the fundus or OCT images. For example, the ROI selection was OD, and glaucoma or health was decided based on OD, similarly, OD and OC, Or RNFL, or full ONH images.

# **Model's Validation Accuracy on internal and external data**

The AI algorithms’ performance was tested either on the same data (Internal validation) or different population data (External validation).

# **Performance metrics**

## **Confusion matrix**

It is known as the error matrix, which visualizes the performance of the algorithms for classification problems in the field of machine learning through a 2 x 2 table with the actual and predicted values.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Actual (Clinical Diagnosis)** | | |
| **Predicted**  **(Index Test-AI)** |  | **Positive** | **Negative** |
| **Positive** | **TP** | **FP** |
| **Negative** | **FN** | **TN** |

Where,

**True Positives (TP):** AI predicted glaucoma true (Actually was glaucoma).

**True Negatives (TN):** AI predicted healthy true (Actually was healthy).

**False Positives (FP):** AI predicted glaucoma but actually was healthy.

**False Negatives (FN):** AI predicted Healthy but actually was glaucoma.

### **Sensitivity or True Positive Rate or Recall**

The proportion of the glaucoma cases correctly classified by AI models.

### **Specificity or True Negative Rate**

The proportion of the healthy cased correctly classified by AI models.

### **Accuracy**

It is the proportion of the number of correctly predicted cases out of the total cases by an algorithm.

=

**Area Under the Receiver Operating Characteristics (AUC)**

AUC is an evaluation metric for classification problems between people with and without disease at various threshold settings. It is a probability curve that plots between sensitivity and the false-positive rate at various threshold values. Higher the AUC, the better the model to distinguish between patients with and without glaucoma.

# **Tool: QUADAS-2**[15](https://paperpile.com/c/EA8xj3/3o6N)

## **State the review question:**

What is the best AI model for discriminating between people with or without glaucoma?

1. **Patients (setting, intended use of index test, presentation, prior testing):** Healthy and glaucoma
2. **Index test(s):** AI-based models for glaucoma diagnosis
3. **Reference standard and target condition:** Clinical investigation (HVF) and manifest glaucoma.

## **Domain 1: Patient Selection**

Risk of bias: could the selection of patients have introduced bias?

**Signaling questions:**

1. Was a consecutive or random sample of patients enrolled?
2. Did the study avoid inappropriate exclusions?

(Exclusions of patients with pathologies; high myopia, disc anomalies, and unreliable data selection (DR))

**Applicability:** are there concerns that the included patients and setting do not match the review question?

**Signaling questions:**

1. Did ophthalmologists make the ground truth (healthy or glaucoma)?

## **Domain 2: Index Test**

Risk of bias: could the conduct or interpretation of the index test have introduced bias?

**Signaling questions:**

1. Were the index test results interpreted without knowledge of the results of the reference standard?
2. Was the AI-model used for the diagnosis of glaucoma from imaging modalities?

**Applicability:** are there concerns that the index test, its conduct, or its interpretation differ from the review question?

**Signaling questions:**

1. Was the dataset separated into training, validation or testing?

## **Domain 3: Reference Standard**

Risk of bias: could the reference standard, its conduct, or its interpretation have introduced bias?

**Signaling questions:**

1. Is the reference standard likely to correctly classify the target condition?
2. Were the reference standard results interpreted without knowledge of the results of the index test?
3. Was the clinical diagnosis of glaucoma based on visual field findings?

**Applicability:** are there concerns that the target condition as defined by the reference standard does not match the question?

**Signaling questions:**

1. Did the patients have a comprehensive ophthalmic examination (IOP, HVF, OCT RNFL)?

## **Domain 4: Flow and Timing**

Risk of bias: could the patient flow have introduced bias?

**Signaling questions:**

1. Was there an appropriate interval between the index test and reference standard?
2. Did all patients receive the same reference standard?
3. Were all patients included in the analysis?

