






## Article

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

Jocelyn J. Drinkwater <sup>1,2</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, WA 6009, Australia; jocelyn.drinkwater@uwa.edu.au

<sup>2</sup> Lions Outback Vision, Lions Eye Institute, 2 Verdun St, Nedlands, WA 6009, Australia; emma.douglas@lei.org.au (E.D.)

<sup>3</sup> Moorfields Eye Hospital NHS Foundation Trust, 11-43 Bath St, London EC1V 9EL, UK; yukun.zhou.19@ucl.ac.uk (Y.Z.); p.keane@ucl.ac.uk (P.A.K.)

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

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

## Abstract

**Background/objective:** Deep learning systems (DLSs) may improve access to screening for diabetic retinopathy (DR), a leading cause of vision loss. Therefore, the aim was to prospectively compare the performance of three DLSs, Google ARDA, Thirona RetCAD™, and EyRIS SELENA+, in the detection of referable DR in a real-world setting. **Methods:** Participants with self-reported 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. **Results:** Single field colour fundus photographs from 188 eyes of 94 participants (51% male, 70% First Nations Australians, and mean  $\pm$  SD age of  $60.3 \pm 12.0$  years) 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%). **Conclusions:** In a small, real-world service evaluation, comprising majority First Nations people from remote Western Australia, DLSs had high sensitivity and specificity for detecting referable DR. A comparative service evaluation can be useful to highlight differences between DLSs, especially in unique settings or with minority populations.

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



Academic Editor: Peter Clifton

Received: 6 October 2025

Revised: 12 November 2025

Accepted: 24 November 2025

Published: 1 December 2025

**Citation:** Drinkwater, J.J.; Li, Q.; Woods, K.; Douglas, E.; Chia, M.; Zhou, Y.; Bartnik, S.; Shah, Y.; Shah, V.; Keane, P.A.; et al. A Real-World Comparison of Three Deep Learning Systems for Diabetic Retinopathy in Remote Australia. *Diabetology* **2025**, *6*, 146. <https://doi.org/10.3390/diabetology6120146>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 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. Screening rates were lower in remote areas [6], which have higher proportions of First Nations people and 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 limited ophthalmology and optometry workforce, significant travel distances to access eye care [2], high staff turnover, and lengthy waits for screening results in primary care settings [10]. Deep learning systems (DLSs), a type of artificial intelligence, can be used for DR screening. They require minimal staff training and in conjunction with automated cameras may overcome these barriers in order to increase access to DR screening in remote regions.

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]. Despite several DLSs for DR being available, few 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 regulatory 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 comprising 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], and no Australian study has directly compared the performance of multiple DLSs on the same population. Therefore, the aim of this study was to prospectively compare the performance of multiple DLSs, to detect referable DR in the Pilbara region, a remote Western Australian real-world setting with a high proportion of First Nations people.

## 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 the 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 taken, centred on the macula, 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<sup>TM</sup> (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.

These DLSs were chosen as RetCAD<sup>TM</sup> [18] and SELENA+ [19] have Australian regulatory approval, yet neither have been validated on First Nations people specifically, and ARDA was chosen as it has previously been validated for this ethnic group [13]. 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 ascertain DR severity. Both RetCAD<sup>TM</sup> and ARDA outputs provided the DR severity grade.

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 an on-

the-spot telehealth consultation was provided for patients with referable disease, or at the patient's request.

