Clinical and Experimental Optometry

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tceo20

Real-world evaluation of RetCAD deep-learning system for the detection of referable diabetic retinopathy and age-related macular degeneration

Joshua R Taylor, Jocelyn Drinkwater, David Cordeiro Sousa, Vaibhav Shah & Angus W Turner

To cite this article: Joshua R Taylor, Jocelyn Drinkwater, David Cordeiro Sousa, Vaibhav Shah & Angus W Turner (12 Aug 2024): Real-world evaluation of RetCAD deep-learning system for the detection of referable diabetic retinopathy and age-related macular degeneration, Clinical and Experimental Optometry, DOI: 10.1080/08164622.2024.2385565

To link to this article: https://doi.org/10.1080/08164622.2024.2385565

Published online: 12 Aug 2024.

CLINICAL AND EXPERIMENTAL OPTOMETRY
https://doi.org/10.1080/08164622.2024.2385565

RESEARCH

Real-world evaluation of RetCAD deep-learning system for the detection of referable diabetic retinopathy and age-related macular degeneration

Joshua R Taylor[a], Jocelyn Drinkwater[b], David Cordeiro Sousa[c], Vaibhav Shah[a] and Angus W Turner[a]

[a]Lions Outback Vision, Lions Eye Institute, Nedlands, Western Australia, Australia; [b]Centre for Ophthalmology and Visual Science, University of Western Australia, Crawley, Western Australia, Australia; [c]Vitreoretinal Unit, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia

ABSTRACT

Clinical Relevance: The challenges of establishing retinal screening programs in rural settings may be mitigated by the emergence of deep-learning systems for early disease detection.

Background: Deep-learning systems have demonstrated promising results in retinal disease detection and may be particularly useful in rural settings where accessibility remains a barrier to equitable service provision. This study aims to evaluate the real-world performance of Thirona RetCAD for the detection of referable diabetic retinopathy and age-related macular degeneration in a rural Australian population.

Methods: Colour fundus images from participants with known diabetic retinopathy or age-related macular degeneration were randomly selected from ophthalmology clinics in four rural Australian centres. Grading was confirmed retrospectively by two retinal specialists. RetCAD produced a quantitative measure (0–100) for DR and AMD severity. The area under the ROC curve (AUC) was calculated. Sensitivity, specificity, and positive and negative predictive values were calculated at a pre-defined cut-point of ≥50.

Results: A total of 150 images from 82 participants were included. The mean age (SD) was 64.0 (12.8) years. Seventy-nine (52.7%) eyes had evidence of referable DR, while 54 (36.0%) had evidence of referable AMD. The AUC for referable DR detection was 0.971 (95% CI 0.950–0.936) with a sensitivity of 86.1% (76.8%–92.0%) and a specificity of 91.6% (82.8%–96.1%) at the pre-defined cut-point. Using the Youden Index method, the optimal cut-point was 41.2 (sensitivity 93.7%, specificity 90.1%). The AUC for the detection of referable AMD was 0.880 (0.824–0.936). At the pre-defined cut-point sensitivity was 88.9% (77.8%–94.8%) and specificity was 66.7% (56.8%–75.3%). The optimal cut-point was 52.6 (sensitivity 87.0%, specificity 75.0%).

Conclusion: RetCAD is comparable with but does not outperform equivalent deep-learning systems for retinal disease detection. RetCAD may be suitable as an automated screening tool in a rural Australian setting.

ARTICLE HISTORY
Received 21 December 2023
Revised 6 July 2024
Accepted 23 July 2024

KEYWORDS
Artificial intelligence; diabetic retinopathy; macular degeneration; rural population

Introduction

Diabetic retinopathy (DR) is a significant and growing public health concern, recognised as a leading cause of blindness amongst working-age adults worldwide.[1] The prevalence of DR amongst those living with diabetes is currently estimated to be 23.7% in non-Indigenous Australians and 30.2% in Indigenous Australians.[2] The growing burden of retinal disease is far from limited to the effects of DR, however. Age-related macular degeneration (AMD) is a leading cause of irreversible blindness in elderly populations and has a weighted prevalence of 14.8% and 13.8% amongst non-Indigenous and Indigenous Australians with visual impairment, respectively.[3]

The value of early detection and treatment of retinal disease through the use of screening programs, at an individual and population level, is well established.[4] However, with the projected increases in retinal disease prevalence, particularly in rural and remote areas that are typically underserviced, the provision of adequate screening models represents a growing challenge in health care.

