

# Assessing Diabetic Retinopathy Staging With AI: A Comparative Analysis Between Pseudocolor and LED Imaging

Maria Vittoria Cicinelli<sup>1,2</sup>, Salvatore Gravina<sup>2</sup>, Carola Rutigliani<sup>2</sup>, Lisa Checchin<sup>1</sup>, Lamberto La Franca<sup>1</sup>, Rosangela Lattanzio<sup>1</sup>, and Francesco Bandello<sup>1,2</sup>

<sup>1</sup> Department of Ophthalmology, IRCCS San Raffaele Hospital, Milan, Italy

<sup>2</sup> School of Medicine, Vita-Salute San Raffaele University, Milan, Italy

**Correspondence:** Maria Vittoria Cicinelli, Department of Ophthalmology, IRCCS San Raffaele Hospital, Via Olgettina 60, Milan 20132, Italy. e-mail: [cicinelli.mariavittoria@hsr.it](mailto:cicinelli.mariavittoria@hsr.it)

**Received:** December 2, 2023

**Accepted:** February 4, 2024

**Published:** March 15, 2024

**Keywords:** diabetic retinopathy; artificial intelligence; diagnosis

**Citation:** Cicinelli MV, Gravina S, Rutigliani C, Checchin L, La Franca L, Lattanzio R, Bandello F. Assessing diabetic retinopathy staging with AI: A comparative analysis between pseudocolor and LED imaging. *Transl Vis Sci Technol.* 2024;13(3):11, <https://doi.org/10.1167/tvst.13.3.11>

**Purpose:** To compare the diagnostic performance of artificial intelligence (AI)-based diabetic retinopathy (DR) staging system across pseudocolor, simulated white light (SWL), and light-emitting diode (LED) camera imaging modalities.

**Methods:** A cross-sectional investigation involved patients with diabetes undergoing imaging with an iCare DRSplus confocal LED camera and an Optos confocal, ultra-widefield pseudocolor camera, with and without SWL. Macula-centered and optic nerve-centered 45 × 45-degree photographs were processed using EyeArt v2.1. Human graders established the ground truth (GT) for DR severity on dilated fundus exams. Sensitivity and weighted Cohen's weighted kappa ( $w\kappa$ ) were calculated. An ordinal generalized linear mixed model identified factors influencing accurate DR staging.

**Results:** The study included 362 eyes from 189 patients. The LED camera excelled in identifying sight-threatening DR stages (sensitivity = 0.83, specificity = 0.95 for proliferative DR) and had the highest agreement with the GT ( $w\kappa$  = 0.71). The addition of SWL to pseudocolor imaging resulted in decreased performance (sensitivity = 0.33, specificity = 0.98 for proliferative DR;  $w\kappa$  = 0.55). Peripheral lesions reduced the likelihood of being staged in the same or higher DR category by 80% ( $P < 0.001$ ).

**Conclusions:** Pseudocolor and LED cameras, although proficient, demonstrated non-interchangeable performance, with the LED camera exhibiting superior accuracy in identifying advanced DR stages. These findings underscore the importance of implementing AI systems trained for ultra-widefield imaging, considering the impact of peripheral lesions on correct DR staging.

**Translational Relevance:** This study underscores the need for artificial intelligence-based systems specifically trained for ultra-widefield imaging in diabetic retinopathy assessment.

## Introduction

Diabetic retinopathy (DR) is a significant global public health concern, demanding continuous advancements in detection and screening strategies.<sup>1</sup> The emergence of artificial intelligence (AI)-driven automated retinal image analysis software (ARIAS) holds great promise for enhancing early detection, streamlining screening initiatives, and promoting patient adherence.<sup>2,3</sup> Among these innovations, EyeArt

v2.1, developed by Eyenuk (Woodland Hills, CA), has secured U.S. Food and Drug Administration (FDA) approval for its automated detection of DR and diabetic macular edema (DME).

Although ARIAS, including EyeArt v2.1, has demonstrated efficacy and comparability with expert graders in the context of conventional color fundus photographs, a critical gap exists in our understanding of its performance across larger fields of view. Ultra-widefield (UWF) imaging, offered by technologies such as those of Optos (Dunfermline, UK), is

being increasingly utilized in clinical and population-based screenings due to its ability to explore the far retinal periphery, a crucial aspect in determining DR severity.<sup>4,5</sup> On the other hand, the landscape of available retinal cameras introduces challenges, including disparities in image resolutions and color discrimination among various manufacturers.<sup>6,7</sup> These discrepancies can potentially influence the performance of ARIAS, especially in referable DR (RDR) detection.

The primary objective of this study was to compare the detection rates of DR and DME using an FDA-approved, true-color, light-emitting diode (LED) camera and pseudocolor UWF imaging, both with and without simulated white light (SWL), in a cohort of patients with diabetes. Additionally, our study aimed to estimate the agreement between these imaging techniques and investigate factors potentially influencing the accurate staging of DR.

## Methods

A cross-sectional study was conducted involving adults ( $\geq 18$  years old) with diabetes mellitus (DM) prospectively enrolled at the Medical Retina Unit of the Department of Ophthalmology at San Raffaele Hospital (Milan, Italy) from April 2022 to January 2023. Participants were referred for screening or management of DR or DME and were included without case selection. Patients with concurrent retinal conditions that might confound DR detection (e.g., age-related macular degeneration, vein occlusion) and those with visible retinal laser scars in ARIAS-analyzed images were excluded. Both eyes of the same patient were included if eligible.

