

Article

Not peer-reviewed version

---

# A Real-World Comparison of Three Deep-Learning Systems for Diabetic Retinopathy in Remote Australia

---

[Jocelyn J. Drinkwater](#) , [Qiang Li](#) , Kerry Woods , Emma Douglas , Mark Chia , Yukun Zhou , Steve Bartnik , [Yachana Shah](#) , Vaibhav Shah , Pearse A. Keane , [Angus W. Turner](#) \*

Posted Date: 13 October 2025

doi: 10.20944/preprints202510.0936.v1

Keywords: diabetic retinopathy; deep learning; First Nations peoples; rural health; diabetes mellitus



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# A Real-World Comparison of Three Deep-Learning Systems for Diabetic Retinopathy in Remote Australia

Jocelyn J. Drinkwater <sup>1</sup>, Qiang Li <sup>2</sup>, Kerry Woods <sup>2</sup>, Emma Douglas <sup>2</sup>, Mark Chia <sup>3,4</sup>, Yukun Zhou <sup>3,4</sup>, Steve Bartnik <sup>2</sup>, Yachana Shah <sup>2</sup>, Vaibhav Shah <sup>2</sup>, Pearse A. Keane <sup>3,4</sup> and Angus W. Turner <sup>1,2,\*</sup>

<sup>1</sup> Centre for Ophthalmology and Visual Science, The University of Western Australia, 2 Verdun St, Nedlands, Western Australia, Australia, 6009

<sup>2</sup> Lions Outback Vision, Lions Eye Institute, 2 Verdun St, Nedlands, Western Australia, Australia, 6009

<sup>3</sup> Moorfields Eye Hospital NHS Foundation Trust, 11-43 Bath St, London, EC1V 9EL, United Kingdom

<sup>4</sup> Institute of Ophthalmology, University College London, 9-11 Bath St, London, EC1V 9LF, United Kingdom

\* Correspondence: angus.turner@uwa.edu.au

## Abstract

**Background/objective:** Deep learning tools may improve access to screening for diabetic retinopathy, a leading cause of vision loss. Therefore, the aim was to prospectively compare the performance of three deep learning systems (DLSs); Google ARDA, Thirona RetCAD™, and EyRIS SELENA+ for detection of referable diabetic retinopathy (DR), in a real-world setting. **Methods:** Participants with diabetes presented to a mobile facility for DR screening in the remote Pilbara region of Western Australia, which has a high proportion of First Nations people. Sensitivity, specificity, and other performance indicators were calculated for each DLS, compared to grading by an ophthalmologist adjudication panel. Cochran's Q with post-hoc Dunn test assessed differences in DLS performance. **Results:** 188 colour fundus photographs were assessed; 39 images had referable DR, 135 had no referable DR and 14 images were ungradable. The sensitivity/specificity of ARDA was 100% (95% CI: 91.03-100%) / 94.81% (89.68-97.47%), RetCAD was 97.37% (86.50-99.53%) / 97.01% (92.58-98.83%) and SELENA+ was 91.67% (78.17-97.13%) / 80.80% (73.02-86.74%). DLS performance remained high in First Nations people. ARDA and RetCAD™ results were not statistically different to the ophthalmologist grading ( $p \geq 0.415$ ). **Conclusions:** In a real-world study comprising majority First Nations people, DLSs had high sensitivity and specificity for detecting referable DR. Implementation may augment DR screening rates, and with appropriate referral and treatment pathways may prevent vision loss and improve health equity.

**Keywords:** diabetic retinopathy; deep learning; First Nations peoples; rural health; diabetes mellitus

## 1. Introduction

Diabetic retinopathy (DR), a common diabetes complication, is a leading cause of vision loss [1,2] expected to affect 160 million people by 2045. [3] Timely detection of disease can prevent 98% of vision loss, hence routine annual or biennial DR screening is typically recommended. [4,5] However, half of First Nations Australians and one-quarter of other Australians did not meet screening recommendations in a national survey and screening rates were lower in remote areas [6] which have higher proportions of First Nations people and increasing diabetes prevalence. [7–9] A recent study reported a diabetes prevalence of 29% in remote First Nations Australians and this was >50% among those  $\geq 45$  years, [8] almost six times higher than the national prevalence. [7] Barriers to DR screening in these remote areas include a lower ophthalmology and optometry workforce, needing to travel significant distances to access eye care, [2] and in primary care, high staff turnover and lengthy waits

for screening results. [10] Deep learning systems (DLSs), a type of artificial intelligence, require minimal staff training and in conjunction with automated cameras may overcome these barriers to provide equitable screening access in remote regions and prevent vision loss.

