ADITUM

Journal of Ophthalmology and Vision Care

Open Access

Research Article

Artificial Intelligence Applications, and Performance Metrics in Ophthalmology: A Systematic Review and Meta-Analysis

Gabriella Bulloch[1,2], Ishith Seth[2*], Zhuoting Zhu[2], Fiona Jane Stapleton[3], Adrian Fung[4], Zachary Tan[2], Hugh R. Taylor[1,2]

[1] Faculty of Science, Medicine and Health, University of Melbourne, Victoria, 3051, Australia
[2] Department of Ophthalmology, Centre of Eye Research Australia, Victoria, 3004, Australia
[3] School of Optometry and Vision Science, University of New South Wales, New South Wales, 2052, Australia
[4] Faculty of Medicine and Health, University of Sydney, New South Wales, 2006, Australia

Article Info

Received: February 23, 2023
Accepted: February 28, 2023
Published: March 03, 2023

***Corresponding author:** Ishith Seth, Department of Ophthalmology, Centre of Eye Research Australia.

Citation: Bulloch G, Seth I, Zhu Z, Fiona J Stapleton, Fung A. (2023) "Artificial Intelligence Applications, and Performance Metrics in Ophthalmology: A Systematic Review and Meta-Analysis", Ophthalmology and Vision Care, 3(1); DOI: http;//doi.org/03.2023/1.10238.

Abstract

Purpose: To evaluate the overall performance of various Artificial intelligence (AI) models in ophthalmology for the diagnosis of various ophthalmic diseases despite of variations in methodology, platforms built, and workflows. AI technologies can revolutionize ophthalmology and vision sciences through automating image analysis.

Methods: A systematic search on EMBASE, Medline (via PubMed), CINHAL, Cochrane Library, Clinicaltrial.gov, Google Scholar, Scopus, and Web of Science was conducted for studies published up to March 2022. Two authors independently screened all titles and abstracts against predefined inclusion and exclusion criteria and extracted data. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was used to assess for risk of bias and applicability. The pooled sensitivity (SE), specificity (SP), accuracy, and area under the curve (AUC) were estimated using a random-effects model with a 95% Confidence Interval (CI). An assessment of publication bias was performed. The protocol of this meta-analysis was published online PROSPERO under registration number CRD42021242593.

Results

Our meta-analysis included a total of 42 studies that met the inclusion criteria. The MESSIDOR database was most frequently used for training and testing among selected studies. Pooled performance of AI algorithms for included ophthalmic disorders were SE=92.93% (95% CI 91.01, 94.86), SP=88.73% (95% CI 83.55, 93.91), accuracy =94.62% (95% CI 91.98, 97.27), and AUC=0.96 (95% CI 0.94, 0.98).

Conclusion

Currently published AI algorithms are highly accurate for diagnosing ophthalmic diseases and have the potential to unlock population-based screening for common eye conditions. The adoption of standardized reporting frameworks and more prospective/randomized control trials are currently required to improve generalizability of AI for clinical practice.

Keywords: artificial intelligence; ophthalmology; predictive analytics; machine learning; deep learning

Introduction

There are 2.2 billion people with visual impairment globally, with almost half of these being preventable or yet to be addressed.[1, 2] Vision impairment without intervention leads to significant morbidity, increases health services demand, and carries a global financial burden of an estimated $244 billion annually.[3] With non-communicable diseases like diabetes and heart disease becoming increasingly common in young populations, retinal pathologies resulting from comorbidities have become more frequent.[4] Similarly, retinopathy of prematurity (ROP), the most common cause of blindness in children worldwide, carries an enormous healthcare burden subserved by limited neonatal intensive care services and late diagnosis.[5] Improving access to eye disease screening is a sensible solution, although as the global population rises, demographics shift towards ageing populations, and clinician availability remains insufficient, these challenges bottleneck eye care services.[6]

practical and financial challenges that inhibit population-based screening and diagnosis of eye diseases.[7] AI utilizes computer-based algorithms and novel software to replicate human intelligence. Its application effectively replaces problem-solving and practical tasks that are otherwise laborious and time-inefficient in domains of society which are bottlenecked by imbalanced service-to-demand ratios.[8] AI technologies have established high efficiency, accuracy, and precision within medicine, and has already demonstrated applications to ophthalmology through data evaluation, segregation, electronic diagnosis, and potential outcome prognosis.[9,10] Machine learning (ML) is a subset of AI that learns automatically from data sets in the absence of explicitly programmed rules.[11] Deep learning (DL) is a subclass of ML and trains itself using multiple layers of neural networks which are adaptable programming units inspired by the structure of human neurons. DL has demonstrated significant potential in classification and feature extraction and has the ability to learn complex representations from raw data to improve pattern recognition.[12, 13] Its image recognition and computer vision have made it a favorable tool for the grading of images, and individual studies show it has improved image analysis for diagnosis and preditction.[13, 14]

AI's accuracy in automated diagnosis, time efficiency, and outcome prediction has enabled desirable applications within healthcare, but for its successful implementation within clinical practice AI needs to ensure its accuracy is not inferior to clinicians.[15] Variations in methodology and platforms built for DL give an overall illusion its clinical validity is not yet warranted. Indeed, various workflows for DL, variations in testing and validation set sizes, fluctuating disease definitions, and the absence of external validation by clinical experts may cloud the establishment of ground truth and diminish its trustworthiness.[14,15] This heterogeneity also complicates a formal evaluation of AI studies and is yet to be accounted for by the integration of specific AI/ML reporting frameworks.[16] Ethical legislation surrounding the use and scrutiny of AI continues to be of concern to healthcare providers, and bench to bedside challenge can only be overcome by conducting studies that assess AI to a high degree of scrutiny not just in performance, but equally in ethics, effectiveness, replicability, and transparency.[17, 18]

Considering the timeliness of AI, this systemic review and meta-analysis investigated and scrutinized the ability of AI to diagnose all ocular disorders that satisfied our search criterion. The advantages and limitations of AI in the management of retinal disorders were also explored.

Methods

This study was performed according to the Preferred Reporting Items for Systematic Reviews and Metanalyses (PRISMA) statement.[19] The protocol of this meta-analysis was published online at the International Prospective Register of Systematic Reviews (PROSPERO) under registration number (CRD42021242593). There were no study restrictions imposed on different populations, races, ethnicity, and origin.

Literature Search

A comprehensive systematic search on EMBASE, Medline (via PubMed), CINHAL, Cochrane Library, Clinicaltrial.gov, Google Scholar, Scopus, and Web of Science was conducted for studies published from January 2009 to March 2022. A variety of all possible keywords like 'artificial intelligence and ophthalmology', 'deep learning and ophthalmology', 'machine learning and ophthalmology', 'convolutional neural network and ophthalmology', 'deep neural network and ophthalmology', 'automated technique and ophthalmology' were listed to avoid any data loss. No age, gender, and population filters were imposed. Two authors independently screened all titles and abstracts against predefined inclusion and exclusion criteria. Any differences in articles selected by the two were discussed with third author to reach a decision regarding inclusion. The reference lists of screened articles were also reviewed for any missed literature.

Inclusion and Exclusion Criteria

The established inclusion criteria were as follows: (1) all published data reporting the use of AI, DL, or ML in ophthalmology, (2) studies that evaluated the sensitivity (SE), specificity (SP), accuracy, and area under ROC curve (AUC) in their study OR any one of the mentioned parameters, (3) studies provided an outcome of AI in ophthalmology for a pathological condition against healthy population sample sets of normal eyes, retinal images, and photographs, (4) studies provided information about databases/methodology used, (5) studies clearly described the type of AI used and detected eye disease, (5) studies published in English. All full-text studies including randomized control trials, original research articles, descriptive and analytic studies (cohort or case-control) were included.