The colour fundus photographs were independently graded by two ophthalmologists (YS and SB) according to the International Clinical Diabetic Retinopathy and Diabetic Macular Oedema 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 was 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]. SELENA+ was developed in Singapore using 76,370 images from their national DR screening programme and internally validated on 70,000 different photographs, also from the screening programme [25]. Its external validation comprised 40,000 photographs from many countries, 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%, an estimated prevalence of referable DR of 20%, and a 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, for eyes and at person level, using the worst graded eye. 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™ in assessing vision-threatening DR was also assessed. To assess for statistically significant differences between the three DLSs, multiple McNemar tests were performed comparing the misclassification (true positives and true negatives versus false positives and false negatives) of both images and people. Post hoc Bonferroni correction for multiple comparisons was applied and results with adjusted  $p$ -values  $< 0.05$  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. The ophthalmologists' grading of the 188 colour fundus photographs determined 135 (71.8%) images that had no referable DR (121 (64.4%) with no DR and 14 (7.4%) with mild NPDR) and 39 (20.7%) images from 23 people that 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. Three people had ungradable images bilaterally. There was substantial agreement between the two main graders ( $\kappa = 0.71$ ).

When analysing the images independently, ARDA had a 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 a high sensitivity and specificity of 97.37% (86.50–99.53%) and 97.01% (92.58–98.83%), respectively. SELENA+ had a sensitivity of 91.67% (78.17–97.13%) and a 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 Australians, the sensitivity was 100% for all DLSs; however, there were only two with referable DR (Table 1). The results at the person level increased sensitivity and for RetCAD™, the specificity was similar, but this decreased for ARDA and SELENA+. The subgroups had smaller sample size that led to less precision, as indicated by the wide confidence intervals, and therefore are interpreted cautiously.

**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 Maestro 2 camera.

Deep Learning System		Google ARDA	Thirona RetCAD™	EyRIS SELENA+
Eyes with diabetes	Number of images 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, eyes	Number of images 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)

Table 1. Cont.

Deep Learning System		Google ARDA	Thirona RetCAD™	EyRIS SELENA+
Other Australians, eyes	Number of images 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)
	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)
People with diabetes	Number of people analysed	91/94 (96.8)	91/94 (96.8)	87/94 (92.6)
	Ungradable people's images (total)	1/94 (1.1)	2/94 (2.1)	7/94 (7.4)
	Ungradable people's images that were gradable by ophthalmologists	0	0	4/7 (57.1)
	Sensitivity **	100 (85.69–100)	100 (85.69–100)	95.45 (78.20–99.19)
	Specificity	92.65 (83.91–96.82)	97.06 (89.90–99.19)	72.31 (60.42–81.71)
	Diagnostic accuracy	94.51 (87.78–97.63)	97.80 (92.34–99.40)	78.16 (68.39–85.55)
	PPV	82.14 (64.41–92.12)	92.00 (75.03–97.78)	53.85 (38.57–68.43)
	NPV	100 (94.25–100)	100 (94.50–100)	97.92 (89.10–99.63)

\* 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; and 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, the First Nations people sample, and the other Australians sample, respectively.

The McNemar tests showed there was no statistically significant difference in the misclassification of images or people between ARDA and RetCAD™ (adjusted  $p > 0.75$ ) but there was a significant difference in misclassification between ARDA and SELENA+ (adjusted  $p \leq 0.002$ ), and RetCAD™ and 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 the detection of vision-threatening DR in eyes, 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 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.

