Rate and Predictors of Misclassification of Active Diabetic Macular Edema as Detected by an Automated Retinal Image Analysis System

Lamberto La Franca · Carola Rutigliani · Lisa Checchin ·

Rosangela Lattanzio · Francesco Bandello · Maria Vittoria Cicinelli

ABSTRACT

Introduction: The aim of this work is to estimate the sensitivity, specificity, and misclassification rate of an automated retinal image analysis system (ARIAS) in diagnosing active diabetic macular edema (DME) and to identify factors associated with true and false positives.

Methods: We conducted a cross-sectional study of prospectively enrolled patients with diabetes mellitus (DM) referred to a tertiary medical retina center for screening or management of DME. All patients underwent two-field fundus photography (macula- and disc-centered) with a true-color confocal camera; images were processed by EyeArt V.2.1.0 (Woodland Hills, CA, USA). Active DME was defined as the presence of intraretinal or subretinal fluid on spectral-domain optical coherence tomography (SD-OCT). Sensitivity and specificity and their 95% confidence intervals (CIs) were calculated. Variables associated with true (i.e., DME labeled as present by ARIAS + fluid on SD-OCT) and false positives (i.e., DME labeled as present by ARIAS + no fluid on SD-OCT) of active DME were explored.

Results: A total of 298 eyes were included; 92 eyes (31%) had active DME. ARIAS sensitivity and specificity were 82.61% (95% CI 72.37–89.60) and 84.47% (95% CI 78.34–89.10). The misclassification rate was 16%. Factors associated with true positives included younger age ($p = 0.01$), shorter DM duration ($p = 0.006$), presence of hard exudates ($p = 0.005$), and microaneurysms ($p = 0.002$). Factors associated with false positives included longer DM duration ($p = 0.01$), worse diabetic retinopathy severity ($p = 0.008$), history of inactivated DME ($p < 0.001$), and presence of hard exudates ($p < 0.001$), microaneurysms ($p < 0.001$), or epiretinal membrane ($p = 0.06$).

Conclusions: The sensitivity of ARIAS was diminished in older patients and those without DME-related fundus lesions, while the specificity was reduced in cases with a history of inactivated DME. ARIAS performed well in screening for naïve DME but is not effective in surveillance inactivated DME.

Keywords: Artificial intelligence; Diabetic retinopathy; Diabetic macular edema; Surveillance

L. La Franca · L. Checchin · R. Lattanzio ·
F. Bandello · M. V. Cicinelli (✉)
Department of Ophthalmology, IRCCS San Raffaele
Scientific Institute, IRCCS Ospedale San Raffaele,
University Vita-Salute, Via Olgettina 60, 20132
Milan, Italy
e-mail: cicinelli.mariavittoria@hsr.it

C. Rutigliani · F. Bandello · M. V. Cicinelli
School of Medicine, Vita-Salute San Raffaele
University, Milan, Italy

Key Summary Points

Why carry out this study?

Diabetic macular edema (DME) is a leading cause of vision loss in people with diabetes.

Automated retinal image analysis systems (ARIAS) using artificial intelligence (AI) have shown promise in screening for DME, but performance in monitoring DME is unknown.

What was learned from the study?

ARIAS' performance in DME detection was lower in subgroups like older patients and those with history of inactive DME.

ARIAS is effective for screening naïve DME cases but has limitations in monitoring inactive DME.

Additional imaging like OCT when using ARIAS for DME surveillance in patients with history of disease is still needed.

INTRODUCTION

Individuals with diabetes mellitus (DM) are prone to various health complications, which can significantly impact their quality of life and increase mortality rates [1]. One of the major threats to vision in patients with uncontrolled or long-standing diabetes is diabetic retinopathy (DR), with diabetic macular edema (DME) being the leading cause of central visual loss in these individuals [2]. The prevalence of DME ranges from 4.2 to 7.9% in patients with type 1 DM and from 1.4 to 12.8% in type 2 DM [3].