The exclusion criteria were as follows: (1) studies that did not measure conclusive performance outcomes, (2) parameters used to analyze data were different from defined ones, (3) incomplete studies, (4) poster or scientific presentations, (5) reviews, meta-analysis, opinion articles, letter to editor, short communications, and case reports.

Outcome Measures

The primary outcome measures assessed the performance of AI in ophthalmology including SE, SP, accuracy, and AUC. Secondary outcomes were not defined in advance.

Data extraction and quality assessment

Data extraction was done twice as per defined inclusion criteria and keywords, to avoid any risk of bias and possibility of missing data.[20] Data extracted include first author's name, publication year, used dataset or methodology information, measured parameters in terms of SE, SP, accuracy, and AUC. We used the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool for assessing the quality of included diagnostic studies. The QUADAS-2 scale comprises four domains: patient selection, index test, reference standard, and flow and timing. The first three domains are used for evaluating the risk of bias in applicability. The overall risk of bias was categorized into three groups (low, high, and unclear risk bias).

Data Synthesis

Statistical analyses were conducted using review manager (RevMan version 5.4, Cochrane collaboration, Oxford, UK). Overall performance measures along with 95% confidence intervals (CIs) range were calculated for all defined primary indicators i.e., sensitivity, specificity, accuracy, and AUC, and were represented by forest plot. A standard error of 0.05 was observed in all tests. Data heterogeneity was checked, and publication bias was assessed using Egger's test. Further, a Youden plot was generated against sensitivity and specificity measures to detect the test accuracy, and Youden's index was calculated.

Results

Study Selection

A total of 1130 potentially eligible records were extracted in the initial data retrieval process. During the screening, 488 records were eliminated due to duplication, and 492 were eliminated based on the study title and abstract. Of the 150 remaining studies reviewed, 90 were excluded for not meeting inclusion criteria, and 18 were conference abstracts or poster presentations. Finally, 42 studies were included in the final meta-analysis.[12,21-62] The process used to search and identify studies is illustrated in Figure 1.

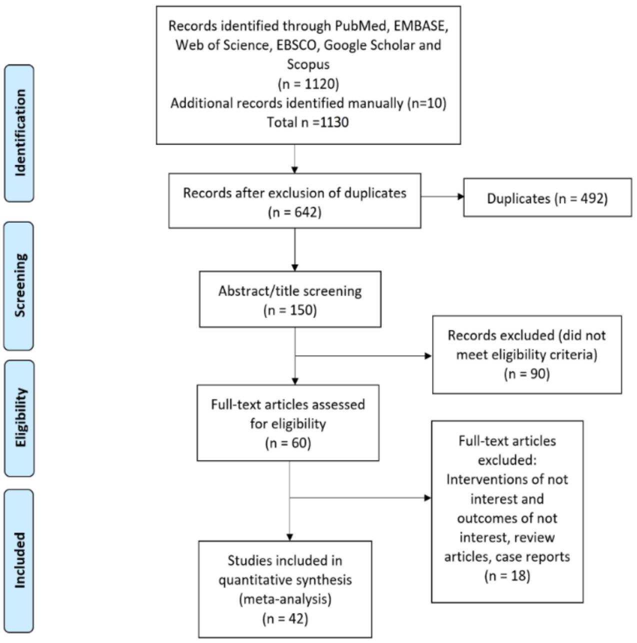

Identification
Screening
Eligibility
Included
Records identified through PubMed, EMBASE, Web of Science, EBSCO, Google Scholar and Scopus
(n = 1120)
Additional records identified manually (n=10)
Total n =1130
Records after exclusion of duplicates
(n = 642)
Duplicates (n = 492)
Abstract/title screening
(n = 150)
Records excluded (did not meet eligibility criteria)
(n = 90)
Full-text articles assessed for eligibility
(n = 60)
Full-text articles excluded:
Interventions of not interest and outcomes of not interest, review articles, case reports
(n = 18)
Studies included in quantitative synthesis (meta-analysis)
(n = 42)

Figure 1: Summary of study selection process Preferred Reporting Items for Systematic Review and Meta-Analyses flow diagram

Study Characteristics

Selected studies were analyzed for the performance of AI in the diagnosis of common eye diseases, dataset/methodology used, AI model tested, validation performed, and reference standards used to assess performance. These findings are described in Supplementary Table 1. Indicators of performance including specificity, sensitivity, accuracy, and AUC are summarized in Table 1.

Table 1: Performance indicators in selected studies

Sl No	Author's Name, year and Reference number	Dataset	Sensitivity (%)	Specificity (%)	Accuracy (%)	Area under ROC Curve
1	Aquino et al., 2009	MESSIDOR	-	-	98.83	-
2	Haloi et al., 2015	MESSIDOR, ROC	97	96	96	0.982
3	Ahmed et al., 2015	MESSIDOR	-	-	97.8	-
4	Liskowski et al., 2016	DRIVE, STARE, CHASE DB	-	-	97	0.99
5	Asoaka et al., 2016	Private: 171	-	-	-	0.926
6	Grinven et al., 2016	MESSIDOR	91.9	91.8	-	0.972
		EyePACS	83.7	85.1	-	0.895
7	Abràmoff M et al., 2016	MESSIDOR -2	96.8	87	-	_
8	Colas E et al., 2016	EyePACS	96.2	66.6	-	0.946
9	Gulshan et al., 2016	EyePACS-1,	90.3	98.1	-	0.991
		MESSIDOR -2	87	98.5	-	0.99
10	Gargeya et al., 2017	EyePACS, MESSIDOR e-Optha2	94	98	-	0.97
11	Quellec et al., 2017	Kaggle,	-	-	-	0.954
		E-Ophtha	-	-	-	0.949
		DIARETDB	-	-	-	0.955
12	Ambrósio R Jr et al., 2017	(RF/LOOCV)	100	100	-	0.996
13	Takahashi et al., 2017	9939 images (Posterior Pole Photographs)	-	-	80	-
14	Tan et al., 2017	CLEOPATRA	-	-	87.58	-
			-	-	71.58	-
15	Oliveira et al., 2018	DRIVE	80.39	98.04	95.76	0.9821
		STARE	83.15	98.58	96.94	0.9905
		CHASE DB1	77.79	98.64	96.53	0.9855