Rapid advances and uptake of automated retinal imaging analysis through the use of artificial intelligence algorithms have mounting potential to overcome the challenges faced in rural population screening.[5–7] Deep-learning systems, a branch of artificial intelligence well suited to image analysis due to their ability to recognise abnormalities with increasing accuracy, have become a focus for the progression of medical artificial intelligence applications.[8] Existing deep-learning system validation studies have demonstrated better performance and potential for reduced screening costs and improved workflow compared to traditional retinal screening processes.[9–11]

Unfortunately, generalisability is a known limitation of deep-learning systems. A growing body of evidence suggests that systems trained on non-diverse and non-representative data may contribute to biased algorithms and creates the risk of inaccurate performance and misdiagnosis in certain population subgroups.[12,13] A recent systematic review by Arora et al.[14] specifically highlights geographical location as an attribute being at the particular risk of harm from under-representation in training datasets.[14] Therefore, rigorous validation in intended settings and subpopulations is paramount prior to widespread implementation.[5,14]

Within this paper, the term 'real-world' is utilised to imply validation using routinely collected health data used for medical assessment and diagnosis, within a specific population sub-group, in which a retinal disease deep-learning system might be employed for screening purposes.[15] There are

CONTACT Angus W Turner angus.turner@uwa.edu.au

apparently only two validation studies that have analysed the performance of retinal disease deep-learning systems in a real-world Western Australian population.[7,16]

Thirona RetCAD (Thirona, Nijmegen, The Netherlands) is a commercially available Therapeutic Goods Administration (TGA) approved medical device software that incorporates a deep learning framework for the detection of DR, AMD and glaucoma from colour fundus images.[17] The software receives the fundus image as input and produces several outputs including an image quality assessment, heatmaps of areas of abnormality and a score for the severity of the ocular diseases. There has been no formal analysis of the Thirona RetCAD deep-learning system in a rural Australian population.

The purpose of this study is to apply RetCAD to a high-risk rural population with either DR or AMD using real-world data to analyse its performance in referable DR and AMD detection and consider its potential for implementation as part of an automated retinal screening program. Furthermore, the study aims to apply evidence-based threshold selection methods to the data to propose disease-specific optimal cut-points for referral.

Materials and methods

Ethics approval was obtained from the University of Western Australia Human Ethics Committee in accordance with the requirements of the National Statement on Ethical Conduct in Human Research.

Single-field colour fundus 45-degree photographs were captured from consecutive patients attending rural ophthalmology clinics between January 2020 and June 2022. Routine retinal imaging was performed by medical practitioners trained in the use of Topcon Maestro Optical Coherence Tomography devices. The locations of the rural clinics were Port Hedland, Karratha, Kalgoorlie and Albany in Western Australia. Based on the real-world nature of the clinics, the pupils of participants were only dilated as necessary.

Images from patients with a pre-existing diagnosis of DR (any) or AMD (any) or both were deidentified and placed into the study dataset. Images were excluded if DR or AMD were unclassifiable due to poor image quality or media opacity. The RetCAD software (v1.3.1) was functionally integrated into the imaging database routinely used for the rural clinics for the purpose of the study. The deep-learning system was run on included images, and the reported outcomes were documented. The RetCAD severity score is reported as a value between 0.00 and 100.00. The algorithm is designed such that those with a score ≥ 50.00 should be referred for further testing, though there is limited evidence regarding the rationale for this cut-off point.

