Advances in Ophthalmology Practice and Research 4 (2024) 164–172

Contents lists available at ScienceDirect

Advances in Ophthalmology Practice and Research

journal homepage: www.journals.elsevier.com/advances-in-ophthalmology-practice-and-research

Review

Ocular image-based deep learning for predicting refractive error: A systematic review

Samantha Min Er Yew [a,b,1], Yibing Chen [c,1], Jocelyn Hui Lin Goh [d], David Ziyou Chen [a,b,e], Marcus Chun Jin Tan [a,b,e], Ching-Yu Cheng [a,b,d,f], Victor Teck Chang Koh [a,b,e], Yih-Chung Tham [a,b,d,f,*]

[a] Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[b] Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[c] School of Chemistry, Chemical Engineering, and Biotechnology, Nanyang Technological University, Singapore
[d] Singapore Eye Research Institute, Singapore National Eye Centre, Singapore
[e] Department of Ophthalmology, National University Hospital, Singapore
[f] Ophthalmology and Visual Sciences (Eye ACP), Duke-NUS Medical School, Singapore

A B S T R A C T

Background: Uncorrected refractive error is a major cause of vision impairment worldwide and its increasing prevalent necessitates effective screening and management strategies. Meanwhile, deep learning, a subset of Artificial Intelligence, has significantly advanced ophthalmological diagnostics by automating tasks that required extensive clinical expertise. Although recent studies have investigated the use of deep learning models for refractive power detection through various imaging techniques, a comprehensive systematic review on this topic is has yet be done. This review aims to summarise and evaluate the performance of ocular image-based deep learning models in predicting refractive errors.

Main text: We search on three databases (PubMed, Scopus, Web of Science) up till June 2023, focusing on deep learning applications in detecting refractive error from ocular images. We included studies that had reported refractive error outcomes, regardless of publication years. We systematically extracted and evaluated the continuous outcomes (sphere, SE, cylinder) and categorical outcomes (myopia), ground truth measurements, ocular imaging modalities, deep learning models, and performance metrics, adhering to PRISMA guidelines. Nine studies were identified and categorised into three groups: retinal photo-based (n = 5), OCT-based (n = 1), and external ocular photo-based (n = 3).

For high myopia prediction, retinal photo-based models achieved AUC between 0.91 and 0.98, sensitivity levels between 85.10% and 97.80%, and specificity levels between 76.40% and 94.50%. For continuous prediction, retinal photo-based models reported MAE ranging from 0.31D to 2.19D, and R^2 between 0.05 and 0.96. The OCT-based model achieved an AUC of 0.79–0.81, sensitivity of 82.30% and 87.20% and specificity of 61.70%–68.90%. For external ocular photo-based models, the AUC ranged from 0.91 to 0.99, sensitivity of 81.13%–84.00% and specificity of 74.00%–86.42%, MAE ranges from 0.07D to 0.18D and accuracy ranges from 81.60% to 96.70%. The reported papers collectively showed promising performances, in particular the retinal photo-based and external eye photo -based DL models.

Conclusions: The integration of deep learning model and ocular imaging for refractive error detection appear promising. However, their real-world clinical utility in current screening workflow have yet been evaluated and would require thoughtful consideration in design and implementation.

1. Introduction

Uncorrected refractive error (URE) is one of the leading causes of vision impairment (VI) globally,[1–3] accounting for 101.2 million cases of moderate to severe visual impairment and 6.8 million cases of blindness.[4] Left untreated, URE reduces quality of life and productivity and increased risk of falls.[5–8] The number of VI due to uncorrected refractive error is expected to rise.[5–8] This escalating trend underscores a growing public health concern, highlighting the need for effective screening and management.[9]

Deep learning, a subset of artificial intelligence, has significantly influenced ophthalmology. It demonstrates high efficacy in diagnosing various eye conditions such as diabetic retinopathy,[10,11] age-related macular degeneration,[12–14] cataract[15,16] and glaucoma.[17] Deep learning has the potential to enhance screening efficiency. Remarkably, their performance often match or even surpass that of expert-level assessments. Additionally, deep learning enhances diagnostic accuracy and efficiency, offering early disease detection. It can potentially facilitate

* Corresponding author. Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.
E-mail address: thamyc@nus.edu.sg (Y.-C. Tham).
[1] Contributed equally as co-first authors.

https://doi.org/10.1016/j.aopr.2024.06.005
Received 12 January 2024; Received in revised form 20 June 2024; Accepted 24 June 2024
Available online 2 July 2024

remote screening and integration with telemedicine, thus expanding access to eye care, particularly in underserved regions. However, in the domain of refractive error detection, although several research groups have explored using deep learning models with various imaging modalities such as fundus images, ultra-widefield fundus images, photorefraction images, ocular appearance images, and optical coherence tomography (OCT) images,[18–26] a comprehensive systematic review of these efforts has not yet been performed.

Therefore, this review evaluates deep learning for detecting refractive errors using ocular images, highlighting its significance and practical utility, and providing insights for future research and clinical applications.

2. Methods

This systematic review was written in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines. The protocol was registered in PROSPERO (reg. CRD42023395772). Institutional review board approval was not required for this study as it involved the use of data extracted from the literature available in the public domain.

2.1. Search strategy and inclusion criteria

In this systematic review, we searched the databases (PubMed, Scopus and Web of Science) from inception till June 13, 2023. Search terms specific to the domains of this systematic review were identified by the team. Firstly, search terms of "Artificial intelligence" and "Deep learning" were used. Secondly, for search terms related to refractive error, we used terms such as "Myopia", "Hyperopia", "Myope", "Hyperope" and "Astigmatism". Lastly, for search terms related to ocular images, we used terms such as "Retinal images", "Retinal", "Retina", "Fundus", "OCT", "Optical Coherence Tomography", and "Ocular images". Boolean operators of "OR" and "AND" were utilised to enhance the scope of studies eligible for screening. Details of the search terms are listed in Table. 1.

