Performance Assessment of an Artificial Intelligence System Used for Case Detection in a Large Community-based Diabetic Retinopathy Screening Program: A Retrospective Diagnostic Accuracy Validation Study

Maria Kristina DG. Eduardo, MD, MBA[1] and Junn R. Pajarillo, MD[2]

[1]Ilocos Training and Regional Medical Center, La Union, Philippines
[2]Department of Ophthalmology and Visual Sciences, Sentro Oftalmologico Jose Rizal, Philippine General Hospital, University of the Philippines Manila, Manila, Philippines

ABSTRACT

Background. In the Philippines, Artificial Intelligence (AI) systems used for diabetic retinopathy (DR) screening are still an emerging technology. There are several AI and Deep Learning Systems (DLS) developed and licensed abroad that are currently undergoing FDA approval in the Philippines. The challenge is to validate these systems on Filipino diabetic cohorts within the uncontrolled environment of a community-based setting before it could be adapted and integrated into routine DR screening programs locally.

Objective. This study aimed to determine the performance of an AI system as a population-based mass screening tool to detect referable cases of DR.

Methods. This study was designed as a retrospective diagnostic accuracy validation study using fundus images obtained from a community-based screening program by the Ilocos Training & Regional Medical Center Eye Mobile Team in the province of La Union. The diagnostic accuracy of the iCare Illume System (Thirona RetCAD + iCare DRS plus) in detecting any stage DR (ADR) and referable DR (RDR) was validated against a reference standard of two-field fundus photographs read by expert human graders (retina specialists). The diagnostic performance was evaluated using various metrics, including Accuracy, Sensitivity (SN), Specificity (SP), Negative Predictive Value (NPV), Positive Predictive Value (PPV), Negative Likelihood Ratio (LR-), and Positive Likelihood Ratios (LR+).

Results. A total of 284 eyes (142 diabetic patients, 717 images) were included in the study. Human expert graders classified 187 eyes (65.85%) as no apparent DR while 77 eyes (27.11%) as ADR. 63 eyes (22.18%) were considered RDR. The iCare Illume AI system achieved an accuracy of 83.21 % and 92.37% for ADR and RDR, respectively. The SN, SP, NPV, PPV, LR-, and LR+ for ADR were 92.11%, 79.57%, 96.1%, 64.81%, 0.10, and 4.50, respectively. The SN, SP, NPV, PPV, LR-, and LR+ for RDR were 79.03%, 96.50%, 93.69%, 87.50%, 0.22, and 22.58, respectively. Technical failure rate was 0.70% (99.3% Imageability).

Conclusion. The iCare Illume system exhibits strong diagnostic accuracy in screening for referable DR in our cohort of Filipino Diabetics. Although its sensitivity is marginally below the UK NICE guideline, the system demonstrates high specificity, negative predictive value, and positive predictive value. This combination highlights

CC BY NC ND

eISSN 2094-9278 (Online)
Published: September 15, 2026
https://doi.org/10.47895/amp.v60i17.13381
Copyright: The Author(s) 2026

Corresponding author: Maria Kristina DG. Eduardo, MD, MBA
Ilocos Training and Regional Medical Center
MacArthur Highway, Barangay Parian,
San Fernando, La Union 2500, Philippines
Email: maki.eduardo@gmail.com
ORCiD: https://orcid.org/0009-0006-7814-6589

its potential as a valuable tool for diabetic retinopathy screening in the Philippines. Future research should focus on refining local AI validation protocols, optimizing data sets of commercially available AI systems, assessing the impact on patient outcomes and disease burden, and evaluating cost-effectiveness.

Keywords: diabetic retinopathy, artificial intelligence, Icare Illume system, Mulat Mata DR Screening Program

INTRODUCTION

Diabetes Mellitus (DM) is estimated to affect 629 million people worldwide by 2045. In the Philippines, 4.3 million (7.1%) of Filipino adults have diabetes, and this number is expected to increase to 7.5 million (8.2%) by 2045.[1] Diabetic Retinopathy (DR), a microvascular complication of Diabetes Mellitus, poses a significant threat to vision and is considered an important public health concern in the Philippines. It is one of the leading causes of visual impairment and blindness in the country.[2]

Traditionally, eyes are screened for DR using various methods, commonly by indirect ophthalmoscopy or by fundus photography with or without pupil dilation. Stereoscopic color fundus photography of seven standard fields has been the basis for the development of the International Clinical Diabetic Retinopathy Severity Scale (DRSS) (Appendix A), which is the classification system most widely used today.[3] However, this method is time-consuming, cumbersome, and uncomfortable for patients. Modified two-field and three-field fundus photography were developed and are now more commonly used for general screening, as they represent a reasonable compromise between diagnostic accuracy, efficiency, and comfort.[4]

The development of DR and its complications follows a predictable pattern and sequential stages. Regular screening at predetermined intervals enables early detection and management, thereby preventing its chronic and often irreversible visual sequelae. Systematic eye screening programs have been implemented in Iceland, the United Kingdom, Singapore, and the United States, with the end goal of reducing the overall incidence and burden of the disease.[5] Despite achieving promising outcomes, these programs continue to incur substantial costs and necessitate a dedicated workforce supported by diabetes and specialized eye care services. Moreover, the success of these programs hinges on the comprehensive networking support provided by the government.

For decades, individual screening for DR has been conducted according to a rigid and repetitive schedule needing multiple examinations throughout the patient's life. However, the escalating prevalence of DR has become disproportionately high compared to the available number of screening ophthalmologists. Consequently, the burden of managing this condition is substantial, and not all patients receive the optimal care they require. It is estimated that 30-50% of patients with diabetes in Southeast Asia remain unscreened.[6]

This immense task of DR screening has sparked an interest in using Artificial Intelligence (AI) and Deep Learning Systems (DLS) as potential solutions to manage the existing and escalating backlog. Landmark work by Gulshan et al. demonstrated the feasibility of using deep learning algorithms to detect diabetic retinopathy from retinal fundus photos.[7] There is a large body of evidence demonstrating the high accuracy and reliability of various AI technologies integrated into screening programs.[8] These advancements have improved patient access to screening while simultaneously reducing the burden of disease and associated costs.

Amidst the competitive landscape of AI system development, the UK National Institute for Clinical Excellence (NICE) recommends that diagnostic screening tests for DR should exhibit a sensitivity of at least 80%, specificity of at least 95%, and a technical failure rate of ≤5% to ensure acceptable performance.[9] Nevertheless, several important technical issues arise that limit its full adaptation into current and mainstream screening practices.[10] One of these challenges is generalizability, as AI platforms are based on big data that is generated from specific populations that might not be applicable to other subpopulations.

In the Philippines, AI systems used for DR screening are still an emerging technology. There are several AI and DLS developed and licensed abroad that are locally available, although none are currently FDA-approved.[11] The challenge is to validate these systems on Filipino diabetic cohorts within the uncontrolled environment of a community-based setting before they can be adapted and integrated into routine DR screening programs.