Treatment for DME typically involves the use of intravitreal anti-vascular endothelial growth factor (VEGF) agents and intravitreal steroids administered at regular intervals until the macula is dry, followed by progressive spacing of treatments [4, 5]. Regardless of the specific treatment received, regular follow-up is crucial,

as DME may persist or recur [6]. Spectral domain-optical coherence tomography (SD-OCT) plays a vital role in the early detection of intraretinal or subretinal fluid, guiding tailored treatment management. Prompt identification and treatment of recurrent DME are essential to prevent vision loss [7].

Recent advancements in computing power, the availability of large datasets, and the accessibility of machine learning and neural network frameworks have greatly improved the automated grading of retinal images [8]. Automated retinal image analysis systems (ARIAS) with artificial intelligence (AI) capabilities have emerged, enabling accurate identification of DR and DME without the need for human graders [9–15]. The diagnostic accuracy of ARIAS for DR and DME detection is comparable to that of expert graders [16], leading to their implementation in nationwide screening programs [17].

Researchers have conjectured the possibility of utilizing ARIAS for the decentralized surveillance of patients with inactive sight-threatening DR or DME [18]. However, it remains unclear whether ARIAS' diagnostic performance would be consistent in such cases, where retinal lesions may be present in the absence of active disease. In this study, our focus was specifically on DME. We aimed to estimate the sensitivity and specificity of ARIAS in detecting active DME, regardless of the patient's treatment status. Additionally, we sought to identify demographic and clinical factors that may influence the performance of ARIAS in DME detection.

METHODS

We conducted a cross-sectional study involving adult patients (≥ 18 years old) with DM who were prospectively recruited at the Medical Retina Unit of the Department of Ophthalmology at San Raffaele Hospital in Milan, Italy, between April 2022 and January 2023. The participants were referred for screening or management of DME. Patients with concomitant retinal conditions that could potentially confound DME detection, such as age-related macular degeneration, were excluded. Both eyes of eligible patients were included in the study.

All investigations were conducted following the principles outlined in the Declaration of Helsinki for research involving human subjects. The study received approval from the San Raffaele Hospital Ethics Committee under the protocol name "OCTA-MIMS" (97/INT/2021), and written informed consent was obtained from all subjects.

Data Collection and Imaging

We collected demographic information, findings from slit-lamp examinations, medical and ocular history of each participant, including the type and duration of DM, glycated hemoglobin (HbA1c) levels, and details of previous ocular treatments. Retinal imaging was performed using a confocal LED fundus camera (iCare DRSplus, Centervue, Padua, Italy) after pupil dilation. The imaging protocol included two retinal photographs covering a field of 45 × 40 degrees, with one centered on the optic disc and the other on the macula. Additionally, ultra-widefield fundus photography (Silverstone, Optos, CA) and SD-OCT (Spectralis, Heidelberg Engineering, Heidelberg, Germany) were performed for each eye.

Imaging Analysis

The retinal images acquired with the LED fundus camera were processed using a commercially available, FDA-approved ARIAS [19] (EyeArt program V2.1.0, Eyenuk Inc, Woodland Hills, CA, USA). This cloud-based software provides disease classification (presence or absence) based on the detection of DR signs and assigns a score for DR severity according to the international clinical diabetic retinopathy (ICDR) severity scale. It also indicates the presence or absence of DME signs.

DR severity was graded based on ultra-widefield fundus photographs, categorized as no DR, non-proliferative DR (NPDR) (mild, moderate, or severe), or proliferative DR (PDR) using the Early Treatment Diabetic Retinopathy Study (ETDRS) scale. The grading of the retinal periphery was facilitated by the use of ultra-widefield fundus photographs. The presence of hard exudates (discrete white-yellow deposits in the posterior pole) and microaneurysms (localized capillary outpouchings in the macular area) was recorded after digitally zooming in the macular region.

