



ARTICLE



Performances of artificial intelligence in detecting pathologic myopia: a systematic review and meta-analysis

Yue Zhang ^{1,2}, Yilin Li³, Jing Liu^{1,2}, Jianing Wang¹, Hui Li¹, Jinrong Zhang¹ and Xiaobing Yu ^{1,2}✉

© The Author(s), under exclusive licence to The Royal College of Ophthalmologists 2023, corrected publication 2023

BACKGROUND/OBJECTIVE: Pathologic myopia (PM) is a major cause of severe visual impairment and blindness, and current applications of artificial intelligence (AI) have covered the diagnosis and classification of PM. This meta-analysis and systematic review aimed to evaluate the overall performance of AI-based models in detecting PM and related complications.

METHODS: We searched PubMed, Scopus, Embase, Web of Science and IEEE Xplore for eligible studies before Dec 20, 2022. The methodological quality of included studies was evaluated using the Quality Assessment for Diagnostic Accuracy Studies (QUADAS-2). We calculated the pooled sensitivity (SEN), specificity (SPE) and the summary area under the curve (AUC) using a random effects model, to evaluate the performance of AI in the detection of PM based on fundus or optical coherence tomography (OCT) images.

RESULTS: 22 studies were included in the systematic review, and 14 of them were included in the quantitative analysis. Of all included studies, SEN and SPE ranged from 80.0% to 98.7% and from 79.5% to 100.0% for PM detection, respectively. For the detection of PM, the summary AUC was 0.99 (95% confidence interval (CI) 0.97 to 0.99), and the pooled SEN and SPE were 0.95 (95% CI 0.92 to 0.96) and 0.97 (95% CI: 0.94 to 0.98), respectively. For the detection of PM-related choroid neovascularization (CNV), the summary AUC was 0.99 (95% CI: 0.97 to 0.99).

CONCLUSION: Our review demonstrated the excellent performance of current AI algorithms in detecting PM and related complications based on fundus and OCT images.

Eye (2023) 37:3565–3573; <https://doi.org/10.1038/s41433-023-02551-7>

INTRODUCTION

Myopia is increasingly prevalent worldwide and has become a serious challenge for public health. The global prevalence of myopia (≥ -0.5 D) is approximately 2 billion and is predicted to be 4.76 billion (49.8% of the world population) by 2050 [1]. At the same time, high myopia (≥ -6.0 D) has become increasingly prevalent in recent decades, especially in Asian countries, and has developed from an earlier age [2]. Pathologic myopia (PM) is a major cause of severe visual impairment and is defined as a special category of myopia associated with excessive axial elongation leading to structural changes in the posterior segment of the eye, such as posterior staphyloma and myopic macular degeneration (MMD), and loss of best corrected visual acuity in International Classification of Diseases 11th Revision (ICD-11) [3]. PM has been estimated to affect 3% of the global population and will lead to great potential productivity loss and a high economic burden on the healthcare system [4]. As a consequence, it is necessary to timely identify PM eyes and prevent the progression of visual impairment.

Recent studies have proven high accuracy, sensitivity (SEN) and specificity (SPE) of artificial intelligence (AI) systems integrated in ophthalmology imaging, especially the subfield of deep learning (DL). Multiple successful algorithms have been developed for

screening and assisted diagnosis of diabetic retinopathy (DR), glaucoma and age-related degeneration (AMD), and myopia [5]. The current applications of AI in myopia cover a variety of aspects, including the diagnosis and classification of PM, prediction of progression and guidance of refractive surgery; meanwhile, imaging modality with which the algorithms were developed has also developed from fundus images to optical coherence tomography (OCT) images.

There are still relatively great variations among studies in development procedures, databases, sample resources and many aspects of methodology. Thus, a detailed assessment of AI performance is needed to quantify the overall accuracy and generalizability and identify the confounding factors of the findings. Recently, meta-analysis and systematic reviews about the diagnostic performance of AI in detecting AMD, glaucoma and diabetic macular oedema (DMO) have been published, while there is still no comprehensive investigation on the performance of AI for the detection of PM [6–8].

Aims of the study

We conducted this meta-analysis and systematic review to evaluate the overall performance of AI-based models in detecting PM and PM-related CNV based on fundus and OCT images, and

¹Department of Ophthalmology, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China. ²Graduate School of Peking Union Medical College, Beijing, China. ³Center for Statistical Sciences, Peking University, Beijing, China. ✉email: yuxiaobing1214@163.com

Received: 7 January 2023 Revised: 30 March 2023 Accepted: 18 April 2023

Published online: 28 April 2023

explore the underlying factors affecting the accuracy and acceptability of algorithms and discuss the limitations and future steps of AI applications in PM.

METHODS

The protocol for this systematic review was registered in PROSPERO (CRD42022379136) and this review was conducted according to the PRISMA statement recommendations.

Search strategy and selection criteria

We searched PubMed, Scopus, Embase, Web of Science and IEEE Xplore for eligible studies published up to Dec 20, 2022, using the combination of search terms associated with PM (e.g., myopia, high myopia and pathologic myopia) crossed with search terms associated with AI (e.g., artificial intelligence, machine learning and deep learning) in the full text. Full search terms were listed in online Supplementary Appendix 1. We also searched the reference lists of included literature to identify potentially eligible studies. The language was limited to English.

Two researchers (HL, JRZ) independently screened the titles and abstracts for eligible literature according to the selection criteria. The eligible studies were further selected with a full-text review after removing duplications. The inclusion criteria were as follows: (1) journal articles or conference papers reporting the primary outcome of the performance of the AI algorithm in the detection of patients with PM; (2) the definition or reference standard for PM were clearly defined; (3) a clear description of the procedure developing algorithms and detailed information about the database were reported; (4) necessary data or evaluation indices were reported to calculate the absolute numbers of true positive (TP), false positive (FP), false negative (FN), and true negative (TN), such as SEN, SPE, accuracy and area under the receiver operating characteristic curve (AUC).

The exclusion criteria were as follows: (1) publication forms of case reports, reviews, comments, letters and editorials; unpublished or ongoing research; (2) studies that detected PM based on imaging methods other than fundus or OCT images; (3) studies that did not report necessary data of the primary outcome.