DLSs can detect DR with sufficient accuracy, have potential for cost-saving and may be of particular value in remote settings with limited health service access. [11,12] In Australia, using a universal DLS-based DR screening program was estimated to save AU\$595.8 million over 40 years [11]. However, few DLSs for DR have been trained or validated on Indigenous peoples internationally. [13–16] Notably, Google Health's Automated Retinal Disease Assessment (ARDA) [17] has outperformed retinal specialist grading in First Nations Australians. [13] ARDA has been Conformité Européenne-marked, but Australian regulatory approval has not yet been sought. Thirona RetCAD [18] and EyRIS SELENA+ [19] have multiple regulatory approvals including Therapeutic Goods Administration approval in Australia, yet neither have been validated on First Nations people specifically. DLSs require validation in the population where they are to be used to ensure adequate performance, especially when comprised of minority populations not sufficiently represented in the training dataset for the DLS development. [20] Data distribution differences between training and validation datasets can potentially lead to a marked drop in DLS performance. [21,22] As such, DLS performance should be assessed in real-world settings, and for First Nations Peoples before implementation. No Australian study has directly compared the performance of multiple DLSs on the same population. A direct real-world comparison of multiple DLSs can be valuable to illustrate differences in performance so the most appropriate tool for that population can be implemented. Therefore, the aim of this study was to prospectively compare the performance of multiple DLSs, to detect referable DR in a remote Western Australian real-world setting.

## 2. Materials and Methods

A mobile service delivered by Lions Outback Vision in consultation with First Nations communities, provided DR screening to people living in the Pilbara region of Western Australia between February and August 2024 with facilities previously described. [23] Those who attended DR screening, with any diabetes type, were eligible to participate. Diabetes status was based on patient self-report.

Participants had single-field colour fundus photographs centered on the macula captured using an automated table-top camera, Topcon Maestro 2 (Topcon Healthcare, Tokyo, Japan), with images uploaded to a viewing platform (Topcon Harmony) where two integrated DLSs, Thirona RetCAD™ (v2.2.0) and EyRIS SELENA+ (v1.7.0) were used to synchronously grade retinal images at the point of care. Images were graded by Google ARDA retrospectively in batches, as software was unavailable to integrate ARDA into the viewing platform for real-time use. All three DLSs are based on convolutional neural networks. [24–26] There was no recalibration nor fine-tuning of these DLSs prior to use in the study, with a categorical result given by each, not a prediction score. SELENA+ output stated whether there was referable DR or non-referable DR, but did not give DR severity. Both RetCAD™ and ARDA outputs gave the DR severity. Due to the real-world nature of this research, participants' pupils were only dilated as necessary. To ensure adequate clinician oversight, when the DR screener was not a clinician, images were reviewed by an on-call doctor at the time of screening and on-the-spot telehealth consultation was provided for patients with referable disease, or at patient request.

The color fundus photographs were independently graded by two ophthalmologists (YS and SB) according to the The International Clinical Diabetic Retinopathy and Diabetic Macular Edema Disease Severity Scale. [27] Any images where there were discrepancies were independently adjudicated by a retinal specialist (VS). This grading formed the reference standard. Referable DR was defined as moderate non-proliferative diabetic retinopathy (NPDR) or worse. Vision-threatening DR was defined as severe non-proliferative DR or proliferative DR.

### 2.1. Development of the DLSs

ARDA was developed on approximately 130,000 images from the US and India and initially validated on approximately 10,000 images from EyePACS-1 and Messidor-2 datasets. [24] RETCAD™ has been validated on multiple datasets including Messidor, Messidor 2 and a private dataset according to their White Paper, yet the initial training set is not clear. [26] SELINA+ was developed in Singapore using 76,370 images from their national DR screening program and internally validated on 70,000 different photographs, also from the screening program. [25] External validation comprised 40,000 photographs from many countries, but mostly comprising people of Chinese ethnicity. [25]

## 2.2. Ethical Approval

The research project was supported by the Pilbara Aboriginal Health Alliance and approved by the Western Australian Aboriginal Health Ethics Committee (HREC1294). All participants included gave written informed consent. People with diabetes were invited to undergo the DLS-assisted DR screening as part of a clinical service regardless of whether they gave consent to use their data for research purposes.

## 2.3. Statistical Analysis

Sample size analysis showed that 173 images would be required, based on an expected sensitivity of 90% and specificity of 85%, estimated prevalence of referable DR of 20% and precision of  $\pm 10\%$ . [28] Data are presented as number (percentage) for categorical variables and mean  $\pm$  standard deviation for normally distributed continuous variables. To assess the performance of the DLSs sensitivity, specificity, diagnostic accuracy, negative predictive value (NPV) and positive predictive value (PPV), excluding images deemed ungradable by the DLSs, were calculated using ophthalmologists' grading as the reference standard. Additionally, the performance of the three DLSs was calculated classifying images deemed ungradable by the DLSs as referable DR. The performance of ARDA and RetCAD™ to assess vision-threatening DR was also assessed.