The "Mulat Mata" (open eyes) DR Screening program of the Vitreoretina Society of the Philippines (VRSP) is an ongoing community screening initiative conducted in the Province of La Union (PLU). This collaborative project involves the VRSP, the Department of Health (DOH), the Local Government Units of La Union, and the Ilocos Training & Regional Medical Center (ITRMC). The objective of the program was to conduct point-of-care community screenings targeting 70% of diabetics within the PLU. Since its inception, the program has screened numerous diabetics. However, the recent introduction and arrival of AI-assisted retinal cameras have significantly enhanced the program's efficiency in terms of enabling it to accommodate a greater patient capacity per screening day, improve patient comfort, reduce overhead costs, and minimize manpower requirements.

This AI system, however, has not been validated in our cohort of Filipino diabetics, hence the need for this study. The potential breakthrough it may generate has the potential to revolutionize the way large-scale screenings for DR are conducted in the Philippines.

METHODS

This study assessed the diagnostic performance of a locally available AI-enabled camera system, the iCare Illume (Thirona RetCad algorithm integrated with the iCare DRSplus), during point-of-care community DR screenings in a Filipino diabetic cohort.

Specifically, the study investigated whether the iCare Illume achieved comparable diagnostic accuracy and validity when compared to a reference standard two-field 45º fundus color photography assessed by expert human graders. Diagnostic acceptability was evaluated using National Institute for Health and Care Excellence (UK NICE) guidelines for DR screening performance.

The primary outcome measures included Accuracy, Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Likelihood Ratios (LRs) for both Any DR (ADR) and Referable DR (RDR).

DR encompasses a spectrum of severity, ranging from mild, non-proliferative forms to more severe proliferative stages. The goal of community-based systematic DR screening programs is to identify Referable Diabetic Retinopathy (RDR), which necessitates referral for further management. For the purposes of this study, RDR was defined as either moderate NPDR or worse stage or with the presence of macular edema (central hard exudates).

In addition, the study evaluated the Technical Failure Rate (TFR) of the AI system in Diabetic Retinopathy (DR) screening. TFR represents the percentage of images that the AI cannot analyze due to technical issues. A high TFR significantly impacts the system's accuracy and reliability, limiting its utility for DR screening in real-world settings where high-volume image processing is crucial. A substantially high number of ungradable images results in an increased number of cases requiring manual review or re-imaging, which consumes valuable time for screening processes, or over-referrals, as these patients are automatically flagged and sent to centers for further evaluation.

While the goal of the study is to test the AI system in a real-world uncontrolled community setting, it is recognized that several factors may affect the quality of data being gathered, such as pupil dilation, external lighting, patient cooperation, and media opacities, among others. In addition, the history of systemic diabetes mellitus is "self-reported," meaning this is based on the patient's or relative's knowledge of his/her disease condition. Verification of diabetic status in unclear cases was done by history, physical examination, and medical records if available alone without the benefit of confirmatory laboratory tests.

The gold standard for DR image grading is based on a seven-field stereophotography technique employed by the modified Airlie House Classification. This method provides the most accurate and consistent grading and is generally regarded as the ground truth. Using a two-field photography method is a recognized limitation of this research, although some studies have demonstrated acceptable sensitivity and specificity (92-96% sensitivity, 89-97% specificity) when compared to the gold standard for the detection of RDR. For screening purposes, a two- or three-field photograph represents a suitable compromise offering reasonable accuracy, process efficiency, and improved patient comfort.[12]

It is also emphasized that the objective of the study is to determine the performance of the AI system as a population-based mass screening tool to detect referable cases of DR. There was no intention to evaluate the AI system as a definitive diagnostic test and medical basis to create a management plan.

Lastly, there are no local FDA guidelines or advisories on the use or approval of commercially available artificial intelligence systems in the Philippines. While the procedure itself poses little to no risk for the patient, its use by the screening team has always been with caution and complete supervision by a team of subspecialists who are present onsite and meticulously examine all output generated by the AI system with their own clinical evaluation.

Before the initiation of the study, the protocol was submitted to the Ilocos Training and Regional Medical Center Ethics Review Board. No patient was approached directly in the study. Patient data were accessed in compliance with standards of privacy and patient confidentiality in all stages of data collection and analysis.[13] De-identification of all patient data and randomization complied with institutional board review protocols.

All clinical data were kept confidential and anonymous in compliance with the Data Processing Guidelines on the Processing of Health Information (DOH Administrative Order No. 2020-0030) of the Philippines, designed to protect the privacy of individuals and to ensure that their health information is used and disclosed in a responsible manner.[14,15]

Randomized gradable ocular fundus photographs of patients included in the sample served as reference. The study was conducted in agreement with the principles of the Declaration of Helsinki.[16]

Study Design

This study was conducted as a comparative chart and image review of diabetic patients screened for DR by the Ilocos Training and Regional Medical Center (ITRMC) Eye Mobile Team as part of the Vitreoretina Society of the Philippines' "Mulat Mata" community-based DR community screening program for the province of La Union (LU). This study included patients seen from June to November 2023 across municipalities of La Union, namely Aringay, San Fernando City, San Gabriel, Bacnotan, Bauang, Bangar, Luna, Caba, and Balaoan. This study workflow is summarized in Figure 1.

Study Population

Eligible participants were aged 18 years or older with diabetes mellitus of any type, duration, or severity based