16	Schmidt-Erfurth U et al., 2018	Coherence Tomography (OCT)	-	-	-	0.68
			-	-	-	0.8
17	Lin et al., 2018	Kaggle	73.24	93.81	86.1	0.92
18	Chakravarty et al., 2018	REFUGE	-	-	-	0.9456
19	Li Z et al., 2018	Private:48000+	95.6	92		0.986
20	Chai Y et al., 2018	Private: 2554	-	-	91.51	-
21	Mitra et al., 2018	MESSIDOR	-	99.14	99.05	-
		EyePACS	-	98.17	98.78	-
22	Liu et al., 2018	HRF, RIM-ONE	86.7	96.5	91.6	0.97
23	Orlando et al., 2018	e-optha	-	-	-	0.8812
		MESSIDOR	-	-	-	0.8932
24	Lam et al., 2018	EyePACS, e-optha	-	-	98	0.95
25	Grassmann et al., 2018	AREDS and KORA	94.3	84.2	-	-
26	Zhang et al., 2018	DRIVE	87.23	96.18	95.04	0.9799
		STARE	76.73	99.01	97.12	0.9882
		CHASE DB1	76.7	99.09	97.7	0.99
27	Zhou W et al., 2018	MESSIDOR	-	-	99.83	-
28	Al-Bander et al., 2018	MESSIDOR and Kaggle	-	-	97	-
29	An G et al., 2019	Machine Learning	-	-	-	0.963
30	Medeiros et al., 2019	Deep-Learning (DL)	-	-	83.7	-
31	Lin, H. et al., 2019	CC Cruiser	-	-	87.4	-
		Kaggle	-	-	70.8	-
32	Zéboulon P et al., 2020	Machine learning algorithm	100	100	99.3	-
33	Varadarajan AV et al., 2020	EyePACS	85	80	-	0.89
34	Ahn H et al., 2020	Artificial intelligence ECcSMOTE II	-	-	99.05	-
35	Rim TH et al., 2020	Deep-Learning (DL) Algorithms RetiSort	-	-	99	-
36	Lee J et al., 2020	Machine learning classifiers	-	-	-	0.881
37	Huang Y-P et al., 2020	VGG19 model	96.6	95.2	96	-
38	Tham YC et al., 2020	ResNet-50	90.7	86.8	-	0.94
39	Son J et al., 2020	IDRiD, e-Ophtha, MESSIDOR	97.2	96.8	-	0.99
40	Li Z et al., 2020	Deep Learning	99.5	99.5	99.5	1.00
41	Dai L et al., 2021	DeepDR	92.8	81.3	-	0.95
42	Luo X et al., 2021	EfficientNet-B3	99.49	97.86	99.02	0.99

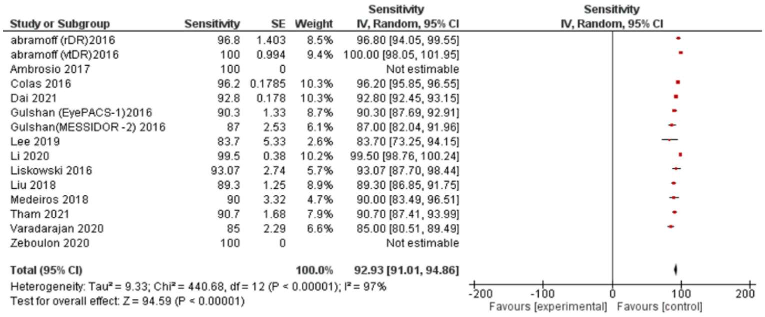
Quality assessment and publication bias

The quality of included studies was assessed using the QUADAS-2 tool and was presented in Supplementary Table 2. For patient selection and index tests, all studies were identified to exhibit a low risk of bias. An unclear risk of bias was found in flow and timing as well as reference standard domains for all included studies. Egger's test for a regression intercept gave a P-value of 1.000, indicating no evidence of publication bias.

Outcome measures

Sensitivity

Of the 42 selected studies, 13 studies, and 15 datasets reported SE performance. The data indicators represent the data as valid with low error and no publication bias. The pooled SE of reviewed publications was 92.93% (95% CI 91.01, 94.86). No studies reported lower than 80% of SE, while Five studies reported 80-90%, Seven studies with >90%, and Three studies with 100 % SE. The I^2 was 97%. (Table 1, Figure 2).



Study or Subgroup	Sensitivity	SE	Weight	Sensitivity IV, Random, 95% CI
abramoff (rDR)2016	96.8	1.403	8.5%	96.80 [94.05, 99.55]
abramoff (vtDR)2016	100	0.994	9.4%	100.00 [98.05, 101.95]
Ambrosio 2017	100	0		Not estimable
Colas 2016	96.2	0.1785	10.3%	96.20 [95.85, 96.55]
Dai 2021	92.8	0.178	10.3%	92.80 [92.45, 93.15]
Gulshan (EyePACS-1)2016	90.3	1.33	8.7%	90.30 [87.69, 92.91]
Gulshan(MESSIDOR -2) 2016	87	2.53	6.1%	87.00 [82.04, 91.96]
Lee 2019	83.7	5.33	2.6%	83.70 [73.25, 94.15]
Li 2020	99.5	0.38	10.2%	99.50 [98.76, 100.24]
Liskowski 2016	93.07	2.74	5.7%	93.07 [87.70, 98.44]
Liu 2018	89.3	1.25	8.9%	89.30 [86.85, 91.75]
Medeiros 2018	90	3.32	4.7%	90.00 [83.49, 96.51]
Tham 2021	90.7	1.68	7.9%	90.70 [87.41, 93.99]
Varadarajan 2020	85	2.29	6.6%	85.00 [80.51, 89.49]
Zeboulon 2020	100	0		Not estimable
Total (95% CI)			**100.0%**	**92.93 [91.01, 94.86]**

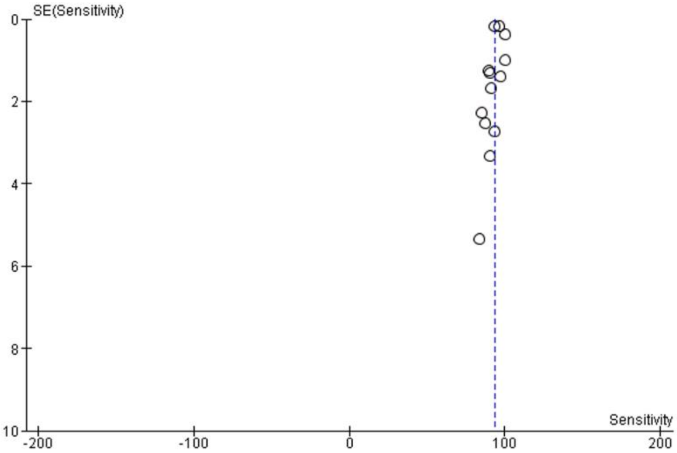
Heterogeneity: $Tau^2 = 9.33$; $Chi^2 = 440.68$, df = 12 ($P < 0.00001$); $I^2 = 97\%$

Test for overall effect: $Z = 94.59$ ($P < 0.00001$)

Figure 2: Forest plot of sensitivity analysis of selected studies

Supplementary Figure 1: Funnel plot of publication bias of sensitivity analysis

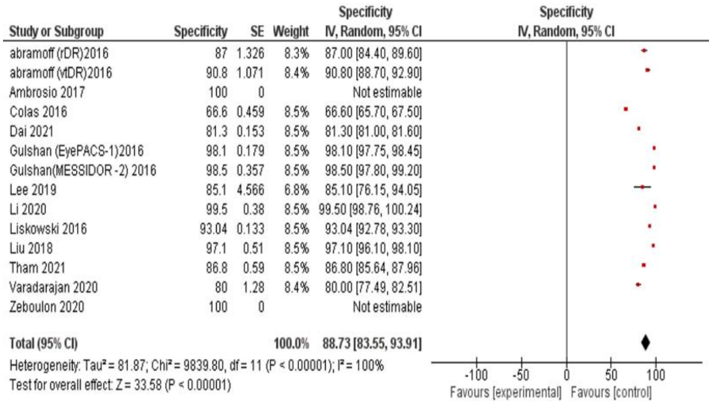
No publication bias was detected, Supplementary figure 1.



Specificity

Twelve studies and 14 datasets reported SP performance. Only one study reported less than 70% SP (66.6%). All remaining studies reported more than 80% SP of AI and 2 studies reported 100% SP (Table 1, Figure 3).