To determine if there was a significant difference between the results of the different DLSs and the reference standard, Cochran's Q, an extension of McNemar's test, was used. [29] Firstly, this was used to compare referable DR versus non-referable DR between the ophthalmology grading and the three DLSs and secondly comparing misclassification (true positives and true negatives versus false positives and false negatives) between the three DLSs. Where this was significant, a post-hoc Dunn test with Bonferroni correction identified significant differences between the different DR grading systems. Results with p-value or adjusted p-value  $< 0.05$  were considered significant.

## 3. Results

Across six communities, there were 94 people included in the study. Of these, 48 (51.1%) were male, 66 (70.2%) identified as First Nations and the mean  $\pm$  SD age was  $60.3 \pm 12.0$  years at the time of screening. Ophthalmologist grading of the 188 colour fundus photographs determined 135 (71.8%) images had no referable DR (121 (64.4%) with no DR and 14 (7.4%) with mild NPDR) and 39 (20.7%) images had referable DR (26 (13.8%) with moderate NPDR, 6 (3.2%) with severe NPDR and 7 (3.7%) with PDR). There were 14 (7.4%) ungradable images, which were excluded from analysis. There was substantial agreement between the two main graders ( $\kappa = 0.71$ ).

ARDA had 100% (95% CI: 91.03-100%) sensitivity overall and therefore 100% NPV, while maintaining specificity of 94.81% (89.68-97.47%) (see Table 1). RetCAD™ had high sensitivity and specificity of 97.37% (86.50-99.53%) and 97.01% (92.58-98.83%), respectively. SELINA+ had sensitivity of 91.67% (78.17-97.13%) and specificity of 80.80% (73.02-86.74%). ARDA had a high proportion of gradable images, giving a DR grade for all the images deemed gradable by the ophthalmologists, only four of the 188 images (2.1%) were ungradable. Some examples of differences in gradability of images by the ophthalmologists and the DLSs are shown in Figure 1. ARDA classified seven images as referable DR that were non-referable. There were two images with referable DR that RetCAD™ classified as non-referable and ungradable. RetCAD™ classified four

images as referable that were non-referable. The diagnostic accuracy for all DLSs was preserved or improved in First Nations Australians (Table 1). For non-First Nations, the sensitivity was 100% for all DLSs, however there were only two with referable DR (Table 1).

There was a significant difference between the DLSs overall for grading for referable DR vs non-referable DR, based on the Cochran's Q test ( $p < 0.001$ ). The ophthalmology grading (reference standard) was not statistically different to RetCAD™ and ARDA grading, nor was there a significant difference between RetCAD™ and ARDA grading (adjusted  $p \geq 0.415$ ). However, the ophthalmology grading, RetCAD™ and ARDA were significantly different to SELENA+ (adjusted  $p \leq 0.002$ ). Additionally, there was an overall difference in misclassification between the DLSs ( $p < 0.001$ ) with post-hoc tests showing no significant differences between RetCAD™ and ARDA but both of these were different to SELENA+ (adjusted  $p < 0.001$ ).

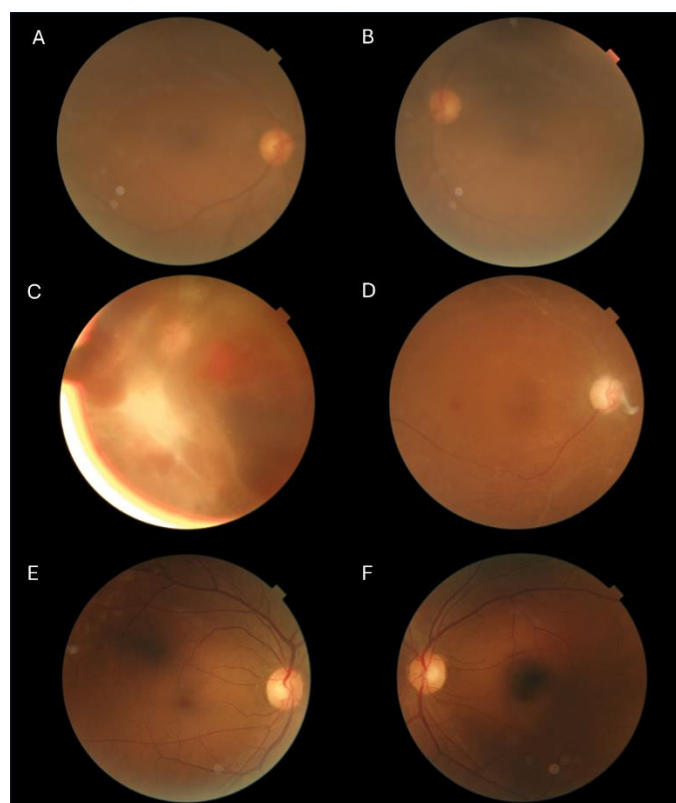
When including the ungradable images from the DLSs as referable DR, there was an increase in sensitivity and decrease in specificity and diagnostic accuracy in both RetCAD™ and SELENA+, as shown in Table 2. When assessing detection of vision-threatening DR in those with diabetes, ARDA had sensitivity of 92.31% (95% CI: 66.69-98.63%) and specificity of 97.52% (93.79-99.03%) and RetCAD™ results were 91.67% (64.61-98.51%) and 96.25% (92.06-98.27%), respectively.