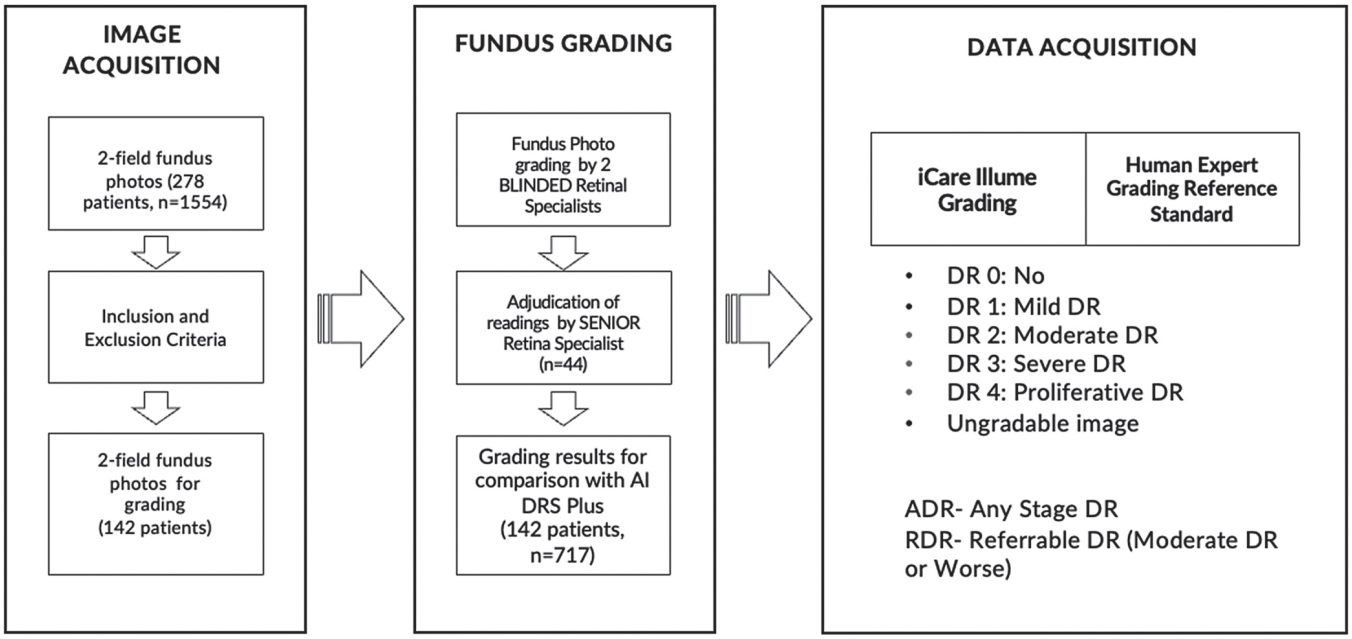


Figure 1. Image Acquisition Workflow, Inclusion and Exclusion Criteria, Fundus Grading and AI Data Acquisition.

on historical data and physical examination findings, with or without laboratory evidence of elevated blood glucose levels. Included participants must have undergone two-field fundus photography in either one or both eyes during the screening program. Patients with a known history of retinal intervention, including intravitreal anti-vascular endothelial growth factor (anti-VEGF) injections, laser ablative therapy, or vitreoretinal surgery, were excluded from the study.

Sample size computation was performed using Slovin's Formula with a 95% confidence interval and 5% margin of error. Based on the 2023 Mulat Mata DR Screening Program database, there were 3,698 eyes identified with diabetic retinopathy in La Union. The computed minimum sample size was 361 eyes, equivalent to approximately 722 fundus photographs when considering bilateral eye imaging analysis. The computations derived were as follows:

$n = N/(1+Ne^2)$
$n = N/(1+Ne^2)$
$n = 3698 / [1 + 3698(0.05)^2]$
$n = 361$ patients * 2 eyes
$N = 722$

$n = 361$ patients * 2 eyes

$N = 722$

Where:
n = sample size
N = population size
e = acceptable margin of error

Image Acquisition and AI-generated Reports

The study used an automated true color confocal fundus imaging system (iCare DRSplus, Finland) (Appendix B) with an integrated cloud-based screening software (iCare Illume, Finland) using the Thirona RetCAD v2.y.z (Thirona Retina B.V., Netherlands) artificial intelligence to generate high-resolution images and DR severity scoring (Appendices D and E). The camera system produces images with a resolution of 3600 x 2910 pixels with a 45° horizontal and 40° vertical field of view, while the AI-generated DR severity score (DR staging) was developed and trained using the international DRSS.

Study participants underwent two (2) field non-mydriatic fundus photography by trained technicians. Images for each eye were centered on the optic disc and on the macula. These were then sent automatically by the software for cloud AI processing and interpretation, and a DR scoring report was sent back (Appendix C). Generated AI grading summary records were de-identified. All personally identifiable information (patient name, date of birth, and examination date) was removed. For anonymized patient tracking, a randomized External Patient ID was used for record keeping.

Interpretation by Expert Human Graders

Two (2) retina specialists independently read the anonymized fundus images. Both graders are not privy to the AI-generated image grading. Disagreements between gradings were adjudicated by a third senior retina specialist with more than 15 years of experience. This method produced a single, unified, expert interpretation for each image, which was considered ground truth and used as reference standards for which the AI grading was compared. All images were

viewed on a uniform single screen of standard size, resolution (2560 x 1440 pixels), and contrast.

Retinal image interpretation was based on the International Clinical Diabetic Retinopathy Severity Scale (ICDRSS).[17] Readers were provided copies of the ICDRSS standard photos for review and reference standard images for each level of DR severity from a different pool of images using the same iCare Illume system were carefully selected and used as study reference standard (Appendix F). This is to ensure consistency between readers. Images were considered ungradable if more than two-thirds cannot be adequately visualized unless the remaining one-third clearly shows the severity of the DR. Inter-rater agreement for reliability testing between the two expert graders was computed using the Kappa coefficient with a 95% confidence interval.

Descriptive statistics were used to summarize demographics. Kappa agreement analysis was used to estimate the level of agreement between human graders.[18] Accuracy, Sensitivity, Specificity, NPV, PPV, and Likelihood Ratios were used to determine the diagnostic performance of the iCare Illume system. Missing values were neither replaced nor estimated. Microsoft Excel and STATA 13.1 were used for data management and analysis.

RESULTS

A total of 278 patients (556 eyes, 1554 images) were screened for DR using the Illume System between June and November 2023 and included in the initial pool of data. After review of charts, the study population was reduced to 142 patients (284 eyes, 717 images) due to the exclusion of several patients with incomplete data or unclear diabetes mellitus status. As a result, the target sample size of 722 eyes was not achieved and could potentially affect the final statistical analysis.

The average age of the study participants was 66 years, with a higher proportion of females (68.3%) compared to males (31.7%). However, due to a significant number of incompletely filled-out data sheets, other contextual factors such as community-specific healthcare access and pertinent demographics were not analyzed.

Inter-Rater Agreement between Human Graders

Table 1 presents a good degree of inter-observer agreement between the two expert human graders, with a kappa coefficient of 0.86 for ADR and 0.70 for RDR. This substantial to high level of agreement indicates consistency in grading using the study's own DRSS reference images, suggesting good reliability.

Distribution across DR Severity Scales

Image evaluation by human experts revealed that DR was absent in 65.85% (187 eyes) while ADR was present in 27.11% (77 eyes). RDR was detected in 22.18% (63 eyes), and 14.08% (40 eyes) were classified as STDR (Table 2). Insufficient image quality prevented adequate grading in 7.04% (20 eyes) of cases.

Diagnostic Accuracy of the iCare Illume

Presence or Absence of Diabetic Retinopathy

The Illume system accurately identified the presence or absence of DR in 83.21% of cases. Sensitivity and Specificity were 92.11% and 79.57%, respectively. The positive

Table 1. Agreement between the Two Human Graders (n=262 eyes)

	Any DR* (ADR)	Referrable DR (RDR)**
Total Agreement	95.09%	90.59%
Kappa Coefficient (95% CI)	0.8593 (0.8122-0.8992)	0.6965 (0.6687-0.7247)
Level of Agreement	High	Substantial

**Any DR (ADR) defined as the presence of DR regardless of severity*

*** RDR, Referable DR – defined as NPDR moderate or worse, DME, ungradable*

Table 2. Distribution of Eyes across Diabetic Retinopathy Severity (N=284*)

No Apparent DR	Mild NPDR	Moderate NPDR	Severe NPDR	PDR	Indeterminate
187 (65.85%)	14 (4.93%)	23 (8.10%)	26 (9.15%)	14 (4.93%)	20 (7.04%)
Any DR (ADR): 77/284 (27.11%)					
Referable DR (RDR): 63/284 (22.18%%)					
Sight-Threatening DR (STDR): 40/284(14.08%)					

** Photographic Diagnosis by Human Grader using DRSS*

NPDR, Non-Proliferative Diabetic Retinopathy; PDR, Proliferative Diabetic Retinopathy; ADR, Any DR; RDR, Referable DR – defined as NPDR moderate or worse, DME, ungradable; STDR, Sight-Threatening DR – defined as severe NPDR or worse

predictive value (PPV) was 64.81%, and the negative predictive value was 96.10% (Table 3). Twenty-two (22) images were classified as indeterminate due to inability to grade them adequately. The positive likelihood ratio (LR+) was 4.5, and the negative likelihood ratio (LR-) was 0.10.