Risk of bias assessment and data extraction

Quality assessment of eligible articles was performed by two reviewers (YZ, HL) independently using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. Any disagreement was resolved by discussion with a senior researcher (XBY) for consensus. The QUADAS-2 tool consists of 4 aspects of assessment: patient selection, index test, reference standard, and flow and timing [9]. All included studies were evaluated for the risk of bias for all 4 aspects and the applicability for the former 3 aspects. The risk of each study was classified into low, high and clear risk of bias, and studies with a high risk of bias or low quality were excluded from our study.

Data were extracted in all eligible full-text studies by two reviewers (YZ, JNW) independently, and the following data were collected if available: the first author; country; publication year; characteristics of datasets (dataset type; total number of images; imaging modality); characteristics of algorithms (types of algorithms, outcome of classification); evaluation indices of the algorithm accuracy derived from internal or external validation datasets, including SEN, SPE, accuracy, AUC, positive likelihood ratio (PLR), negative likelihood ratio (NLR) and diagnostic odds ratio (DOR). The results from different validation datasets in the same study were considered independent data. If the data of evaluation indices were insufficient to calculate the 2-by-2 table for the outcomes of validation, the study was not included in the meta-analysis but only for literature review.

We defined PM as eyes with maculopathy based on fundus images not less than category two or with "plus" features according to Meta-analysis for Pathologic Myopia (META-PM) study classification [3, 10]. Another system considering the three most crucial myopic lesions was the atrophy, traction, and neovascularization (ATN) grading system based on OCT images (Supplementary Appendix 2) [11]. In particular, as choroidal neovascularization (CNV) is a leading cause of vision impairment of PM and should receive timely referrals, we further evaluated the performance of the included algorithms in the detection of CNV in PM eyes (if available).

Statistical analysis

We used the RevMan 5.3 platform (Cochrane Collaboration, Denmark) to conduct quality assessment for all included studies. Next, Stata version

17.0 MP (StataCorp) was applied to perform all the analyses, and a 2-tailed $P < 0.05$ was considered statistically significant. We applied random-effects models to combine the included studies. The pooled quantitative analysis of indicators for diagnostic performance was performed, including SEN, SPE, PLR, NLR, and DOR, with results shown in forest plots with 95% confidence interval (CI). The I^2 statistic was used to assess heterogeneity among studies (25–49%: low heterogeneity; 50–74%: moderate heterogeneity; and more than 75%: high heterogeneity). To avoid threshold effects, we adopted a hierarchical summary receiver operating characteristic (HSROC) model to assess the relationship between SEN and SPE and plotted the summary receiver operating characteristic (SROC) curves with 95% CIs and prediction regions.

Meta-regression was performed to investigate the reasons for the heterogeneity among studies. For subgroup analysis, the following covariates were considered: research regions (developing countries and developed countries); different types of validation datasets (internal and external validation datasets); imaging modalities (fundus and OCT images); types of datasets (public and hospital datasets); and total number of images (<5000 and ≥ 5000). The direction of subgroup analysis was not a priori due to unclear hypothesized direction. Furthermore, we conducted sensitivity analysis to estimate the robustness and reliability of our analysis and assessed publication bias with Deek's plot. The sensitivity analysis was the repeat of the primary meta-analysis, after excluding studies with high or unclear risk of bias.

RESULTS

Search results and study characteristics

Initially, our literature search identified 1036 studies, and 587 studies were screened after the removal of duplicated records. Figure 1 shows the flowchart of the literature eligibility process. Finally, 22 studies were included for systematic review [12–33], and 14 of them were included for quantitative meta-analysis [12, 13, 16, 19–23, 26, 27, 29–31, 33].

The characteristics of all eligible studies are summarized in Table 1. In total, 348,861 fundus images and 22,560 OCT images were used for training, testing and validation. Of all included studies, SEN and SPE ranged from 80.0% to 98.7% and from 79.5% to 100.0% for PM detection, respectively. Two categories (PM and non-PM) were exported as the primary outcome in 14 studies (63.6%); 5 categories (META-PM) of PM were exported in 4 studies (18.2%); and 3 categories (ATN) of PM were exported in 1 study (4.6%). The remaining 6 studies (27.3%) identified specific PM-related lesions (CNV, myopic traction maculopathy, retinal detachment, etc).

Most studies ($n = 20$, 90.9%) applied convolutional neural network (CNN) to develop algorithms, of which 12 studies used ResNet. There was also 1 study using support vector machine (SVM) and 1 study using Adaboost. 16 studies (72.7%) obtained images from hospitals, and 6 studies (27.3%) from public databases, of which the PathologicAL Myopia (PALM) database was the most frequently adopted public database ($n = 4$, 18.2%).

Risk of bias assessment and publication bias

We assessed the quality of all included studies using the QUADAS-2 tool, and the results are presented in Supplementary Appendix 3. 7 studies (31.8%) were graded as having a low risk of bias in all 4 domains [16, 18, 20, 26–28, 33]. 12 studies (54.5%) had at least one domain with applicability concerns [14, 15, 17, 19, 21, 22, 24, 25, 29–32]. For patient selection, 12 studies (54.5%) were graded as having an unclear risk of bias because of the lack of a clear description of public datasets, and 12 studies (54.5%) had unclear applicability concerns due to unavailable composition information. For the index test, most studies ($n = 16$, 72.7%) had a low risk of bias and concern of applicability, and only 6 studies (27.3%) were graded with an unclear risk of bias due to underlying data overlap among datasets. For the reference standard, the risk of bias and concern of applicability were low in all included studies. Finally, for the flow and timing domain, 8 studies (36.4%) had unclear risk of bias considering the unclear