The study adhered to the tenets of the Declaration of Helsinki, and ethical approval was obtained from the local ethics committee (protocol: OCTA-MIMS). Written informed consent was provided by all subjects.

## Data Collection and Imaging

Demographic information, slit-lamp examination findings, and comprehensive medical and ocular history were collected for each participant, including DM type and duration, glycosylated hemoglobin (HbA1c, %), and prior ocular treatments. Imaging involved two  $45 \times 45$ -degree retinal photographs (centered on the optic disc and macula) captured by an iCare DRsplus confocal LED camera (CenterVue, Padua, Italy) after pupil dilation. Additionally, imaging with an Optos Silverstone confocal UWF fundus camera was performed, with pseudocolor and

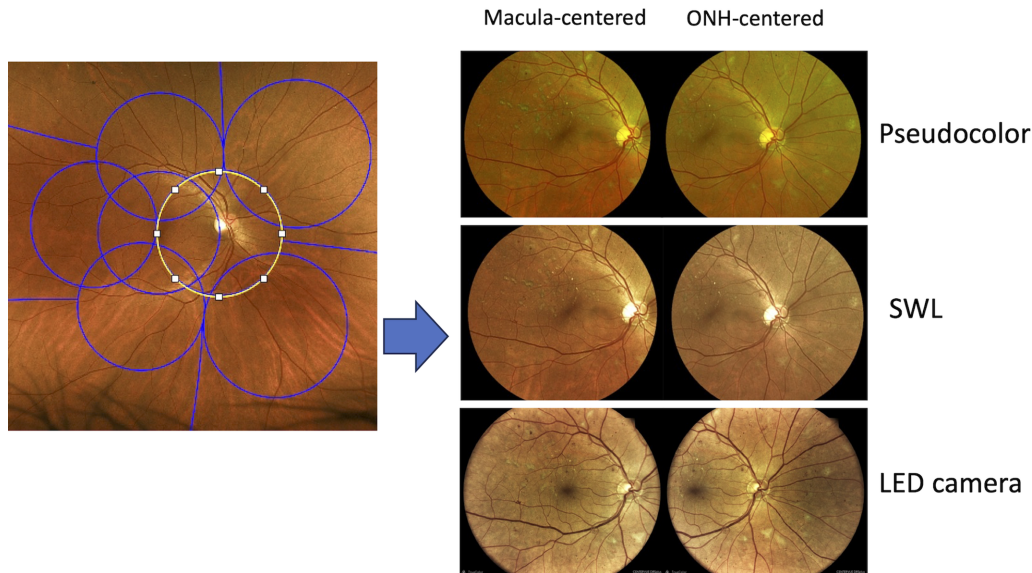
SWL versions of the same picture automatically being saved. Macula-centered optical coherence tomography (OCT) raster imaging (SPECTRALIS OCT; Heidelberg Engineering, Heidelberg, Germany) was also obtained.

## Imaging Processing

From each UWF photo, two 45-degree circular images were extracted, one centered on the macula and one on the optic nerve, to align with LED camera-acquired images. Precise delineation of these regions of interest (ROIs) utilized the Early Treatment Diabetic Retinopathy Study (ETDRS) map as a guide during the UWF image cropping process, and they were overlaid using the embedded mask tool in the Optos software (Fig. 1). Two distinct images were exported in JPEG format for each original UWF image—one with the ETDRS map overlay and one without. After ensuring consistent zoom levels, these images were imported into ImageJ (National Institutes of Health, Bethesda, MD) for processing. Circular ROIs overlapping with optic nerve ETDRS field 1 were selected and enlarged by 1.5 times using the Edit > Selection > Enlarge option to standardize ROI sizes to 45 degrees. This facilitated a robust comparative analysis between UWF and LED camera-acquired photographs, a process repeated for macular ETDRS field 2 and both pseudocolor and SWL images. Images were saved with a minimum resolution of  $1300 \times 1300$  pixels to meet the ARIAS program criteria.

## Imaging Analysis

For each eye, a set of six images (pseudocolor macula, pseudocolor optic nerve, SWL macula, SWL optic nerve, LED camera macula, and LED camera optic nerve) was imported into the FDA-approved ARIAS (Eyenuk EyeArt v2.1.0). This cloud-based platform provided disease or no-disease outcomes based on the presence of DR signs, along with a score for DR severity on an international clinical diabetic retinopathy (ICDR) scale: no diabetic retinopathy (no DR), mild non-proliferative DR (mild NPDR), moderate NPDR, severe NPDR, and proliferative DR (PDR).<sup>8</sup> A continuous DR score, ranging from 0 to 5 on the ICDR scale, was also generated, and indications of the presence or absence of DME signs were provided. For the ground truth (GT), dilated fundus exams were independently performed by two trained retinal specialists (SG, MVC), and DR was graded using the same ICDR scale. Any retinopathy more severe than mild DR or with DME was labeled as RDR.