Presence or Absence of Referable DR (RDR)

The Illume system accurately identified the presence or absence of RDR in 92.37% of cases. Sensitivity and Specificity were 79.03% and 96.50%, respectively. The positive predictive value (PPV) was 87.50%, and the negative predictive value (NPV) was 93.69% (Table 4). Twenty-two (22) images were classified as indeterminate due to inadequate grading. LR+ was 22.58 and LR- was 0.22.

Technical Failure Rate

The TFR of the AI DRS system was low at 0.70% (2/284 images). This is within the acceptable TFR set by the UK NICE of ≤5%.[9]

Table 3. Diagnostic Accuracy of AI DRSplus in Detecting the Presence or Absence of Diabetic Retinopathy (N = 262 eyes)*

		Reference Standard (Human Grader)		Total
		RDR** detected	RDR** not detected	
AI iCare Illume	*RDR** detected*	70	38	108
	*RDR** not detected*	6	148	154
	Total	76	186	262***
Sensitivity (UK NICE: 80%)**	92.11% (95% CI: 84.10 - 96.60%)	***Positive LR***	4.51 (95% CI: 3.38 to 6.02)	
Specificity (UK NICE: 95%)**	79.57% (95% CI: 73.1 to 85.00%)	***Negative LR***	0.10 (95% CI: 0.05 to 0.21)	
PPV	64.81% (95% CI: 55.00 to 73.50%)	***Prevalence***	29.01 % (95% CI: 23.51 to 34.90%)	
NPV	96.10% (95% CI: 91.80 to 98.30%)	***Accuracy***	83.21 % (95% CI: 78.10 to 87.40%)	

** Reference standard: Clinical Grading using Two (2) field Fundus Photography by Human Grader*

*** UK NICE Performance Guidelines[9]*

**** 22 indeterminate or insufficient images (combined AI and human)*

DR, Diabetic Retinopathy; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio

Table 4. Diagnostic Accuracy of AI DRSplus in Detecting the Presence or Absence of Referable Diabetic Retinopathy (N = 262 eyes)*

		Reference Standard (Human Grader)		Total
		RDR** detected	RDR** not detected	
AI iCare Illume	*RDR** detected*	49	7	56
	*RDR** not detected*	13	193	206
	Total	62	200	262****
Sensitivity (UK NICE: 80%)**	79.03% (95% CI: 66.69 to 87.97%)	***Positive LR***	22.58 (95% CI: 10.96 to 46.40)	
Specificity (UK NICE: 95%)**	96.50% (95% CI: 92.98 to 98.30%)	***Negative LR***	0.22 (95% CI: 0.13 to 0.35)	
PPV	87.50% (95% CI: 75.54 to 94.39%)	***Prevalence***	23.66 % (95% CI: 19.0 to 29.3%)	
NPV	93.69% (95% CI: 89.53 to 96.32%)	***Accuracy***	92.37 % (95% CI: 88.41 to 95.15%)	

** Reference standard: Clinical Grading using Two (2) field Fundus Photography by Human Grader*

*** Referable DR (RDR) defined as moderate NPDR or worse or DME*

**** UK NICE Performance Guidelines[9]*

*****22 indeterminate or insufficient image in either AI or human*

DR, Diabetic Retinopathy; STDR, Sight Threatening Diabetic Retinopathy; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio.

DISCUSSION

In Southeast Asia, a significant proportion of individuals with diabetes remain undiagnosed with DR.[6] Despite the widespread adoption of AI for routine DR screening in various countries globally, coupled with its proven cost-effectiveness, its application in the Philippines remains largely unexplored.[19] This study sought to assess the diagnostic accuracy of the iCare Illume system in detecting any stage of DR (ADR) and referable DR (RDR) among Filipino diabetics in a real-world community setting. The findings of this study are pivotal in determining the feasibility of integrating the iCare Illume system into routine DR screening programs in the Philippines.

This study is subject to several limitations and potential sources of bias. The reduced sample size may have reduced statistical power, resulting in wider confidence intervals and increased risk of Type II error. Selection bias may also be present, as participants were drawn from voluntary community-based DR screening activities in La Union, potentially limiting the generalizability to the broader Filipino diabetic population in terms of health-seeking behavior, socioeconomic status, and access to healthcare. Verification bias is another consideration, since the reference standard utilized expert grading of two-field fundus photographs rather than the gold standard seven-field stereoscopic fundus photography defined by the Early Treatment Diabetic Retinopathy Study (ETDRS), which may lead to under-detection of peripheral lesions and could influence estimates of diagnostic accuracy. Spectrum bias may likewise exist because the screened population may not fully represent the complete range of DR severity encountered in routine clinical practice, particularly among patients with advanced disease who may have already sought specialist care.

To minimize variability and measurement error, standardized image acquisition protocols, validated DR grading systems, and masked expert graders were utilized throughout the study. Two-field fundus photography has been shown in prior studies to be an acceptable and pragmatic alternative for large-scale screening programs due to its efficiency and patient acceptability, where it is tolerated for mass community screenings. Despite limitations, the study was designed to reflect the real-world community screening conditions relevant to the practical implementation of AI-assisted DR screening systems.

Validity of the iCare Illume (Thirona RetCAD AI) Grading

Any Stage of Diabetic Retinopathy (ADR)

The Illume system demonstrated reasonable accuracy when the threshold level used is ADR. Sensitivity and negative predictive values (ADR prevalence rate of 29.0%) were both high. However, specificity and positive predictive values were low, falling below the minimum standards established by the UK NICE. These data suggest that there was a high number of false positives but a low number of false negatives. In clinical practice, this would imply a low incidence of missed ADR cases, but it could lead to excessive referrals, potentially resulting in unnecessary clinic visits. Therefore, it is imperative to investigate the underlying reasons for these findings to enhance efficiency.

Subgroup analysis of the six eyes where human graders identified ADR but not detected by AI (false negatives) was conducted to gain insights into this undesirable outcome. Diagnoses by human graders were mild DR (n=2), moderate DR (n=3), and proliferative DR (n=1). This corresponds to six missed referrals by AI, one of which is sight-threatening. This area warrants attention and highlights the potential challenges of generalizability when employing foreign-developed AI systems. To address these concerns, developing local and/or enhancing commercially available but foreign deep learning systems by incorporating large numbers of local data images into their training and validation datasets may be beneficial. Furthermore, implementing additional system checks, such as regular image grading by human experts on select data images, can ensure accuracy and quality control.