SD-OCT scans were used to assess the presence of active DME [18], defined as intraretinal and/or subretinal fluid involving the central subfield zone (center-involving DME), not involving the central subfield zone (non-center involving DME), or a combination of both (diffuse DME). The grading was performed by two trained readers (L.L.F. and C.R.). Additionally, the presence of an epiretinal membrane (ERM) was evaluated, and the central macular thickness (CMT) was automatically measured using the Spectralis software. Cases where assessment was not feasible were treated as missing values.

Inactivated DME was characterized by a documented history of the condition, evidenced through prior treatments or OCT scans, coupled with the absence of DME in the current OCT scan.

Statistical Analysis

Statistical analyses were conducted using R version 4.2.2 and SAS v9.3 (SAS Institute Inc, Cary, NC, USA). The sample size calculation was based on assuming a sensitivity of 90% and a DME prevalence of 18% [20] using the epi.ssdxsesp function from the R epiR package. The sample size of 301 eyes provided 95% confidence to estimate the sensitivity within 0.08 of the true population value [21].

Demographic and clinical characteristics were summarized as means ± standard deviations (SD) or frequencies and proportions (%). Logistic regression models, with the patient identification number as the random term to account for eye inclusion from some patients, were used to compare features between eyes with active DME and those without active DME. The disease agreement between the two eyes of the same patient was assessed using the kappa statistic (κ), which measures the ratio between the observed proportion of agreement and the proportion of agreement expected by chance.

Kappa values range from + 1 (perfect agreement) to -1 (perfect disagreement) [22].

A generalized estimating equations (GEEs) approach from the GENMOD SAS procedure was employed to estimate the sensitivity and specificity of ARIAS. The ARIAS results served as the outcome variable, and the presence of active DME on SD-OCT was the predictor [23]. The 95% confidence intervals (CIs) were calculated, accounting for potential inter-eye correlations [23]. The positive predictive value (PPV) and negative predictive value (NPV) were determined based on the observed prevalence of active DME in the study group.

Logistic regression models were used to identify factors associated with true positives (i.e., DME labeled as present by ARIAS + intraretinal and/or subretinal fluid on SD-OCT) in the subset of eyes with active DME on SD-OCT. Inter-eye correlations were not corrected in the models as the analysis focused on eye-level ARIAS performance. Similar analyses were conducted to identify factors associated with false positives (i.e., DME labeled as present by ARIAS + no intraretinal and/or subretinal fluid on SD-OCT) in the subset of eyes without active DME on SD-OCT. The optimal cut-point for numerical covariates was determined using the cutpointr R package. Odds ratios (OR), 95% CI, and corresponding p-values were reported. The level of alpha = 0.10 was used to assess clinical significance.

RESULTS

A total of 399 eyes from 205 patients with DM were initially collected. After excluding eyes with retinal co-morbidities (24 eyes) and those with low-quality or missing SD-OCT scans (58 eyes), a total of 298 eyes from 154 patients were included in the study. Among these, 144 patients contributed both eyes to the study. Inter-observer agreement was 100%.

Prevalence and Clinical Characteristics of DME

Out of the included eyes, 73 (24%) had a history of DME, and 50 eyes had received previous intravitreal treatments. Sixty-two eyes had active DME on SD-OCT at the study visit, while 11 had inactivated DME. An additional 30 eyes were first diagnosed with DME during the study visit, resulting in a total of 92 eyes (31%) from 64 patients with active DME on SD-OCT. There was a high agreement (79%) between the two eyes of patients regarding the presence or absence of DME, with a kappa value of 0.51 (95% CI 0.35–0.66) (Table 1).