As part of our inclusion criteria, included studies had to be available in full text, published in English and reported refractive error outcomes. The year of publication was not restricted. We excluded studies that were reported as surveys, case reports, editorials, opinions. Studies which only reported findings on detection of pathological myopia disease, were also excluded.

2.2. Data extraction and analysis

Screening of reports and data extraction were performed and cross-checked by SYME and CYB. If there were disagreements between YSME and CYB, senior author (TYC) was consulted. Data extracted included study setting details (first author, year of publication), study population (sample size of training database, internal database, and external database), input modalities [fundus images, ultra-widefield fundus images, photorefraction images, external ocular appearance images and, OCT images], output modalities (e.g. spherical equivalent (SE), sphere, cylinder, classification of myopia, high myopia, etc), types of deep learning model used (name of neural network), types of models (e.g. binary, regression, and classification), and study results (performances of the reported deep learning algorithm). From the included studies, the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, mean absolute error (MAE), and coefficient of determination (R^2) were summarised in Tables 2-4.

During the initial literature search, a total of 417 records were retrieved. 111 were from PubMed, 163 from Scopus and 143 were from Web of Science. Among the 417 records identified, 116 duplicate records were removed, leaving 301 articles for review based on the titles and abstract. We excluded 114 articles from this process, leaving 187 articles to be assessed for eligibility. We further excluded 178 articles due to it not relating to refractive error outcomes. After further review, 9 studies were ultimately included in this systematic review paper (Fig. 1).

2.3. Risk of bias

We tailored the QUADAS-2 tool specifically for this review to assess the quality of the included studies. Data extraction and risk of bias assessment were independently conducted by two authors (YSME and CYB), with discrepancies reviewed and resolved by a senior author (TYC) to ensure the integrity and reliability of the assessment process.

3. Results

Of the 9 included studies, 3 were retinal photo-based deep learning algorithms, 2 were based on ultra-wide field fundus photo, 1 was based

Table 1
Search Strategies Deployed in this Systematic Review.

Databases		PubMed	Scopus	Web of Science
#1	("Deep learning" OR "Artificial Intelligence")	n = 91131	n = 695535	n = 584596
#2	("Refractive error" OR "Myopia" OR Hyperopia" OR "Myope" OR "Hyperope" OR "Astigmatism")	n = 34087	n = 68668	n = 42242
#3	("Eye-related images" OR "Retinal images" OR "Retinal" OR "Retina" OR "Fundus" OR "OCT" OR "Optical Coherence Tomography" OR "Ocular images")	n = 2948981	n = 343166	n = 1868911
#4	#1 and #2 and #3	n = 111	n = 163	n = 143

Table 2
Summary of the characteristic of the included studies.

	Author, Year	AI Output	GT Measurement	GT Definition	Input Modalities	Deep Learning Model Used	Models type
1	Tan et al., 2021[20]	High myopia	Non-cycloplegic autorefraction and subjective refraction	High myopia: SE of −6.00D or higher and/or AL of 26.0 mm or more	Retinal fundus images	ResNet-101	Binary
2	Varadarajan et al., 2018[21]	SE Sphere Cylinder	Non-cycloplegic autorefraction and subjective refraction	NA	Retinal fundus images	Combined ResNet and Soft-attention	Regression
3	Zou et al., 2022[26]	Sphere Cylinder	Cycloplegic subjective refraction	NA	Retinal fundus images	Fusion model-based deep learning system (FMDLS)	FMDLS (using eigenvector without age parameter)
4	Shi et al., 2021[19]	SE	Cycloplegic refraction	NA	Ultra-widefield fundus images	Attention Dense block that combines the advantages of dense connection and Residual Squeeze-and-Excitation (Res-SE_1)	Regression
5	Yang et al., 2022[23]	SE	Non-cycloplegic subjective refraction	NA	Ultra-widefield fundus images	ResNet-50, InceptionV3 Inception-ResNet-v2	Regression
6	Yoo et al., 2022[25]	SE High myopia and moderate myopia and wore	Non-cycloplegic subjective refraction	High myopia: SE ≤ −6.00 D Moderate myopia and worse: SE ≤ −3.00D	OCT images	ResNet50, inceptionV3, and VGG16 as feature extractors.	Regression and Multiclass-classification
7	Chun et al., 2020[18]	Different classes of refractive error	Cycloplegic retinoscopy	7 classes of refractive error: High myopia: ≤−5.00D Moderate myopia: >−5.00D and ≤−3.00D Mild myopia: >−3.00D and ≤−0.50D Emmetrope: <−0.50D and <+0.50D Mild hyperopia: ≥+0.50D and <+3.00D Moderate hyperopia: ≥+3.00D and <+5.00D High hyperopia: ≥+5.00D	Photorefraction images, acquired using a smartphone	ResNet-18	Multiclass-Classification
8	Xu et al., 2022[22]	SE Sphere Cylinder High Myopia	Non-cycloplegic autorefraction		Photorefraction images, acquired using a Basler ace 2 camera	Refractive Error Detection Network (REDNet), combining CNN and recurrent CNN.	Regression and Binary
9	Yang et al., 2020[24]	Myopia	Non-cycloplegic autorefraction	Myopia is defined as SE ≤ −0.50D	External eye photographs, with 3 angles taken (from the side, 45-degree angle front, and front), acquired using smartphones and a digital single-lens reflex (DSLR) camera	VGG-Face model	Binary

AI = artificial intelligence. AL = axial length. GT = ground truth. SE = spherical equivalent. OCT = ocular coherence tomography.