Each colour fundus photograph was individually graded by two retinal specialists (DS and VS) for the presence of referable DR, referable AMD, or both. Discrepancies were analysed by a third ophthalmologist (AT) for arbitration. Referable DR was defined as the presence of moderate non-proliferative DR or worse based on the International Clinical Diabetic Retinopathy (ICDR) severity scale,[18] meaning evidence of DR with more than just microaneurysms (i.e. any intraretinal haemorrhages, venous beading, prominent intraretinal microvascular abnormalities, or signs of proliferative DR). Referable AMD was defined as intermediate AMD or worse as per Beckman classification in Table 1 (i.e. presence of large drusen (>125 µm) and/or AMD pigmentary abnormalities).[19] These definitions were chosen for consistency with existing Thirona RetCAD validation papers and reflect routine clinical practice in Australia, which allows for effective deep-learning system performance comparison in this real-world rural Australian population.[17,20–22]

Data are presented as mean (SD) or number (%). Area under the receiver operating characteristics curve (AUC) were calculated. Sensitivity, specificity, positive predictive value, negative predictive value and diagnostic accuracy for referable disease were calculated at the pre-defined cut-point of ≥50. A recommended threshold value for each disease was identified using the Youden Index method, where the optimal cut-point is the point maximising the Youden function, the difference between true positive rate and false positive rate over all possible cut-point values. This method was chosen for its widespread use for threshold selection in medical and biological sciences.[23] An additional cut-point was identified at the point where sensitivity and specificity were approximately equal.[24]

Results

The demographic information of study participants is summarised in Table 2. One hundred and fifty eyes, from 82 people, were randomly selected for analysis. Forty-two (51%) of participants were female. Participant age was normally distributed with a mean (SD) age of 64.0 (12.8) years. Following grading by the two retinal specialists, 36 images required arbitration by the third grader. Seventy-nine eyes (52.7%) were graded as having evidence of referable DR,

Table 1. Summary of non-referable and referable disease characteristics for DR and AMD, as per ICDR and Beckman classification, respectively.

	Observable findings	Non-referable/referable
DR classification		
No retinopathy	No abnormalities	Non-referable
Mild non-proliferative DR	Microaneurysms only	
Moderate non-proliferative DR	More than just microaneurysms but less than severe NPDR	Referable
Severe non-proliferative DR	Any of the following: – >20 intraretinal haemorrhages in each of 4 quadrants – Venous beading in ≥2 quadrants – Prominent IRMA in ≥1 quadrant	
Proliferative DR	Neovascularisation and/or vitreous or pre-retinal haemorrhage	
AMD classification		
Normal ageing changes	Small drusen <63 µm and no AMD pigmentary abnormalities	Non-referable
Early AMD	Medium drusen >63 µm and ≤125 µm and no AMD pigmentary abnormalities	
Intermediate AMD	Large drusen >125 µm and/or AMD pigmentary abnormalities	Referable
Late AMD	Neovascular AMD and/or geographic atrophy	

DR = Diabetic retinopathy; AMD = age-related macular degeneration; IRMA = intraretinal microvascular abnormalities.

Table 2. Baseline characteristics of study participants.

Characteristic	
Male/Female, n	40/42
Age, years, mean (SD)	64.0 (12.8)
Moderate DR or worse, n (%)	79 (52.7)
Intermediate AMD or worse, n (%)	54 (36.0)
Neither moderate DR or worse nor intermediate AMD or worse, n (%)	17 (11.3)

DR = Diabetic retinopathy; AMD = age-related macular degeneration.
Included images/eyes = 150.

while 54 eyes (36.0%) had referable AMD. Seventeen eyes were graded as having neither referable DR nor referable AMD. The mean (SD) quality score for the images was 51.7 (28.8) with a range of 0.2–99.7.

Detailed analysis results of referable DR and AMD detection at various cut-points are reported in Tables 3 and 4. The AUC for referable DR detection was 0.971 (95% CI 0.950–0.936, $p < 0.001$) (Figure 1). At the pre-defined cut-point of ≥50, RetCAD detected referable DR with a sensitivity of 86.1% (76.8%–92.0%), a specificity of 91.6% (82.8%–96.1%) and a diagnostic accuracy of 88.7% (82.6%–92.8%). The optimal cut-point through Youden index analysis was ≥41.2, producing a sensitivity of 93.7% (86.0%–97.3%) and a specificity of 90.1% (81.0%–95.1%). The diagnostic accuracy was 92.0% (84.2%–95.4%).