**Figure 1.** UWF image processing workflow. Two 45-degree circular images were extracted from each UWF photo, one centered on the macula and the other on the optic nerve. Precise delineation of these ROIs involved leveraging the ETDRS map as a guiding reference during the UWF image cropping process (left image). Circular ROIs, corresponding to optic nerve ETDRS field 1 and macular ETDRS field 2, were selected and enlarged by 1.5 times. This process was iterated for both pseudocolor and SWL images. The resultant images were aligned with LED camera-acquired images.

The readers also annotated the presence of microaneurysms, hard exudates, or an epiretinal membrane in the macula and predominantly peripheral lesions (PPLs) in the periphery, defined as lesions mostly or only present outside the seven ETDRS fields.<sup>9</sup> Spectral-domain OCT was analyzed for sub- or intraretinal fluid in the central macula. The agreement between the two readers was calculated with Cohen's unweighted kappa statistic ( $\kappa$ ). In uncertain or discordant cases, a third senior reader (RL) adjudicated the final grade.

## Statistical Analysis

Statistical analysis was conducted using R 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria). Descriptive statistics were utilized to summarize patient demographics. As a primary outcome, heatmap representations and confusion matrices were generated to visualize the distribution of AI grading discrepancies between each imaging modality against the GT. These matrices, representing the distribution of true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) cases, were instrumental in calculating sensitivity, specificity, and overall accuracy for each modality. The accuracy was quantified using the formula:

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$

A Cohen's weighted kappa statistic ( $w\kappa$ ) was computed to evaluate the agreement between each pair of measures (GT vs. pseudocolor, SWL, or LED imaging), with interpretation guided by the Landis and Koch guidelines.<sup>10</sup> The sample size was determined based on the prior estimate of ARIAS diagnostic test sensitivity (sensitivity = 91.3%, according to Bhaskaranand et al.<sup>11</sup>) and considering a prevalence of any DR of 63% in our sample. The sample size was calculated to ensure that we could be 95% confident that our estimates of sensitivity were within a margin of 0.07 of the true sensitivity value for the population. This precision reflects the allowable error range in our estimation process. The calculated sample size was  $n = 250$ .

The second objective involved assessing the agreement among the LED, pseudocolor, and SWL imaging using an ordinal generalized linear mixed model (GLMM) with crossed random effects. This GLMM incorporated crossed random effects for eye identification numbers and the technique used. The mean bias in quantitative continuous DR severity score among the three modalities was calculated using Bland-Altman plots. Similar statistics were performed with the presence of RDR and DME as the main outcomes.

As a third outcome, the study investigated factors influencing the performance of AI-based diagnostic techniques. The same GLMM with crossed random effects was used but now included the GT measure as a covariate. The interpretation of the coefficients in

the ordinal logistic regression pertained to the odds of receiving an equal or higher DR staging while holding the GT fixed. All values were computed based on the results for each eye (eye level). Standard error (SE) and 95% confidence intervals (CIs) were reported as measures of uncertainty.

## Results

Four hundred eyes from 210 patients were imaged; 38 eyes were excluded (24 had ungradable ARIAS readings, 11 had concurrent retinal diseases, and three had macular laser scars). A total of 362 eyes from 189 patients were included in the study. The mean age was  $59.2 \pm 15.2$  years, with a balanced distribution between genders (144 females, 40%). The average duration of diabetes was  $16.7 \pm 12.0$  months, and the mean HbA1c level was  $7.16\% \pm 1.26\%$ . The study population included mainly patients with type 2 DM (241 eyes, 67%). The distribution of DR severity stages according to the ICDR scale is presented in Table 1. The agreement between the two readers was almost perfect ( $\kappa = 0.95$ ; 95% CI, 0.92–0.97), with only 14 cases requiring adjudication by a third reviewer.

**Table 1.** Distribution of Demographic and Clinical Characteristics of the Patients

	Overall (N = 362 Eyes)
Age (y)	
Mean $\pm$ SD	59.2 $\pm$ 15.2
Median (min, max)	61.0 (18.0, 90.0)
Female gender, n (%)	144 (40)
DM type 2, n (%)	241 (67)
DM duration (mo)	
Mean $\pm$ SD	16.7 $\pm$ 12.0
Median (min, max)	15.0 (1.00, 45.0)
HbA1c (%)	
Mean $\pm$ SD	7.16 $\pm$ 1.26
Median (min, max)	7.00 (5.00, 14.0)
Pseudophakia, n (%)	61 (17)
DR grading, n (%)	
No DR	131 (36)
Mild NPDR	47 (13)
Moderate NPDR	77 (21)
Severe NPDR	80 (22)
PDR	27 (8)
DME, n (%)	99 (27)
RDR, n (%)	192 (57)

## Detection Rate of DR

The overall diagnostic performance of the AI algorithm in detecting RDR was the highest for pseudocolor (sensitivity = 0.77, specificity = 0.92) and LED (sensitivity = 0.75, specificity = 0.91) and was lowest for SWL (sensitivity = 0.70, specificity = 0.85). The overall accuracy values for RDR detection were 0.86 (95% CI, 0.82–0.90) for pseudocolor, 0.85 (95% CI, 0.80–0.89) for LED, and 0.79 (95% CI, 0.74–0.83) for SWL.