Figure 3: Forest plot of specificity analysis of selected studies

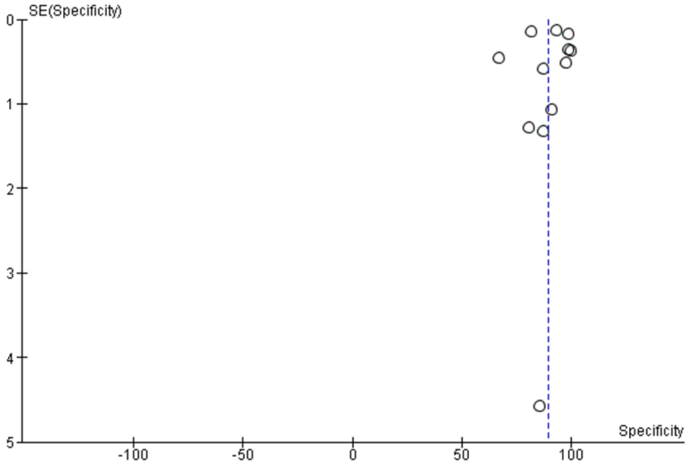


Study or Subgroup	Specificity	SE	Weight	Specificity IV, Random, 95% CI
abramoff (rDR)2016	87	1.326	8.3%	87.00 [84.40, 89.60]
abramoff (vtDR)2016	90.8	1.071	8.4%	90.80 [88.70, 92.90]
Ambrosio 2017	100	0		Not estimable
Colas 2016	66.6	0.459	8.5%	66.60 [65.70, 67.50]
Dai 2021	81.3	0.153	8.5%	81.30 [81.00, 81.60]
Gulshan (EyePACS-1)2016	98.1	0.179	8.5%	98.10 [97.75, 98.45]
Gulshan(MESSIDOR -2) 2016	98.5	0.357	8.5%	98.50 [97.80, 99.20]
Lee 2019	85.1	4.566	6.8%	85.10 [76.15, 94.05]
Li 2020	99.5	0.38	8.5%	99.50 [98.76, 100.24]
Liskowski 2016	93.04	0.133	8.5%	93.04 [92.78, 93.30]
Liu 2018	97.1	0.51	8.5%	97.10 [96.10, 98.10]
Tham 2021	86.8	0.59	8.5%	86.80 [85.64, 87.96]
Varadarajan 2020	80	1.28	8.4%	80.00 [77.49, 82.51]
Zeboulon 2020	100	0		Not estimable
Total (95% CI)			**100.0%**	**88.73 [83.55, 93.91]**

Heterogeneity: $Tau^2 = 81.87$; $Chi^2 = 9839.80$, df = 11 ($P < 0.00001$); $I^2 = 100\%$
Test for overall effect: Z = 33.58 ($P < 0.00001$)

The overall SP of reviewed publications was 88.73% (95% CI 83.55, 93.91). The I^2 of included studies was 100%. No publication bias was detected, Supplementary figure 2.

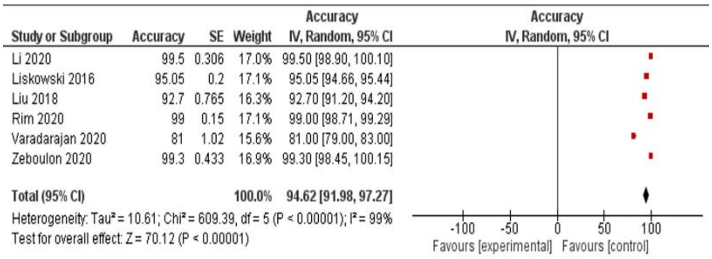
Supplementary Figure 2: Funnel plot of publication bias of specificity analysis



Accuracy

Accuracy is the second most reported performance indicator, which was reported in Six studies among 42 selected studies. Only one study reported less than 85% accuracy, five studies reported less than 90% accuracy. The overall accuracy of AI among selected studies was 94.62% (95% CI 91.98, 97.27), (Table 1, Figure 4).

Figure 4: Forest plot of accuracy measurement of selected studies

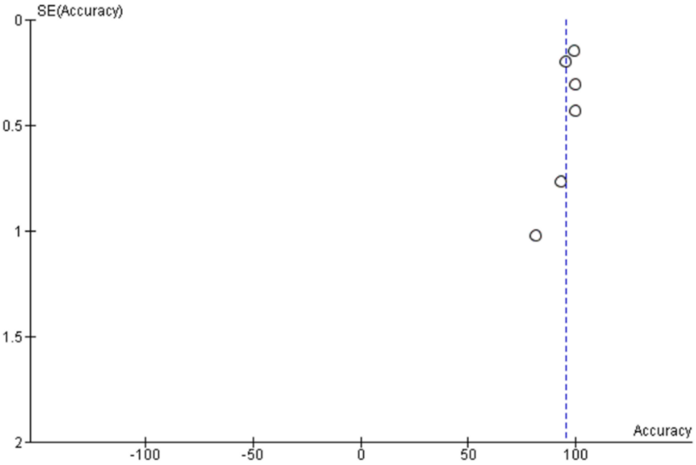


Study or Subgroup	Accuracy	SE	Weight	Accuracy IV, Random, 95% CI
Li 2020	99.5	0.306	17.0%	99.50 [98.90, 100.10]
Liskowski 2016	95.05	0.2	17.1%	95.05 [94.66, 95.44]
Liu 2018	92.7	0.765	16.3%	92.70 [91.20, 94.20]
Rim 2020	99	0.15	17.1%	99.00 [98.71, 99.29]
Varadarajan 2020	81	1.02	15.6%	81.00 [79.00, 83.00]
Zeboulon 2020	99.3	0.433	16.9%	99.30 [98.45, 100.15]
Total (95% CI)			**100.0%**	**94.62 [91.98, 97.27]**

Heterogeneity: $Tau^2 = 10.61$; $Chi^2 = 609.39$, df = 5 ($P < 0.00001$); $I^2 = 99\%$
Test for overall effect: Z = 70.12 ($P < 0.00001$)

The I^2 of included studies was 100%. No publication bias was detected, Supplementary figure 3.

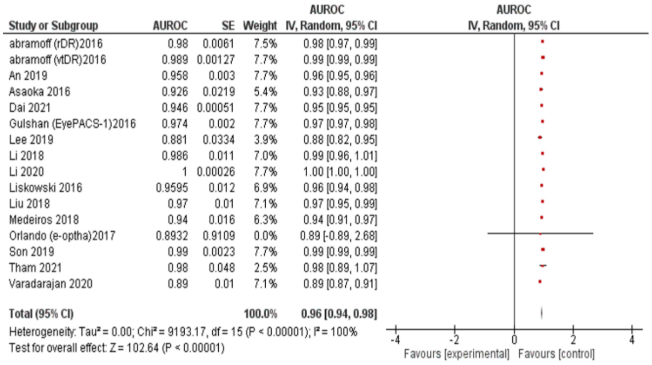
Supplementary Figure 3: Funnel plot of publication bias of accuracy analysis



Calculation of area under the ROC Curve (AUC)

AUC is the most reported indicator of AI performance, which represents a combined performance measure across all viable classification thresholds. AUC values between 0.8 to 0.9 are considered excellent, and more than 0.9 considered to be outstanding.[63] The overall AUC of studies was 0.96 (95% CI 0.94 – 0.98). All reported results fall in the excellent and outstanding categories. The I^2 of included studies was 100% (Table 1, Figure 5).