The AUC for referable AMD detection was 0.880 (0.824–0.936, $p < 0.001$) (Figure 2). At the pre-defined cut-point of ≥50, referable AMD detection sensitivity was 88.9% (77.8%–94.8%) and specificity was 66.7% (56.8%–75.3%). Diagnostic accuracy was 74.7% (67.2%–81.0%). Applying the Youden index, yielded an optimal cut-point of ≥52.6. At this point, referable disease detection sensitivity was 87.0% (75.6%–93.6%) and specificity 75.0% (65.5%–82.6%), with a diagnostic accuracy of 79.3% (72.2%–85.0%).

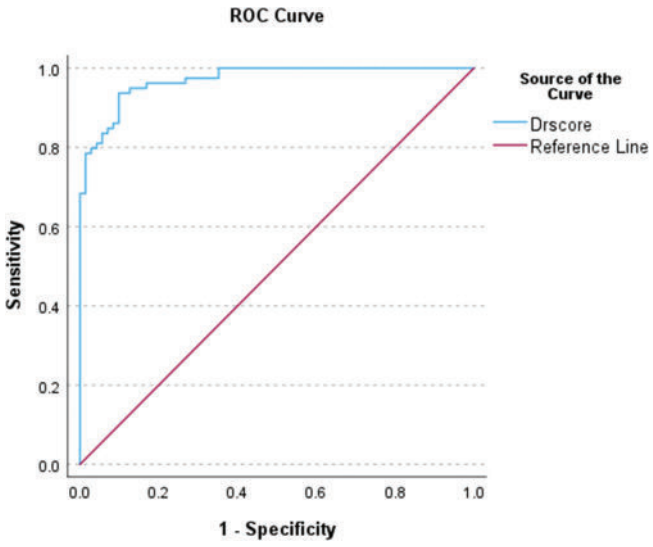

Figure 1. Area under the ROC curve (AUC) for referable DR detection.

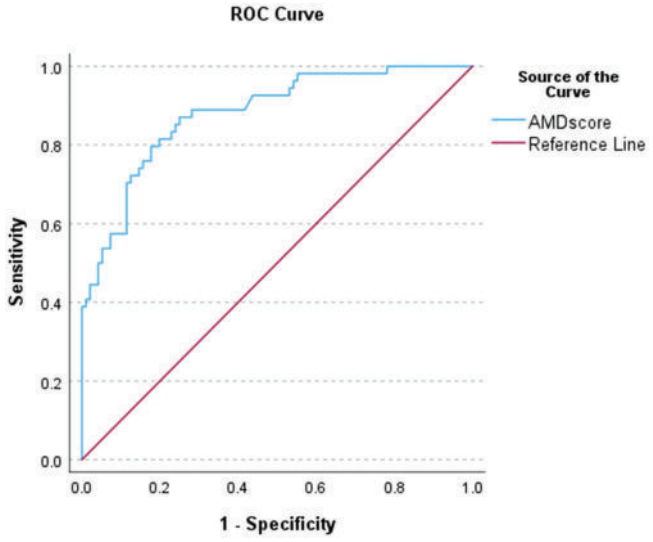

Figure 2. Area under the ROC curve (AUC) for referable AMD detection.

Discussion

Using retinal images acquired in real-world rural eye clinics in Western Australia, the RetCAD software was able to identify those with referable DR with a reasonable level of

Table 3. Thirona RetCAD performance for referable DR detection at various cut-points.

Cut-point for referable DR	41.2[†]	44.7[‡]	50[§]
Sensitivity (%)	93.7 (86.0-97.3)	89.9 (81.3-94.8)	86.1 (76.8-92.0)
Specificity (%)	90.1 (81.0-95.1)	90.1 (81.0-95.1)	91.6 (82.8-96.1)
PPV (%)	91.4 (83.2-95.8)	91.0 (82.6-95.6)	91.9 (83.4-96.2)
NPV (%)	92.8 (84.1-96.9)	88.9 (79.6-94.3)	85.5 (75.9-91.7)
Diagnostic accuracy (%)	92.0 (86.5-95.4)	90.0 (84.2-93.9)	88.7 (82.6-92.8)

DR = Diabetic retinopathy; PPV = positive predictive value, NPV = negative predictive value.
[†]Youden index cut-point.
[‡]Sensitivity = specificity cut-point.
[§]Pre-defined cut-point.