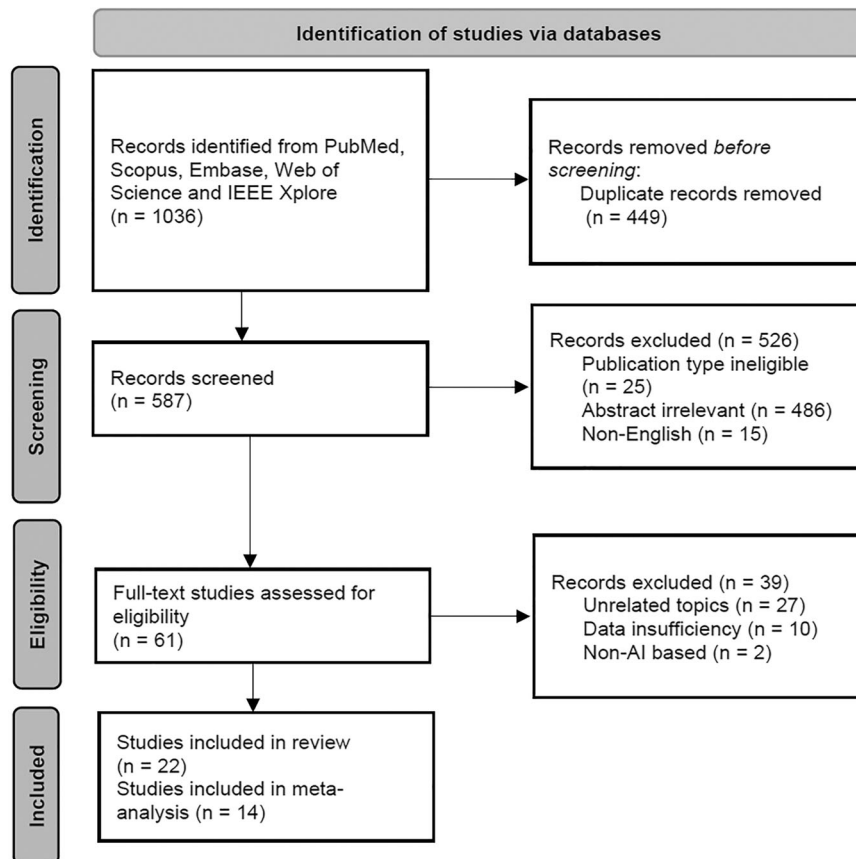


Fig. 1 PRISMA 2020 flow diagram of study selection. Flow diagram showing the study selection process and reasons for exclusion.

construction procedure of public datasets. No publication bias existed ($P=0.10$) by Deek's funnel plot asymmetry test, shown in Supplementary Appendix 4. Furthermore, we assessed and confirmed the low credibility of our meta-analysis using Instrument for assessing the Credibility of Effect Modification Analyses (ICEMAN), shown in Supplementary Appendix 5.

Meta-analysis for the performance of AI in PM and PM-CNV detection

For the detection of PM, the forest plots of SEN, SPE and 95% CIs for the included studies are shown in Fig. 2A, B [13, 16, 19, 20, 22, 23, 27, 29–31]. Using the HSROC model, we obtained the SROC curve with a 95% confidence region and prediction region (Fig. 2C). The summary AUC was 0.99 (95% CI: 0.97 to 0.99), and the pooled SEN, SPE, PLR, NLR, and DOR were 0.95 (95% CI: 0.92 to 0.96), 0.97 (95% CI: 0.94 to 0.98), 28.1 (95% CI: 15.8 to 50.2), 0.06 (95% CI: 0.04 to 0.08), and 495 (95% CI: 243 to 1008), respectively. For the detection of PM-CNV, the forest plots for the included studies and the SROC curve plot are shown in Fig. 3 [12, 13, 21, 26, 33]. The summary AUC was 0.99 (95% CI: 0.97 to 0.99), and the pooled SEN, SPE, PLR, NLR, and DOR were 0.94 (95% CI: 0.90 to 0.97), 0.96 (95% CI: 0.94 to 0.98), 25.9 (95% CI: 16.1 to 41.7), 0.06 (95% CI: 0.03 to 0.10), and 435 (95% CI: 220 to 860), respectively.

Heterogeneity analysis and meta-regression analysis

Since high heterogeneity ($I^2 > 50$) was found in our forest plots when assessing the SEN and SPE for the detection of PM, we performed meta-regression to explore the potential reasons for heterogeneity. Through our analysis, the DOR was not correlated with any factors as follows: research regions ($P=0.15$); different types of validation datasets ($P=0.23$); imaging modalities

($P=0.78$); types of datasets ($P=0.36$); total number of images ($P=0.07$).

Subgroup analysis

The results of subgroup analysis are summarized in Table 2. We found imaging modalities and resources of data had no significant contributions to the diagnostic performance. For different types of validation datasets, there was a better performance in the internal dataset (SEN = 0.95, 95% CI: 0.94–0.96; SPE = 0.97, 95% CI: 0.96–0.99; AUC = 0.99, 95% CI: 0.97–1.00) than external dataset (SEN = 0.93, 95% CI: 0.92–0.95; SPE = 0.96, 95% CI: 0.94–0.97; AUC = 0.99, 95% CI: 0.98–0.99). For research regions, we found a better performance in developed countries (SEN = 0.96, 95% CI: 0.93–0.98; SPE = 0.98, 95% CI: 0.97–0.99; AUC = 0.99, 95% CI: 0.97–0.99) than developing countries (SEN = 0.94, 95% CI: 0.90–0.95; SPE = 0.96, 95% CI: 0.93–0.98; AUC = 0.98, 95% CI: 0.97–0.99). For different total sizes of data, a better performance was detected in data larger than 5000 (SEN = 0.96, 95% CI: 0.95–0.98; SPE = 0.97, 95% CI: 0.96–0.99; AUC = 0.99, 95% CI: 0.97–0.99) than smaller than 5000 (SEN = 0.93, 95% CI: 0.91–0.95; SPE = 0.96, 95% CI: 0.94–0.98; AUC = 0.98, 95% CI: 0.98–0.99).

Sensitivity analysis

The sensitivity analysis is the repeat of the primary meta-analysis. We excluded 5 studies without sufficient information about the division of datasets or in-depth details of clinical data resources [19, 22, 29–31]. Then, the pooled SEN was 0.94 (95% CI: 0.90 to 0.97), and the pooled SPE was 0.96 (95% CI: 0.95 to 0.98) for the detection of PM. The results were similar to our main findings; hence, there was no evidence that our main outcome was influenced by which studies were included.

Table 1. Characteristics of the included studies for review.