on OCT image, and 3 were based on external ocular images. All of the 9 studies evaluated the performance of deep learning models in predicting refractive error, however, the outputs differ slightly. Of these, 4 studies investigated the use of deep learning algorithms to predict continuous outcomes such as SE power, sphere, cylinder and axis. 3 studies explored the predictive capabilities of these models for binary or categorical outcomes (e.g. high myopia, myopia and, classification of myopia and hyperopia). The remaining 2 studies assessed the efficacy of various deep learning models in predicting both binary/categorical and continuous outcomes. The characteristics and datasets used in these 9 included studies were summarised in Tables 2 and 3

3.1. Retinal photo-based deep learning algorithms for detection of refractive error

It is known that refractive error is associated with retinal fundus changes (e.g. tessellated fundus, myopia-related fundus changes etc.).[27,28] Based on this rationale, several retinal photo-based deep learning algorithms were developed for automated detection of refractive error. In our literature review, 3 deep learning studies reported the use of standard fundus images (30–50° imaging field of view) as input modalities to detect refractive error. These results are summarised in Table 4.

Table 3
Source of dataset and number of images reported in the included studies.

	Author, Year	Train (data source, images)	Internal Test (data source, images)	External Test (data source, images)
1	Tan et al., 2021[20]	**Total:** 13751 SEED: 13482 SNEC: 269	**Total:** 5840 SEED: 5750 SNEC: 90	***SE of -6.00D or worse*** **Total:** 206080 Test Dataset 1: BES: 5673 Test Dataset 2: HES: 11985 Test Dataset 3: CIEMS: 6,552, UEMS: 7781, and CGMH: 776 Test Dataset 5: UKBB: 173313 ***AL of 26.0 mm or more*** **Total:** 17658 Test Dataset 1: BES: 5673 Test Dataset 2: HES: 11985
2	Varadarajan et al., 2018[21]	**Total:** 226870 UKBB: 96081 AREDS: 130789 UKBB: 96081	**Total:** 31155 UKBB: 23520 AREDS: 7635 UKBB: 23520	NA NA
3	Zou et al., 2022[26]	Tianjin Eye Hospital of Nankai University, China: 7086	Tianjin Eye Hospital of Nankai University, China: 787	NA
4	Shi et al., 2021[19]	China Aier Eye Hospital & Southern Medical University, China: 22692	China Aier Eye Hospital & Southern Medical University, China: 600	NA
5	Yang et al., 2022[23]	Eye & ENT Hospital of Fudan University, China: 790	Eye & ENT Hospital of Fudan University, China: 197	Eye & ENT Hospital of Fudan University, China: 133
6	Yoo et al., 2022[25]	B&VIIT Eye Center, Korea: 688	B&VIIT Eye Center, Korea: 248	NA
7	Chun et al., 2020[18]	Samsung Medical Center, Korea: 213	Samsung Medical Center, Korea: 61	NA
8	Xu et al., 2022[22]	Unknown source: 3907	Unknown source: 1000	NA
9	Yang et al., 2020[24]	Myopia Artificial Intelligence Program, China: 1763	Myopia Artificial Intelligence Program, China: 587	100

SEED = Singapore Epidemiology of Eye Disease. SNEC = Singapore National Eye Center. SE = Spherical Equivalent. BES = Beijing Eye Study. HES = Handan Eye Study. CIEMS = Central India Eye and Medical Study. UEMS = Ural Eye and Medical Study. CGMH = Chang Gung Memorial Hospital. UKBB = United Kingdom BioBank. AL = axial length. AREDS = Age-Related Eye Disease Study.

Using retinal photos, Tan et al. (2021) developed ResNet-101 binary model to detect presence of high myopia.[20] In this study, high myopia was defined based on SE of −6.00D or worse or based on axial length (AL) of ≥26 mm. For SE-defined high myopia detection, using an internal test set of 5840 images, the reported system achieved an AUC of 0.98, sensitivity of 91.30% and specificity of 94.50%. In 4 external test sets consisting of a total 206,080 images, the model achieved AUCs ranging from 0.91 to 0.97, sensitivity values of 85.30%–92.70%, and specificity values of 80.80%–95.50%. For detection of AL-defined high myopia, based on the same internal test set, the model achieved an AUC of 0.95, sensitivity of 85.10% and specificity of 91.30%. In two external test sets of total 17568 images, the model achieved AUC of 0.96–0.97, sensitivity of 94.4%–97.8%, and specificity of 76.4%–88.1%.

In another study, Varadarajan et al. (2018) developed Combined ResNet and Soft-attention regression model to predict refractive error power (SE, sphere, and cylinder).[21] For SE power prediction, based on 2 internal test sets consisting of 31155 images, the model achieved MAE of 0.56D to 0.91D, and R^2 of 0.69–0.90. For sphere power prediction, using an internal test set of 23520 images, the model achieved MAE of 0.63D and R^2 of 0.88. For cylinder power prediction, using the same internal test set of 23520 images, the model achieved MAE of 0.43D and R^2 of 0.05. However, there was no external validation performed in this study.

Zou et al. (2022) developed a fusion model-based deep learning system (FMDLS) to predict cycloplegic sphere and cylinder power.[26] The FMDLS was built using both a regression and classification model. For sphere power prediction, using an internal test set of 787 images, the model achieved a MAE of 0.63D and R^2 of 0.66. For cylinder power prediction, using the same internal test set, the model achieved a MAE of 0.31D and R^2 of 0.65. External validation was not performed in this study.