The distribution of grading discrepancies between AI and GT across pseudocolor, SWL, and LED imaging is illustrated in Figure 2. Pseudocolor and SWL tended to overestimate DR severity for no-DR cases (pseudocolor sensitivity = 0.63, specificity = 0.92; SWL sensitivity = 0.53, specificity = 0.90; LED sensitivity = 0.75, specificity = 0.91). SWL had the lowest performance for detecting severe NPDR (pseudocolor sensitivity = 0.73, specificity = 0.93; SWL sensitivity = 0.51, specificity = 0.93; LED sensitivity = 0.72, specificity = 0.91), and the LED camera had the highest performance in detecting PDR (pseudocolor sensitivity = 0.78, specificity = 0.96; SWL sensitivity = 0.33, specificity = 0.98; LED sensitivity = 0.83, specificity = 0.95) (Supplementary Fig. S1).

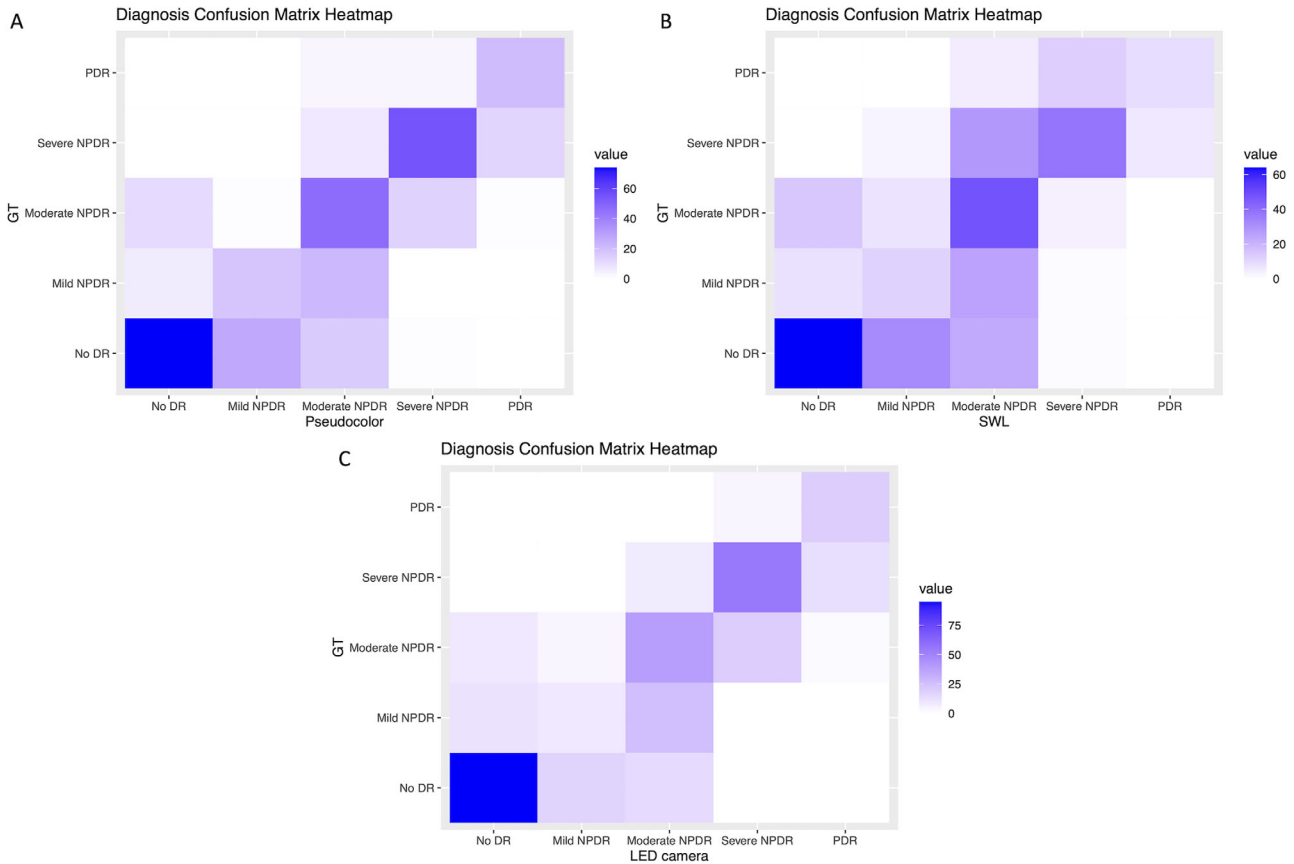
The agreement between AI and GT grading across different DR severity stages was substantial for pseudocolor ( $\kappa = 0.69$ , asymptotic SE = 0.03) and LED ( $\kappa = 0.71$ , asymptotic SE = 0.02) and moderate for SWL ( $\kappa = 0.55$ , asymptotic SE = 0.03).