First Author	Country	Publication year	Dataset Type	Total Number	Imaging Modality	Algorithm	Outcome of Classification	Grading System	SEN	SPE	Accuracy	AUC
Sogawa [12]	Japan	2020	Department of Ophthalmology, Tsukazaki Hospital	910	OCT images	CNN	Non-CNV/ CNV	NA	0.906	0.942	NA	0.970
Lu and Ren [13]	China	2021	Ophthalmic Clinics of Hospitals in Zhejiang Province	32,010	Fundus images	CNN	Non-PM/PM Category 0 Category 1 Category 2 Category 3 Category 4 CNV	META- PM	0.946 ⁱ 0.987 ^e NA NA NA NA 0.973 ⁱ 0.970 ⁱ 0.915 ^e	0.992 ⁱ 0.945 ^e NA NA NA NA NA 0.970 ⁱ 0.915 ^e	0.984 NA 0.977 0.997 0.967 0.984 0.952 NA NA	0.995 NA 0.997 0.997 0.967 0.984 0.952 NA NA
Wan [14]	China	2021	The Affiliated Eye Hospital of Nanjing Medical University	858	Fundus images	CNN	Low-risk HM High-risk HM	NA	1.000 0.9524	0.9787 1.0000	0.990 0.9900	0.9968 0.9964
Tang [15]	China	2022	Ophthalmic Clinics of PUMCH and Beijing Tongren Hospital	1395	Fundus images	CNN	Category 0 Category 1 Category 2 Category 3 Category 4	META- PM	0.9286 0.9778 0.8977 0.7917 0.5000	0.9952 0.9459 0.9667 0.9766 0.9870	0.9370 0.9370 0.9370 0.9370 0.9370	0.9980 0.9980 0.9980 0.9980 0.9980
Li and Wang [16]	China	2022	Ophthalmology Clinics of Six Hospitals	57,148	Fundus images	CNN	Non-PM/PM	NA	0.964 ⁱ 0.933 ^e 0.910 ^e	0.992 ⁱ 0.996 ^e 0.987 ^e	0.965 NA NA	0.997 NA NA
Rauf [17]	Pakistan	2021	PALM	400	Fundus images	CNN	Non-PM/PM	NA	NA	NA	NA	0.9845
Ye [18]	China	2021	Two Affiliated Eye Hospitals of WMU (Wenzhou and Hangzhou)	2342	OCT images	CNN	MTM DSM BM Defect	NA	0.928 0.745 0.889	0.905 0.940 0.848	NA NA NA	0.974 0.955 0.938
Du [19]	Japan	2021	Tokyo High Myopia Clinic; PALM; SEED	7020	Fundus images	CNN	Non-PM/PM Category 2 Category 3 Category 4	META- PM	0.8879 0.8444 0.8722 0.8510	0.9583 0.9450 0.9602 0.9834	0.9208 0.9018 0.9528 0.9750	NA 0.970 0.978 0.982
Park [20]	Korea	2022	Incheon St. Mary's Hospital and Seoul St. Mary's Hospital	367	3D OCT images	CNN	Non-PM/PM	NA	0.93	0.96	0.95	0.98
Li and Feng [21]	China	2022	Zhongshan Ophthalmic Center	5505	OCT images	CNN	Retinoschisis CNV	NA	0.900 0.952	0.905 0.957	NA NA	0.961 0.994
Lu and Zhou [22]	China	2021	The First Affiliated Hospital of School of Medicine, Zhejiang University	16,428	Fundus images	CNN	Non-PM/PM Category 0 Category 1 Category 2 Category 3 Category 4	META- PM	0.977 ^j 0.973 ^e NA NA NA NA	0.972 ^j 0.905 ^e NA NA NA NA	0.977 NA 0.988 0.993 0.953 0.967	0.993 NA 0.999 0.998 0.953 0.972
Kim [23]	Korea	2021	Incheon St. Mary's Hospital and Seoul St. Mary's Hospital	860	OCT images	SVM	Non-PM/PM	NA	0.8000	0.9358	0.9147	0.8679

Table 1. continued

First Author	Country	Publication year	Dataset Type	Total Number	Imaging Modality	Algorithm	Outcome of Classification	Grading System	SEN	SPE	Accuracy	AUC
Hemelings [24]	Belgium	2021	PALM	1200	Fundus images	CNN	Non-PM/PM	NA	NA	NA	NA	0.9867
Cui [25]	China	2021	PALM	1200	Fundus images	CNN	Non-PM/PM	NA	NA	NA	0.9725	NA
Wu [26]	China	2022	Zhongshan Ophthalmic Center	1853	Fundus images and OCT images	CNN	Atrophy Traction CNV	ATN	0.9216 0.7245 0.8594	0.9148 0.9658 0.9812	0.9238 0.8534 0.9421	0.969 0.895 0.936
Tan [27]	Singapore	2021	SEED; SNEC-HMC	226,686	Fundus images	CNN	Non-HM/HM Non-PM/PM	NA	0.913 0.914 ⁱ 0.968 ^e 0.984 ^e 0.942 ^e	0.945 0.942 ⁱ 0.873 ^e 0.855 ^e 0.959 ^e	NA NA NA NA NA	0.978 0.975 NA NA NA
Du [28]	Japan	2022	Tokyo High Myopia Clinic	9176	OCT images	CNN	MNV MTM DSM	ATN	NA NA NA	NA NA NA	NA NA NA	0.985 0.946 0.978
Pathan [29]	India	2020	PALM	400	Fundus images	AdaBoost	Non-PM/PM	NA	0.903	1.000	0.950	NA
Dai [30]	China	2020	Private retinal fundus dataset from a regional hospital	1251	Fundus images	CNN	Non-PM/PM	NA	0.8352	0.795	0.8182	NA
Himami [31]	Indonesia	2021	PALM; ODJR	612	Fundus images	CNN	Non-PM/PM	NA	0.97	0.93	1.00	NA
Kalyanasundaram [32]	India	2020	PALM	400	Fundus images	CNN	Non-PM/PM	NA	NA	NA	0.9808	NA
He [33]	China	2022	The Eye Centers of Three Affiliated Hospitals of Universities in China	3400	OCT images	CNN	Macular-schisis Full-thickness MH CNV MH with RD	ATN	0.9412 0.9107 0.9912 0.9130	0.9879 0.9965 0.9842 0.9906	NA NA NA NA	0.991 0.962 0.997 0.988

SEN sensitivity, SPE specificity, AUC area under the receiver operating characteristic curve, META-PM Meta-analysis for Pathologic Myopia, ATN atrophy, traction, and neovascularization, PALM Pathological Myopia, SEED Singapore Epidemiology of Eye Diseases, CNN convolutional neural network, PM pathologic myopia, OCT optical coherence tomography, MNV macular neovascularization, MTM myopic traction maculopathy, DSM dome-shaped macula, MH macular hole, CNV choroidal neovascularization, RD retinal detachment, SNEC-HMC Singapore National Eye Center High Myopia Clinic, HM high myopia, SVM support vector machine, BM Bruch membrane, ODJR Ocular Disease Intelligent Recognition, NA not available.