**Table 1.** The performance of Google ARDA, Thirona RetCAD™ and EyRIS SELENA+ deep learning systems compared to ophthalmologist grading for referable diabetic retinopathy (DR)\* using 188 images taken on the Topcon Maestro2 camera .

Deep learning system:	Google ARDA	Thirona RetCAD™	EyRIS SELENA+	
People with diabetes	Number analysed	174/188 (92.6)	172/188 (91.5)	161/188 (85.6)
	Ungradable images (total)	4/188 (2.1)	9/188 (4.8)	24/188 (12.8)
	Ungradable images that were gradable by ophthalmologists	0/4 (0)	2/9 (22.2)	13/24 (54.2)
	Sensitivity	100 (91.03-100)	97.37 (86.50-99.53)	91.67 (78.17-97.13)
	Specificity	94.81 (89.68-97.47)	97.01 (92.58-98.83)	80.80 (73.02-86.74)
	Diagnostic accuracy	95.98 (91.93-98.04)	97.09 (93.38-98.75)	83.23 (76.70-88.21)
	PPV	84.78 (71.78-92.43)	90.24 (77.45-96.14)	57.89 (44.98-69.81)
	NPV	100 (97.09-100)	99.24 (95.80-99.87)	97.12 (91.86-99.01)
First Nations with diabetes	Number analysed	122/132 (92.4)	120/132 (90.9)	111/132 (84.1)
	Ungradable images (total)	4/132 (3.0)	7/132 (5.3)	19/132 (14.4)
	Ungradable images that were gradable by ophthalmologists	0/4 (0)	2/7(28.6)	11/19 (57.9)
	Sensitivity	100 (90.59-100)	97.22 (85.83-99.51)	91.18 (77.04-96.95)
	Specificity	96.47 (90.13-98.79)	97.62 (91.73-99.34)	89.61 (80.82-94.64)
	Diagnostic accuracy	97.54 (93.02-99.16)	97.50 (92.91-99.15)	90.09 (83.12-94.38)
	PPV	92.50 (80.14-97.42)	94.59 (82.30-98.50)	79.49 (64.47-89.22)
	NPV	100 (95.52-100)	98.80 (93.49-99.79)	95.83 (88.45-98.57)
Other Australians	Number analysed	52/56 (92.9)	52/56 (92.9)	50/56 (89.3)
	Ungradable images (total)	0/56 (0)	2/56 (3.6)	5/56 (8.9)

with diabetes	Ungradable images that were gradable by ophthalmologists	0 (0)	0/2 (0)	2/5 (40.0)
	Sensitivity**	100 (34.24-100)	100 (34.24-100)	100 (34.24-100)
	Specificity	92.00 (81.16-96.85)	96.00 (86.54-98.90)	66.67 (52.54-78.32)
	Diagnostic accuracy	92.31 (81.83-96.97)	96.15 (87.02-98.94)	68.00 (54.19-79.24)
	PPV	33.33 (9.68-70.00)	50.00 (15.00-85.00)	11.11 (3.10-32.80)
	NPV	100 (92.29-100)	100 (92.59-100)	100 (89.28-100)

\*Referable DR was defined as moderate non-proliferative DR or worse and images deemed ungradable by the DLSs were excluded from the respective analysis.\*\*There were only two images with referable DR that were of other Australian ethnicity. Data presented as number (percentage) or percentage (95% confidence interval); PPV: positive predictive value; NPV: negative predictive value. When retinal images were graded by the ophthalmologists there were 14 (7.4%), 10 (7.6%) and 4 (7.1%) ungradable images in the whole sample, First Nations people and other Australians, respectively.



**Figure 1.** A sample of study images showing gradability of images by the ophthalmologists and deep learning systems. A&B: Two retinal photographs ungradable by ophthalmologists and all DLSs. C: Photograph ungradable by Thirona RetCAD™ and EyRIS SELENA+, graded as referable DR by ophthalmologists and Google ARDA. D: Photograph ungradable by EyRIS SELENA+, graded as referable DR by ophthalmologists, Google ARDA and Thirona RetCAD™. E&F: Two photographs ungradable by EyRIS SELENA+ and both graded as no referable DR by ophthalmologists, Google ARDA and Thirona RetCAD™.

**Table 2.** Table showing performance of Google ARDA, Thirona RetCAD™ and EyRIS SELENA+ deep learning systems compared to ophthalmologist grading for referable (including ungradable) diabetic retinopathy, using 174 images of people with diabetes .