Notably, there were 38 eyes where human graders did not find ADR but it was detected by AI (presumed false positive). A plausible explanation for this is that the two-field photography technique employed by human graders may not be sufficiently sensitive to capture an image of more peripheral DR lesions or DR with extremely subtle signs. Patient callback and review of these 38 eyes are warranted to ascertain the accuracy of these AI reports. These limitations also underscore the need for future multi-center validation studies with larger sample sizes and the implementation of gold-standard imaging protocols to enhance the robustness and applicability of the results.

Referable Diabetic Retinopathy (RDR)

The Illume system exhibited high accuracy when the threshold level employed was RDR. Specificity, positive predictive value, and negative predictive value (RDR prevalence rate of 23.66%) were all high, although the sensitivity measure fell slightly below UK NICE standards. These findings suggest a low incidence of over-referrals (false positives) but a substantial number of missed referrals (false negatives).

Subgroup analysis conducted for the 13 missed RDR cases (false negatives) by AI revealed that all 13 eyes were classified by human graders as moderate DR, while AI reports indicated diagnoses of no DR (n=3) and mild DR (n = 10). This discrepancy resulted in 13 missed referrals when AI was utilized.

While this discrepancy may be attributed to true grading inaccuracies by AI, the authors hypothesize that these errors are attributable to the subjective nature of image evaluation by humans and the tendency to classify images favoring more severe grading to minimize the risk of missed referrals.

Table 5. Comparison of Diagnostic Accuracies of AI DRSplus*: Any Stage Diabetic Retinopathy** versus Referable Diabetic Retinopathy*** (N = 262 eyes)*

	Any Diabetic Retinopathy (ADR)	Referable Diabetic Retinopathy (RDR)
Accuracy	83.21 % (95% CI: 78.10 to 87.40%)	92.37 % (95% CI: 88.41 to 95.15)
Sensitivity (UK NICE: 80%)*	92.11% (95% CI: 84.10 to 96.60%)	79.03% (95% CI: 66.69 to 87.97)
Specificity (UK NICE: 95%)*	79.57% (95% CI: 73.1 to 85.00%)	96.50% (95% CI: 92.98 to 98.30)
PPV	64.81% (95% CI: 55.00 to 73.50%)	87.50% (95% CI: 75.54 to 94.39)
NPV	96.10% (95% CI: 91.80 to 98.30%)	93.69% (95% CI: 89.53 to 96.32)

** Reference standard: Clinical Grading using Two (2) field Fundus Photography by Human Grader*

*** Any DR (ADR) defined as the presence of DR regardless of severity*

**** Referable DR (RDR) defined as moderate NPDR or worse*

***** UK NICE Performance Guidelines[9]*

DR, Diabetic Retinopathy; RDR, Referral Diabetic Retinopathy; PPV, positive predictive value; NPV, negative predictive value

This is particularly pertinent in the context of two-field photography where areas not captured by the camera pose uncertainty. Consequently, we encounter the challenge of establishing ground truth in terms of DR severity grading, where human bias and inconsistency play significant roles. To address this issue, more robust data can be obtained by employing the current gold standard of 7-field stereoscopic color fundus photography, as traditionally employed by the modified Airlie House Classification System.

Comparing the diagnostic performance of the Illume AI across threshold referral levels (Table 5), it is evident that employing RDR as the threshold in our cohort of diabetic patients yields more accurate assessments, closely aligning with the UK NICE standards. This is consistent with most DR screening protocols that utilize RDR as the referral threshold. While the primary objective of screening is to identify with a high degree of accuracy most patients at elevated risk for developing DR complications, it is important to maintain an efficient and cost-effective system by minimizing excessive referrals.[20]

It is important to note that while the diagnostic performance of the Illume system when set at RDR threshold performs reasonably well in our cohort of diabetic patients, there is a significant decrease in sensitivity when compared to foreign data obtained using the Thirona RetCAD and other commercially available AI systems employed in real-world settings. At the RDR threshold, our study achieved a sensitivity of 79.03% and a specificity of 96.5%, respectively, while published literature reported a sensitivity and specificity ranging from 91-100% and 94-98%, respectively.[8,21] Again, this discrepancy in sensitivity values may be attributed to issues of generalizability when applied to our Filipino cohort or may be due to potential limitations arising from our study protocol. Further investigation is necessary to determine the factors that might have contributed to this discrepancy.

In addition to the performance metrics, the technical failure rate of the Illume system in terms of image and report generation was found to be 0.7% (99.3% imageability). This value falls below the 5% threshold permitted by the UK NICE guidelines despite the uncontrolled community setting in which the images were captured. This provides further evidence on the Illume system's image processing capabilities.

Overall, the study underscores the substantial accuracy of the Illume AI technology in detecting RDR within our cohort of Filipino patients. Furthermore, it demonstrates the potential practical applicability of this technology for community DR screening in resource-constrained settings like the Philippines.

Several key areas for future research and development of AI-based diabetic retinopathy (DR) screening systems have emerged. Firstly, while the AI system demonstrated acceptable sensitivity and specificity, further optimization through enhanced algorithms and expanded, diverse training datasets is needed.[22] Secondly, a clear framework is essential to align AI outputs with evidence-based practice guidelines, ensuring effective integration into clinical workflows and resource optimization.[23] Lastly, studies on cost-effectiveness and minimization are needed to provide evidence on the potential benefits of integrating this AI system into existing programs in the Philippines. These recommendations are crucial to fully realize the potential of AI in DR screening and ensure its successful integration into clinical practice.

CONCLUSION

The iCare Illume system exhibits strong diagnostic accuracy in screening for referable DR in our cohort of Filipino Diabetics. Although its sensitivity is marginally below the UK NICE guideline, the system demonstrates high specificity, negative predictive value, and positive predictive value. This combination highlights its potential as a valuable

tool for diabetic retinopathy screening in the Philippines. Future research should focus on refining local AI validation protocols, optimizing data sets of commercially available AI systems, assessing the impact on patient outcomes and disease burden, and evaluating cost-effectiveness.

Acknowledgments

The authors thank the Vitreoretinal Society of the Philippines, Ilocos Training and Regional Medical Center, Dr. Junn Rojas Pajarillo, Dr. Chester Pataray, and Dr. Karen Rabara-Cabrera.

Statement of Authorship

Both authors certified fulfillment of ICMJE authorship criteria.

Author Disclosure

Both authors declared no conflicts of interest.

Funding Source

None.