## Agreement Among Pseudocolor, SWL, and LED Imaging

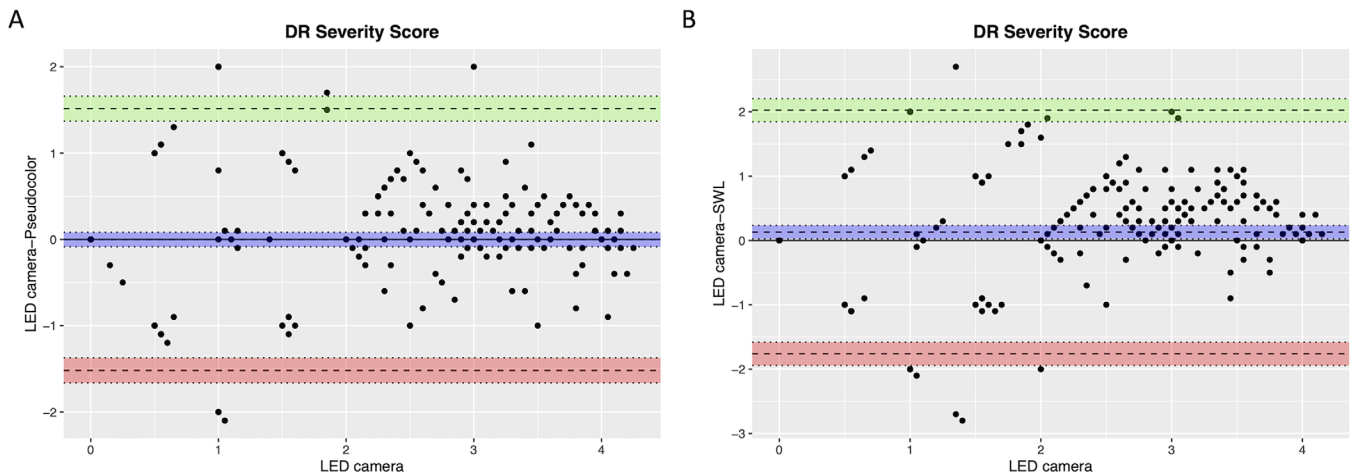
The model-based kappa agreement rate among the three instruments was substantial ( $\kappa = 0.65$ , SE = 0.01). The Bland–Altman graph, plotting the continuous-scale DR severity scores, showed a mean bias of  $-0.002$  (95% CI, 0.08 to  $-0.09$ ) between the LED camera and the pseudocolor, with higher agreement for more severe NPDR and PDR and lower agreement for no-DR and mild DR (Fig. 3A). A similar trend was seen comparing LED with SWL, with a higher bias (bias = 0.13; 95% CI, 0.24–0.03) and wider limits of agreement (Fig. 3B).

## Agreement for DME

The prevalence of DME on OCT was 27% (99 eyes). The overall accuracy for DME detection was 0.79 (95% CI, 0.74–0.84) for pseudocolor, 0.79 (95% CI, 0.74–0.84) for SWL, and 0.84 (95% CI, 0.79–0.88) for



**Figure 2.** Heatmaps of DR severity distribution. This figure presents heatmaps offering a visual representation of the distribution of AI-based grading discrepancies across various imaging modalities (pseudocolor, SWL, and LED) compared to the GT. The color gradients in the heatmaps correspond to the magnitude of specific DR severity levels.



**Figure 3.** Bland–Altman analysis of DR severity scores. **(A)** The Bland–Altman graph shows continuous-scale DR severity scores, revealing a mean bias of  $-0.002$  (95% CI, 0.08 to  $-0.09$ ) between the LED camera and pseudocolor imaging. Notably, higher agreement was observed for more severe DR, whereas lower agreement is evident in cases with no DR and mild DR. **(B)** Comparison of the LED camera with SWL indicates a higher bias (bias = 0.13; 95% CI, 0.24–0.03) and wider limits of agreement. This graph visually represents similar variations in agreement across different DR severity levels.

**Table 2.** Factors Influencing the Accuracy of AI-Based Diagnostic Techniques in DR Severity Staging

Factor	OR	95% CI	P
Age (for 10-yr increase)	1.34	0.81–1.69	0.3
Female vs. male gender	1.24	0.33–3.02	0.1
Type 2 DM vs. type 1 DM	1.01	0.48–8.50	0.9
Duration of DM (for 10 yr)	0.90	0.61–1.60	0.2
Microaneurysms vs. none	2.34	1.51–3.64	<0.001*
Exudates vs. none	2.66	1.78–3.96	<0.001*
PPLs vs. no PPLs	0.20	0.13–0.31	<0.001*
ERM vs. no ERM	1.03	0.59–1.79	0.9
Pseudophakia vs. phakia	1.16	0.78–1.71	0.5

The coefficients in the ordinal logistic regression represent the odds of receiving a DR staging equal to or higher while keeping the ground truth fixed.

\*Statistically significant.

LED. The model-based kappa agreement rate between the three modalities was only moderate ( $\kappa = 0.58$ , SE = 0.01).

### Factors Associated With AI-Based Diagnostic Performance

Factors associated with varying performance of AI-based diagnostic techniques in diagnosing and rating DR severity were explored. The presence of macular microaneurysms (odds ratio [OR] = 2.34;  $P < 0.001$ ) and hard exudates (OR = 2.66;  $P < 0.001$ ) was associated with a higher chance of being staged as equal or more severe DR. Conversely, the presence of PPLs (OR = 0.20;  $P < 0.001$ ) was associated with an 80% reduction in the chance of being staged in the same or higher category. Age, gender, type, and duration of DM; history of cataract surgery; and the presence of epiretinal membrane did not impact the chance of DR recognition (Table 2).