**Tables**

# **Table 1. PRISMA-DTA Checklist**

|  |  |  |  |
| --- | --- | --- | --- |
| **Section/topic** | **#** | **PRISMA-DTA Checklist Item** | **Reported on page #** |
| **TITLE / ABSTRACT** | | |  |
| Title | 1 | Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies. | 1 |
| Abstract | 2 | Abstract: See PRISMA-DTA for abstracts. | 2 |
| **INTRODUCTION** | | |  |
| Rationale | 3 | Describe the rationale for the review in the context of what is already known. | 3 |
| Clinical role of index test | D1 | State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design). | 3 |
| Objectives | 4 | Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s). | 4 |
| **METHODS** | | |  |
| Protocol and registration | 5 | Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number. | 4 |
| Eligibility criteria | 6 | Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. | 5 |
| Information sources | 7 | Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched. | 5 |
| Search | 8 | Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated. | 6 |
| Study selection | 9 | State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis). | 7 |
| Data collection process | 10 | Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators. | 7 |
| Definitions for data extraction | 11 | Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting). | 7 |
| Risk of bias and applicability | 12 | Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question. | 7 |
| Diagnostic accuracy measures | 13 | State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion). | 7 |
| Synthesis of results | 14 | Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards | 8 |
| Meta-analysis | D2 | Report the statistical methods used for meta-analyses, if performed. | 8 |
| Additional analyses | 16 | Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified. | 8 |
| **RESULTS** | | |  |
| Study selection | 17 | Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram. | 9 |
| Study characteristics | 18 | For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources | 9 |
| Risk of bias and applicability | 19 | Present evaluation of risk of bias and concerns regarding applicability for each study. | 10 |
| Results of individual studies | 20 | For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot. | 10 |
| Synthesis of results | 21 | Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals. | 10 |
| Additional analysis | 23 | Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events). | 12 |
| **DISCUSSION** | | |  |
| Summary of evidence | 24 | Summarize the main findings including the strength of evidence. | 13 |
| Limitations | 25 | Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research). | 16 |
| Conclusions | 26 | Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test). | 17 |
| **FUNDING** | | |  |
| Funding | 27 | For the systematic review, describe the sources of funding and other support and the role of the funders. | 1 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2. Electronic databases with a search strategy:**  **2.1. Database(s): Embase 1974 to August 07, 2020**   |  |  |  | | --- | --- | --- | | **Search Strategy:** | **Searches** | **Results** | | 1 | exp glaucoma/di [Diagnosis] | 10930 | | 2 | Optical coherence tomography.af. | 67742 | | 3 | Retinal nerve fiber layer.af. | 9354 | | 4 | Optic disc.af. | 11831 | | 5 | Cup to disc ratio.af. | 1600 | | 6 | Visual field.af. | 53413 | | 7 | 1 or 2 or 3 or 4 or 5 or 6 | 130578 | | 8 | exp artificial intelligence/ | 40486 | | 9 | Machine learning.af. | 51525 | | 10 | Deep learning.af. | 14100 | | 11 | Transfer learning.af. | 1401 | | 12 | Convolutional Neural Network.af. | 8113 | | 13 | computer-aided diagnosis.af. | 4294 | | 14 | 8 or 9 or 10 or 11 or 12 or 13 | 100674 | | 15 | 7 and 14 | 1235 |     **2.2. Database(s): Ovid MEDLINE(R) ALL 1946 to August 07, 2020**   |  |  |  | | --- | --- | --- | | **Search Strategy:** | **Searches** | **Results** | | 1 | exp glaucoma/di [Diagnosis] | 11295 | | 2 | Optical coherence tomography.af. | 36657 | | 3 | Retinal nerve fiber layer.af. | 5020 | | 4 | Optic disc.af. | 8616 | | 5 | Cup to disc ratio.af. | 1194 | | 6 | Visual field.af. | 30453 | | 7 | 1 or 2 or 3 or 4 or 5 or 6 | 77201 | | 8 | exp artificial intelligence/ | 98261 | | 9 | Machine learning.af. | 36393 | | 10 | Deep learning.af. | 11109 | | 11 | Transfer learning.af. | 1169 | | 12 | Convolutional Neural Network.af. | 4734 | | 13 | computer-aided diagnosis.af. | 2965 | | 14 | 8 or 9 or 10 or 11 or 12 or 13 | 127035 | | 15 | 7 and 14 | 796 |       **2.3. Web of Science Core Collection up to August 07, 2020**   |  |  |  | | --- | --- | --- | | **Search Strategy:** | **Searches** | **Results** | | 1 | TITLE:  (Glaucoma)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 34174 | | 2 | TITLE:  (optical coherence tomography)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 23255 | | 3 | TITLE:  (retinal nerve fiber layer)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 2850 | | 4 | TITLE:  (optic disc)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 2881 | | 5 | TITLE:  (cup to disc ratio)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 134 | | 6 | TITLE:  (visual field)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 9628 | | 7 | #6 OR #5 OR #4 OR #3 OR #2 OR #1  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 28240 | | 8 | TITLE:  (Artificial intelligence)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 12625 | | 9 | TITLE:  (Machine learning)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 35718 | | 10 | TITLE:  (Deep learning)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 22166 | | 11 | TITLE:  (Transfer learning)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 5012 | | 12 | TITLE:  (Convolutional Neural Network)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 12489 | | 13 | TITLE:  (computer aided diagnosis)  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 1962 | | 14 | #13 OR #12 OR #11 OR #10 OR #9 OR #8  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 87589 | | 15 | #14 AND #7  Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years | 357 |       **2.4. Scopus up to August 07, 2020**   |  |  | | --- | --- | | **Search Strategy:** | **Results** | | ( TITLE ( "Glaucoma"  OR  "retinal nerve fiber layer"  OR  "optic disc"  OR  "cup to disc ratio"  OR  "visual field"  OR  "Optical coherence tomography " ) )  AND  ( TITLE ( "artificial intelligence"  OR  "machine learning"  OR  "deep learning"  OR  "transfer learning"  OR  "Convolutional Neural Network"  OR  "computer aided diagnosis" ) ) | 311 | |