Deep learning system:	Google ARDA	Thirona RetCAD™	EyRIS SELENA+
Sensitivity	100 (91.03-100)	97.44 (86.82-99.55)	92.31 (79.68-97.35)

Specificity	94.81 (89.68-97.47)	96.30 (91.62-98.41)	74.81 (66.88-81.38)
Diagnostic accuracy	95.98 (91.93-98.04)	96.55 (92.68-98.41)	78.74 (72.07-84.16)
PPV	84.78 (71.78-92.43)	88.37 (77.52-94.93)	51.43 (39.95-62.75)
NPV	100 (97.09-100)	99.24 (95.80-99.87)	97.12 (91.86-99.01)

Data presented as percentage (95% confidence interval); PPV: positive predictive value; NPV: negative predictive value.

#### 4. Discussion

In a remote Australian population with a high proportion of First Nations people, ARDA correctly identified all patients with referable DR. This is the first Australian study to compare multiple DLSs for DR on the same real-world study population and showed that without any modification or finetuning to the DLSs, RetCAD™ and ARDA detected referable DR and vision-threatening DR with both sensitivity and specificity  $\geq 94.8\%$ . These results were statistically significantly different to SELENA+. While the accuracy was generally high across all three DLSs, there was variation in sensitivity, specificity and gradability underlining the importance of validating DLSs prior to implementation to determine which may be the most appropriate for that setting. Our study highlights that DLSs can be successfully used for DR screening prospectively in a remote Australian setting, providing further evidence to support implementation. Utilising this technology may improve access to DR screening in remote settings, improve health equity and reduce the risk of vision loss. [4,30,31]

ARDA and RetCAD™ had similar results to other international validation studies, yet SELENA+ had lower performance. ARDA had a sensitivity/specificity of 100% (95%CI: 91.03-100%)/94.81% (89.68-97.47%) which is improved or consistent with previous validation studies from India and Thailand that reported sensitivities between 89-98% and specificities between 92-95%. [32,33] In the current study, RetCAD™ had a sensitivity/specificity of 97.37% (86.50-99.53%)/97.01% (92.58-98.83%), which is more similar to studies evaluating later versions of RetCAD™, with sensitivities between 95-97% and specificities 92-94% (version 2) [34,35] than earlier versions with sensitivities of 84-86% and specificities of 92-93% (version 1), [36,37] but none assessed the same version (v2.2.0) used in this study. SELENA+ had a sensitivity of 91.67% (78.17-97.13%) which is comparable to other studies, however the specificity of 80.80% (73.02-86.74%) was lower than previously reported. The primary validation of SELENA+ for referable DR yielded a sensitivity of 90.5% and specificity of 91.6%. [25] SELENA+ was also validated on 4504 retinal images of Zambians with a sensitivity of 92.25% and specificity of 89.04% for referable DR, including macular oedema. [38] Of note, the dataset used to train SELENA+ comprised a large proportion of Chinese people, [25] while our study had a large population of First Nations Australians, as such, ethnic bias may explain differences in performance.

Australian studies validating the performance of DLSs have shown reasonably good performance. Optain (previously Eyetelligence), DAPHNE and CSIRO's Dr Grader have reported both sensitivity and specificity over 90%. [14,16,39] However, Optain had lower performance when only assessing First Nations peoples. [15,16] Dr Grader also showed 100% sensitivity but there were only two cases of referable DR in the sample and specificity was lower than Thirona and ARDA. [39] Similar to the current study, DAPHNE was validated on 393 images from First Nations peoples living in remote Australia, yet it used different outcomes of any DR and proliferative DR, [14] which may not be as useful for a population screening tool, where those with moderate NPDR or worse are typically referred to ophthalmology. ARDA has previously been validated on an urban First Nations population, using 1682 images, with a sensitivity of 98.0% and specificity of 95.1% for more than mild DR. [13] ARDA has now been validated in both urban and remote First Nations populations and has high sensitivity, suggesting fewer positive cases would be missed if used for a DR screening program.

Establishment of screening programs requires consideration of diagnostic sensitivity and specificity trade-off. [40] A higher sensitivity would mean that fewer true-positive referable DR cases would be missed, but that more false-positive cases would be referred to ophthalmology. Conversely an increase in specificity would result in fewer unnecessary referrals, but higher number of missed

referrable DR cases. Although ARDA had highest sensitivity and RetCAD™ had highest specificity, both these DLSs had high sensitivity and high specificity, with similar diagnostic accuracy. When determining which tool to use the population needs should be considered. Minimising false-positives may be preferred for rural or remote areas for patients facing significant travel and costs to access ophthalmology services and many patients for few clinicians. [41] For those who attend regular annual or biennial DR screening, it is unlikely that a false-negative would result in vision loss prior to the next screening. [42] However, where recommended DR screening adherence is low, such as in remote Australian settings, [6] it is possible that missing a case of referrable DR could result in vision loss before the next opportunity for DR screening, especially in the context of other risk factors, for example suboptimal glycaemic control. [42] In these cases, ARDA may be better suited. In practice, a hybrid system, incorporating human 'over-reading' prior to a specialist review may mitigate unnecessary referral burden. [43]