The demographic and clinical characteristics of the study eyes are presented in Table 2. Patients with active DME were older on average compared to those without DME ($p = 0.03$). They were also more likely to have type 2 DM ($p = 0.007$) and higher HbA1c levels ($p < 0.001$). The presence of cataracts or prior cataract surgery was more common in eyes with active DME ($p = 0.001$). The severity of DR was significantly worse in eyes with DME, with higher proportions of moderate NPDR, severe NPDR, or PDR ($p < 0.001$). Fundus examination revealed a higher frequency of hard exudates ($p < 0.001$), microaneurysms ($p < 0.001$), and ERM ($p = 0.07$) in eyes with active DME. The CMT was also significantly higher in the active DME group ($p < 0.001$).

SD-OCT and AI Grading Agreement

Among the 298 paired SD-OCT and ARIAS gradings, there was an 84% agreement on DME status (Table 3). The rate of false positives was 11%, while the rate of false negatives was 5%.

ARIAS Performance Indices

In the entire eye pool, the sensitivity of ARIAS for detecting active DME was 82.61% (95% CI 72.37–89.60), with a specificity of 84.47% (95% CI 78.34–89.10%). The PPV was 70.37% (95% CI 60.82–78.77), and the NPV was 91.58% (95% CI 86.68–95.11). The overall test accuracy was 84%, with a misclassification rate of 16%. Excluding eyes with a history of DME and no active DME at the study visit (as it would routinely happen in a screening setting) increased the specificity and sensitivity of ARIAS to 87.69% and 86%, respectively. Figure 1 shows

Table 1 Inter-eye agreement in active DME status from SD-OCT examination ($N = 144$ patients contributing with both eyes to the study pool)

		Right eye		Total
		Active DME absent	Active DME present	
Left eye	Active DME absent	86	20	106
	Active DME present	10	28	38
	Total	96	48	144

DME diabetic macular edema, *SD-OCT* spectral domain-optical coherence tomography

four example cases: a true positive, a true negative, a false positive, and a false negative.

Factors Associated with True Positives on ARIAS Test

The analysis of factors associated with true positives was conducted in the subset of 92 eyes with active DME on SD-OCT. Younger age (60.6 ± 11.1 vs. 68.4 ± 11.6 years, OR = 0.94 for each year, 95% CI 0.88–0.99, $p = 0.01$), shorter duration of DM (15 ± 11.2 vs. 26.9 ± 9.46 years, OR = 0.92 for each year, 95% CI 0.85–0.97, $p = 0.006$), advanced DR stage (53% vs. 19% with severe NPDR or PDR; OR = 8.89 vs. mild NPDR, 95% CI 1.91–80.9, $p = 0.045$), presence of hard exudates (95% vs. 69%, OR = 8.18 if present, 95% CI 1.90–37.9, $p = 0.005$), and microaneurysms (97% vs. 69%, OR = 16.8 if present, 95% CI 3.21–128.2, $p = 0.002$) were associated with a higher chance of active DME detection by ARIAS. Previous diagnosis of DME and previous DME treatments did not affect ARIAS detection performance (Table 4).

In analyzing the distribution regarding the true positives and false negatives of DME detection by ARIAS, it is notable that among the categories of No DR, Mild NPDR, Moderate NPDR, Severe NPDR, and PDR, a total of 11 cases (eight from Moderate NPDR and three from Severe NPDR/PDR) would still be classified as referable retinopathy. Only four cases in the No DR/Mild NPDR category were completely missed.