ⁱResults validated in the internal dataset.

^eResults validated in the external dataset.

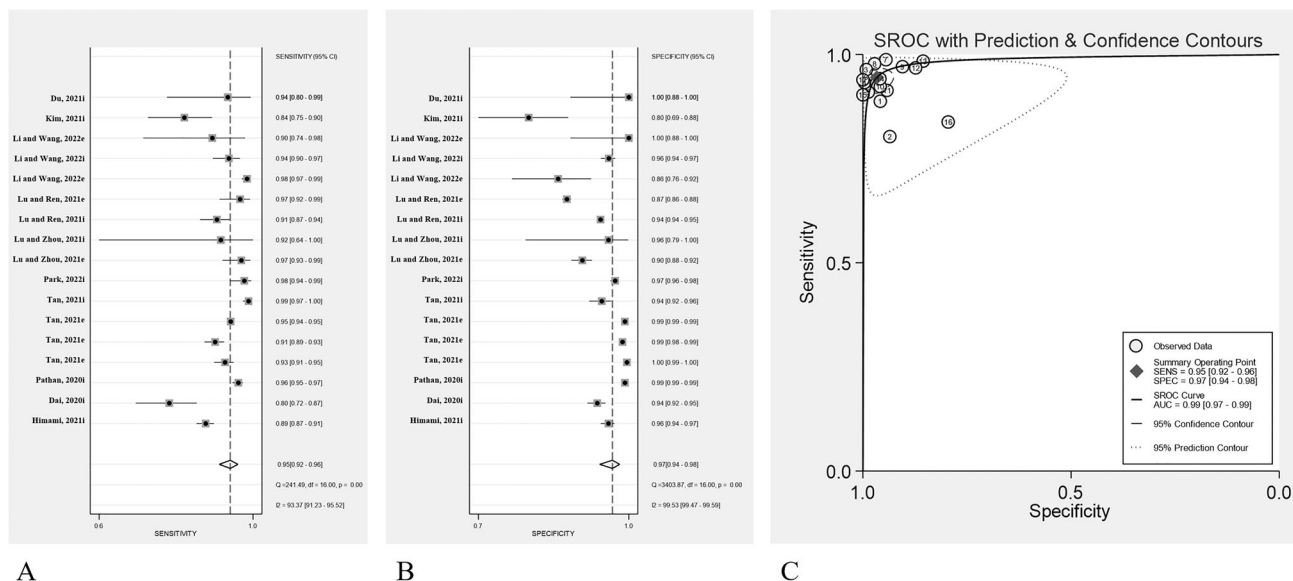


Fig. 2 Forest plots and summary receiver operating characteristic (SROC) curves for the performance of artificial intelligence for the detection of pathologic myopia. A The pooled sensitivity was 0.95 (95% CI: 0.92 to 0.96). **B** The pooled specificity was 0.97 (95% CI: 0.94 to 0.98). **C** The pooled area under the summary receiver operating characteristic (SROC) curve was 0.99 (95% CI: 0.97 to 0.99).

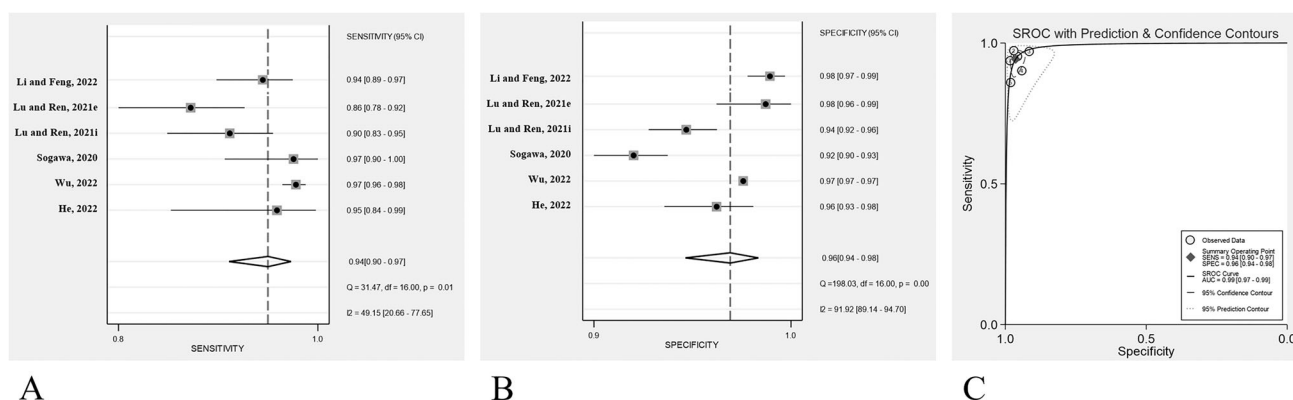


Fig. 3 Forest plots and summary receiver operating characteristics (SROC) curve for the performance of artificial intelligence for the detection of pathologic myopia-related choroidal neovascularization. A The pooled sensitivity was 0.94 (95% CI: 0.90 to 0.97). **B** The pooled specificity was 0.96 (95% CI: 0.94 to 0.98). **C** The pooled area under the summary receiver operating characteristic (SROC) curve was 0.99 (95% CI: 0.97 to 0.99).