Figure 5: Forest plot of area under receiver operating characteristics curve of selected studies



Study or Subgroup	AUROC	SE	Weight	AUROC IV, Random, 95% CI
abramoff (rDR)2016	0.98	0.0061	7.5%	0.98 [0.97, 0.99]
abramoff (vtDR)2016	0.989	0.00127	7.7%	0.99 [0.99, 0.99]
An 2019	0.958	0.003	7.7%	0.96 [0.95, 0.96]
Asaoka 2016	0.926	0.0219	5.4%	0.93 [0.88, 0.97]
Dai 2021	0.946	0.00051	7.7%	0.95 [0.95, 0.95]
Gulshan (EyePACS-1)2016	0.974	0.002	7.7%	0.97 [0.97, 0.98]
Lee 2019	0.881	0.0334	3.9%	0.88 [0.82, 0.95]
Li 2018	0.986	0.011	7.0%	0.99 [0.96, 1.01]
Li 2020	1	0.00026	7.7%	1.00 [1.00, 1.00]
Liskowski 2016	0.9595	0.012	6.9%	0.96 [0.94, 0.98]
Liu 2018	0.97	0.01	7.1%	0.97 [0.95, 0.99]
Medeiros 2018	0.94	0.016	6.3%	0.94 [0.91, 0.97]
Orlando (e-optha)2017	0.8932	0.9109	0.0%	0.89 [-0.89, 2.68]
Son 2019	0.99	0.0023	7.7%	0.99 [0.99, 0.99]
Tham 2021	0.98	0.048	2.5%	0.98 [0.89, 1.07]
Varadarajan 2020	0.89	0.01	7.1%	0.89 [0.87, 0.91]
Total (95% CI)			**100.0%**	**0.96 [0.94, 0.98]**

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 9193.17$, df = 15 ($P < 0.00001$); $I^2 = 100\%$
Test for overall effect: $Z = 102.64$ ($P < 0.00001$)

Test of Accuracy
Youden's Index is a combined measure of SE and SP for indexing test accuracy (Supplementary Figure 5).

Supplementary Figure 5: Youden's plot of sensitivity and specificity of selected studies

The maximum value of Youden's index is 1 indicating a perfect test, while the minimum value possible is 0 when the test has no diagnostic value. This study recorded a Youden's index of 0.85, indicating a high accuracy across studies.

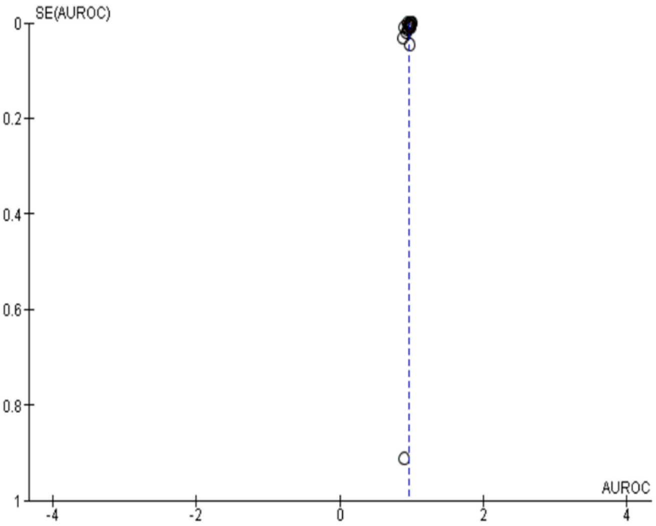
Discussion

This systematic review and meta-analysis included studies up to March 2022 and demonstrated AI achieved a high performance for the recognition of eye diseases. Performance in sensitivity was 93%, specificity 88.73%, accuracy 94.62%, and AUC 0.96. Of the various AI platforms used, MESSIDOR database was most frequently utilized for training and testing amongst our selected studies. A Youden's index of 0.85 indicated that pooled estimates were of high accuracy. These results suggest AI technologies may assist in the diagnosis of eye diseases and improve access to screening and management of these conditions.[64]

AI in ophthalmology has gained popularity over the last decade, evidenced by the plethora of publications accrued and innovations since acquired.[65] Our study set ranged from 2009-2022 and despite the impressive performance, significant methodological deficits were noted amongst the 42 included studies. After QUADAS assessment it was clear most retrospective studies included data labels and quality reference standards which were not primarily intended for the purpose of measuring AI's utility and performance. As a result, the translation of these results to real-world outcomes may be limited. Furthermore, smaller training sets in the learning phase of some DL models may lead to poor prediction accuracy due to overfitting, or rather them being insufficient representatives of disease.[66] If this occurs, accuracy suffers in the long term and the model may miss important features of a disease or illness. External validation is therefore critical to the evaluation of a model's accuracy and ensuring a ground truth is addressed. Unfortunately, many studies did not perform an external validation of the model in a separate test dataset which is crucial to ensuring real-world clinical performance of DL, eliminating bias, and ensuring diagnosis/prediction accuracy of the model.[13] Low evaluative measures limit the validity of performance parameters particularly AUC, and hints to the presence of unpublished publication bias. Robustly reported evaluative measures would have provided an objective standard and needs to be confronted in future publications to ensure reliability of AI and enable comparison between DL models.[29, 42, 43, 49, 51]

No publication bias was detected, Supplementary figure 4.

Supplementary Figure 4: Funnel plot of publication bias of calculation of area under the ROC curve



In the past, AI has lacked directives that would remove heterogeneity between studies and allow its transition from bench to bedside. To improve its integrity, guidelines including the SPIRIT-AI[67], CONSORT-AI[68], STARD-AI[15], and TRIPOD-AI[69] statements are currently being developed and phased into studies under published guidance from the EQUATOR network. SPIRIT-AI[67] and CONSORT-AI[68] statements are the first international standards for clinical trials of AI systems and aim to improve the standard of design and delivery by providing completeness of intention and transparency of reporting. In contrast, STARD-AI[15] and TRIPOD-AI[69] are specific to studies investigating diagnostic accuracy and prognostic modelling, and aim to improve methodology reporting and outcome measures, and to standardise nomenclature. This is a significant step for AI because reporting of clinical prediction models is currently poor and a threat to its clinical integrity. The implementation of these protocols will be a welcomed improvement to the field of AI to ensure its performance and safety in clinical practice.

In addition, few studies were randomized control trials (RCT) or prospective studies, while a large proportion were retrospective cohort studies. Using retrospective data is certainly a convenient and less costly way of testing AI's accuracy which, by design, inherently demands large quantities of data to optimize its neural networks. An unfortunate side effect is the current literature does not compare AI's performance against experts in a clinical, real time setting. Hence there exists a large potential for the development of prospective studies and RCTs which may compare the applicability of these algorithms in clinical practice. If undertaken, it cannot be underestimated how important these applications can be for rural and developing areas particularly. AI techniques can smoothen the multistage process of screening, staging and treatment decision for a condition, thereby sharing the burden of clinical experts and providing a greater population outreach.[70] In Australia, remote and regional communities are associated with less frequent eye checkups and are even lower in First Nations people.[71] Indigenous Australians also present later to eye health professionals for a presenting problem relative to Non-Indigenous Australians, predisposing them to preventative ocular diseases.[71] As a result, there have historically been higher instances of pterygium, cataract, ocular trauma, and glaucoma in rural populations.[72] Here, AI has the potential utility of providing basic eye health reports and information to remote regions that lack access to consistent eye health-care services. Moreover, AI systems will be invaluable to Indigenous communities particularly by negating many current concerns pertaining to Non-Indigenous clinicians, interpreters, cultural and language barriers which currently contribute to health gaps between Indigenous and Non-Indigenous Australians.[73, 74] To date, only few AI studies have been the subject of prospective cohort studies but the preliminary results have so far been positive.[73-77]