Factors Associated with False Positives on ARIAS Test

The analysis of factors associated with false positives was conducted in the subset of 206 eyes without active DME on SD-OCT. Longer DM duration (23.4 ± 12.6 vs. 16.1 ± 12.4 years, OR = 1.05 for each year, 95% CI 1.01–1.08, $p = 0.01$), worse DR severity (63% vs. 20% with moderate NPDR, OR = 5.90 vs. mild NPDR, 95% CI 1.81–26.8; 25% vs. 2% with severe NPDR or PDR; OR = 20.7 vs. mild NPDR, 95% CI 4.25–132.5, $p < 0.001$), a history of DME (28% vs. 2%, OR = 19.4 if present, 95% CI 5.18–94.2, $p < 0.001$), and the presence of hard exudates (44% vs. 5%, OR = 16 if present, 95% CI 6.07–45.4, $p < 0.001$), microaneurysms (91% vs. 30%, OR = 22.5 if present, 95% CI 7.57–96.9, $p < 0.001$), or ERM (13% vs. 4%, OR = 3.39 if present, 95% CI 0.84–12.0, $p = 0.06$) were associated with a higher chance of false positive results by ARIAS. A higher CMT was also associated with false positives, with an optimal cut-point of 291 μm (53% vs. 28%, OR = 2.97 if CMT $\geq$ 291 μm, 95% CI 1.38–6.49, $p = 0.005$) (Table 5).

DISCUSSION

Traditionally, DR and DME screening and monitoring have relied on dilated fundus examinations and OCT conducted by trained ophthalmologists. However, the advent of newer imaging modalities, including stereoscopic imaging, nonmydriatic cameras, and mobile phone fundus cameras, has

Table 2 Demographic and clinical characteristics of the study eyes stratified by the presence of active DME

	Active DME absent (N = 206)	Active DME present (N = 92)	Overall (N = 298)	p value
Age (years)				0.03
Mean (SD)	57.3 (16.0)	62.0 (11.5)	58.7 (14.9)	
Gender				0.2
Male	129 (62.6%)	51 (55.4%)	180 (60.4%)	
Female	77 (37.4%)	41 (44.6%)	118 (39.6%)	
DM type[a]				0.007
Type 1	51 (27.9%)	8 (10.1%)	59 (22.5%)	
Type 2	132 (72.1%)	71 (89.9%)	203 (77.5%)	
Duration of DM (years)				0.9
Mean (SD)	17.2 (12.7)	16.8 (11.7)	17.1 (12.4)	
HbA1c (%)				< 0.001
Mean (SD)	7.02 (1.02)	7.66 (1.75)	7.21 (1.31)	
Lens status				0.001
Transparent	77 (44.3%)	12 (16.2%)	89 (35.9%)	
Cataract	75 (43.1%)	37 (50.0%)	112 (45.2%)	
IOL	22 (12.6%)	25 (33.8%)	47 (19.0%)	
DR				< 0.001
Absent	103 (50.0%)	2 (2.2%)	105 (35.2%)	
Present	103 (50.0%)	90 (97.8%)	193 (64.8%)	
DR severity				< 0.001
No DR	103 (50.5%)	2 (2.3%)	105 (36.2%)	
Mild NPDR	34 (16.7%)	5 (5.8%)	39 (13.4%)	
Moderate NPDR	55 (27.0%)	36 (41.9%)	91 (31.4%)	
Severe NPDR	10 (4.9%)	20 (23.3%)	30 (10.3%)	
PDR	2 (1.0%)	23 (26.7%)	25 (8.6%)	
DME type				N/A
No DME	206 (100%)	–	206 (69.1%)	
Center-involving DME	–	13 (14.1%)	13 (4.4%)	
Non center-involving DME	–	15 (16.3%)	15 (5.0%)	
Diffuse	–	64 (69.6%)	64 (21.5%)	
Hard exudates[a]				< 0.001

Table 2 continued

	Active DME absent (N = 206)	Active DME present (N = 92)	Overall (N = 298)	p value
Absent	183 (89.3%)	9 (9.8%)	192 (64.6%)	
Present	22 (10.7%)	83 (90.2%)	105 (35.4%)	
Microaneurysms[a]				< 0.001
Absent	124 (60.5%)	7 (7.6%)	131 (44.1%)	
Present	81 (39.5%)	85 (92.4%)	166 (55.9%)	
ERM[a]				0.07
Absent	194 (94.6%)	77 (84.6%)	271 (91.6%)	
Present	11 (5.4%)	14 (15.4%)	25 (8.4%)	
CMT (μm)				< 0.001
Mean (SD)	286 (39.3)	374 (111)	313 (80.8)	