### Discussion

Automated retinal image analysis systems have demonstrated remarkable potential in enhancing the efficiency, cost-efficiency, and accessibility of DR screening programs.<sup>12</sup> The deployment of AI algorithms has been associated with high sensitivity and specificity in detecting RDR based on conventional color fundus photographs.<sup>13–15</sup> However, the integration of AI models into other existing camera systems, especially UWF imaging, poses significant technical challenges that demand careful considerations,<sup>16</sup> and the perfor-

mance of these algorithms across different imaging modalities remains an area of limited understanding. Our study aimed to bridge this gap by comparing the performance of the AI algorithm across LED imaging and pseudocolor modalities, both with and without SWL.

The Optos technology employs low-powered laser wavelengths (green at 532 nm and red at 633 nm) that scan simultaneously. This approach allows for a detailed review of the retinal substructures, with the green laser focusing on the sensory retina and the red laser scanning deeper from the retinal pigment epithelium to the choroid. However, it is important to note that this technology produces pseudocolor images that are a representation rather than a direct capture of the retinal coloration. On the other hand, the iCare DRSplus confocal system uses a white LED fundus camera that captures a well-balanced color image. This balance is attributed to the barycenter position of the images, which is typically located near the center of the RGB chromaticity space.<sup>6</sup> The Optos system, through its use of SWL, attempts to reconstruct the white light image from its laser-based scans. Although this allows for detailed structural analysis, especially in deeper layers due to the specific wavelengths used, the resulting images may not fully replicate the color accuracy and richness seen in images taken with traditional white light sources.

Overall, we found that the AI algorithm EyeArt v2.1 exhibited robust performance in RDR detection across both pseudocolor and LED imaging, with substantial agreement between the different modalities. However, our analysis revealed differences when disaggregating data according to DR severity. The LED camera excelled in detecting advanced stages of DR (namely, severe NPDR and PDR). Conversely, pseudocolor imaging tended to overestimate DR severity in cases with no DR or mild DR, emphasizing the need for careful interpretation in the clinical context.

Although our study reported high diagnostic performance values, they were comparatively lower than previously reported. For example, a large prospective validation study in the United Kingdom, involving over 30,000 patients and using EyeArt v2.1, reported a sensitivity of 95.7% and a specificity of 54.0% for triaging RDR.<sup>17</sup> Another prospective study validated EyeArt in multiple primary care centers in the United States for detecting more than mild DR, finding a sensitivity of 95.5%.<sup>18</sup> However, it is essential to note that all of these prior studies used conventional fundus photography as the GT, whereas we utilized dilated fundus examination to establish DR severity levels. Dilated fundus exam, with a wider field of view of the

retina, likely includes peripheral lesions often missed by traditional retinal imaging.<sup>19</sup> This suggestion aligns with previous studies indicating only fair agreement between AI-based systems coupled with traditional fundus cameras and human-graded UWF fundus imaging.<sup>20</sup> When Olvera-Barrios et al.<sup>21</sup> compared DR grades derived from two imaging platforms, one providing a single 60-degree and the other providing a standard 45-degree field of view, they found that the wider angle camera identified a larger number of patients with referable disease that were not detected with standard imaging.

Comparisons with previous studies also underscore the impact of image resolution and chromaticity on AI-based DR recognition. Sarao et al.<sup>22</sup> compared the diagnostic performance of EyeArt between a conventional flash fundus camera and a white LED confocal scanner and found superior diagnostic performance of the latter. The unique characteristics of LED imaging, including improved color discrimination and higher resolution, may contribute to its superiority in identifying subtle retinal changes in different DR stages. More recent studies have corroborated the use of LED cameras for AI-based DR screening. Olvera-Barrios et al.<sup>15</sup> compared white LED cameras with traditional fundus cameras in 1257 patients with diabetes from the UK National Health Service Diabetic Eye Screening Programme and reported similar diagnostic accuracy between the two modalities. Similarly, Wongchaisuwat et al.<sup>23</sup> prospectively validated the use of a white LED camera for DR screening, as they achieved a sensitivity of 92% and a specificity of 84% for LED photographs in detecting RDR.

Limited studies have explored the accuracy of pseudocolor imaging in AI-based DR recognition, and our study adds insights to this growing body of knowledge. Wang and colleagues<sup>24</sup> evaluated the effectiveness of EyeArt software for detecting RDR in Asian Indian patients using pseudocolor images. The authors observed good sensitivity (90.3% on an eye level) but moderate specificity (53.6% on an eye level), possibly related to the fact that pseudo red/green coloration may have misled the algorithm in detecting RDR. Both the Optos and LED cameras use a confocal system that blocks light reflected from out-of-focus layers, but the LED camera provides color images with richer color content. Moreover, although the LED camera has an average barycenter position (red, 0.412; green, 0.314; blue, 0.275) closest to the center of the chromaticity space, pseudocolor images show a complete absence of the blue channel.<sup>7</sup> An additional imaging tool, SWL (488–633 nm), simulates white light on R/G images. To date, a comparison among pseudocolor, SWL, and LED imaging has been missing. Our data showed

that the addition of SWL to pseudocolor imaging had a negative impact on the AI algorithm performance. The decrease in overall accuracy, sensitivity, and specificity suggests that incorporating SWL may introduce confounding factors that affect the ability of the algorithm to accurately grade DR severity. Further investigation is warranted to understand the specific challenges that SWL poses and to optimize AI algorithms for this imaging modality.