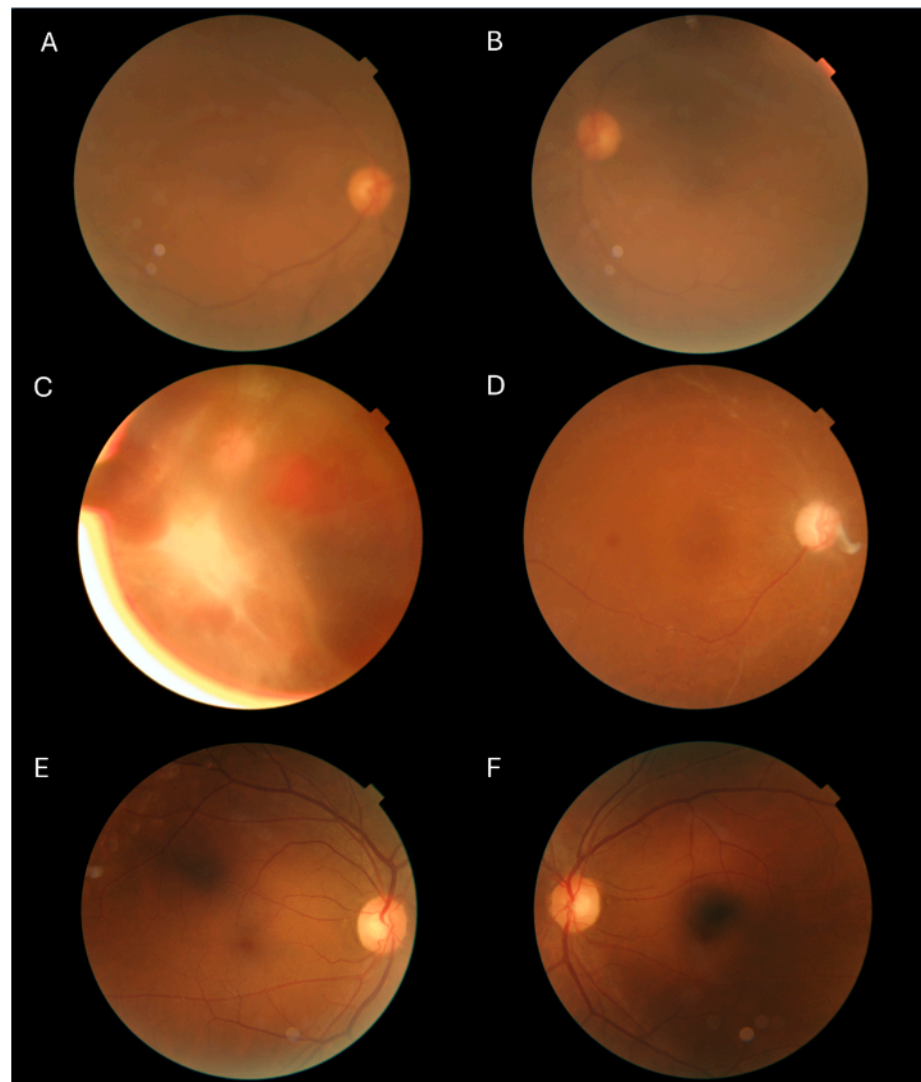
Deep Learning System	Google ARDA	Thirona RetCAD™	EyRIS SELENA+
Sensitivity	100 (91.03–100)	97.44 (86.82–99.55)	92.31 (79.68–97.35)
Specificity	94.81 (89.68–97.47)	96.30 (91.62–98.41)	74.81 (66.88–81.38)



Table 2. Cont.

Deep Learning System	Google ARDA	Thirona RetCAD™	EyRIS SELENA+
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.



**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 DLs. (C): Photograph ungradable by Thirona RetCAD™ and EyRIS SELENA+, graded as referable DR by ophthalmologists and Google ARDA. (D): Photograph ungradable by EyRIS SELENA+, but graded as referable DR by ophthalmologists, Google ARDA, and Thirona RetCAD™. (E,F): Two photographs ungradable by EyRIS SELENA+ and both graded as non-referable DR by ophthalmologists, Google ARDA, and Thirona RetCAD™.

#### 4. Discussion

In a small 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 DLs for DR on the same real-world study population, and it showed that without any modification or fine-tuning to the DLs, RetCAD™ and ARDA detected referable DR and vision-threatening DR with both sensitivity and specificity >90%.

RetCAD<sup>TM</sup> and ARDA had statistically significantly less misclassification compared to SELENA+. While the accuracy was generally high across all three DLs, there were variations in sensitivity, specificity, and gradability, underlining the importance of validating DLs prior to implementation to determine which may be the most appropriate for that setting. Our study highlights that DLs 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 and, with appropriate referral and treatment pathways, may reduce the risk of vision loss [4,29,30].