An ungradable image in a real-world setting would likely result in a referral to ophthalmology or a repeat visit, which decreased DLS specificity and leads to greater unnecessary referrals. Ideally a DLS would be able to grade an image that is of sufficient quality for an ophthalmologist to grade, yet this nuance is not always reported. In the current study, ARDA was able to grade all images deemed gradable by the ophthalmologists, while two were ungradable by RetCAD™. In contrast, almost 13% of all images were ungradable by SELENA+ and 54% of these ungradable images were gradable by ophthalmologists. Proportion of gradable images, and gradability in comparison to a human grader should also be considered prior to implementing DLS-assisted DR screening.

#### 4.1. Strengths and Limitations

The strengths of this study include the real-world design, comparison of multiple DLSs on a population with a high proportion of First Nations people and images graded by an ophthalmologist adjudication panel. By assessing three DLSs simultaneously, the diagnostic performance metrics demonstrate pertinent differences to consider prior to implementation. However, limitations of the study include that diabetes status relied on self-report and consistent with other similar studies, diabetes type was not ascertained, [13,16,24,35,36,39] which may introduce misclassification bias. As expected with the real-world nature of the study, there was inconsistent dilation, which may have contributed to more ungradable images. Furthermore, by using single-field retinal images it is possible that some DR was not detected. This study did not specifically assess diabetic macular oedema from the retinal photographs, as this is not reported by RetCAD™ or SELENA+. While we had sufficient sample size to analyse performance of the DLSs in First Nations, only two people of other Australian ethnicity had referable DR. A greater sample size would have resulted in smaller confidence interval widths. The Pilbara region of Australia is unique, hence results may not be generalisable to other settings.

#### 4.2. Conclusions

In a remote real-world study comprising majority First Nations people, DLSs had high sensitivity and specificity for detecting referable DR. Direct real-world comparison is useful to guide implementation, particularly when the population comprises minority groups. As DLS are not widely adopted, research focussed on addressing implementation barriers is needed. Point-of-care DR screening using DLSs may improve DR screening rates and potentially improve health equity and prevent vision loss when paired with appropriate referral and treatment pathways.

**Author Contributions:** JJD assisted with project design, writing the protocol, data analysis, data interpretation and wrote the first draft of the manuscript. QL assisted with writing the protocol, data collection, data interpretation. KW ensured the project was culturally appropriate, conducted stakeholder meetings, assisted with study design and data interpretation. ED assisted with project design, writing the protocol, research coordination and data interpretation. YZ assisted with data interpretation. SB, YS and VS assisted with diabetic retinopathy grading and data interpretation. PAK assisted with study design and data interpretation. AWT was

the principal investigator on the study, assisted with design, stakeholder engagement, securing funding, coordination, and data interpretation. All authors reviewed and edited the manuscript.

**Funding:** This activity was supported by the Western Australian Future Health Research and Innovation Fund, Grant ID DoH202310626/1. The funders had no role in study design, data collection, analysis or reporting.

**Data Availability Statement:** Data available on reasonable request from the corresponding author. Data are not publicly available due to confidentiality restrictions.

**Acknowledgments:** We would like to acknowledge the contribution of Aboriginal communities in the Pilbara for welcoming our staff into their communities, attending consultation meetings and participating in the project. We specifically acknowledge Jigalong, Wakathuni, Bellary Springs, Bindi Bindi as well as the Aboriginal communities within Newman, Tom Price, Paraburdoo, Onslow. We would like to acknowledge the unwavering local support of our WA / Pilbara regional partners: Puntukurnu Aboriginal Medical Service, Mawarnkarra Health Service, Royal Flying Doctor Service, Nintirri Centre, Karratha Central Healthcare, Aboriginal Health Council of Western Australia, Diabetes WA, WA Country Health Service, IBN Group, Panaceum Pilbara, Jigalong Community Inc., Ashburton Aboriginal Corporation, Gumala Aboriginal Corporation and Yinhawangka Aboriginal Corporation. We would like to equally acknowledge the in-kind support of our technology partners: Thirona RetCADTM, EyRIS, Topcon Healthcare and Optomed. The DLSs were provided without cost and these partners did not have any involvement with the study design, data collection, analysis or reporting.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