The EyeArt system stands out as a highly validated AI technology designed for the autonomous detection of RDR, specifically indicated for use with Canon CR-2 AF and CR-2 Plus AF digital retinal cameras (Canon, Tokyo, Japan) and the Topcon TRC-NW400 non-mydratic retinal camera (Topcon, Tokyo, Japan).<sup>25</sup> Given its comparable performance with the iCare DRSpplus, this study potentially provides preliminary support for the validation of EyeArt with DRSpplus. The findings presented herein could serve as a foundation for establishing the effectiveness of EyeArt in conjunction with DRSpplus, validating its diagnostic capabilities across different imaging platforms. Additionally, it might be worthwhile to consider implementing EyeArt training with DRSpplus images. The potential improvement in performance if DRSpplus images were used for EyeArt training is a theoretical consideration that should be explored, as it could further emphasize its utility with diverse fundus cameras.

As an ancillary analysis, we explored factors influencing the diagnostic performance of AI, revealing that the presence of macular microaneurysms and hard exudates increased the likelihood of DR being staged as equal or more severe DR. On the other hand, the presence of PPLs was associated with a reduced chance of correct staging. Notably, substantial diabetic retinal pathology may exist in the retinal periphery, extending beyond the seven ETDRS standard fields.<sup>9</sup> PPLs potentially signal higher risk for DR progression, irrespective of baseline DR severity score.<sup>26</sup> Moreover, studies using UWF fluorescein angiography have highlighted a link between PPLs and peripheral nonperfusion in diabetic eyes.<sup>27</sup> Our findings emphasize the importance of UWF imaging in detecting PPLs that might be overlooked by standard field imaging. Given the rising interest in incorporating UWF imaging into large-population screening programs,<sup>28</sup> our study advocates for the inclusion of UWF imaging and the development of AI models specifically trained on this modality.

Limitations of this study include reliance on data from a single center, which may limit the generalizability of our findings across diverse populations. Additionally, the hospital-based sample may not fully

represent the broader community due to the specific patient demographics encountered in a tertiary referral center. This setting contributed to an uneven distribution of subgroups within the various severity levels of DR, potentially impacting the uniformity of our statistical analyses. Nevertheless, to minimize selection bias, our study design included all consecutive patients who met the inclusion criteria over the study period, resulting in subgroup sizes that naturally reflect the prevalence rates of each DR severity stage within our patient population.

Using pseudocolor imaging without the blue channel might impact the performance of the algorithm, and the recent implementation of blue light in Optos systems could potentially improve the ARIAS diagnostic performance. Furthermore, our use of the DRSpplus LED camera, providing 45-degree images, differs from prior studies that have used 60-degree images,<sup>21–23</sup> making direct comparisons challenging. All patients in our study were referred to the retina clinic, and pupil dilation plausibly increased imageability. Nonetheless, EyeArt demonstrated consistent sensitivities and specificities for RDR detection in both dilated and non-dilated pupils.<sup>18</sup> The impact of image quality, artifacts, and other confounding factors on algorithm performance was not systematically investigated. Finally, although the weighted ordinal kappa coefficient provides a nuanced measure of agreement for our ordinal data, it is important to acknowledge that the smaller sample size, particularly in the PDR subgroup, may have introduced variability and affected the robustness of this statistic, necessitating cautious interpretation of these specific results within the broader context of our study.

In conclusion, our study contributes valuable insights into the performance of an AI-based DR staging system across different imaging modalities. The AI-based system exhibited commendable performance on both LED camera true color and pseudocolor camera images. However, these modalities were not interchangeable, as the LED camera outperformed with regard to identifying more severe and sight-threatening stages of DR. The addition of SWL to pseudocolor imaging resulted in lower performance. The findings highlight the importance of tailoring AI algorithms to the specific features of each imaging platform, considering factors such as color discrimination, image resolution, and the presence of SWL. The implementation of AI systems trained on UWF imaging may further enhance the accuracy of DR staging by capturing subtle changes in the retinal periphery, significantly contributing to accurate prognosis and referral decisions.

## Acknowledgments

Supported by Eyenuk (Woodland Hills, CA) and CenterVue (Padua, Italy) for providing the necessary resources to conduct this research.

Disclosure: **M.V. Cicinelli**, None; **S. Gravina**, None; **C. Rutigliani**, None; **L. Checchin**, None; **L. La Franca**, None; **R. Lattanzio**, Abbvie (C), Bayer (C), Novartis (C), SIFI (C); **F. Bandello**, Allergan (C), Bayer (C), Boehringer Ingelheim (C), Fidia (C), Novartis (C), NTC Pharma (C), Oxurion (C), Roche (C), Sanofi-Aventis (C), SIFI (C), ZEISS (C)