ARDA and RetCAD<sup>TM</sup> 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 is consistent with previous validation studies from India and Thailand that reported sensitivities between 89 and 98% and specificities between 92 and 95% [31,32]. In the current study, RetCAD<sup>TM</sup> 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<sup>TM</sup>, with sensitivities between 95 and 97% and specificities between 92 and 94% (version 2) [33,34], than to earlier versions, with sensitivities of 84–86% and specificities of 92–93% (version 1) [35,36], 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 [37]. Notably, 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 and, as such, ethnic bias may explain the differences in performance.

Australian studies validating the performance of DLs have shown reasonably good performance. Optain (previously Eyetelligence), DAPHNE, and CSIRO's Dr Grader have reported both a sensitivity and specificity over 90% [14,16,38]. However, Optain had lower performance when only assessing First Nations people [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 [38]. Similarly to the current study, DAPHNE was validated on 393 images from First Nations people living in remote Australia, yet it had 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 programme.

The establishment of screening programmes requires the consideration of the diagnostic sensitivity and specificity trade-off [39]. 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 a higher number of missed referable DR cases. Interestingly, RetCAD had the same sensitivity and higher specificity than ARDA at the person-level analysis, but when assessing images, ARDA had the highest sensitivity and RetCAD<sup>TM</sup> had the highest specificity. When determining which tool to use, population needs should be considered. Minimising false positives may be preferred for rural or remote areas for patients facing significant travel and cost barriers in accessing ophthalmology services and

in areas with many patients and few clinicians [40]. 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 [41]. However, where recommended DR screening adherence is low, such as in remote Australian settings [6], it is possible that missing a case of a referable DR could result in vision loss before the next opportunity for DR screening, especially in the context of other risk factors, such as suboptimal glycaemic control [41]. 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 [42].

An ungradable image in a real-world setting would likely result in a referral to ophthalmology or a repeat visit, which decreases 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<sup>TM</sup>. In contrast, almost 13% of all images were ungradable by SELENA+ and 54% of these ungradable images were gradable by ophthalmologists. The proportion of gradable images and the 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, and the comparison of multiple DLSs in a population with a high proportion of First Nations people to 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 [13,16,24,34,35,38], which may introduce misclassification bias. As expected, due to the real-world nature of the study, dilation was inconsistent, 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<sup>TM</sup> or SELENA+. As one camera was used, these results may not be generalisable to other imaging systems. The sample of those with referable DR was small, limiting our statistical power. We did not have the sample size to adjust for other confounding factors. Greater sample size would have resulted in smaller confidence interval widths. Nevertheless, the data is important and highlights complexities of conducting research in remote First Nations communities. The real-world nature of the study, using the commercially available versions of the DLS, meant raw scores were not available so we were unable to calculate some parameters such as receiver operating curves and attention maps. The Pilbara region of Australia is unique; hence, results may not be generalisable to other settings.

#### 4.2. Conclusions

In this descriptive, real-world evaluation, comprising mainly First Nations people, DLSs had high sensitivity and specificity in detecting referable DR. However, due to the unique study setting, with a relatively small sample size, our results are population-specific and not widely generalisable and, as such, should be interpreted with caution. Regardless, a direct real-world comparison can be useful to highlight differences between these tools, particularly when the population comprises minority groups.