On the other hand, there were 2 studies which used ultra-widefield fundus images as input modalities to detect refractive error (Table 2). Ultra-widefield fundus provides a wider imaging field of view (up to 200°) and contains more peripheral fundus information (e.g. peripheral degenerative changes in myopes). Shi et al. (2021) used Attention Dense Block regression model to predict cycloplegic SE.[19] For SE power prediction, based on an internal test set of 600 images, the model achieved a MAE of 1.12D. There was no external validation in this study.

In addition, using ultra-wide field fundus images, Yang et al. (2022) reported the performances of ResNet-50, Inception-v3 and Inception-ResNet-v2 models in predicting SE power.[23] Based on an internal test set of 197 images, the ResNet-50 model achieved an MAE of 1.72D and R^2 of 0.96; the Inception-v3 achieved model achieved an MAE of 1.75D and R^2 of 0.96; and the Inception-ResNet-v2 achieved an MAE of 1.76D and R^2 of 0.96. Based on an external test set of 133 images, the ResNet-50 model achieved MAE of 1.94D and R^2 of 0.93; the Inception-v3 model achieved a MAE of 1.79D and R^2 of 0.92; and the Inception-ResNet-v2 model achieved a MAE of 2.19D and R^2 of 0.93.

3.2. OCT-based deep learning algorithms for detection of refractive error

Past studies had shown that degree of refractive error can influence OCT measurement.[29–31] Based on this rationale, using OCT images as input modalities, Yoo et al. (2022) developed deep learning algorithms to detect refractive error (Table 2). Using horizontal and vertical OCT B-scan images as inputs, Yoo et al. (2022) reported the performances of developed CNN models which included the ResNet50, InceptionV3, and VGG16 in predicting SE power, and classifying moderate and high myopia.[25]

For SE power prediction, using an internal test set of 248 horizontal OCT B-scan images, the ResNet50 model achieved a MAE of 2.66D and R^2 of 0.35, the InceptionV3 model achieved a MAE of 2.68D and R^2 of 0.34; and the VGG16 model achieved a MAE of 2.71D and R^2 of 0.34. Similarly, based on 248 vertical OCT B-scan images, ResNet50 achieved MAE of 2.75D and R^2 of 0.33; InceptionV3 achieved MAE of 2.76D and R^2 of 0.32; and VGG16 achieved MAE of 2.79D and R^2 of 0.32. For detection of moderate myopia and worse, using an internal test set of 248 horizontal OCT images, ResNet50 achieved an AUC of 0.79, accuracy of 79.80%, sensitivity of 82.30%, and specificity of 68.90%. Based on the same internal test set, for high myopia detection, the ResNet50 model achieved

Table 4
Results of deep learning performance in detecting refractive error reported in the included studies.

	Author, Year	Outcome	Performance											
			Internal						External					
			AUC	Se (%)	Sp (%)	MAE (D)	R^2	Acc (%)	AUC	Se (%)	Sp (%)	MAE (D)	R^2	Acc (%)
1	Tan et al., 2021[20]	Binary: SE of −6.00D or higher	0.98	91.3	94.5	NA	NA	NA	0.91–0.97[a]	85.3–92.7[a]	80.8–95.5[a]	NA	NA	NA
		Binary: AL of 26.0 mm or more	0.95	85.1	91.3	NA	NA	NA	0.96–0.97[b]	94.4–97.8[b]	76.4–88.1[b]	NA	NA	NA
2	Varadarajan et al., 2018[21]	Continuous: SE (Train on UKBB and AREDS)	NA	NA	NA	0.56–0.91[c]	0.69–0.90[c]	NA	NA	NA	NA	NA	NA	NA
		Continuous: Sphere (Train on UKBB only)	NA	NA	NA	0.63	0.88	NA	NA	NA	NA	NA	NA	NA
		Continuous: Cylinder (Train on UKBB only)	NA	NA	NA	0.43	0.05	NA	NA	NA	NA	NA	NA	NA
3	Zou et al., 2022[26]	Continuous: Sphere	NA	NA	NA	0.63	0.66	NA	NA	NA	NA	NA	NA	NA
		Continuous: Cylinder	NA	NA	NA	0.31	0.65	NA	NA	NA	NA	NA	NA	NA
4	Shi et al., 2021[19]	Continuous: SE	NA	NA	NA	1.12	–	NA	NA	NA	NA	NA	NA	NA
5	Yang et al., 2022[23]	Continuous: SE (Model: ResNet-50)	NA	NA	NA	1.72	0.96	NA	NA	NA	NA	1.94	0.93	NA
		Continuous: SE (Model: Inception-v3)	NA	NA	NA	1.75	0.96	NA	NA	NA	NA	1.79	0.92	NA
		Continuous: SE (Model: Inception-ResNet-v2)	NA	NA	NA	1.76	0.96	NA	NA	NA	NA	2.19	0.93	NA
6	Yoo et al., 2022[25]	Continuous: SE (Model: ResNet50)	NA	NA	NA	2.66–2.75[d]	0.33–0.35[d]	NA	NA	NA	NA	NA	NA	NA
		Continuous: SE (Model: Inception-v3)	NA	NA	NA	2.68–2.76[d]	0.32–0.34[d]	NA	NA	NA	NA	NA	NA	NA
		Continuous: SE (Model: VGG16)	NA	NA	NA	2.71–2.79[d]	0.32–0.34[d]	NA	NA	NA	NA	NA	NA	NA
		Binary: Moderate and worse myopia (Model: ResNet50)	0.79	82.3	68.9	NA	NA	79.8	NA	NA	NA	NA	NA	NA
		Binary: High myopia (Model: ResNet50)	0.81	87.2	61.7	NA	NA	71.4	NA	NA	NA	NA	NA	NA
7	Chun et al., 2020[18]	Multiclassification: 7 different classes of refractive error[e]	NA	NA	NA	NA	NA	75.0–83.3[f]	NA	NA	NA	NA	NA	NA
8	Xu et al., 2022[22]	Binary: High myopia	0.99	97.19	96.97	NA	NA	97.13	NA	NA	NA	NA	NA	NA
		Continuous: Sphere	NA	NA	NA	0.17	–	89.5	NA	NA	NA	NA	NA	NA
		Continuous: Cylinder	NA	NA	NA	0.07	–	96.7	NA	NA	NA	NA	NA	NA
		Continuous: SE	NA	NA	NA	0.18	–	89.38	NA	NA	NA	NA	NA	NA
9	Yang et al., 2020[24]	Binary: Myopia	0.93	81.13	86.42	NA	NA	NA	0.91	84.0	74.0	NA	NA	NA