# **Table 3. Test of heterogeneity for subgroup studies**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Studies** | **I2** | **tau2** | **Cochran's Q** | **P-value** |
| **1. Overall DA of AI-models** | 98.5% | 2.44 | 4305.32 | 0 |
| **Subgroup Analysis** |  |  |  |  |
| 2. **All Fundus Images** | 98.3% | 2.01 | 3252.57 | 0 |
| 2.1. Mono fundus images | 98.2% | 3.26 | 2091.98 | 0 |
| 2.2. Stereo fundus images | 0 | 0 | 4.90 | < 0.43 |
| 2.3. Mixed fundus images | 78.2% | 0.11 | 45.88 | < 0.01 |
| **3. OCT images** | 85.7% | 0.79 | 62.78 | < 0.01 |
| **4. Applied Techniques** |  |  |  |  |
| 4.1. ML with segmentation | 85% | 2.09 | 79.92 | < 0.01 |
| 4.2. ML with non-segmentation | 94.7% | 3.23 | 187.78 | < 0.01 |
| 4.3. DL with segmentation | 97.2% | 3.00 | 393.94 | < 0.01 |
| 4.4. DL with non-segmentation | 99.1% | 2.32 | 3164.32 | < 0.01 |
| **5. Region of interest (ROI)** |  |  |  |  |
| 5.1. Full ONH images | 97.4% | 1.88 | 1132.49 | < 0.01 |
| 5.2. OD | 97.7% | 7.33 | 385.20 | < 0.01 |
| 5.3. OD and OC | 93.0% | 1.74 | 243.92 | < 0.01 |
| 5.4. RNFL | 99.4% | 3.57 | 1151.41 | < 0.01 |
| **6. Validation accuracy** |  |  |  |  |
| 6.1. Internal | 98.4% | 2.14 | 3572.98 | 0 |
| 6.2. External | 96.2% | 4.20 | 210.97 | < 0.01 |
| **7. Electronic Databases** |  |  |  |  |
| 7.1. Publicly available | 95.7% | 2.38 | 492.63 | < 0.01 |
| 7.2. Private data | 98.8% | 2.43 | 3400.59 | 0 |
| 7.3. Mixed data | 88.1% | 1.12 | 16.83 | < 0.01 |

# 

# **Table 4. HSROC parameters**

|  |  |  |  |
| --- | --- | --- | --- |
| **Studies** | **Theta** | **Lambda** | **beta** |
| **1. Overall DA of AI-models** | -0.108 | 4.753 | 0.044 |
| **Subgroup Analysis** |  |  |  |
| 2. **All Fundus Images** | -0.046 | 4.717 | 0.049 |
| 2.1. Mono fundus images | 0.075 | 4.926 | 0.097 |
| 2.2. Stereo fundus images | -3.197 | 7.903 | -1.856 |
| 2.3. Mixed fundus images | -0.625 | 3.697 | -0.586 |
| **3. OCT images** | -5.421 | 11.922 | -2.635 |
| **4. Applied Techniques** |  |  |  |
| 4.1. ML with segmentation | 0.059 | 4.449 | 0.204 |
| 4.2. ML with non-segmentation | 0.335 | 5.250 | 0.303 |
| 4.3. DL with segmentation | 0.025 | 4.488 | 0.180 |
| 4.4. DL with non-segmentation | -0.272 | 4.822 | -0.046 |
| **5. Region of interest (ROI)** |  |  |  |
| 5.1. Full ONH images | -0.092 | 5.257 | 0.013 |
| 5.2. OD | -0.014 | 4.889 | 0.013 |
| 5.3. OD and OC | 0.157 | 4.128 | 0.283 |
| 5.4. RNFL | 0.041 | 3.782 | 0.597 |
| **6. Validation accuracy** |  |  |  |
| 6.1. Internal | -0.114 | 4.940 | 0.030 |
| 6.2. External | 0.587 | 3.684 | 0.827 |
| **7. Electronic Databases** |  |  |  |
| 7.1. Publicly available | -0.478 | 4.219 | -0.353 |
| 7.2. Private data | -0.125 | 5.089 | 0.051 |
| 7.3. Mixed data | 0.933 | 3.758 | 1.001 |