Prediction tools are another innovative and potentially positive application for AI, especially in comorbid populations where retinopathies and retinal structure abnormalities are associated with comorbidities. In contrast to classical prediction models that rely on cross-sectional data and are prone to overfitting, AI techniques can incorporate longitudinal data which optimizes prediction over time without human intervention.[78-81] Despite these unique characteristics, prediction models are not yet optimized and are subject to the same scrutiny as other AI technologies within medicine. Arcadu et al predicted DR worsening in 529 patients at 6, 12, and 24 months with an overall AUC of 0.68 using deep CNN (DCNN) and random forest aggregation.[82] Schmidt-Erfurth et al successfully predicted a 2-year progression of intermediate AMD to choroidal neovascularization or geographic atrophy in 495 eyes with AUC= 0.68 & 0.80, respectively.[36] Lastly, a recent systematic review summarized the ability of various DL models to isolate and predict geographic atrophy progression, an end-stage feature of chronic AMD, reported a low R^2value of 0.32 in the studies that predicted progression.[83] An online survey of clinicians on the use of AI in ophthalmology, dermatology, radiology and radiation oncology revealed improved access to disease screening as the greatest perceived advantage to the use of AI.[84] Whilst AI shows potential utility in ophthalmology for disease prediction and progression, its reliability remains to be optimized and is an ongoing area of emerging research.

Currently, regulatory agencies such as the United States Food and Drug Administration (USFDA) and Therapeutic Goods Australia (TGA) loop AI/ML under the umbrella of software as a medical device (SaMD)[85, 86] when they are being approved for therapeutic use. In 2018 the first DL system in ophthalmology to be cleared by USFDA was IDx-DR for automated diagnosis of more-than-referable DR.[87] IDx-DR utilizes AI as a fundus image analyzer and provides diagnosis and referral to a specialist if a pathology is detected. In 2020 EyeArt also achieved USFDA clearance for detecting clinical DR and vision-threatening DR retinopathy in adults with diabetes.[88] Both technologies received 510(k) clearance by the U.S. Food & Drug Administration for DR meaning the technologies demonstrate themselves as safe and effective compared to a similar, legally marketed algorithm. These clearances are landmark occurrences for AI/ML in Ophthalmology because pathways by governing regulatory bodies are evolving entities with stringent criteria needed to prove risk and functionality. Despite being proven safe in comparison with other marketable technologies, the more political challenge remains in its adoption to clinical practice whilst physicians and patients still lack confidence and trust.[89] Therefore, whilst reporting criteria and regulatory bodies are hurdles which may be succeeded by improvements previously outlined, AI/ML will face an uphill popularity battle before earning a place at the clinician's desk.

The present meta-analysis is one of few systematic reviews and meta-analysis interpreting the performance of AI in ophthalmology. It comprises a total of 42 studies based on 23 different databases, and our results suggest that AI has immense potential in ophthalmology for image interpretation. The breadth of studies selected encompasses performance across various important pathological conditions in ophthalmology, highlighting the generalizability of AI for image analysis.

Despite this, our findings have several limitations. Firstly, because our aim was broadly defined and lent itself to a pooled analysis, AI performance according to ocular pathology was not investigated. Even so, our analysis shows that across most studies there is a high sensitivity, specificity, and AUC. Secondly, we excluded studies that did not report performance indicators like sensitivity, specificity, accuracy, and AUC. This limited the scope of eye diseases available for analysis. Thirdly many studies have various methodological deficits as detailed earlier, making their reported diagnostic accuracy potentially unreliable, and our pooled accuracy potentially an overestimation of true accuracy in real-world practice. Fourthly we were not able to assess the selected studies against any standard reporting framework as AI-specific reporting guidelines including SPIRIT-AI,[67] CONSORT-AI[68], STARD-AI[15], and TRIPOD-AI[69] are not widely adopted by current literature. Lastly, our study aimed to have a comprehensive overview of AI in ophthalmology, it was beyond our scope to statistically compare between different imaging modalities, thus leading us to accept their innate differences. Despite this, the imaging modalities used are all diagnostically accepted means of screening for eye diseases.

A primary concern after analyzing the chosen studies was their significant heterogeneity. This may limit the generalizability of AI performance. Reporting standards for ML related studies across the globe are currently unsatisfactory, and with AI being of great ethical concern there needs to be a governing force for regulating its use in research and clinical practice. The universal adoption of SPIRIT-AI, CONSORT-AI, TRIPOD-AI, and STARD-AI frameworks in future studies will eliminate inconsistencies, homogenize means of data reporting and ensure data reproducibility.[90] While the integration of AI into healthcare is likely to be widespread in the future it remains a current challenge for clinicians and patients to fully trust the potential of AI/DL. Therefore, guidelines to regulate AI use and build appropriate ethical legislation to safeguard concerns pertaining to medical error, control over AI, and patient data protection need to be developed in a timely fashion.

The following would benefit for streamlining AI into ethical legislation; 1) encouraging the external validation of DL and AI systems from clinicians or experienced graders to reach a ground truth, 2) guiding methods for determining appropriate training and test set size and 3) mandatory reporting of sensitivity, specificity, accuracy, and AUC. By adopting the recently constructed frameworks for reporting AI, studies will become more reliable in their relatedness to clinical practice and deemed more trustworthy by design.

AI is a rapidly evolving field with immense potential in healthcare. This study has demonstrated AI has high and, in some cases, excellent performance in the field of ophthalmology. These technologies may soon play an increasingly significant role in the diagnosis and treatment of ocular pathologies. The adoption of standardized reporting frameworks and more prospective/randomized control trials are currently required to improve generalizability of AI for clinical practice.

Short running title: Artificial Intelligence and ophthalmology.

Conflict of interest: The authors have no conflicts of interest to disclose.

Funding sources: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector. The authors have no financial relationships relevant to this article to disclose.

Acknowledgment: None

Data statement: The data that support the findings of this study are available on request from the corresponding author.