Table 4. Thirona RetCAD performance for referable AMD detection at various cut-points.

Cut-point for referable AMD	50[†]	52.6[‡]	56.7[§]
Sensitivity (%)	88.9 (77.8-94.8)	87.0 (75.6-93.6)	79.6 (67.1-88.2)
Specificity (%)	66.7 (56.8-75.3)	75.0 (65.5-82.6)	80.2 (71.1-87.0)
PPV (%)	60.0 (49.1-70.0)	66.2 (54.6-76.1)	69.4 (57.0-79.4)
NPV (%)	91.4 (82.5-96.0)	91.1 (82.8-95.6)	87.5 (79.0-92.9)
Diagnostic accuracy (%)	74.7 (67.2-81.0)	79.3 (72.2-85.0)	80.0 (72.9-85.6)

AMD = Age-related macular degeneration; PPV = positive predictive value, NPV = negative predictive value.
[‡]Youden index cut-point.
[§]Sensitivity = specificity cut-point.
[†]Pre-defined cut-point.

accuracy; however, the deep-learning system did not work as well for referable AMD. In recent years, multiple artificial intelligence systems for retinal disease detection have become available, including the approval of several systems by the Therapeutic Goods Administration. Validation of these systems for DR detection has estimated a sensitivity range between 87% and 98% and a specificity range between 86% and 97%.[6]

Similarly, AMD detection estimates for sensitivity are between 90% and 100% and specificity between 70% and 93%.[6] The results of this study indicate that RetCAD performs comparably with most other TGA-approved deep-learning systems for retinal disease detection in a real-world population, at the identified Youden Index cut-point.

The RetCAD software has been validated in several studies, predominately on publicly available fundus image datasets.[17,20–22] González-Gonzalo et al.[20] applied RetCAD (v.1.3.0) to Messidor, a publicly available dataset of 1,200 colour fundus images. Two-thirds of these images were taken on dilated pupils. The detection of referable DR yielded an AUC of 0.975 with a sensitivity of 92.0% and a specificity of 92.1%. A second image dataset, AREDS (133,821 images), was used to assess referable AMD detection, yielding an AUC of 0.927 with a sensitivity of 85.8% and a specificity of 86.0%. A joint detection analysis was also performed on a combined DR-AMD dataset.

Sánchez-Gutiérrez et al.[17] assessed the performance of RetCAD (v.1.3.1) for DR detection, applied to a dataset of 7,195 non-mydriatic fundus images against a reference standard set by a human expert. The study used a cut-point of ≥50 to indicate referable DR. The analysis yielded an AUC of 0.988, a sensitivity of 90.5% and a specificity of 97.1%. The study predicted a workload reduction of 96% at a cost of six false negatives. Skevas et al.[21] ran a prospective study applying RetCAD (v.1.3.1) to 1,245 fundus photos captured on patients attending an ophthalmology day clinic. Images were graded in parallel by an expert reference examiner. All images were taken without pupil dilation; however, only images with clear media allowing sharp fundus photography were included. Again, the evaluation used a cut-point of ≥50 to indicate referable DR or AMD. The study demonstrated an AUC of 0.961 and 0.964 for DR and AMD detection, respectively. Sensitivity and specificity values were 83.9/93.3% for DR and 98.2/79.1% for AMD.

The most recent evaluation by Meredith et al.[22] applied RetCAD (v. 2.1.0) to a dataset of 9,817 fundus images captured from consenting patients attending routine Diabetic Eye Screening Programme (DESP) appointments in London, UK. At the pre-defined cut-point for referable DR detection, the sensitivity was 95.4% and specificity 92.0%. The AUC was 0.979. Given the use of an updated version of RetCAD, utilising different referable disease cut-points, this study has been excluded from the comparisons here.