DME diabetic macular edema, *SD* standard deviation, *DM* diabetes mellitus, *HbA1c* glycated hemoglobin, *IOL* intraocular lens, *DR* diabetic retinopathy, *NPDR* non-proliferative diabetic retinopathy, *PDR* proliferative diabetic retinopathy, *ERM* epiretinal membrane, *CMT* central macular thickness

[a]This variable contains missing data. Percentages are calculated on the available sample

Table 3 Cross-tabulation of active DME status from SD-OCT and ARIAS at the eye-level (N = 298 eyes from 154 patients)

Eye-level comparison		SD-OCT		Total
		Active DME absent	Active DME present	
ARIAS	Active DME absent	174 (58%)	16 (5%)	190 (68%)
	Active DME present	32 (11%)	76 (26%)	108 (36%)
	Total	206 (69%)	92 (31%)	298 (100%)

DME diabetic macular edema, *SD-OCT* spectral domain-optical coherence tomography, *ARIAS* automated retinal image analysis systems

revolutionized the landscape of eyecare for patients with DR. These technologies facilitate decentralized care delivery through telemedicine [24] and virtual clinics [25], enabling ophthalmologists to assess patient conditions remotely by examining fundus or OCT images, thereby reducing the need for in-person consultations.

While these imaging techniques have been helpful, picture grading by humans is subjective, time-consuming, and requires specialized training. Moreover, despite the increasing accessibility of OCT, its availability remains limited in many primary care settings, especially in resource-constrained areas. To address the increasing workload for eyecare services due to the growing diabetic population and uneven resource distribution, automatic algorithms like ARIAS have been developed for the recognition of DR and DME. These algorithms have demonstrated efficiency, cost-effectiveness, and reproducibility in screening for sight-threatening DR on a large scale [16, 26]. Among them, the EyeArt software has shown a sensitivity of
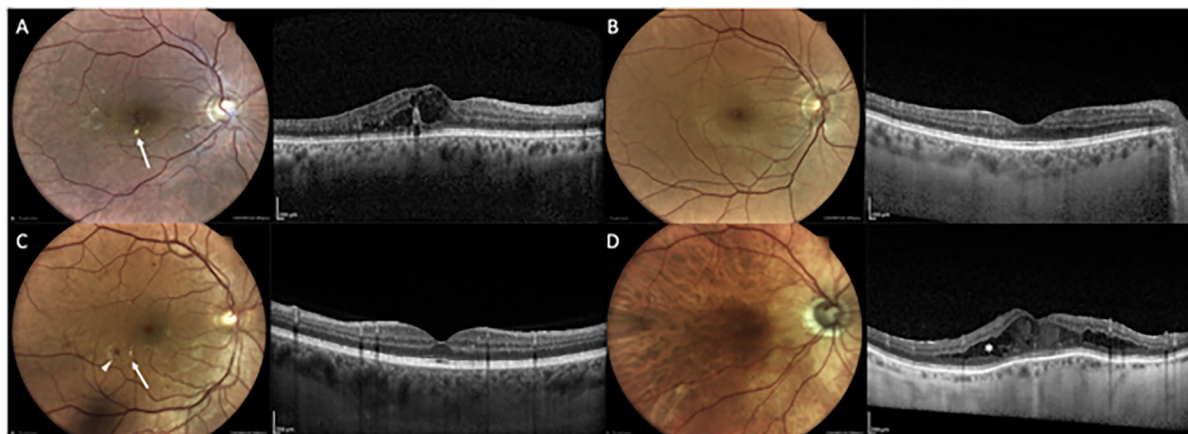
Fig. 1 A series of cases that were found to be true positive (**A**), true negative (**B**), false positive (**C**), and false negative (**D**) for active diabetic macular edema (DME) detection, respectively

over 96% in diagnosing referable DR [17, 27–29], regardless of various factors such as age, ethnicity, dilation status of the lens or pupil, and the type of fundus camera used [30]. ARIAS tools have been validated and approved for screening purposes, with previous studies conducted in treatment-naïve patients with DR. However, the performance of ARIAS in detecting inactivated sight-threatening DR, such as reabsorbed DME or inactive PDR, remains to be determined.