References

1. World Health Organization. Blindness and vision impairment. Available from https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment
2. Lamoureux E, Pesudovs K. Vision-specific quality-of-life research: a need to improve the quality. Am J Ophthalmol. 2011 Feb;151(2):195-7.e2.
3. GBD 2019 Blindness and Vision Impairment Collaborators; Vision Loss Expert Group of the Global Burden of Disease Study. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Glob Health. 2021 Feb;9(2):e144-e160. Epub 2020 Dec 1. Erratum in: Lancet Glob Health. 2021 Apr;9(4):e408.
4. Chopra R, Chander A, Jacob JJ. Ocular associations of metabolic syndrome. Indian J Endocrinol Metab. 2012;16 Suppl 1(Suppl1):S6-S11. doi:10.4103/2230-8210.94244
5. Limburg and Hans. Prevalence and Causes of Blindness in Children in Vietnam. Ophthalmology, Volume 119, Issue 2, 355 – 361.
6. Eye Health. Available from https://www.aihw.gov.au/reports/eye-health/eye-health/contents/how-common-is-visual-impairment.
7. Scheetz, J., Koca, D., McGuinness, M. Real-world artificial intelligence-based opportunistic screening for diabetic retinopathy in endocrinology and indigenous healthcare settings in Australia. Sci Rep 11, 15808 (2021).
8. Tan Z, Scheetz J, He M. Artificial Intelligence in Ophthalmology: Accuracy, Challenges, and Clinical Application. Asia Pac J Ophthalmol (Phila). 2019 May-Jun;8(3):197-199.
9. Tai MC. The impact of artificial intelligence on human society and bioethics. Tzu Chi Med J. 2020;32(4):339-343.
10. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artificial Intelligence in Healthcare. 2020;25-60.
11. Sindhu V, Nivedha S, Prakash M. An Empirical Science Research on Bioinformatics in Machine Learning. Journal of Mechanics of Continua and Mathematical Sciences 2020;7:86-94
12. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ2019 7:e7702
13. Moraru AD, Costin D, Moraru RL, Branisteanu DC. Artificial intelligence and deep learning in ophthalmology - present and future (Review). Exp Ther Med. 2020 Oct;20(4):3469-3473.
14. Sivaraman, A., Savoy, F., & Rajalakshmi, R. Insights into the growing popularity of artificial intelligence in ophthalmology. Indian journal of ophthalmology 2020, 68(7), 1339–1346.
15. Aggarwal, R., Sounderajah, V., Martin, G. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. npj Digit. Med. 4, 65 (2021).
16. Sounderajah V, Ashrafian H, Aggarwal R, De Fauw J, Denniston AK, Greaves F et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. Nat Med. 2020 Jun;26(6):807-808.
17. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness.
18. Lee A, Taylor P, Kalpathy-Cramer J and Tufail A: Machine learning has arrived! Ophthalmology 124: 1726-1728, 2017
19. Swartz MK. The PRISMA statement: a guideline for systematic reviews and meta-analyses. J Pediatr Health Care 2011;25:1–2.
20. Higgins JP, Altman DG, Gøtzsche PC. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.
21. Aquino A, Gegundez ME, Marin D. Automated optic disc detection in retinal images of patients with diabetic retinopathy and risk of macular edema. International Journal of Biological & Life Sciences, 2009;3(12) 353–358.
22. Haloi M. Improved microaneurysm detection using deep neural net works. arXiv preprint arXiv 2015;1505.04424.

23. Ahmed MI, Amin MA. High speed detection of optical disc in retinal fundus image. SIViP 2015;9: 77–85.
24. Liskowski P, Krawiec K. Segmenting Retinal Blood Vessels With Deep Neural Networks. IEEE Trans Med Imaging. 2016 Nov;35(11):2369-2380.
25. Asaoka R, Murata H, Iwase A, Araie M. Detecting Preperimetric Glaucoma with Standard Automated Perimetry Using a Deep Learning Classifier. Ophthalmology. 2016 Sep;123(9):1974-80.
26. MJJP van Grinsven, B van Ginneken, CB Hoyng, T Theelen, and CI S´anchez. Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fun dus images. IEEE Transactions on Medical Imaging, 2016; 35(5):1273–1284.
27. Tan JH, Fujita H, Sivaprasad S, Bhandary SV, Rao AK, Chua KC, et al. Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. Information Sciences, 2017;420:66–76.
28. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, Niemeijer M. Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning. Invest Ophthalmol Vis Sci. 2016 Oct 1;57(13):5200-5206.
29. Colas E, Besse A, Orgogozo A, Schmauch B, Meric N, Besse E. Deep learning approach for diabetic retinopathy screening. Acta Ophthalmologica, 2016;94.
30. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016 Dec 13;316(22):2402-2410.
31. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. Ophthalmology. 2017 Jul;124(7):962-969.
32. Quellec G, Charrière K, Boudi Y, Cochener B, Lamard M. Deep image mining for diabetic retinopathy screening. Med Image Anal. 2017 Jul;39:178-193.
33. Ambrósio R Jr, Lopes BT, Faria-Correia F, Salomão MQ, Bühren J, Roberts CJ et al. Integration of Scheimpflug-Based Corneal Tomography and Biomechanical Assessments for Enhancing Ectasia Detection. J Refract Surg. 2017 Jul 1;33(7):434-443.
34. Takahashi H, Tampo H, Arai Y, Inoue Y, Kawashima H. Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy. PLoS One. 2017 Jun 22;12(6):e0179790.
35. Oliveira AFM, Pereira SRM, Silva CAB. Retinal vessel segmentation based on fully convolutional neural networks. Expert Systems with Applications, 2018;112:229–242.
36. Schmidt-Erfurth U, Waldstein SM, Klimscha S, Sadeghipour A, Hu X, Gerendas BS et al. Prediction of Individual Disease Conversion in Early AMD Using Artificial Intelligence. Invest Ophthalmol Vis Sci. 2018 Jul 2;59(8):3199-3208.
37. Lin GM, Chen MJ, Yeh CH, Lin YY, Kuo HY, Lin MH et al. Transforming Retinal Photographs to Entropy Images in Deep Learning to Improve Automated Detection for Diabetic Retinopathy. J Ophthalmol. 2018 Sep 10;2018:2159702.
38. Chakravarty A, Sivswamy J. A deep learning based joint segmentation and classification framework for glaucoma assesment in retinal color fundus images. arXiv preprint arXiv:1808.01355, 2018.
39. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. Ophthalmology. 2018 Aug;125(8):1199-1206.
40. Chai Y, Liu H, Xu J. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. Knowledge-Based Systems, 2018;161:147–156.
41. Mitra A, Banerjee PS, Roy S, Roy S, Setua SK. The region of interest localization for glaucoma analysis from retinal fundus image using deep learning. Computer Methods and Programs in Biomedicine, 2018;165:25–35.
42. Liu S, Graham SL, Schulz A, Kalloniatis M, Zangerl B, Cai W et al. A deep learning-based algorithm identifies glaucomatous discs using monoscopic fundus photographs. Ophthalmology Glaucoma, 2018;1(1):15–22.
43. Orlando JI, Prokofyeva E, Del Fresno M, Blaschko MB. An ensemble deep learning based approach for red lesion detection in fundus images. Comput Methods Programs Biomed. 2018 Jan;153:115-127.
44. Lam C, Yu C, Huang L, Rubin D. Retinal Lesion Detection With Deep Learning Using Image Patches. Invest Ophthalmol Vis Sci. 2018 Jan 1;59(1):590-596.
45. Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B et al. A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography. Ophthalmology. 2018 Sep;125(9):1410-1420.
46. Zhang Y, Chung A. Deep supervision with additional labels for retinal vessel segmentation task. arXiv preprint arXiv 2018;1806.02132.
47. Zhou W, Wu H, Wu C, Yu X, Yi Y. Automatic Optic Disc Detection in Color Retinal Images by Local Feature Spectrum Analysis. Comput Math Methods Med. 2018 Jun 14;2018:1942582.
48. B. Al-Bander, W. Al-Nuaimy, B. M. Williams, and Y. Zheng, “Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc,” Biomedical Signal Processing and Control, vol. 40, pp. 91–101, 2018.
49. An G, Omodaka K, Hashimoto K, Tsuda S, Shiga Y, Takada N et al. Glaucoma Diagnosis with Machine Learning Based on Optical Coherence Tomography and Color Fundus Images. J Healthc Eng. 2019 Feb 18;2019:4061313.
50. Medeiros FA, Jammal AA, Thompson AC. From Machine to Machine: An OCT-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs. Ophthalmology. 2019 Apr;126(4):513-521.
51. Lin, H. “Diagnostic Efficacy and Therapeutic Decision-making Capacity of an Artificial Intelligence Platform for Childhood Cataracts in Eye Clinics: A Multicentre Randomized Controlled Trial.” EClinicalMedicine 9 (2019): 52 - 59.
52. Zéboulon P, Debellemanière G, Bouvet M, Gatinel D. Corneal Topography Raw Data Classification Using a Convolutional Neural Network. Am J Ophthalmol. 2020 Nov;219:33-39.
53. Varadarajan AV, Bavishi P, Ruamviboonsuk P, Chotcomwongse P, Venugopalan S, Narayanaswamy A et al. Predicting optical coherence tomography-derived diabetic

macular edema grades from fundus photographs using deep learning. Nat Commun. 2020 Jan 8;11(1):130.