**Author Contributions:** J.J.D. assisted with project design, writing the protocol, data analysis, and data interpretation, and wrote the first draft of the manuscript. Q.L. assisted with writing the protocol, data collection, and data interpretation. K.W. ensured the project was culturally appropriate, conducted



stakeholder meetings, and assisted with study design and data interpretation. E.D. assisted with project design, writing the protocol, research coordination, and data interpretation. Y.Z. and M.C. assisted with data interpretation. S.B., Y.S. and V.S. assisted with diabetic retinopathy grading and data interpretation. P.A.K. assisted with study design and data interpretation. A.W.T. was the principal investigator on the study, and assisted with design, stakeholder engagement, securing funding, coordination, and data interpretation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work has been supported by the Western Australian Department of Health, through the Future Health Research and Innovation Fund and keystone partners RIO Tinto and BHP, Grant ID DoH202310626/1. The funders had no role in study design, data collection, analysis, or reporting.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Western Australian Aboriginal Health Ethics Committee (HREC1294) on 17 November 2023.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

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

**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 DLs 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, R.R.A.; Jonas, J.B.; Flaxman, S.R.; Keeffe, J.; Leasher, J.; Naidoo, K.; Parodi, M.B.; Pesudovs, K.; Price, H.; A White, R.; 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*, 629–638. [CrossRef] [PubMed]
2. Australian Institute of Health and Welfare. *Indigenous Eye Health Measures 2021*. AIHW 261; Australian Institute of Health and Welfare: Canberra, Australia, 2021. Available online: <https://www.aihw.gov.au/reports/indigenous-australians/indigenous-eye-health-measures-2021> (accessed on 20 January 2025).
3. Teo, Z.L.; Tham, Y.-C.; Yu, M.; Chee, M.L.; Rim, T.H.; Cheung, N.; Bikbov, M.M.; Wang, Y.X.; Tang, Y.; Lu, Y.; et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis. *Ophthalmology* **2021**, *128*, 1580–1591. [CrossRef] [PubMed]
4. Wong, T.Y.; Sun, J.; Kawasaki, R.; Ruamviboonsuk, P.; Gupta, N.; Lansingh, V.C.; Maia, M.; Mathenge, W.; Moreker, S.; Muqit, M.M.; 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*, 1608–1622. [CrossRef]
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: Canberra, Australia, 2008; pp. 1–183.
6. Foreman, J.; Keel, S.; Xie, J.; Van Wijngaarden, P.; Taylor, H.R.; Dirani, M. Adherence to diabetic eye examination guidelines in Australia: The National Eye Health Survey. *Med. J. Aust.* **2017**, *206*, 402–406. [CrossRef]
7. Australian Institute of Health and Welfare. *Diabetes: Australian Facts*; Australian Institute of Health and Welfare: Canberra, Australia, 2023. Available online: <https://www.aihw.gov.au/reports/diabetes/diabetes> (accessed on 26 August 2024).