To alleviate the burden of frequent ophthalmic visits, the EMERALD study in the UK explored whether nonmedical staff in community-based settings could follow up with patients having inactivated sight-threatening DR. The study found that nonmedical graders had a sensitivity of 97% and a specificity of 31% in detecting active DME compared to face-to-face encounters with ophthalmologists [18]. The authors suggested implementing this alternative pathway in standard care to increase hospital capacity and improve cost-effectiveness. If ARIAS could detect the reactivation of PDR or recurrence of DME, it would further support the adoption of less burdensome surveillance strategies for the healthcare system.

This present study investigated the diagnostic performance of ARIAS in detecting active DME in a cohort of patients with diabetes referred to a tertiary medical retina center for either DME screening or management. The patients in the study were heterogeneous in terms of DR stage, DME history, and treatment status. ARIAS demonstrated excellent sensitivity and specificity for DME, as previously reported [28]. However, its diagnostic performance decreased when patients with a history of inactivated DME were included in the study. Some eyes were misclassified as not having DME despite the presence of intraretinal or subretinal fluid on SD-OCT (false negatives), while others were classified as having DME even though their macula was dry (false positives).

Detecting active DME in non-stereoscopic fundus images is inherently challenging and relies on surrogate biomarkers like hard exudates, microaneurysms, or hemorrhages. Most true positive cases had either hard exudates or microaneurysms in the macular area. However, 17% of eyes with active DME lacked these biomarkers, indicating a weak correlation between surrogate biomarkers and the presence of fluid on SD-OCT [31]. Older age and longer disease duration were predictors of lower sensitivity since fundus lesions suggestive of DME tend to spontaneously reduce over time [32], leading to higher false negative rates. In our analysis of the true positives and false negatives in DME detection by ARIAS across various retinopathy categories, it is significant to note that 11 cases (eight in Moderate NPDR and three in Severe NPDR/PDR) were still correctly identified as referable retinopathy.

ARIAS demonstrated a specificity of 84%. Longer disease duration and worse DR severity

Table 4 Distribution of demographic and clinical characteristics associated with true positives and false negatives on ARIAS test

	True positives (N = 76)	False negatives (N = 16)	p value
Age (years)			0.01
Mean (SD)	60.6 (11.1)	68.4 (11.6)	
Gender			0.5
Male	41 (53.9%)	10 (62.5%)	
Female	35 (46.1%)	6 (37.5%)	
DM type[a]			0.7
Type 1	7 (10.8%)	1 (7.1%)	
Type 2	58 (89.2%)	13 (92.9%)	
Duration of DM (years)			0.006
Mean (SD)	15.0 (11.2)	26.9 (9.46)	
HbA1c (%)			0.1
Mean (SD)	7.81 (1.80)	6.90 (1.29)	
Lens status			> 0.1
Transparent	12 (19.7%)	0 (0%)	
Cataract	30 (49.2%)	7 (53.8%)	
IOL	19 (31.1%)	6 (46.2%)	
DR			0.9
Absent	0 (0%)	2 (12.5%)	
Present	76 (100%)	14 (87.5%)	
DR severity			0.045
No DR	0 (0%)	2 (13.3%)	
Mild NPDR	3 (4.2%)	2 (13.3%)	
Moderate NPDR	28 (39.4%)	8 (53.3%)	
Severe NPDR	18 (25.4%)	2 (13.3%)	
PDR	22 (31.0%)	1 (6.7%)	
History of DME			0.9
No	21 (29.2%)	5 (31.3%)	
Yes	51 (70.8%)	11 (68.8%)	
Previous DME treatments			0.9
No	31 (43.1%)	7 (43.8%)	
Yes	41 (56.9%)	9 (56.3%)	
DME type			> 0.1