54. Ahn H. Artificial intelligence method to classify ophthalmic emergency severity based on symptoms: a validation study. BMJ Open 2020;10:e037161.
55. Rim TH, Soh ZD, Tham YC, Yang HHS, Lee G, Kim Y et al. Deep Learning for Automated Sorting of Retinal Photographs. Ophthalmol Retina. 2020 Aug;4(8):793-800.
56. Lee J, Kim YK, Jeoung JW, Ha A, Kim YW, Park KH. Machine learning classifiers-based prediction of normal-tension glaucoma progression in young myopic patients. Jpn J Ophthalmol. 2020 Jan;64(1):68-76.
57. Huang Y-P, Vadloori S, Chu H-C, Kang EY-C, Wu W-C, Kusaka S, et al. Deep Learning Models for Automated Diagnosis of Retinopathy of Prematurity in Preterm Infants. Electronics. 2020; 9(9):1444.
58. Tham YC, Anees A, Zhang L, Goh JHL, Rim TH, Nusinovici S et al. Referral for disease-related visual impairment using retinal photograph-based deep learning: a proof-of-concept, model development study. Lancet Digit Heal. 2021; 3:e29–e40.
59. Son J, Shin JY, Kim HD, Jung KH, Park KH, Park S. Development and Validation of Deep Learning Models for Screening Multiple Abnormal Findings in Retinal Fundus Images. Ophthalmology 2020;127:85–94.
60. Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen C et al. Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. Commun Biol. 2020; 3:15.
61. Dai L, Wu L, Li H, Cai C, Wu Q, Kong H et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. Nature communications 2021;253 12(1):1–11.
62. Luo X, Li J, Chen M, Yang X, Li X. Ophthalmic disease detection via deep learning with a novel mixture loss function. IEEE J Biomed Health Inform.
63. Hosmer DW, Lemeshow S. Applied Logistic Regression, 2nd Ed. Chapter 5, John Wiley and Sons, New York, NY (2000), pp. 160-164
64. Heidary F, Gharebaghi R. Ideas to assist the underprivileged dispossessed individuals. Med Hypothesis DiscovInnovOphthalmol 2012;1(3):43-44.
65. Boudry C, Al Hajj H, Arnould L, Mouriaux F. Analysis of international publication trends in artificial intelligence in ophthalmology. Graefes Arch Clin Exp Ophthalmol. 2022 May;260(5):1779-1788. doi: 10.1007/s00417-021-05511-7.
66. Park SH, Kressel HY. Connecting Technological Innovation in Artificial Intelligence to Real-world Medical Practice through Rigorous Clinical Validation: What Peer-reviewed Medical Journals Could Do. J Korean Med Sci. 2018;33(22):e152.
67. Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. BMJ. 2020.
68. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, The SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health. 2020.
69. Collins GS, Dhiman P, Andaur Navarro CL. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008.
70. Benet D, Pellicer-Valero OJ. Artificial intelligence: the unstoppable revolution in ophthalmology. Surv Ophthalmol. 2022 Jan-Feb;67(1):252-270.
71. Xie J, Keel J, Taylor H. R, Dirani M. Utilization of eye health-care services in Australia: the National Eye Health Survey. Clinical & experimental ophthalmology 2018, 46(3), 213–221.
72. Simmons AC, McCarty D, Khan C.A, Taylor H. R. Eye health in rural Australia. Clinical & experimental ophthalmology 2002, 30(5), 316-321
73. Yashadhana, A, Fields, T, Blitner, G, Stanley, R, Zwi, A. B. Trust, culture and communication: determinants of eye health and care among Indigenous people with diabetes in Australia. BMJ global health 2022, 5(1), e001999.
74. Australia. Department of the Prime Minister and Cabinet. Closing the Gap: Prime Minister's Report 2017. Australia. Department of the Prime Minister and Cabinet.
75. Zhang Y, Shi J, Peng Y. Artificial intelligence-enabled screening for diabetic retinopathy: a real-world, multicenter and prospective study. BMJ Open Diabetes Research and Care 2020;8:e001596.
76. Lee AY, Yanagihara RT, Lee CS, Blazes M, Jung HC, Chee Y. Diabetes Care May 2021, 44 (5) 1168-1175;
77. Heydon P, Egan C, Bolter L. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. British Journal of Ophthalmology 2021;105:723-728.
78. Konerman MA, Beste LA, Van T, Liu B, Zhang X, Zhu J. Machine learning models to predict disease progression among veterans with hepatitis C virus. PLoS ONE 2019; 14(1): e0208141.
79. Wongvibulsin S, Wu K.C, Zeger S.L. Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis. BMC Med Res Methodol 20, 1 (2020).
80. Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevska O, written on behalf of AME Big-Data Clinical Trial Collaborative Group. Predictive analytics with gradient boosting in clinical medicine. Annals of translational medicine, 2019; 7(7), 152.
81. O'Byrne C, Abbas A, Korot E, Keane PA. Automated deep learning in ophthalmology: AI that can build AI. Curr Opin Ophthalmol. 2021 Sep 1;32(5):406-412.
82. Arcadu F, Benmansour F, Maunz A. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. NPJ Digit Med 2019; 2: 1–9.
83. Arslan J, Samarasinghe G, Benke K, Sowmya A, Wu Z, Guymer R. Artificial Intelligence Algorithms for Analysis of Geographic Atrophy: A Review and Evaluation. Translational vision science & technology, 2020; 9(2), 57.
84. Scheetz J, Rothschild P, McGuinness M. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep 11, 5193 (2021).
85. FDA Releases Artificial Intelligence/Machine Learning Action Plan. Available at https://www.fda.gov/news-

events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan . Accessed on April 10, 2022

86. Consultation: Regulation of software, including Software as a Medical Device (SaMD). Available at https://www.tga.gov.au/sites/default/files/consultation-regulation-software-including-software-medical-device-samd.pdf . Accessed on April 10, 2022
87. US Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems, 2018. [Cited July 18, 2022]. Available from: https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye
88. Eyewire News. Eyenuk Announces FDA Clearance for EyeArt Autonomous AI System for Diabetic Retinopathy Screening, 2020. [Cited July 18, 2022]. Available from https://eyewire.news/articles/eyenuk-announces-fda-clearance-for-eyeart-autonomous-ai-system-for-diabetic-retinopathy-screening/
89. Tseng R, Gunasekeran DV, Tan SSH. Considerations for Artificial Intelligence Real-World Implementation in Ophthalmology: Providers' and Patients' Perspectives, Asia-Pacific Journal of Ophthalmology: May-June 2021 - Volume 10 - Issue 3 - p 299-306
90. Wei N, Shihao Z, Zhaoran W. Updates in deep learning research in ophthalmology. Clin Sci (Lond) 29 October 2021; 135 (20): 2357–2376.