AUC= Area Under Curve. Se= Sensitivity. Sp = Specificity. MAE = Mean Absolute Error. R^2= Coefficient of determination.
[a] = Range of performance across the 4 datasets.
[b] = Range of performance across the 2 datasets.
[c] = Range of performance across UKBB and AREDS.
[d] = Range of performance across horizontal and vertical OCT B-scan.
[e] = The 7 different classes are ≤ −5.00D, > −5.00D and ≤ −3.00D, > −3.00D and ≤ −0.50D, < −0.50D and < +0.50D, ≥ +0.50D and <+3.00D, ≥ +3.00D and < +5.00D, ≥ +5.00D.
[f] = The accuracy range in the 7 different classes.

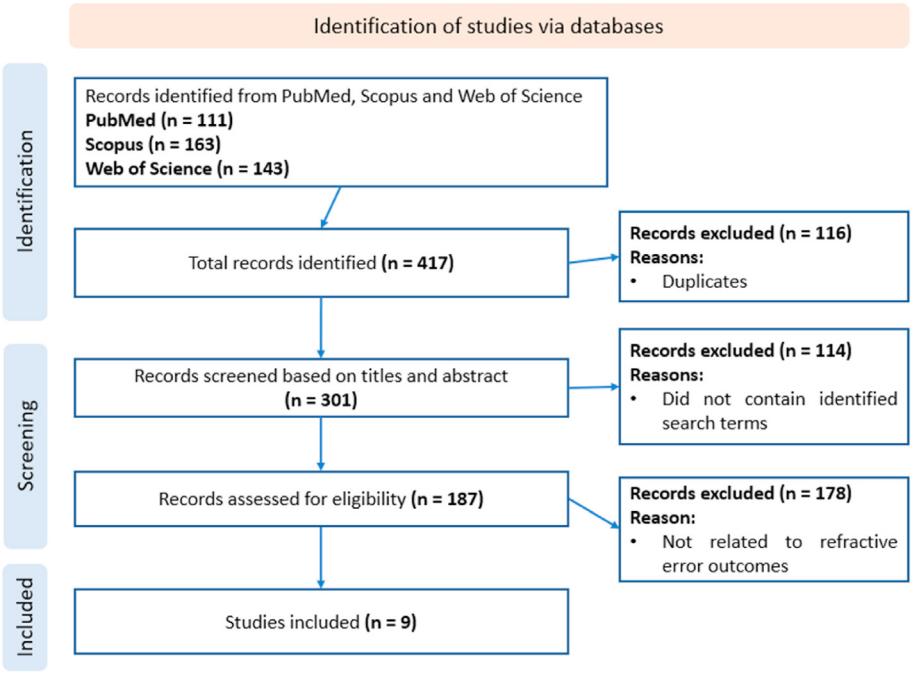

Fig. 1. PRISMA flow diagram.

an AUC of 0.81, accuracy of 71.40%, sensitivity of 87.20%, and, specificity of 61.70%. There was no external validation in this study.

3.3. External ocular photo-based deep learning algorithms for detection of refractive error

In addition, these were 3 studies which used external eye images (photorefraction images and ocular appearance images) as input modalities to detect refractive error. These results are shown in Table 4.

Chun et al. (2020) captured external eye photographs using a smartphone coupled with flash illumination. The acquired photos showed the type and position of ocular reflectance images known as photorefraction images. Using these photorefraction images as inputs, Chun et al. (2020) developed a ResNet-18 multiclass-classification model to detect cycloplegic refractive error.[18] In this study, refractive error was further classified into 7 classes. High myopia was defined as ≤ -5.00D; moderate myopia as > -5.00D to ≤ -3.00D; mild myopia as > -3.00D to ≤ -0.50D; emmetrope as < -0.50D to $< +0.50$D; mild hyperopia defined as $\geq +0.50$D to $< +3.00$D; moderate hyperopia as $\geq +3.00$D to $< +5.00$D; and high hyperopia as $\geq +5.00$D. For all classes of refractive error, internal test set of 61 images were used. Based on a 5-fold cross-validation method, the system achieved 80.0% accuracy for high myopia, 77.8% accuracy for moderate myopia, 82.0% accuracy for mild myopia, 83.3% accuracy for emmetrope, 82.8% accuracy for mild hyperopia, 79.3% accuracy for moderate hyperopia, and 75.0% accuracy for high hyperopia. The overall accuracy of the multiclass-classification model reached 81.6% and mean accuracy reached 80.03%. However, no external validation was performed in this study.

On the other hand, using the Basler ace 2 camera device with infrared illumination, Xu et al. (2022) acquired photorefraction images and developed a REDNet model to predict sphere and cylinder power, and detect presence of high myopia (SE < -6.00D).[22] Using an internal test of 1000 photorefraction images, REDNet-sphere power prediction achieved a MAE of 0.17D and accuracy of 89.5%, REDNet-cylinder power prediction achieved a MAE of 0.07D and accuracy of 96.7% and REDNet-SE prediction achieved a MAE of 0.18D and accuracy of 89.38%. Using the same internal test, for the detection of high myopia, the model achieved AUC of 0.99, accuracy of 97.13%, specificity of 97.19%, and sensitivity of 96.97%.