REFERENCES

1. IDF Diabetes Atlas. Philippine Diabetes Report 2000-2045 [Internet]. 2021 [cited 2025 Apr 5]. Available from: https://diabetesatlas.org/data/en/country/157/ph.html.
2. Philippine Eye Research Institute, National Institutes of Health, University of the Philippines Manila. Philippine Eye Disease Study [Internet]. 2018 [cited 2025 Apr 5]. Available from: https://www.peri.ph/philippine-eye-disease-study.
3. Wilkinson CP, Ferris FL 3rd, Klein RE, Lee PP, Agardh CD, Davis M, et al. Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology. 2003 Sep;110(9):1677-82. doi: 10.1016/S0161-6420(03)00475-5. PMID: 13129861.
4. Goh JKH, Cheung CY, Sim SS, Tan PC, Tan GSW, Wong TY. Retinal Imaging Techniques for Diabetic Retinopathy Screening. J Diabetes Sci Technol. 2016 Feb;10(2):282-294. doi: 10.1177/1932296816629491 PMID: 26830491; PMCID: PMC4773981.
5. Huemer J, Wagner SK, Sim DA. The Evolution of Diabetic Retinopathy Screening Programmes: A Chronology of Retinal Photography from 35 mm Slides to Artificial Intelligence. Clin Ophthalmol. 2020 Jul; 14:2021-35. doi: 10.2147/OPTH.S261629. PMID: 32764868; PMCID: PMC7381763.
6. Yin J, Ngiam KY, Teo HH. Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review. J Med Internet Res. 2021 Apr;23(4):e25759. doi: 10.2196/25759. PMID: 33885365; PMCID: PMC8103304.
7. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016 Dec;316(22):2402-10. doi: 10.1001/jama.2016.17216. PMID: 27898976.
8. Wewetzer L, Held LA, Steinhäuser J. Diagnostic performance of deep-learning-based screening methods for diabetic retinopathy in primary care-A meta-analysis. PLoS One. 2021 Aug;16(8):e0255034. doi: 10.1371/journal.pone.0255034. PMID: 34375355; PMCID: PMC8354436.
9. Uy H, Fielding C, Hohlfeld A, Ochodo E, Opare A, Mukonda E, et al. Diagnostic test accuracy of artificial intelligence in screening for referable diabetic retinopathy in real-world settings: A systematic review and meta-analysis. PLoS Glob Public Health. 2023 Sep;3(9): e0002160. doi: 10.1371/journal.pgph.0002160. PMID: 37729122; PMCID: PMC10511145.
10. Tayapad JB, Bengzon AU, Valero SO, Arroyo MH, Papa RTM, Fortuna EJS, et al. Implementation and pilot data on diabetic retinopathy in a teleophthalmology program at a multispecialty primary care clinic. Philipp J Ophthalmol. 2014 Dec;39(2):90-3. Available from: https://paojournal.com/article/implementation-and-pilot-data-on-diabetic-retinopathy-in-a-teleophthalmology-program-at-a-multispecialty-primary-care-clinic/.
11. Alejo GPI, Pinzon JC, Silva PS. Advances in AI screening for diabetic retinopathy and DME [Internet]. 2024 [cited 2025 Apr 5]. Available from: https://www.retina-specialist.com/article/advances-in-ai-screening-for-diabetic-retinopathy-and-dme.
12. Boucher MC, Gresset JA, Angioi K, Olivier S. Effectiveness and safety of screening for diabetic retinopathy with two nonmydriatic digital images compared with the seven standard stereoscopic photographic fields. Can J Ophthalmol. 2003 Dec;38(7):557-568. doi: 10.1016/s0008-4182(03)80109-6 PMID: 14740797.
13. Department of Health. Administrative Order No. 2007-0022: Health Privacy Code [Internet]. 2007 [cited 2025 Apr 5]. Available from: https://www.scribd.com/document/395339808/ao2007-0022.
14. Department of Health. Administrative Order No. 2020-0030: Data Processing Guidelines on Health Information [Internet]. 2020 [cited 2025 Apr 5]. Available from: https://law.upd.edu.ph/wp-content/uploads/2020/07/DOH-AO-No-2020-0030.pdf.
15. Evans NG, Wenner DM, Cohen IG, Purves D, Chiang MF, Ting DSW, et al. Emerging Ethical Considerations for the Use of Artificial Intelligence in Ophthalmology. Ophthalmol Sci. 2022 Mar; 2(2):100141. doi: 10.1016/j.xops.2022.100141 PMID: 36249707; PMCID: PMC9560632.
16. Goodyear MD, Krleza-Jeric K, Lemmens T. The Declaration of Helsinki. BMJ. 2007 Sep;335(7621):624-5. doi: 10.1136/bmj.39339.610000.BE. PMID: 17901471; PMCID: PMC1995496.
17. International Council of Ophthalmology. International clinical diabetic retinopathy disease severity scale, detailed table [Internet]. [cited 2025 Apr 5]. Available from: http://www.icoph.org/resources/45/International-Clinical-Diabetic-Retinopathy-Disease-Severity-Scale-Detailed-Table-.html.
18. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med (Zagreb). 2012 Oct;22(3):276-82. PMID: 23092060 PMCID: PMC3900052.
19. Department of Science and Technology. PCHRD-Funded Study Marks the First Use of AI in PH Clinical Setting for Ophthalmology [Internet]. 2021 Sep 21 [cited 2025 Apr 5]. Available from: https://www.pchrd.dost.gov.ph/news_and_updates/pchrd-funded-study-marks-the-first-use-of-ai-in-ph-clinical-setting-for-ophthalmology/.
20. World Health Organization. Diabetic Retinopathy Screening: a short guide [Internet]. 2020 [cited 2025 Apr 5]. Available from: https://iris.who.int/bitstream/handle/10665/336660/9789289055321-eng.pdf.
21. Piatti A, Giorda CB, Romeo F. Diabetic Retinopathy Screening with Artificial Intelligence: Preliminary Experience in Italian Healthcare System [Internet]. 2025 [cited 2025 Apr 5]. Available from: https://www.icare-world.com/publications/diabetic-retinopathy-screening-with-artificial-intelligence/.
22. Sorrentino FS, Gardini L, Fontana L, Musa M, Gabai A, Maniaci A, et al. Novel approaches for early detection of retinal diseases using artificial intelligence. J Pers Med. 2024 Jun;14(7):690. doi: 10.3390/jpm14070690. PMID: 39063944; PMCID: PMC11278069.
23. Liu Y, Rajamanickam VP, Parikh RS, Loomis SJ, Kloek CE, Kim LA, et al. Diabetic Retinopathy Assessment Variability Among Eye Care Providers in an Urban Teleophthalmology Program. Telemed J E Health. 2019 Apr;25(4):301-307. doi: 10.1089/tmj.2018.0019 PMID: 30040526; PMCID: PMC6482899.

APPENDICES

Appendix A. Standard Diabetic Retinopathy Grading

Grade	IDRSS Classification*	DR Category
DR0	No DR detected	Non-referrable
DR1	Mild DR detected	Non-referrable
DR2	Moderate DR detected	Referrable
DR3	Severe DR detected	Referrable
DR4	Proliferative DR detected	Referrable

** International Council of Ophthalmology. International Clinical Diabetic Retinopathy Disease Severity Scale (IDRSS)*[18]

The iCare DRSplus

Appendix B. iCare DRSplus with Thirona RetCAD Software Description

The AI machine, DRSplus was used. The DRSplus is intended for the acquisition of colored images of the retina without the use of a mydriatic agent. The device includes a dedicated AI software application (Thirona RetCAD) and operates as a stand-alone unit. The Thirona RetCAD is an MDR class IIa and CE 0344 certified Artificial Intelligence software that uses Deep Learning.