# **Table 5. Quality assessment of included studies (n=66)—QUADAS-2 tool**

Table

Description automatically generated

# **Figure 1. Funnel plot**

**Chart, scatter chart

Description automatically generated**

# **Summary Receiver Operating Characteristic (SROC) Curve**

|  |
| --- |
| Chart, scatter chart  Description automatically generated |

## **Figure 2.1. SROC Curve for Overall Diagnostic Performance of AI**

|  |
| --- |
| Chart, scatter chart  Description automatically generatedChart  Description automatically generated with low confidenceChart  Description automatically generatedChart, scatter chart  Description automatically generatedA picture containing chart  Description automatically generated |

## **Figure 2.2. SROC Curve for Imaging Modalities: Fundus+OCT images**

|  |
| --- |
| Diagram  Description automatically generated with low confidenceChart  Description automatically generated with low confidenceChart  Description automatically generatedChart, scatter chart  Description automatically generated |

## **Figure 2.3. SROC Curve for Applied Techniques: Segmentation and Non-segmentation**

|  |
| --- |
| Chart, scatter chart  Description automatically generatedChart, scatter chart  Description automatically generatedChart, scatter chart  Description automatically generatedChart  Description automatically generated |

## **Figure 2.4. SROC Curve for Region of Interest (ROI) Selection: ONH Images, OD, OD and OC, and RNFL**

|  |
| --- |
| Chart, scatter chart  Description automatically generatedChart  Description automatically generated |

## **Figure 2.5. SROC Curve for Validation Accuracy: External and Internal Data**

|  |
| --- |
| Chart  Description automatically generated with low confidenceChart, scatter chart  Description automatically generatedA picture containing chart  Description automatically generated |