1. Bourne RRA, Jonas JB, Flaxman SR, et al. Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe: 1990–2010. *Br J Ophthalmol.* 2014;98(5): 629-638. <https://doi.org/10.1136/bjophthalmol-2013-304033>
2. Australian Institute of Health and Welfare. Indigenous eye health measures 2021. AIHW 261. Canberra: Australian Institute of Health and Welfare; 2021. Accessed 20 January 2025. <https://www.aihw.gov.au/reports/indigenous-australians/indigenous-eye-health-measures-2021>
3. Teo ZL, Tham YC, Yu M, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis. *Ophthalmology.* 2021;128(11): 1580-1591. <https://doi.org/10.1016/j.ophtha.2021.04.027>
4. Wong TY, Sun J, Kawasaki R, et al. Guidelines on diabetic eye care: The International Council of Ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings. *Ophthalmology.* 2018;125(10): 1608-1622. <https://doi.org/10.1016/j.ophtha.2018.04.007>
5. Mitchell P, Foran S, Wong T, Chua B, Patel I, Ojaimi E. Guidelines for the management of diabetic retinopathy. National Health and Medical Research Council; 2008:1-183.
6. Foreman J, Keel S, Xie J, Van Wijngaarden P, Taylor HR, Dirani M. Adherence to diabetic eye examination guidelines in Australia: The National Eye Health Survey. *Med J Aust.* 2017;206(9): 402-406. <https://doi.org/10.5694/mja16.00989>
7. Australian Institute of Health and Welfare. Diabetes: Australian facts. Canberra: Australian Institute of Health and Welfare; 2023. Accessed 26 August 2024. <https://www.aihw.gov.au/reports/diabetes/diabetes>
8. White C, Seear K, Anderson L, Griffiths E. Use of a primary care dataset to describe 'the real picture' of diabetes in Kimberley Aboriginal communities. *Journal of the Australian Indigenous HealthInfoNet.* 2024;5(1). <https://doi.org/10.14221/2653-3219.1035>
9. Australian Bureau of Statistics. 2021 census Aboriginal and/or Torres Strait Islander people quickstats. Australian Bureau of Statistics; 2021. Accessed 25 August 2024. <https://abs.gov.au/census/find-census-data/quickstats/2021/IQS51002>
10. Khou V, Khan MA, Jiang IW, Katalinic P, Agar A, Zangerl B. Evaluation of the initial implementation of a nationwide diabetic retinopathy screening programme in primary care: A multimethod study. *BMJ Open.* 2021;11(8): e044805. <https://doi.org/10.1136/bmjopen-2020-044805>

11. Hu W, Joseph S, Li R, et al. Population impact and cost-effectiveness of artificial intelligence-based diabetic retinopathy screening in people living with diabetes in Australia: A cost effectiveness analysis. *eClinicalMedicine*. 2024;67. <https://doi.org/10.1016/j.eclinm.2023.102387>
12. Joseph S, Selvaraj J, Mani I, et al. Diagnostic accuracy of artificial intelligence-based automated diabetic retinopathy screening in real-world settings: A systematic review and meta-analysis. *Am J Ophthalmol*. 2024;263: 214-230. <https://doi.org/10.1016/j.ajo.2024.02.012>
13. Chia MA, Hersch F, Sayres R, et al. Validation of a deep learning system for the detection of diabetic retinopathy in Indigenous Australians. *Brit J Ophthalmol*. 2023: bjo-2022-322237. <https://doi.org/10.1136/bjo-2022-322237>
14. Quinn N, Brazionis L, Zhu B, et al. Facilitating diabetic retinopathy screening using automated retinal image analysis in underresourced settings. *Diabetic Med*. 2021;38(9): e14582. <https://doi.org/10.1111/dme.14582>
15. Scheetz J, Koca D, McGuinness M, et al. Real-world artificial intelligence-based opportunistic screening for diabetic retinopathy in endocrinology and Indigenous healthcare settings in Australia. *Sci Rep*. 2021;11(1): 15808. <https://doi.org/10.1038/s41598-021-94178-5>
16. Li Z, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. *Diabetes Care*. 2018;41(12): 2509-2516. <https://doi.org/10.2337/dc18-0147>
17. Google Health.. Using AI to prevent blindness. Accessed 19 February 2025. <https://health.google/caregivers/arda/>
18. Thirona Retina. Artificial intelligence for high performance eye disease screening. Accessed 19 February 2025. <https://retcad.eu/>
19. EyRIS. Revolutionizing the detection of eye diseases. Accessed 19 February 2025. <https://www.eyris.io/index.cfm>
20. Arora A, Alderman JE, Palmer J, et al. The value of standards for health datasets in artificial intelligence-based applications. *Nat Med*. 2023;29(11): 2929-2938. <https://doi.org/10.1038/s41591-023-02608-w>
21. Ktena I, Wiles O, Albuquerque I, et al. Generative models improve fairness of medical classifiers under distribution shifts. *Nat Med*. 2024;30(4): 1166-1173. <https://doi.org/10.1038/s41591-024-02838-6>
22. Yang Y, Zhang H, Gichoya JW, Katabi D, Ghassemi M. The limits of fair medical imaging AI in real-world generalization. *Nat Med*. 2024;30(10): 2838-2848. <https://doi.org/10.1038/s41591-024-03113-4>
23. Li Q, Drinkwater JJ, Woods K, Douglas E, Ramirez A, Turner AW. Implementation of a new, mobile diabetic retinopathy screening model incorporating artificial intelligence in remote Western Australia. *Aust J Rural Health*. 2025;33(2): e70031. <https://doi.org/10.1111/ajr.70031>
24. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22): 2402-2410. <https://doi.org/10.1001/jama.2016.17216>
25. Ting DSW, Cheung CY-L, 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*. 2017;318(22): 2211-2223. <https://doi.org/10.1001/jama.2017.18152>
26. RetCAD version 2.2 White paper. Accessed May 12, 2025. <https://delft.care/retcad/>. 2023.
27. Wilkinson CP, Ferris FL, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110(9): 1677-1682. [https://doi.org/10.1016/S0161-6420\(03\)00475-5](https://doi.org/10.1016/S0161-6420(03)00475-5)
28. Arifin WN. Sample size calculator (web). <http://wnarifin.github.io>.
29. Al-Jawabreh A, Dumaidi K, Erekat S, et al. A comparison of the efficiency of three sampling methods for use in the molecular and conventional diagnosis of cutaneous leishmaniasis. *Acta Tropica*. 2018;182: 173-177. <https://doi.org/10.1016/j.actatropica.2018.03.001>
30. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: An assailable barrier to equitable digital health care. *Lancet Digit Health*. 2021;3(4): e260-e265. [https://doi.org/10.1016/S2589-7500\(20\)30317-4](https://doi.org/10.1016/S2589-7500(20)30317-4)