Deep learning is a type of artificial intelligence (AI) that uses artificial neural networks to learn from data. Neural networks are inspired by the structure and function of the human brain, and they can learn complex patterns from large amounts of data24. It has been shown to be very effective at screening for diabetic retinopathy. In some studies, deep learning models have been shown to be as accurate as ophthalmologists at detecting the disease. This is particularly significant because diabetic retinopathy screening can be time-consuming and expensive, and deep learning models can help to automate the process and make it more accessible to patients.

In diabetic retinopathy screening, deep learning is used to train AI models to detect signs of the disease in retinal images. These models are trained on large datasets of manually-labeled retinal images from human experts.

Once the AI model has been trained, it can be used to automatically screen new retinal images for signs of the disease. The model analyzes each image and assigns it a score based on its likelihood of having diabetic retinopathy. If the score is above a certain threshold, the image is flagged for review by an ophthalmologist.

The DRSplus provides colored images of the retina with a field of view of 45° x 40°, in fully automatic mode.

Appendix C. ILLUME and RetCAD Analysis Process

Images are then sent to the Cloud for analysis by the RetCAD Artificial Intelligence, under ILLUME, a unique combination of specialized screening software which includes an automated workflow, enabling interpretation of high-quality detailed images without dilation, and the future option to share referrals also outside of the source organization A collated data of human-specialist graded images are used to enable simultaneous grading of DR based on obtained fundus photos from the DRSplus.

iCare ILLUME

Summary

EXAMINATION DATE: 23/06/2023 5:51 AM

PATIENT:

Abra Kadabra

DOB:	GENDER:	PATIENT ID:	EXTERNAL ID:
-	F	PAT-d57b09d9-b91a-49d5-af61-7df04eeb702e	8cMwSwaSucuS

ASSESSMENT:
Diabetic Retinopathy (DR), Age-Related Macular Degeneration (AMD)

SCREENING OUTCOME:
No abnormalities are detected.

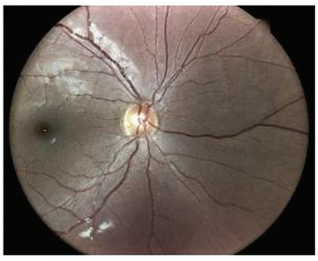
RECOMMENDATION:
No follow-up action is advised.

The summarized clinical data and recommendation for patients on this page are extracted from the Thirona RetCAD™ artificial intelligence product (MDR class IIa and CE 0344 certified). The original RetCAD™ is shown below.

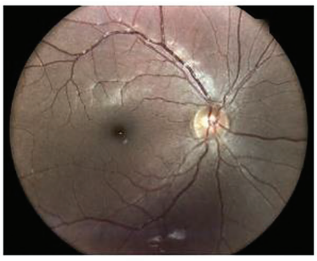
Please note that iCare's recommendation is to get images for both central and nasal orientations of left and right eyes when possible.

RIGHT EYE (OD)

EXAM QUALITY: good



Nasal



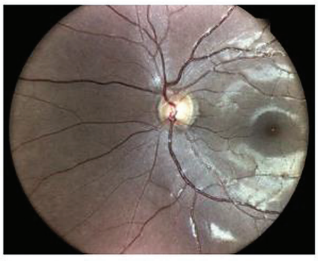
Central Nasal

DR ASSESSMENT:
No DR detected

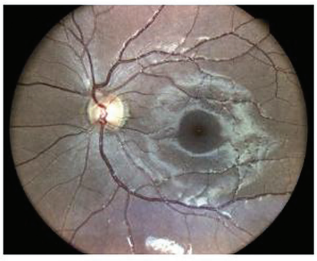
AMD ASSESSMENT:
No AMD detected

LEFT EYE (OS)