DISCUSSION

We compared and analysed the results and characteristics of published studies and addressed the gaps in the current meta-analysis in the field of the application of AI in PM. Through our review, AI technology has the potential to benefit the detection and management procedure of PM patients in real-world settings, similar to other eye diseases. By estimation, CNV occurred in approximately 5–11% of eyes with high myopia, and early detection and interventions for high-risk lesions in PM patients are necessary to prevent underlying progression [34]. Through our review, AI models based on fundus or OCT images both achieved acceptable accuracy in the detection of CNV. According to published literature, junior to senior retinal specialists can achieve the accuracy of 0.999 for the detection of PM [35]. Despite the relatively lower accuracy compared to traditional clinical examinations, the utilization of AI can maximize the detection rates using a convenient method. Apart from CNV, several published algorithms also can identify complications in PM patients, for example, the extraction and segmentation of peripapillary atrophy, automatic quantitative analysis of fundus tessellation and automatic segmentation and measurement of the choroid layer [24, 36, 37].

These advances can help efficiently quantify large amounts of data and assist in detecting subtle differences that are difficult for ophthalmologists.

Through subgroup meta-analysis, there was no significant relationship between the diagnostic accuracy and imaging modalities or resources of databases, while there were significant relationships between the diagnostic performance and the scale of databases, the types of validation sets and the countries where the study was conducted affected. Compared to fundus images, advances in OCT can help detect more characteristics, such as macular-schisis and dome-shaped macula. With more studies included in the future, it would be more meaningful to compare the performances of AI in detail based on fundus and OCT imaging. We also believe the diagnostic performance would be further improved with the combination of AI technology and advanced imaging modalities such as ultra-widefield fundus images or swept-source OCT angiography (OCTA).

In contrast, AI algorithms demonstrated better performance in internal validation datasets than external datasets. Such a lack of reliability suggests that it is necessary to improve the generalization and robustness under different environments through a

Table 2. Subgroup analysis for the performance of AI algorithms for the detection of PM.

Subgroup variables	No. of studies	SEN (95%CI)	SPE (95%CI)	P value for interaction	AUC (95%CI)	LR + (95%CI)	LR- (95%CI)	DOR (95%CI)	I ² statistic, %
Imaging modality									
Fundus images	14	0.95 (0.92–0.97)	0.97 (0.97–0.99)	0.781	0.99 (0.98–1.00)	34.0 (14.8–47.1)	0.03 (0.01–0.08)	895 (279–1675)	99.3
OCT images	3	0.94 (0.94–0.96)	0.98 (0.96–0.99)		0.98 (0.97–0.99)	25.2 (11.6–151.9)	0.09 (0.06–0.13)	282 (43–10,760)	56.4
Types of validation dataset									
Internal	10	0.95 (0.94–0.96)	0.97 (0.96–0.99)	0.002*	0.99 (0.97–1.00)	30.7 (15.7–50.2)	0.04 (0.01–0.06)	573 (157–862)	98.1
External	7	0.93 (0.92–0.95)	0.96 (0.94–0.97)		0.99 (0.98–0.99)	21.9 (9.8–34.5)	0.08 (0.06–0.09)	417 (25–1306)	85.6
Resources of data									
Public	3	0.95 (0.94–0.97)	0.98 (0.97–0.98)	0.479	0.98 (0.96–0.99)	25.2 (13.9–33.7)	0.07 (0.05–0.09)	256 (19–13,849)	60.8
Hospital	14	0.95 (0.93–0.95)	0.97 (0.95–0.98)		0.99 (0.98–0.99)	41.8 (4.1–285.9)	0.02 (0.01–0.08)	731 (549–1354)	99.7
Regions									
Developed countries	7	0.96 (0.93–0.98)	0.98 (0.97–0.99)	0.034*	0.99 (0.97–0.99)	34.8 (25.7–47.2)	0.05 (0.05–0.06)	853 (203–4528)	89.6
Developing countries	10	0.94 (0.90–0.95)	0.96 (0.93–0.98)		0.98 (0.97–0.99)	17.8 (12.1–26.4)	0.08 (0.06–0.11)	290 (124–459)	97.9
Total number of images									
<5000	6	0.93 (0.91–0.95)	0.96 (0.94–0.98)	0.008*	0.98 (0.98–0.99)	21.3 (15.5–69.3)	0.07 (0.04–0.09)	289 (35–647)	86.9
≥5000	11	0.96 (0.95–0.98)	0.97 (0.96–0.99)		0.99 (0.97–0.99)	33.6 (14.1–50.7)	0.02 (0.01–0.03)	924 (340–1512)	99.0

AI artificial intelligence, PM pathologic myopia, SEN sensitivity, SPE specificity, AUC area under the receiver operating characteristic curve, LR + positive likelihood ratio, LR- negative likelihood ratio, DOR diagnostic odds ratio, OCT optical coherence tomography.

*P < 0.05.

variety of methods, such as training and testing the model widely in different populations or devices [38]. Another point of interest is that we found studies in developed countries showed higher accuracy than developing countries. This might be related to insufficient capacity to conduct high-quality studies in low- and middle-income countries (LMICs). However, the Southeast Asia, South Asia, and East Asia regions bear the greatest potential burden as a proportion of the economy associated with visual impairment resulting from uncorrected myopia and MMD globally [39]. More importantly, it might be difficult for healthcare systems in these countries to cope with a relatively greater burden, especially during the COVID-19 pandemic period. At the same time, detailed health economic evaluation for the application of AI-assisted models in real-world settings is required to identify the priorities and strategies of implementation.

More studies with high quality are necessary to enhance reliability to unleash greater potential in real-world settings. First, we can incorporate the data from multimodality images into future AI systems to build a screening system that can detect more dimensional characteristics in PM patients. Next, the establishment of longitudinal medical records for patients can help explore morphological characteristic parameters closely related to the progression of PM. Predicting potential risk of developing PM from school-aged myopia can provide evidence for precise individualized interventions. Additionally, the algorithms developed with the integration of more information, such as genomic readouts and metabolomics from patients, will increase the diagnostic or predictive power.