## References

1. GBD 2019 Blindness and Vision Impairment Collaborators; Vision Loss Expert Group of the Global Burden of Disease Study. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Glob Health*. 2021;9(2):e144–e160.
2. Nguyen HV, Tan GSW, Tapp RJ, et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. *Ophthalmology*. 2016;123(12):2571–2580.
3. Liu J, Gibson E, Ramchal S, et al. Diabetic retinopathy screening with automated retinal image analysis in a primary care setting improves adherence to ophthalmic care. *Ophthalmol Retina*. 2021;5(1):71–77.
4. Byberg S, Vistisen D, Diaz L, et al. Optos wide-field imaging versus conventional camera imaging in Danish patients with type 2 diabetes. *Acta Ophthalmol*. 2019;97(8):815–820.
5. Silva PS, Cavallerano JD, Sun JK, Noble J, Aiello LM, Aiello LP. Nonmydriatic ultrawide field retinal imaging compared with dilated standard 7-field 35-mm photography and retinal specialist examination for evaluation of diabetic retinopathy. *Am J Ophthalmol*. 2012;154(3):549–559.e2.
6. Sarao V, Veritti D, Borrelli E, Sarda SVR, Poletti E, Lanzetta P. A comparison between a white LED confocal imaging system and a conventional flash fundus camera using chromaticity analysis. *BMC Ophthalmol*. 2019;19(1):231.
7. Fantaguzzi F, Servillo A, Sacconi R, Tombolini B, Bandello F, Querques G. Comparison of peripheral extension, acquisition time, and image chromaticity in wide-field imaging systems. *Ophthalmol Retina*. 2021;5(1):71–77.

- maturity of Optos, Clarus, and EIDON systems. *Graefes Arch Clin Exp Ophthalmol.* 2023;261(5):1289–1297.
8. Wilkinson CP, Ferris FL, 3rd, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology.* 2003;110(9):1677–1682.
  9. Silva PS, Cavallerano JD, Sun JK, Soliman AZ, Aiello LM, Aiello LP. Peripheral lesions identified by mydriatic ultrawide field imaging: distribution and potential impact on diabetic retinopathy severity. *Ophthalmology.* 2013;120(12):2587–2595.
  10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174.
  11. Bhaskaranand M, Ramachandra C, Bhat S, et al. The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes. *Diabetes Technol Ther.* 2019;21(11):635–643.
  12. Tufail A, Rudisill C, Egan C, et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology.* 2017;124(3):343–351.
  13. Srinivasan R, Surya J, Ruamviboonsuk P, Chotcomwongse P, Raman R. Influence of different types of retinal cameras on the performance of deep learning algorithms in diabetic retinopathy screening. *Life (Basel).* 2022;12(10):1610.
  14. Tufail A, Kapetanakis VV, Salas-Vega S, et al. An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health Technol Assess.* 2016;20(92):1–72.
  15. Olvera-Barrios A, Heeren TF, Balaskas K, et al. Diagnostic accuracy of diabetic retinopathy grading by an artificial intelligence-enabled algorithm compared with a human standard for wide-field true-colour confocal scanning and standard digital retinal images. *Br J Ophthalmol.* 2021;105(2):265–270.
  16. Grzybowski A, Singhanetr P, Nanegrungsunk O, Ruamviboonsuk P. Artificial intelligence for diabetic retinopathy screening using color retinal photographs: from development to deployment. *Ophthalmol Ther.* 2023;12(3):1419–1437.
  17. Heydon P, Egan C, Bolter L, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30,000 patients. *Br J Ophthalmol.* 2021;105(5):723–728.
  18. Ipp E, Liljenquist D, Bode B, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referable and vision-threatening diabetic retinopathy. *JAMA Netw Open.* 2021;4(11):e2134254.
  19. Price LD, Au S, Chong NV. Optomap ultrawide field imaging identifies additional retinal abnormalities in patients with diabetic retinopathy. *Clin Ophthalmol.* 2015;9:527–531.
  20. Sedova A, Hajdu D, Datlinger F, et al. Comparison of early diabetic retinopathy staging in asymptomatic patients between autonomous AI-based screening and human-graded ultrawidefield colour fundus images. *Eye (Lond).* 2022;36(3):510–516.
  21. Olvera-Barrios A, Heeren TF, Balaskas K, et al. Comparison of true-colour wide-field confocal scanner imaging with standard fundus photography for diabetic retinopathy screening. *Br J Ophthalmol.* 2020;104(11):1579–1584.
  22. Sarao V, Veritti D, Lanzetta P. Automated diabetic retinopathy detection with two different retinal imaging devices using artificial intelligence: a comparison study. *Graefes Arch Clin Exp Ophthalmol.* 2020;258(12):2647–2654.
  23. Wongchaisuwat N, Trinavarat A, Rodanant N, et al. In-person verification of deep learning algorithm for diabetic retinopathy screening using different techniques across fundus image devices. *Transl Vis Sci Technol.* 2021;10(13):17.
  24. Wang K, Jayadev C, Nittala MG, et al. Automated detection of diabetic retinopathy lesions on ultrawidefield pseudocolour images. *Acta Ophthalmol.* 2018;96(2):e168–e173.
  25. Eyenuk. EyeArt AI eye screening system. Available at: <https://www.eyenuk.com/en/products/eyeart/>. Accessed March 8, 2023.
  26. Silva PS, Cavallerano JD, Haddad NM, et al. Peripheral lesions identified on ultrawide field imaging predict increased risk of diabetic retinopathy progression over 4 years. *Ophthalmology.* 2015;122(5):949–956.
  27. Silva PS, Dela Cruz AJ, Ledesma MG, et al. Diabetic retinopathy severity and peripheral lesions are associated with nonperfusion on ultrawide field angiography. *Ophthalmology.* 2015;122(12):2465–2472.
  28. Cui T, Lin D, Yu S, et al. Deep learning performance of ultra-widefield fundus imaging for screening retinal lesions in rural locales. *JAMA Ophthalmol.* 2023;141(11):1045–1051.