8. White, C.S.; Seear, K.; Anderson, L.; Griffiths, E. Use of a primary care dataset to describe ‘the real picture’ of diabetes in Kimberley Aboriginal communities. *J. Aust. Indig. Heal.* **2024**, *5*, 4. [CrossRef]
9. Australian Bureau of Statistics. *2021 Census Aboriginal and/or Torres Strait Islander People Quickstats*; Australian Bureau of Statistics: Canberra, Australia, 2021. Available online: <https://abs.gov.au/census/find-census-data/quickstats/2021/IQS51002> (accessed on 25 August 2024).
10. Khou, V.; Khan, M.A.; Jiang, I.W.; 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*, e044805. [CrossRef]
11. Hu, W.; Joseph, S.; Li, R.; Woods, E.; Sun, J.; Shen, M.; Jan, C.L.; Zhu, Z.; He, M.; Zhang, L. 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*, 102387. [CrossRef]
12. Joseph, S.; Selvaraj, J.; Mani, I.; Kumaragurupari, T.; Shang, X.; Mudgil, P.; Ravilla, T.; He, M. Diagnostic accuracy of artificial intelligence-based automated diabetic retinopathy screening in real-world settings: A systematic review and meta-analysis. *Arch. Ophthalmol.* **2024**, *263*, 214–230. [CrossRef]
13. Chia, M.A.; Hersch, F.; Sayres, R.; Bavishi, P.; Tiwari, R.; Keane, P.A.; Turner, A.W. Validation of a deep learning system for the detection of diabetic retinopathy in Indigenous Australians. *Br. J. Ophthalmol.* **2023**, *108*, 268–273. [CrossRef]
14. Quinn, N.; Brazionis, L.; Zhu, B.; Ryan, C.; D’alio, R.; Tang, H.L.; Peto, T.; Jenkins, A. The Centre of Research Excellence in Diabetic Retinopathy Study, TEAMSnet Study Groups. Facilitating diabetic retinopathy screening using automated retinal image analysis in underresourced settings. *Diabet. Med.* **2021**, *38*, e14582. [CrossRef]
15. Scheetz, J.; Koca, D.; McGuinness, M.; Holloway, E.; Tan, Z.; Zhu, Z.; O’Day, R.; Sandhu, S.; MacIsaac, R.J.; Gilfillan, C.; et al. Real-world artificial intelligence-based opportunistic screening for diabetic retinopathy in endocrinology and indigenous healthcare settings in Australia. *Sci. Rep.* **2021**, *11*, 15808. [CrossRef]
16. Li, Z.; Keel, S.; Liu, C.; He, Y.; Meng, W.; Scheetz, J.; Lee, P.Y.; Shaw, J.; Ting, D.; Wong, T.Y.; 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*, 2509–2516. [CrossRef]
17. Google Health. Using AI to Prevent Blindness. Available online: <https://health.google/caregivers/arda/> (accessed on 19 February 2025).
18. Thirona Retina. Artificial Intelligence for High Performance Eye Disease Screening. Available online: <https://retcad.eu/> (accessed on 19 February 2025).
19. EyRIS. Revolutionizing the Detection of Eye Diseases. Available online: <https://www.eyris.io/index.cfm> (accessed on 19 February 2025).
20. Arora, A.; Alderman, J.E.; Palmer, J.; Ganapathi, S.; Laws, E.; McCradden, M.D.; Oakden-Rayner, L.; Pfohl, S.R.; Ghassemi, M.; McKay, F.; et al. The value of standards for health datasets in artificial intelligence-based applications. *Nat. Med.* **2023**, *29*, 2929–2938. [CrossRef]
21. Ktena, I.; Wiles, O.; Albuquerque, I.; Rebuffi, S.-A.; Tanno, R.; Roy, A.G.; Azizi, S.; Belgrave, D.; Kohli, P.; Cemgil, T.; et al. Generative models improve fairness of medical classifiers under distribution shifts. *Nat. Med.* **2024**, *30*, 1166–1173. [CrossRef]
22. Yang, Y.; Zhang, H.; Gichoya, J.W.; Katabi, D.; Ghassemi, M. The limits of fair medical imaging AI in real-world generalization. *Nat. Med.* **2024**, *30*, 2838–2848. [CrossRef]
23. Li, Q.; Drinkwater, J.J.; Woods, K.; Douglas, E.; Ramirez, A.; Turner, A.W. Implementation of a new, mobile diabetic retinopathy screening model incorporating artificial intelligence in remote western australia. *Aust. J. Rural. Heal.* **2025**, *33*, e70031. [CrossRef]
24. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **2016**, *316*, 2402–2410. [CrossRef] [PubMed]
25. Ting, D.S.W.; Cheung, C.Y.-L.; Lim, G.; Tan, G.S.W.; Quang, N.D.; Gan, A.; Hamzah, H.; Garcia-Franco, R.; Yeo, I.Y.S.; Lee, S.Y.; 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*, 2211–2223. [CrossRef] [PubMed]
26. RetCAD, Version 2.2; White Paper, 2023. Available online: <https://delft.care/retcad/> (accessed on 12 May 2025).
27. Wilkinson, C.; Ferris, F.L.; Klein, R.; Lee, P.P.; Agardh, C.D.; Davis, M.; Dills, D.; Kampik, A.; Pararajasegaram, R.; Verdager, J.T. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* **2003**, *110*, 1677–1682. [CrossRef] [PubMed]
28. Arifin, W.N. Sample Size Calculator (Web). Available online: <http://wnarifin.github.io> (accessed on 29 July 2024).
29. Ibrahim, H.; Liu, X.; Zariffa, N.; Morris, A.D.; Denniston, A.K. Health data poverty: An assailable barrier to equitable digital health care. *Lancet Digit. Health* **2021**, *3*, e260–e265. [CrossRef] [PubMed]
30. Wolf, R.M.; Channa, R.; Liu, T.Y.A.; Zehra, A.; Bromberger, L.; Patel, D.; Ananthakrishnan, A.; Brown, E.A.; Prichett, L.; Lehmann, H.P.; et al. Autonomous artificial intelligence increases screening and follow-up for diabetic retinopathy in youth: The ACCESS randomized control trial. *Nat. Commun.* **2024**, *15*, 421. [CrossRef]