In addition, Yang et al. (2020) utilised external eye images taken from 3 different angles (from the side, 45-degree angle front, and front) and developed a VGG-Face model for detection of myopia (defined as SE of ≤ -0.50D).[24] Using an internal test set of 587 images, the model achieved an AUC of 0.93, sensitivity of 81.13%, and specificity of 86.42%. Based on an external test set of 100 images the model achieved an AUC of 0.91, sensitivity of 84.0%, and specificity of 74.0%.

3.4. Risk of bias

Supplementary Fig. 1 shows a summary of the quality assessment. In the patient selection domain, one of the nine studies were assessed as having an unclear risk of bias due to unspecified sources of photorefraction images used in algorithm development. The remaining studies clearly described their data sources and exclusion criteria, resulting in a low risk of bias. Applicability concerns for patient selection were low across all studies.

In the domain of index tests, all the studies were evaluated as having a low risk of bias and low applicability concerns because the tests used were appropriate and relevant for the study conditions. In the reference standard domain, all studies provided clear and adequate ground truth labels for refractive errors. These labels were based on established diagnostic criteria, indicating a low risk of bias and low applicability concerns. Similarly, in the flow and timing domain, all studies were considered to have a low risk of bias and applicability concerns, as all subjects received the reference standard and were included in the analysis.

Overall, the quality assessment indicates that the studies included in this review generally maintain high methodological standards with low risks of bias and low concerns regarding the applicability of their findings.

4. Discussion

4.1. Main findings

Through rigorous evaluation of the currently available literature, we

have analysed and summarised its performance. Overall, majority of the aforementioned papers reported promising performances, with standard retinal fundus images and external eye photo-based deep learning models producing better performances as compared to other imaging modalities such as wide-field retinal imaging and OCT.

We noted that the ultra-widefield-regression-based predictive accuracy (Shi et al., 2021: MAE: 1.12D for SE; Yang et al., 2022: 1.72 to 1.76D for SE) is poorer compared to standard-regression-based retinal images (Varadarajan et al., 2020: MAE: 0.56 to 0.91D for SE, MAE: 0.63D for Sphere; Zou et al., 2022: MAE: 0.63D for Sphere). Several factors could explain this discrepancy. Firstly, we postulated that ultra-widefield fundus imaging captures a broader field of view ultra-widefield fundus imaging captures a broader field of view, the peripheral regions tend to have more artefacts, which may inherently produce more noise, thus resulting in higher MAE value. Yang et al. (2022) encountered artefacts such as shadows and colour casts, which are common in ultra-widefield imaging. In contrast, Shi et al. (2021) cropped their images, resulting in a loss of resolution. These issues introduce noise, leading to poorer results. Secondly, the models using ultra-widefield images had smaller training sample sizes (790–22692 images) compared to those using standard images (7086 to 226870 images), which may have contributed to the poorer performance of ultra-widefield-based deep learning models. Moreover, we observed that Yang et al. (2022) reported high MAE values, yet also showed high R^2 values ($\geq$0.92), which seems contradictory. It is unlikely for a model with large prediction errors (high MAE) to also explain a large portion of the variability in the data (high R^2). Typically, high MAE indicates large prediction errors, which would correspond to lower R^2 values, thus further analysis is needed to elucidate this aspect in the study. Consequently, the results should be interpreted with caution.

The OCT-based deep learning model study by Yoo et al. (2022) reported the poorest performance in detecting refractive error. Critically, axial length, a key factor strongly associated with myopia,[32] was not considered in their analysis. This significant exclusion may impact the model's accuracy. Incorporating axial length into model development or post-analysis may improve predictive accuracy.

Lastly, deep learning models based on external eye photos, using photorefraction images[18,22] (taken with infrared cameras and smartphones) and anterior eye images[24] (taken with various types of cameras, including smartphones), have good results, Notably, Xu et al. (2022) reported exceptional performance in predicting refractive error, achieving a high AUC of 0.99 and a low MAE ranging from 0.07D to 0.18D. Although anterior eye images are not commonly used in clinical settings, they offer the convenience of home acquisition. The widespread use of smartphones could potentially improve accessibility and affordability of vision care, especially in resource-limited settings. This method is also beneficial for young patients and those with mobility issues. Nonetheless, photorefraction imaging requires precise distances, angles, and specific infrared illumination, which may hinder widespread adoption.

4.2. Strengths

To our knowledge, this is the first systematic review to comprehensively evaluate the performance of various ocular image-based deep learning models in detecting refractive error. We identified appropriate search terms and carefully reviewed and extracted the relevant literature. We tailored the QUADAS-2 tool specifically for this study. Data extraction and risk of bias assessment were performed by two review authors (YSME and CYB), with any discrepancies reviewed by a senior author (TYC), thereby ensuring a robust and unbiased evaluation process. Through the QUADAS-2 assessment, all included studies demonstrated a low risk of bias, indicating high methodological quality. All studies had appropriate patient selection, index test, reference standard, and flow and timing, ensuring the reliability of their findings.

4.3. Limitations

As summarised in , the definition of myopia and the measurement of refractive error (e.g., with or without cycloplegia) vary across the models. Additionally, the performance metrics (e.g., AUC, MAE, etc.) for refractive error predictions differ among the nine included studies. Due to these discrepancies, direct comparison of these deep learning models is challenging. Some studies also have relatively smaller sample sizes in both training and testing datasets, which may affect the robustness of their algorithms. Furthermore, although all the studies did conduct internal test, the lack of external validation shows uncertainty in the generalization of models. Therefore, further validation work is needed to evaluate their performance.