A summary of results compared to the existing validation studies at the pre-defined cut-point of 50 is outlined in Table 5. Compared to existing studies, the results obtained in this study indicate a comparable detection capability of referable DR in a real-world rural population.[17,20–22] However, Sánchez-Gutiérrez et al.[17] reported a higher specificity (91.6% vs. 97.1%).

For referable AMD detection, compared to Skevas et al.,[21] the present study produced significantly lower sensitivity and specificity values at the pre-defined cut-point (88.9% vs. 98.2%, 66.7% vs. 79.1%). Reasons for this discrepancy may include a smaller sample size in the present study. In comparison with the AREDS arm of González-Gonzalo et al.,[20] only specificity was lower to a statistically significant degree (66.7% vs. 86.0%).

The significant range of quality score (0.2–99.7) of the images, despite the removal of poor-quality images for the purpose of gradability, indicates the automated image quality metric may not correspond with true gradability. It is difficult to predict the image features that the algorithm defines as low quality, and whether these features can be effectively navigated by human graders in a screening and referral setting. Further analysis of diagnostic performance adjusted for image quality score is warranted.

According to the Australian Department of Health Population Based Screening Framework,[25] adapted from the original screening principles framework of the World Health Organisation, there are several key criteria in which a proposed screening test is required to meet before implementation at the population level. These include a test that is both sensitive and specific, has relatively high positive and negative predictive values, is validated and safe, and is acceptable to the target population.[25]

One focus of RetCAD as a proposed screening tool in this study is the trade-off between sensitivity and specificity. It is important to consider the real-world consequences of accepted sensitivity and specificity values in screening test implementation. The ramifications of over-referral include excess healthcare costs, increased specialist wait periods and even the potential for public mistrust in the retinal screening and other national screening programs. Prioritisation of higher specificity in screening program implementation ensures the minimisation of false positives and therefore less unnecessary specialist referrals. Alternatively, emphasis on higher sensitivity ensures lower rates of false negatives and therefore fewer patients being missed who require specialist assessment.

An ideal screening tool maximises sensitivity and specificity to as close to maximum (100%) as possible; however, due

Table 5. Comparison to existing RetCAD validation studies at pre-defined cut-point of ≥50.

		This study	González-Gonzalo et al.[20]	Sánchez-Gutiérrez et al.[17]	Skevas et al.[21]
DR	n	150	1,200	7,195	1,245
	AUC	0.971	0.975	0.988	0.961
	Sensitivity (%)	86.1 (76.8–92.0)	92.0 (89.3–97.2)	90.5†	83.9†
	Specificity (%)	91.6 (82.8–96.1)	92.1 (88.6–95.2)	97.1†	93.3†
AMD	n	150	133,821		1,245
	AUC	0.880	0.927		0.964
	Sensitivity (%)	88.9 (77.8–94.8)	85.8 (84.6–86.2)		98.2†
	Specificity (%)	66.7 (56.8–75.3)	86.0 (85.7–87.4)		79.1†

DR = diabetic retinopathy; AMD = age-related macular degeneration; AUC = Area under receiver operating characteristic curve.
†Confidence intervals not reported.

to the difficult nature of perfect diagnostic accuracy, a trade-off in sensitivity vs. specificity may be required in screening program design. One of the benefits of the RetCAD deep-learning system for retinal disease screening is that the cut-point value for referable disease can be manipulated based on the needs of the program. For both retinal disease groups, the sensitivity is higher than the specificity at the Youden index cut-point, although the difference is not significant in both cases. Implementing the Youden index cut-points into a screening program would therefore emphasise lower false negative results to ensure capture and referral for more patients with referable disease, at the expense of higher rates of unnecessary referrals.