31. Ruamviboonsuk, P.; Tiwari, R.; Sayres, R.; Nganthavee, V.; Hemarat, K.; Kongprayoon, A.; Raman, R.; Levinstein, B.; Liu, Y.; Schaekermann, M.; 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*, e235–e244. [\[CrossRef\]](#)
32. Gulshan, V.; Rajan, R.P.; Widner, K.; Wu, D.; Wubbels, P.; Rhodes, T.; Whitehouse, K.; Coram, M.; Corrado, G.; Ramasamy, K.; et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmol.* **2019**, *137*, 987–993. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Piatti, A.; Rui, C.; Gazzina, S.; Tartaglino, B.; Romeo, F.; Manti, R.; Doglio, M.; Nada, E.; Giorda, C. Diabetic retinopathy screening with confocal fundus camera and artificial intelligence—Assisted grading. *Eur. J. Ophthalmol.* **2024**, *35*, 679–688. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Meredith, S.; van Grinsven, M.; Engelberts, J.; Clarke, D.; Prior, V.; Vodrey, J.; Hammond, A.; Muhammed, R.; Kirby, P. Performance of an artificial intelligence automated system for diabetic eye screening in a large English population. *Diabet. Med.* **2023**, *40*, e15055. [\[CrossRef\]](#)
35. Taylor, J.R.; Drinkwater, J.; Sousa, D.C.; Shah, V.; Turner, A.W. Real-world evaluation of RetCAD deep-learning system for the detection of referable diabetic retinopathy and age-related macular degeneration. *Clin. Exp. Optom.* **2024**, *108*, 601–606. [\[CrossRef\]](#) [\[PubMed\]](#)
36. 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*, 1985–1993. [\[CrossRef\]](#)
37. Bellemo, V.; Lim, Z.W.; Lim, G.; Nguyen, Q.D.; Xie, Y.; Yip, M.Y.T.; Hamzah, H.; Ho, J.; Lee, X.Q.; Hsu, W.; et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: A clinical validation study. *Lancet Digit. Health* **2019**, *1*, e35–e44. [\[CrossRef\]](#)
38. 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 Netw. Open* **2018**, *1*, e182665. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Chu, K. An introduction to sensitivity, specificity, predictive values and likelihood ratios. *Emerg. Med.* **1999**, *11*, 175–181. [\[CrossRef\]](#)
40. 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*, e3422415. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Drinkwater, J.J.; Kalantary, A.; Turner, A.W. A systematic review of diabetic retinopathy screening intervals. *Acta Ophthalmol.* **2023**, *102*, e473–e484. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Ta, A.W.A.; Goh, H.L.; Ang, C.; Koh, L.Y.; Poon, K.; Miller, S.M. Two Singapore public healthcare AI applications for national screening programs and other examples. *Health Care Sci.* **2022**, *1*, 41–57. [\[CrossRef\]](#) [\[PubMed\]](#)

**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.