We should state that there exist several limitations in our meta-analysis and review. First, our study only confirmed the diagnostic power of AI in the detection of PM, but it is still unknown whether AI algorithms have overall good performance for grading PM according to different category systems. Second, there was high heterogeneity among the included studies due to the varying study designs, imaging modalities, algorithm characteristics and threshold effects. Third, the definition of pathological myopia was still controversial, META-PM based on fundus figures only, and ATN classification combined fundus figures and OCT, whereas both ignored the existence of posterior staphyloma. Fourth, some included studies were published in the journals of AI or computer science, and few clinical details were reported. Thus, there were unknown risks of bias in the selection of patients and the patient flow. Moreover, it has been reported that the QUADAS-2 tool might underestimate the risk of bias of the included studies [9]. Fifth, some included studies used the duplicated database (PALM database), and there were overlapping data in our pooled meta-analysis with underlying implications. Sixth, as in many AI-based studies especially in big image databases, the sampling mechanisms are unclear, and many diagnostic studies were case-control, meaning that diseased and non-diseased subjects were recruited based on different criteria. Last, current DL algorithms lack the interpretability for their detection outcome, which is called the “black box phenomenon”. The improvement of interpretability will help ophthalmologists identify probable structural features related to better diagnostic performance.

In conclusion, our review demonstrated the excellent performance of current AI algorithms in detecting PM patients based on fundus and OCT images, and AI-assisted automated screening systems are promising for ameliorating increasing demands in clinical settings. To the best of our knowledge, this was the first published meta-analysis for the assessment of AI algorithms applied in PM and PM-related CNV quantitatively. Nevertheless, to provide substantial benefits in regular clinical practice under different conditions, we still need to conduct continuous innovative research with newly developed algorithms and larger-scale databases.

SUMMARY

What was known before

- Current applications of artificial intelligence (AI) in ophthalmic diseases have covered a variety of aspects with good performance, including the diagnosis and classification of pathologic myopia (PM).
- There are still relatively great variations among studies in development procedures, databases, sample resources, and many aspects of methodology.

What this study adds

- Our study demonstrated the excellent performance of current AI algorithms in detecting PM patients based on fundus and OCT images, and this was the first published meta-analysis for the assessment of AI algorithms applied in PM quantitatively.
- For the detection of PM, the summary area under the receiver operating characteristic curve (AUC) was 0.99 (95% confidence interval (CI): 0.97 to 0.99), and the pooled sensitivity and specificity were 0.95 (95% CI: 0.92 to 0.96) and 0.97 (95% CI: 0.94 to 0.98), respectively.
- It provides crucial evidence for the application of AI-assisted automated screening systems to ameliorate increasing demands in the healthcare system.

DATA AVAILABILITY

Data are available from the corresponding author on reasonable request.

REFERENCES

1. Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology*. 2016;123:1036–42.
2. World Health Organization. The impact of myopia and high myopia: report of the Joint World Health Organization–Brien Holden Vision Institute Global Scientific Meeting on Myopia, University of New South Wales, Sydney, Australia. Geneva: World Health Organization, 2015.
3. Flitcroft DI, He M, Jonas JB, Jong M, Naidoo K, Ohno-Matsui K, et al. IMI—defining and classifying myopia: a proposed set of standards for clinical and epidemiologic studies. *Investig Ophthalmol Vis Sci*. 2019;60:M20.
4. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med*. 2021;4:65.
5. Du R, Ohno-Matsui K. Novel uses and challenges of artificial intelligence in diagnosing and managing eyes with high myopia and pathologic myopia. *Diagn Basel Switz*. 2022;12:1210.
6. Dong L, Yang Q, Zhang RH, Wei WB. Artificial intelligence for the detection of age-related macular degeneration in color fundus photographs: a systematic review and meta-analysis. *EClinicalMedicine*. 2021;35:100875.
7. Wu J-H, Nishida T, Weinreb RN, Lin J-W. Performances of machine learning in detecting glaucoma using fundus and retinal optical coherence tomography images: a meta-analysis. *Am J Ophthalmol*. 2022;237:1–12.
8. Li H-Y, Wang D-X, Dong L, Wei W-B. Deep learning algorithms for detection of diabetic macular edema in OCT images: a systematic review and meta-analysis. *Eur J Ophthalmol*. 2022;112067212210947:1–13.
9. Whiting PF. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529.
10. Ohno-Matsui K. Definition of Pathologic Myopia (PM). In: Ohno-Matsui K, editor. *Atlas of Pathologic Myopia* [Internet]. Singapore: Springer Singapore; 2020 [cited 2022 Oct 22]. p. 3–6. Available from: https://doi.org/10.1007/978-981-15-4261-9_1.
11. Ruiz-Medrano J, Montero JA, Flores-Moreno I, Arias L, García-Layana A, Ruiz-Moreno JM. Myopic maculopathy: current status and proposal for a new classification and grading system (ATN). *Prog Retin Eye Res*. 2019;69:80–115.
12. Sogawa T, Tabuchi H, Nagasato D, Masumoto H, Ikuno Y, Ohsugi H, et al. Accuracy of a deep convolutional neural network in the detection of myopic macular