The resolution of input images is critical for algorithm performance, with higher-quality images yielding better results.[33] However, in a community or primary care setting, images may be taken with different cameras, resulting in varying image quality. Most of the cameras mentioned in the included papers are traditional non-portable fundus camera, raising concerns about the generalizability of these algorithms. Future work should evaluate the utility in real-world settings, such as community screenings, primary care, and tertiary eye hospitals. In this regard, further refinement is needed to ensure that algorithm's performance can remain intact even when images are acquired from different camera types. Addressing these issues would help to further enable real-world deployment in the future.

4.4. Future outlook

In light of these findings, potential future implementation of refractive error-based deep learning model may concentrate on utilising standard retinal photos as image inputs. Ocular imaging techniques play a fundamental role in the diagnosis and management of a wide range of eye conditions. Standard retinal fundus imaging is widely used in global vision screening programs for diabetic retinopathy, notably in countries like the United Kingdom and the United States.[34] This paves the way for opportunistic screening. Within environments where retinal imaging infrastructure is already in place, the same images originally intended for use can be harnessed for the assessment of refractive errors using deep learning algorithms. This complementary approach can potentially enhance screening effectiveness, particularly in distinguishing between visual impairments arising from refractive errors or underlying eye diseases. This could help improve screening efficiency, provide early detection and treatment, and improve quality of life. Conversely, OCT imaging demonstrated the least effective performance, followed by ultra-widefield imaging, and both are less accessible for screening. Additionally, external eye imaging is not commonly practiced in clinics. Therefore, standard retinal imaging, being more widely used, could be more viable for opportunistic screening.

4.5. Remaining gaps and challenges

The practicality of integrating retinal images and deep learning algorithm for refractive error prediction warrants careful consideration. Traditional methods such as auto-refraction and visual acuity tests are affordable and user-friendly. However, challenges persist, particularly in rural areas, due to the need for well-lit spaces, good communication skills, and skilled clinical staff to perform these tests. A streamlined approach involving a single, user-friendly device, such as a portable fundus camera incorporated with deep learning, may improve screening efficiency. This approach requires only one operator and could be effectively adapted for opportunistic screening of refractive errors. Overall, while deep learning models are capable of detecting refractive errors, it is prudent to explore the practicality and effectiveness of this innovative approach in a real clinical setting to fully ascertain its potential benefits.

5. Conclusions

Findings from this review offers a thorough insight into the capabilities of various ocular image-based deep learning algorithms for detecting refractive errors. The performance of each ocular imaging method varies. Retinal imaging demonstrated superior performance, emerging as more viable choice for opportunistic screening.

From a technical standpoint, further research is necessary to validate these algorithms in broader populations and in practical screening environments, where image acquisition often presents greater challenges. Exploring various types of deep learning models could also address potential biases arising from imbalanced datasets.

Overall, while the integration of deep learning model and ocular imaging for refractive error detection appear promising, especially for opportunistic screenings, their real-world clinical utility and implementation still demand further thoughtful design and implementation planning.

Study approval

Not Applicable.

Author contributions

YSME, CYB, and TYC conceptualized and designed the study. YSME and CYB collected the data. YSME, CYB, and TYC analysed the data. YSME, CYB, and TYC drafted the manuscript, with equal contributions from YSME and CYB as first authors. All authors critically summarised the findings and approved the final version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aopr.2024.06.005.