## **Figure 2.6. SROC Curve for Electronic Databases: Publicly Available and Private**

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# **Forest Plots**

## 

Table

Description automatically generated

**1.1. Overall Diagnostic Performance of AI in Glaucoma Detection: DOR**

A picture containing chart

Description automatically generated

**1.2. Overall Diagnostic Performance of AI in Glaucoma Detection: Sensitivity**

Chart

Description automatically generated with medium confidence

**1.3. Overall Diagnostic Performance of AI in Glaucoma Detection: Specificity**

# **Forest Plots 2.0: Imaging Modalities: Fundus and OCT images**

Table

Description automatically generated

**2.1. Forest Plot for Imaging Modalities: DOR**

A picture containing table

Description automatically generated

**2.2. Forest Plot for Imaging Modalities: Sensitivity**

**A picture containing diagram

Description automatically generated**

**2.3. Forest Plot for Imaging Modalities: Specificity**

**Forest Plots 3.0: Applied Techniques: Segmentation and non-segmentation**

**Table

Description automatically generated**

**3.1. Forest Plot for Applied Techniques: DOR**

**A picture containing text

Description automatically generated**

**3.2. Forest Plot for Applied Techniques: Sensitivity**

**A picture containing diagram

Description automatically generated**

**3.3.Forest Plot for Applied Techniques: Specificity**

## **Forest Plots 4.0: Region of Interest (ROI) Selection: ONH Images, OD, OD and OC, and RNFL**

Table

Description automatically generated

**4.1. Forest Plot for Region of Interest (ROI) Selection: DOR**

A picture containing table

Description automatically generated

**4.2. Forest Plot for Region of Interest (ROI) Selection: Sensitivity**

A picture containing table

Description automatically generated

**4.3. Forest Plot for Region of Interest (ROI) Selection: Specificity**

# **Forest Plots 5.0: Validation Accuracy: External and Internal Data**

## Table Description automatically generated

## **5.1. Forest Plot for Validation Accuracy: DOR**

A picture containing table

Description automatically generated

**5.2. Forest Plot for Validation Accuracy: Sensitivity**

A picture containing table

Description automatically generated

**5.3. Forest Plot for Validation Accuracy: Specificity**

## **Forest Plots 6.0: Electronic Databases: Publicly Available and Private**

**Table

Description automatically generated**

**6.1. Forest Plot for Electronic Databases: DOR**

**A picture containing table

Description automatically generated**

**6.2. Forest Plot for Electronic Databases: Sensitivity**

## A picture containing graphical user interface Description automatically generated

## **6.3. Forest Plot for Electronic Databases: Specificity**

# **References:**

1. [Bussel II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol*. 2014;98 Suppl 2(Suppl 2). doi:](http://paperpile.com/b/EA8xj3/DIHWE)[10.1136/bjophthalmol-2013-304326](http://dx.doi.org/10.1136/bjophthalmol-2013-304326)

2. [Chauhan BC, Garway-Heath DF, Goñi FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol*. 2008;92(4). doi:](http://paperpile.com/b/EA8xj3/Nc0fB)[10.1136/bjo.2007.135012](http://dx.doi.org/10.1136/bjo.2007.135012)

3. [Kass MA, Heuer DK, Higginbotham EJ, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthal*. 2002;120(6). doi:](http://paperpile.com/b/EA8xj3/KRJRd)[10.1001/archopht.120.6.701](http://dx.doi.org/10.1001/archopht.120.6.701)

4. [Mowatt G, Burr JM, Cook JA, et al. Screening tests for detecting open-angle glaucoma: systematic review and meta-analysis. *Invest Ophthalmol Vis Sci*. 2008;49(12). doi:](http://paperpile.com/b/EA8xj3/JhamM)[10.1167/iovs.07-1501](http://dx.doi.org/10.1167/iovs.07-1501)

5. [Fortune B. Optical coherence tomography evaluation of the optic nerve head neuro-retinal rim in glaucoma. *Clin Exp Optom*. 2019;102(3):286-290.](http://paperpile.com/b/EA8xj3/UOwKm)

6. [Anton A, Fallon M, Cots F, et al. Cost and detection rate of glaucoma screening with imaging devices in a primary care center. *Clin Ophthalmol*. 2017;11:337-346.](http://paperpile.com/b/EA8xj3/WdttG)

7. [Kim M, Yun J, Cho Y, et al. Deep Learning in Medical Imaging. *Neurospine*. 2019;16(4):657.](http://paperpile.com/b/EA8xj3/DuV4G)

8. [Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016;316(22). doi:](http://paperpile.com/b/EA8xj3/3kNAK)[10.1001/jama.2016.17216](http://dx.doi.org/10.1001/jama.2016.17216)

9. [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*. September 2019. doi:](http://paperpile.com/b/EA8xj3/nfnIn)[10.1001/jamaophthalmol.2019.3501](http://dx.doi.org/10.1001/jamaophthalmol.2019.3501)

10. [Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639). doi:](http://paperpile.com/b/EA8xj3/ZTF9U)[10.1038/nature21056](http://dx.doi.org/10.1038/nature21056)

11. [Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digital Medicine*. 2020;3(1):1-8.](http://paperpile.com/b/EA8xj3/NDJW6)

12. [van der Heijden AA, Abramoff, Verbraak F, van Hecke MV, Liem A, Nijpels G. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmol*. 2018;96(1). doi:](http://paperpile.com/b/EA8xj3/i8ipG)[10.1111/aos.13613](http://dx.doi.org/10.1111/aos.13613)

13. [Kim Y-G, Kim S, Cho CE, et al. Effectiveness of transfer learning for enhancing tumor classification with a convolutional neural network on frozen sections. *Sci Rep*. 2020;10(1):1-9.](http://paperpile.com/b/EA8xj3/cXho2)

14. [Bock R, Meier J, Nyúl LG, Hornegger J, Michelson G. Glaucoma risk index: automated glaucoma detection from color fundus images. *Med Image Anal*. 2010;14(3). doi:](http://paperpile.com/b/EA8xj3/UNNyl)[10.1016/j.media.2009.12.006](http://dx.doi.org/10.1016/j.media.2009.12.006)

15. [University of Bristol. QUADAS-2.](http://paperpile.com/b/EA8xj3/3o6N) <https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/.> [Accessed February 23, 2021.](http://paperpile.com/b/EA8xj3/3o6N)

# 