- diseases using swept-source optical coherence tomography. *PloS One*. 2020;15:e0227240.
13. Lu L, Ren P, Tang X, Yang M, Yuan M, Yu W, et al. AI-model for identifying pathologic myopia based on deep learning algorithms of myopic maculopathy classification and "plus" lesion detection in fundus images. *Front Cell Dev Biol*. 2021;9:719262.
 14. Wan C, Li H, Cao G-F, Jiang Q, Yang W-H. An artificial intelligent risk classification method of high myopia based on fundus images. *J Clin Med*. 2021;10:4488.
 15. Tang J, Yuan M, Tian K, Wang Y, Wang D, Yang J, et al. An artificial-intelligence-based automated grading and lesions segmentation system for myopic maculopathy based on color fundus photographs. *Transl Vis Sci Technol*. 2022;11:16.
 16. Li J, Wang L, Gao Y, Liang Q, Chen L, Sun X, et al. Automated detection of myopic maculopathy from color fundus photographs using deep convolutional neural networks. *Eye Vis*. 2022;9:13.
 17. Rauf N, Gilani SO, Waris A. Automatic detection of pathological myopia using machine learning. *Sci Rep*. 2021;11:16570.
 18. Ye X, Wang J, Chen Y, Lv Z, He S, Mao J, et al. Automatic screening and identifying myopic maculopathy on optical coherence tomography images using deep learning. *Transl Vis Sci Technol*. 2021;10:10.
 19. Du R, Xie S, Fang Y, Igarashi-Yokoi T, Moriyama M, Ogata S, et al. Deep learning approach for automated detection of myopic maculopathy and pathologic myopia in fundus images. *Ophthalmol*. 2021;5:1235–44.
 20. Park S-J, Ko T, Park C-K, Kim Y-C, Choi I-Y. Deep learning model based on 3D optical coherence tomography images for the automated detection of pathologic myopia. *Diagn Basel Switz*. 2022;12:742.
 21. Li Y, Feng W, Zhao X, Liu B, Zhang Y, Chi W, et al. Development and validation of a deep learning system to screen vision-threatening conditions in high myopia using optical coherence tomography images. *Br J Ophthalmol*. 2022;106:633–9.
 22. Lu L, Zhou E, Yu W, Chen B, Ren P, Lu Q, et al. Development of deep learning-based detecting systems for pathologic myopia using retinal fundus images. *Commun Biol*. 2021;4:1225.
 23. Kim YC, Chang DJ, Park SJ, Choi IY, Gong YS, Kim H-A, et al. Machine learning prediction of pathologic myopia using tomographic elevation of the posterior sclera. *Sci Rep*. 2021;11:6950.
 24. Hemelings R, Elen B, Blaschko MB, Jacob J, Stalmans I, De Boever P. Pathological myopia classification with simultaneous lesion segmentation using deep learning. *Comput Methods Prog Biomed*. 2021;199:105920.
 25. Cui J, Zhang X, Xiong F, Chen C-L. Pathological myopia image recognition strategy based on data augmentation and model fusion. Lu H-C, editor. *J Healthc Eng*. 2021;2021:1–15.
 26. Wu Z, Cai W, Xie H, Chen S, Wang Y, Lei B, et al. Predicting optical coherence tomography-derived high myopia grades from fundus photographs using deep learning. *Front Med*. 2022;9:842680.
 27. Tan T-E, Anees A, Chen C, Li S, Xu X, Li Z, et al. Retinal photograph-based deep learning algorithms for myopia and a blockchain platform to facilitate artificial intelligence medical research: a retrospective multicohort study. *Lancet Digit Health*. 2021;3:e317–29.
 28. Du R, Xie S, Fang Y, Hagino S, Yamamoto S, et al. Validation of soft labels in developing deep learning algorithms for detecting lesions of myopic maculopathy from optical coherence tomographic images. *Asia Pac J Ophthalmol*. 2022;11:227–36.
 29. Pathan S, Siddalingaswamy PC, Dsouza N. Automated detection of pathological and non-pathological myopia using retinal features and dynamic ensemble of classifiers. *Telecommun Radio Eng*. 2020;79:1857–67.
 30. Dai S, Chen L, Lei T, Zhou C, Wen Y. Automatic detection of pathological myopia and high myopia on fundus images. In: 2020 IEEE International Conference on Multimedia and Expo (ICME) [Internet]. London, United Kingdom: IEEE; 2020 [cited 2022 Oct 10]. p. 1–6. Available from: <https://ieeexplore.ieee.org/document/9102787/>
 31. Himami ZR, Bustamam A, Anki P. Deep learning in image classification using dense networks and residual networks for pathologic myopia detection. In: 2021 International Conference on Artificial Intelligence and Big Data Analytics [Internet]. Bandung, Indonesia: IEEE; 2021 [cited 2022 Oct 10]. p. 1–6. Available from: <https://ieeexplore.ieee.org/document/9689744/>
 32. Kalyanasundaram A. Detection of pathological myopia using convolutional neural network. *Int J Psychosoc Rehabil*. 2020;24:2310–7.
 33. He X, Ren P, Lu L, Tang X, Wang J, Yang Z, et al. Development of a deep learning algorithm for myopic maculopathy classification based on OCT images using transfer learning. *Front Public Health*. 2022;10:1005700.
 34. Neelam K, Cheung CMG, Ohno-Matsui K, Lai TYY, Wong TY. Choroidal neovascularization in pathological myopia. *Prog Retin Eye Res*. 2012;31:495–525.
 35. Dong L, He W, Zhang R, Ge Z, Wang YX, Zhou J, et al. Artificial intelligence for screening of multiple retinal and optic nerve diseases. *JAMA Netw Open*. 2022;5:e229960.
 36. Du Y, Chen Q, Fan Y, Zhu J, He J, Zou H, et al. Automatic identification of myopic maculopathy related imaging features in optic disc region via machine learning methods. *J Transl Med*. 2021;19:167.
 37. Shao L, Zhang QL, Long TF, Dong L, Zhang C, Da Zhou W, et al. Quantitative assessment of fundus tessellated density and associated factors in fundus images using artificial intelligence. *Transl Vis Sci Technol*. 2021;10:23.
 38. Li J-PO, Liu H, Ting DSJ, Jeon S, Chan RVP, Kim JE, et al. Digital technology, telemedicine and artificial intelligence in ophthalmology: a global perspective. *Prog Retin Eye Res*. 2021;82:100900.
 39. Naidoo KS, Fricke TR, Frick KD, Jong M, Naduvilath TJ, Resnikoff S, et al. Potential lost productivity resulting from the global burden of myopia: systematic review, meta-analysis, and modeling. *Ophthalmology*. 2019;126:338–46.

AUTHOR CONTRIBUTIONS

YZ was responsible for the conceptualization of the research topic, designing and writing the protocol, conducting the database, writing, and editing the paper. YL was responsible for analyzing the data using statistical software and drawing the figures. JNW, HL, and JRZ were responsible for the screening of the studies, conducting the risk of bias assessment, curating the data. JL was responsible for the conceptualization of the research topic. XBY was responsible for the validation of results, review and editing the paper.

FUNDING

This study was supported by National High Level Hospital Clinical Research Funding (BJ-2022-104).

COMPETING INTERESTS

The authors declare no competing interests.

CONSENT FOR PUBLICATION

All listed authors consent to the submission.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41433-023-02551-7>.

Correspondence and requests for materials should be addressed to Xiaobing Yu.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.