References

1. Blindness GBD, Vision Impairment C, Vision Loss Expert Group of the Global Burden of Disease S. 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 Global Health*. Feb 2021;9(2):e144–e160. https://doi.org/10.1016/S2214-109X(20)30489-7.
2. Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis. *Lancet Global Health*. Dec 2017;5(12):e1221–e1234. https://doi.org/10.1016/S2214-109X(17)30393-5.
3. Keel S, Li Z, Scheetz J, et al. Development and validation of a deep-learning algorithm for the detection of neovascular age-related macular degeneration from colour fundus photographs. *Clin Exp Ophthalmol*. Nov 2019;47(8):1009–1018. https://doi.org/10.1111/ceo.13575.
4. Naidoo KS, Leasher J, Bourne RR, et al. Global vision impairment and blindness due to uncorrected refractive error, 1990-2010. *Optom Vis Sci*. Mar 2016;93(3):227–234. https://doi.org/10.1097/opx.0000000000000796.
5. Honavar SG. The burden of uncorrected refractive error. *Indian J Ophthalmol*. May 2019;67(5):577–578. https://doi.org/10.4103/ijo.IJO_762_19.
6. Dai W, Tham Y-C, Chee M-L, et al. Falls and recurrent falls among adults in A multi-ethnic asian population: the Singapore Epidemiology of eye diseases study. *Sci Rep*. 2018/05/15 2018;8(1):7575. https://doi.org/10.1038/s41598-018-25894-8.
7. Lamoreux EL, Chong E, Wang JJ, et al. Visual impairment, causes of vision loss, and falls: the Singapore Malay eye study. *Invest Ophthalmol Vis Sci*. 2008;49(2):528–533. https://doi.org/10.1167/iovs.07-1036.
8. Mehta J, Czanner G, Harding S, Newsham D, Robinson J. Visual risk factors for falls in older adults: a case-control study. *BMC Geriatr*. 2022/02/17 2022;22(1):134. https://doi.org/10.1186/s12877-022-02784-3.
9. Organization WH. *Report of the 2030 Targets on Effective Coverage of Eye Care*. 2022.
10. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA Network*. Dec 13 2016;316(22):2402–2410. https://doi.org/10.1001/jama.2016.17216.
11. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA Network*. Dec 12 2017;318(22):2211–2223. https://doi.org/10.1001/jama.2017.18152.
12. Grassmann F, Mengelkamp J, Brandl C, 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*. Sep 2018;125(9):1410–1420. https://doi.org/10.1016/j.ophtha.2018.02.037.
13. *World Report on Vision*. Geneva: World Health Organization; 2019.
14. Peng Y, Dharssi S, Chen Q, et al. DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. *Ophthalmology*. Apr 2019;126(4):565–575. https://doi.org/10.1016/j.ophtha.2018.11.015.
15. Tham Y, Goh JHL, Anees A, et al. Detecting visually significant cataract using retinal photograph-based deep learning. *Nature Aging*. 2022. https://doi.org/10.1038/s43587-022-00171-6.
16. Tham YC, Anees A, Zhang L, et al. Referral for disease-related visual impairment using retinal photograph-based deep learning: a proof-of-concept, model development study. *Lancet Digit Health*. Jan 2021;3(1):e29–e40. https://doi.org/10.1016/S2589-7500(20)30271-5.
17. Liu H, Li L, Wormstone IM, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. *JAMA Ophthalmol*. Dec 1 2019;137(12):1353–1360. https://doi.org/10.1001/jamaophthalmol.2019.3501.
18. Chun J, Kim Y, Shin KY, et al. Deep learning-based prediction of refractive error using photorefraction images captured by a smartphone: model development and validation study. *JMIR Med Inform*. May 5 2020;8(5):e16225. https://doi.org/10.2196/16225.
19. Shi Z, Wang T, Huang Z, Xie F, Song G. A method for the automatic detection of myopia in Optos fundus images based on deep learning. *Int J Numer Method Biomed Eng*. 2021;37(6):e3460. https://doi.org/10.1002/cnm.3460.
20. Tan TE, Anees A, Chen C, 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*. May 2021;3(5):e317–e329. https://doi.org/10.1016/S2589-7500(21)00055-8.
21. Varadarajan AV, Poplin R, Blumer K, et al. Deep learning for predicting refractive error from retinal fundus images. *Invest Ophthalmol Vis Sci*. Jun 1 2018;59(7):2861–2868. https://doi.org/10.1167/iovs.18-23887.
22. Xu D, Ding S, Zheng T, et al. Deep learning for predicting refractive error from multiple photorefraction images. *Biomed Eng Online*. Aug 8 2022;21(1):55. https://doi.org/10.1186/s12938-022-01025-3.
23. Yang D, Li M, Li W, et al. Prediction of refractive error based on ultrawide field images with deep learning models in myopia patients. *Front Med*. 2022;9:834281. https://doi.org/10.3389/fmed.2022.834281.
24. Yang Y, Li R, Lin D, et al. Automatic identification of myopia based on ocular appearance images using deep learning. *Ann Transl Med*. Jun 2020;8(11):705. https://doi.org/10.21037/atm.2019.12.39.
25. Yoo TK, Ryu IH, Kim JK, Lee IS. Deep learning for predicting uncorrected refractive error using posterior segment optical coherence tomography images. *Eye*. 2022;36(10):1959–1965. https://doi.org/10.1038/s41433-021-01795-5.
26. Zou H, Shi S, Yang X, et al. Identification of ocular refraction based on deep learning algorithm as a novel retinoscopy method. *Biomed Eng Online*. 2022;21(1):87. https://doi.org/10.1186/s12938-022-01057-9.
27. Atchison DA, Pritchard N, Schmid KL, Scott DH, Jones CE, Pope JM. Shape of the retinal surface in emmetropia and myopia. *Invest Ophthalmol Vis Sci*. Aug 2005;46(8):2698–2707. https://doi.org/10.1167/iovs.04-1506.
28. Chang L, Pan CW, Ohno-Matsui K, et al. Myopia-related fundus changes in Singapore adults with high myopia. *Am J Ophthalmol*. Jun 2013;155(6):991–999 e1. https://doi.org/10.1016/j.ajo.2013.01.016.
29. Shimada N, Ohno-Matsui K, Nishimuta A, Tokoro T, Mochizuki M. Peripapillary changes detected by optical coherence tomography in eyes with high myopia. *Ophthalmology*. Nov 2007;114(11):2070–2076. https://doi.org/10.1016/j.ophtha.2007.01.016.
30. Ostrin LA, Yuzuriha J, Wildsoet CF. Refractive error and ocular parameters: comparison of two SD-OCT systems. *Optom Vis Sci*. Apr 2015;92(4):437–446. https://doi.org/10.1097/OPX.0000000000000559.
31. Oner V, Tas M, Turkcu FM, Alakus MF, Iscan Y, Yazici AT. Evaluation of peripapillary retinal nerve fiber layer thickness of myopic and hyperopic patients: a controlled

study by Stratus optical coherence tomography. *Curr Eye Res*. Jan 2013;38(1): 102–107. https://doi.org/10.3109/02713683.2012.715714.
32. Jonas JB, Bikbov MM, Wang YX, Jonas RA, Panda-Jonas S. Anatomic peculiarities associated with axial elongation of the myopic eye. *J Clin Med*. Feb 7 2023;12(4). https://doi.org/10.3390/jcm12041317.
33. Thambawita V, Strumke I, Hicks SA, Halvorsen P, Parasa S, Riegler MA. Impact of image resolution on deep learning performance in endoscopy image classification: an experimental study using a large dataset of endoscopic images. *Diagnostics*. Nov 24 2021;11(12). https://doi.org/10.3390/diagnostics11122183.
34. Das T, Takkar B, Sivaprasad S, et al. Recently updated global diabetic retinopathy screening guidelines: commonalities, differences, and future possibilities. *Eye*. 2021/10/01 2021;35(10):2685–2698. https://doi.org/10.1038/s41433-021-01572-4.