EXAM QUALITY: good



Nasal

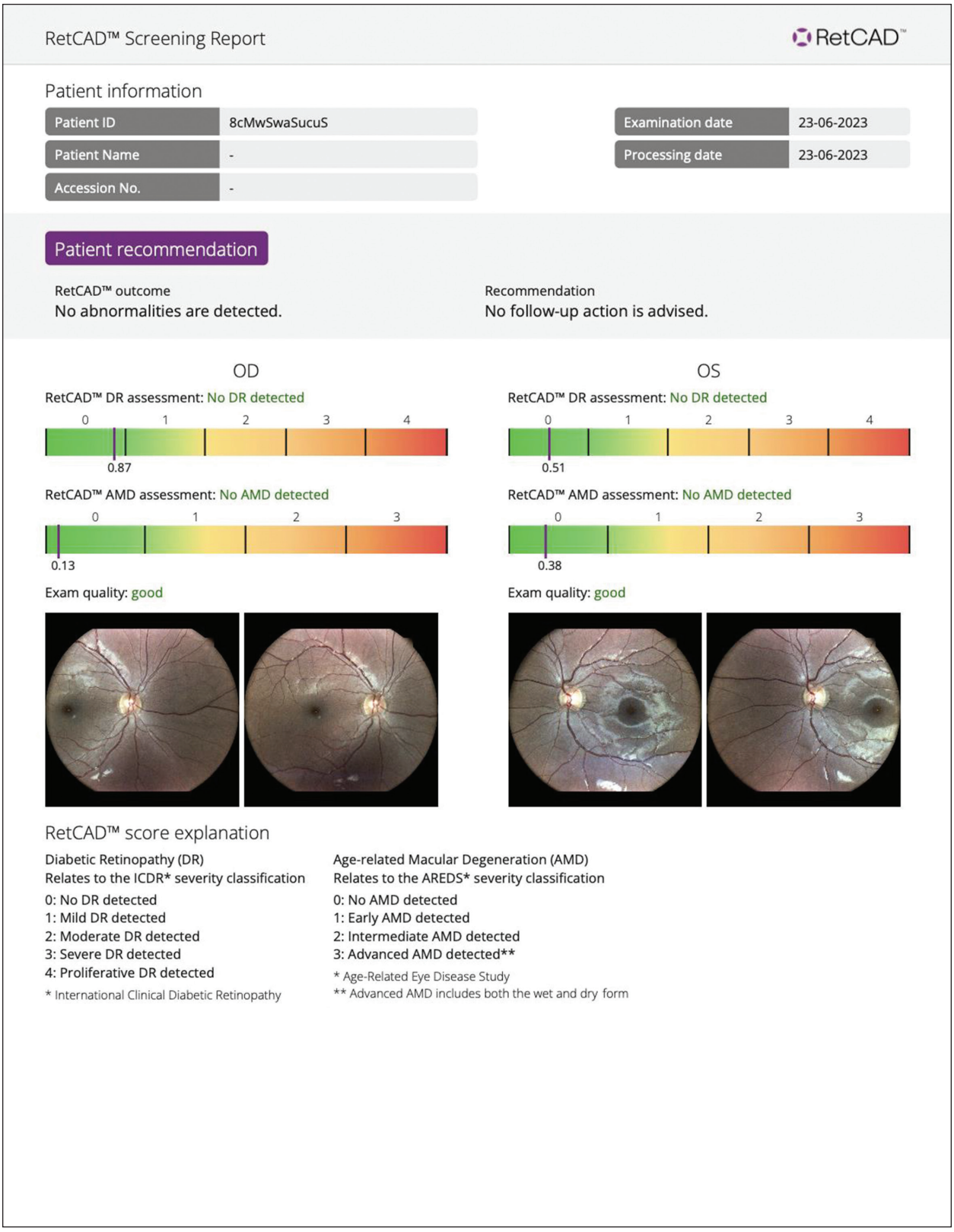


Central Nasal

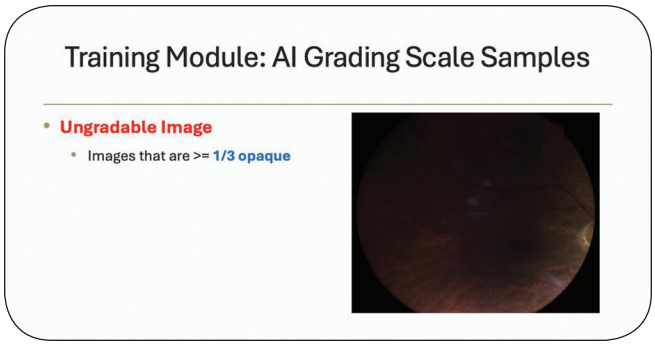
DR ASSESSMENT:
No DR detected

AMD ASSESSMENT:
No AMD detected

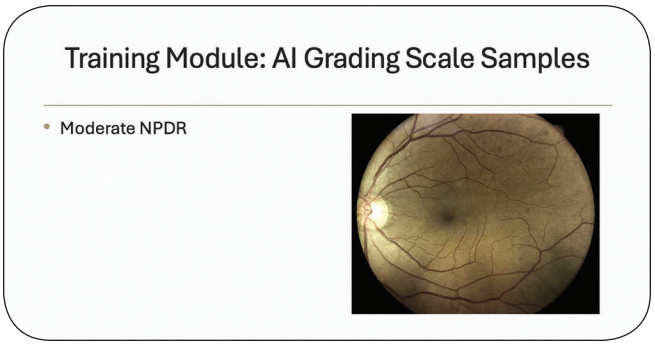
Appendix D. Sample RetCAD Screening Report (Page 1/2).



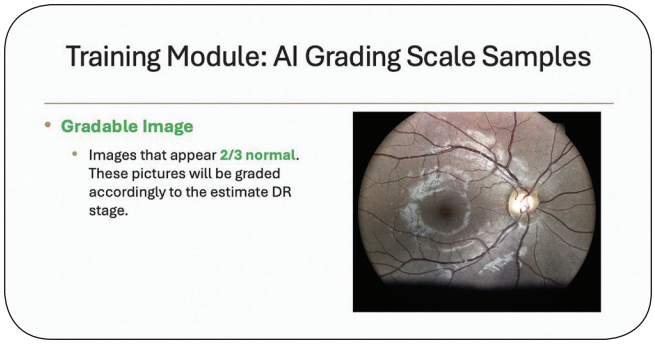
Appendix E. Sample RetCAD Screening Report (Page 2/2).



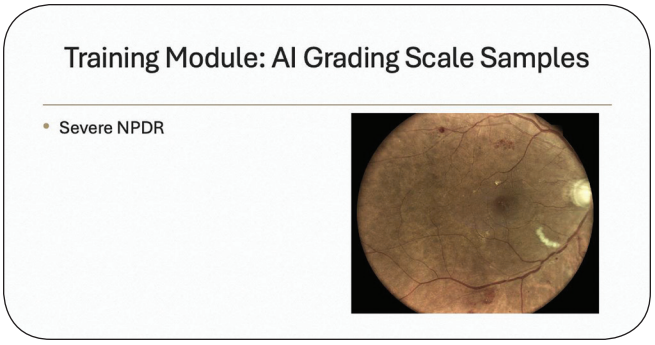
F1: Sample Ungradable Image



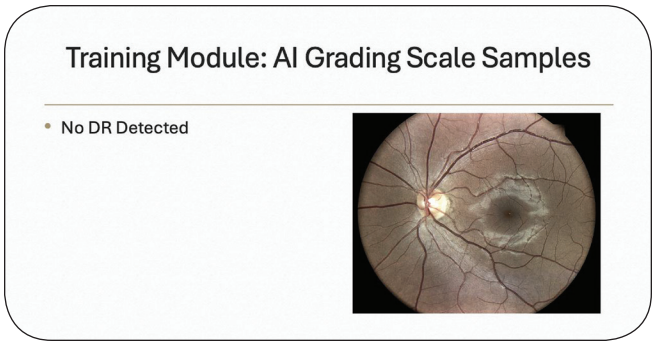
F5: Sample "Moderate NPDR" Image



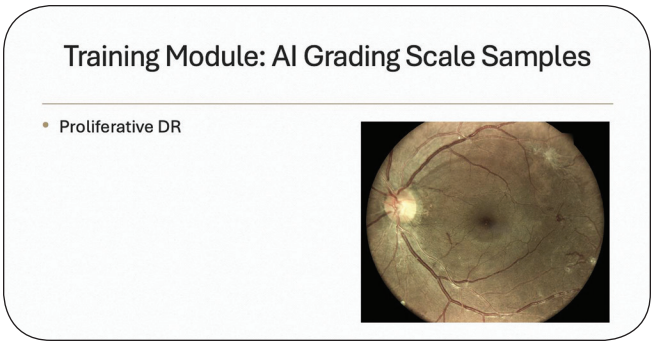
F2: Sample Gradable Image



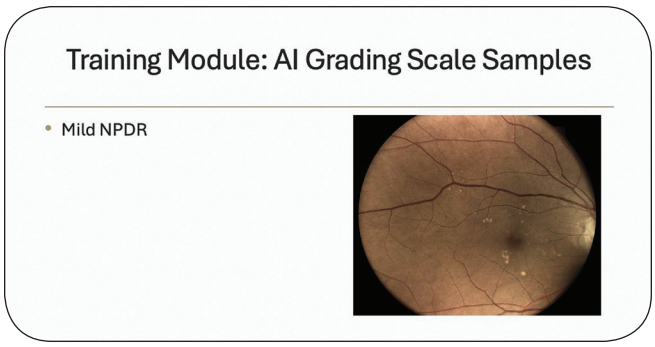
F6: Sample "Severe NPDR" Image



F3: "No DR Detected" Image



F7: Sample "Proliferative NPDR" Image



F4: Sample "Mild NPDR" Image

Appendix F. Sample Images of AI Grading Training Module.