31. Wolf RM, Channa R, Liu TYA, et al. Autonomous artificial intelligence increases screening and follow-up for diabetic retinopathy in youth: The ACCESS randomized control trial. *Nat Commun.* 2024;15(1): 421. <https://doi.org/10.1038/s41467-023-44676-z>
32. Ruamviboonsuk P, Tiwari R, Sayres R, et al. Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: A prospective interventional cohort study. *Lancet Digit Health.* 2022;4(4): e235-e244. [https://doi.org/10.1016/S2589-7500\(22\)00017-6](https://doi.org/10.1016/S2589-7500(22)00017-6)
33. Gulshan V, Rajan RP, Widner K, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmol.* 2019;137(9): 987-993. <https://doi.org/10.1001/jamaophthalmol.2019.2004>
34. Piatti A, Rui C, Gazzina S, et al. Diabetic retinopathy screening with confocal fundus camera and artificial intelligence - assisted grading. *Eu J Ophthalmol.* 2024: 11206721241272229. <https://doi.org/10.1177/11206721241272229>
35. Meredith S, van Grinsven M, Engelberts J, et al. Performance of an artificial intelligence automated system for diabetic eye screening in a large English population. *Diabet Med.* 2023;40(6): e15055. <https://doi.org/10.1111/dme.15055>
36. Taylor JR, Drinkwater J, Sousa DC, Shah V, Turner AW. Real-world evaluation of retcad deep-learning system for the detection of referable diabetic retinopathy and age-related macular degeneration. *Clin Exp Optom.* 2024: 1-6. <https://doi.org/10.1080/08164622.2024.2385565>
37. Skevas C, Weindler H, Levering M, Engelberts J, van Grinsven M, Katz T. 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(12): 1985-1993. <https://doi.org/10.18240/ijo.2022.12.14>
38. Bellemo V, Lim ZW, Lim G, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: A clinical validation study. *Lancet Digital Health.* 2019;1(1): e35-e44. [https://doi.org/10.1016/S2589-7500\(19\)30004-4](https://doi.org/10.1016/S2589-7500(19)30004-4)
39. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney M-L, Mehrotra A. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. *JAMA Network Open.* 2018;1(5): e182665-e182665. <https://doi.org/10.1001/jamanetworkopen.2018.2665>
40. Chu K. An introduction to sensitivity, specificity, predictive values and likelihood ratios. *Emerg Med.* 1999;11(3): 175-181. <https://doi.org/10.1046/j.1442-2026.1999.00041.x>
41. Cheung R, Ly A. A survey of eyecare affordability among patients seen in collaborative care in Australia and factors contributing to cost barriers. *Public Health Res Pract.* 2024;34(2):e3422415. <http://dx.doi.org/https://doi.org/10.17061/phrp3422415>
42. Drinkwater JJ, Kalantary A, Turner AW. A systematic review of diabetic retinopathy screening intervals. *Acta Ophthalmologica.* 2024;102(4): e473-e484. <https://doi.org/10.1111/aos.15788>
43. Ta AWA, Goh HL, Ang C, Koh LY, Poon K, Miller SM. Two Singapore public healthcare AI applications for national screening programs and other examples. *Health Care Sci.* 2022;1(2): 41-57. <https://doi.org/10.1002/hcs2.10>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.