The authors suggest an implemented screening program cut-off score of 45 for referable DR and 55 for referable AMD. There are two reasons for which these values have been suggested. First, for ease of consistency across screening centres, simple rounded values promote easier recall and minimise resistance to protocol adherence.[26] Second, in both retinal disease groups these values narrow the difference between sensitivity and specificity by reducing the former and increasing the latter. This is deemed an acceptable trade-off because the primary aim is to detect the disease in its moderate/intermediate state, prior to approaching vision-threatening severity. It is anticipated that in most cases the small excess of false negatives will be captured and referred at their next screening appointment before any significant disease progression has occurred. Higher specificity also reduces the wait times and burden on the health system by minimising unnecessary specialist referrals.

The strengths of the present study include the use of real-world data from a rural Australian setting with the use of colour fundus images obtained from consecutive DR/AMD patients attending rural ophthalmology clinics. This is apparently the first study to analyse RetCAD performance at various cut-points and use evidence-based AUC analysis to make logical suggestions for optimal referable threshold scores.

There are relevant limitations of this study, including a relatively small sample size compared to existing validation studies, and the retrospective nature of the study possibly misrepresenting the challenges of prospective cohort screening. The inclusion of images from patients with pre-existing diseases only may not represent the true nature of a screening program within a target population, thus creating a biased sample. Furthermore, the study population had a higher prevalence of both referable DR and referable AMD compared to the general population, which could result in a higher sensitivity and lower specificity.[27] However, given the purposes of the analysis of the ability of RetCAD to detect disease severity warranting referral, the findings of this study are still useful in advising screening program suitability.

Further performance analysis is warranted in rural populations with risk factors for retinal disease, without a pre-existing diagnosis. Unfortunately, data on those excluded due to poor-quality images was not available. Furthermore, the use of de-identified clinic data means that available information for sub-group analysis was limited, which may have enriched the evaluation.

Further population sub-group analysis is likely to be beneficial given the known limitation in the performance of deep-learning systems when applied to populations external to their development.[5,14] Image quality analysis could further guide decision-making by healthcare workers in regard to the need for environment manipulation, dilation and future screening intervals at the time of image capture.

Conclusion

This study demonstrates that the Thirona RetCAD deep-learning system for referable DR detection performs comparably to other validation studies when applied to a real-world rural Australian population. However, the deep-learning system may be less accurate in this setting for referable AMD detection, compared to other studies. Early inter-tool comparisons suggest that RetCAD competes with but does not outperform other equivalent deep-learning systems for retinal disease detection. Further prospective analysis is required on population sub-groups and calibration for image quality.

Acknowledgments

The authors received no financial support for the research, authorship, and/or publication of this article. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Joshua R Taylor http://orcid.org/0000-0001-8659-5743
Jocelyn Drinkwater http://orcid.org/0000-0001-5155-5469
David Cordeiro Sousa http://orcid.org/0000-0002-7680-0325
Angus W Turner http://orcid.org/0000-0001-5949-7451

References

1. Cheung N, Mitchell P, Wong T. Diabetic retinopathy. Lancet 2010; 376: 124–136. doi: 10.1016/S0140-6736(09)62124-3.
2. Chia MA, Taylor JR, Stuart KV et al. Prevalence of diabetic retinopathy in Indigenous and non-Indigenous Australians: a systematic review and meta-analysis. Ophthalmology 2023; 130: 56–67. doi: 10.1016/j.ophtha.2022.07.024.
3. Keel S, Xie J, Foreman J et al. Prevalence of age-related macular degeneration in Australia: the Australian National Eye Health survey. JAMA Ophthalmol 2017; 135: 1242–1249. doi: 10.1001/jamaophthalmol.2017.4182.
4. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. BMJ Open 2014; 4: e004015. doi: 10.1136/bmjopen-2013-004015.
5. Chia MA, Turner AW. Benefits of integrating telemedicine and artificial intelligence into outreach eye care: stepwise approach and future directions. Front Med 2022; 9. doi: 10.3389/fmed.2022.835804.
6. Kiburg KV, Turner A, He M. Telemedicine and delivery of ophthalmic care in rural and remote communities: drawing from Australian experience. Clin Exp Ophthalmol 2022; 50: 793–800. doi: 10.1111/ceo.14147.
7. Chia MA, Hersch F, Sayres R et al. Validation of a deep learning system for the detection of diabetic retinopathy in Indigenous Australians. Br J Ophthalmol 2024; 108: 268–273. doi: 10.1136/bjo-2022-322237.

8. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436–444. doi: 10.1038/nature14539.
9. Islam MM, Yang HC, Poly TN et al. Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: a systematic review and meta-analysis. Comput Methods Programs Biomed 2020; 191: 105320. doi: 10.1016/j.cmpb.2020.105320.
10. Bellemo V, Lim G, Rim TH et al. Artificial intelligence screening for diabetic retinopathy: the real-world emerging application. Curr Diab Rep 2019; 19. doi: 10.1007/s11892-019-1189-3.
11. Nielsen KB, Lautrup ML, Andersen JKH et al. Deep learning-based algorithms in screening of diabetic retinopathy: a systematic review of diagnostic performance. Ophthalmol Retina 2019; 3: 294–304. doi: 10.1016/j.oret.2018.10.014.
12. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mac Intell 2019; 1: 389–399. doi: 10.1038/s42256-019-0088-2.
13. Zou J, Schiebinger L. Ensuring that biomedical AI benefits diverse populations. EBioMedicine 2021; 67. doi: 10.1016/j.ebiom.2021.103358.
14. Arora A, Alderman JE, Palmer J et al. The value of standards for health datasets in artificial intelligence-based applications. Nat Med 2023; 29: 2929–2938. doi: 10.1038/s41591-023-02608-w.
15. Chodankar D. Introduction to real-world evidence studies. Perspect Clin Res 2021; 12: 171–174. doi: 10.4103/picr.picr_62_21.
16. Kanagasingam Y, Xiao D, Vignarajan J et al. Evaluation of artificial intelligence–based grading of diabetic retinopathy in primary care. JAMA Netw Open 2018; 1. doi: 10.1001/jamanetworkopen.2018.2665.
17. Sánchez-Gutiérrez V, Hernández-Martínez P, Muñoz-Negrete FJ et al. Performance of a deep learning system for detection of referable diabetic retinopathy in real clinical settings. arXiv Preprint arXIV:220505554. 2022.
18. Abràmoff MD, Folk JC, Han DP et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol 2013; 131: 351–357. doi: 10.1001/jamaophthalmol.2013.1743.
19. Ferris FL III, Wilkinson C, Bird A et al. Clinical classification of age-related macular degeneration. Ophthalmology 2013; 120: 844–851. doi: 10.1016/j.ophtha.2012.10.036.
20. Gonzalez-Gonzalo C, Sanchez-Gutierrez V, Hernandez-Martinez P et al. Evaluation of a deep learning system for the joint automated detection of diabetic retinopathy and age-related macular degeneration. Acta Ophthalmol 2020; 98: 368–377. doi: 10.1111/aos.14306.
21. Skevas C, Weindler H, Levering M et al. Simultaneous screening and classification of diabetic retinopathy and age-related macular degeneration based on fundus photos—a prospective analysis of the RetCAD system. Int J Ophthalmol 2022; 15: 1985–1993. doi: 10.18240/ijo.2022.12.14.
22. Meredith S, Grinsven MV, Engelberts J et al. Performance of an artificial intelligence automated system for diabetic eye screening in a large English population. Diabet Med 2023; 40. doi: 10.1111/dme.15055.
23. Yin J, Tian L. Joint confidence region estimation for area under ROC curve and Youden index. Stat Med 2013; 33: 985–1000. doi: 10.1002/sim.5992.
24. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochem Med 2016; 26: 297–307. doi: 10.11613/BM.2016.034.
25. Department of Health. Population based screening framework. Australian Government; 2018. 2–13 p.
26. Weigl B, Boyle D, Santos TDL et al. Simplicity of use: a critical feature for widespread adoption of diagnostic technologies in low-resource settings. Expert Rev Med Devices 2009; 6: 461–464. doi: 10.1586/erd.09.31.
27. Murad MH, Lin L, Chu H et al. The association of sensitivity and specificity with disease prevalence: analysis of 6909 studies of diagnostic test accuracy. Can Med Assoc J 2023; 195: 925–931. doi: 10.1503/cmaj.221802.