Table 4 continued

	True positives ($N = 76$)	**False negatives ($N = 16$)**	**p value**
No DME	0 (0%)	0 (0%)	
Center-involving DME	10 (13.2%)	3 (18.8%)	
Non-center-involving DME	14 (18.4%)	1 (6.3%)	
Diffuse	52 (68.4%)	12 (75.0%)	
Hard exudates[a]			0.005
Absent	4 (5.3%)	5 (31.3%)	
Present	72 (94.7%)	11 (68.8%)	
Microaneurysms[a]			0.002
Absent	2 (2.6%)	5 (31.3%)	
Present	74 (97.4%)	11 (68.8%)	
ERM[a]			0.3
Absent	62 (82.7%)	15 (93.8%)	
Present	13 (17.3%)	1 (6.3%)	
CMT (μm)			0.4
Mean (SD)	378 (114)	355 (100)	

DME diabetic macular edema, *SD* standard deviation, *DM* diabetes mellitus, *HbA1c* glycated hemoglobin, *IOL* intraocular lens, *DR* diabetic retinopathy, *NPDR* non-proliferative diabetic retinopathy, *PDR* proliferative diabetic retinopathy, *ERM* epiretinal membrane, *CMT* central macular thickness

[a]This variable contains missing data. Percentages are calculated on the available sample

were associated with higher rates of false positives, where chronic damage to the macula, such as macular ischemia [33], might be misinterpreted as active DME. The presence of microaneurysms in the fovea and hard exudates were common in false positive eyes. Additionally, subclinical macular thickening with CMT higher than the optimal cut-point of 291 μm was associated with a high false positive rate by ARIAS for DME recognition [34]. Further research is needed to explore other macular features that may act as confounders for DME and evaluate their impact on ARIAS performance [35].

As OCT technology becomes increasingly available, the integration of AI for the analysis of OCT images may significantly enhance diagnostic precision. AI's ability to detect subtle changes indicative of early DME—changes that might be missed in manual or fundus examinations—exemplifies its potential to augment traditional diagnostic methods. The widespread availability of OCT scanners is undeniably beneficial, and when paired with AI, this duo can transform DME screening into a more efficient and accessible process. Nonetheless, while embracing this technological synergy, it is imperative to rigorously assess AI performance across varied clinical settings. This ensures that AI acts as a supportive tool, enriching rather than supplanting the invaluable expertise of healthcare professionals.

The study has some limitations, including its cross-sectional design, which precludes assessing ARIAS repeatability and consistency over time. The reported PPV and NPV for DME detection should be interpreted carefully as they are influenced by disease prevalence. Biases

Table 5 Distribution of demographic and clinical characteristics associated with true negatives and false positives on ARIAS test

	False positives (N = 32)	True negatives (N = 174)	p value
Age (years)			0.9
Mean (SD)	57.1 (14.7)	57.3 (16.3)	
Gender			0.7
Male	21 (65.6%)	108 (62.1%)	
Female	11 (34.4%)	66 (37.9%)	
DM type[a]			0.3
Type 1	10 (35.7%)	41 (26.5%)	
Type 2	18 (64.3%)	114 (73.5%)	
Duration of DM (years)			0.01
Mean (SD)	23.4 (12.6)	16.1 (12.4)	
Lens status			> 0.1
Transparent	12 (46.2%)	65 (43.9%)	
Cataract	8 (30.8%)	67 (45.3%)	
IOL	6 (23.1%)	16 (10.8%)	
DR			0.002
Absent	1 (3.1%)	102 (58.6%)	
Present	31 (96.9%)	72 (41.4%)	
DR severity			< 0.001
No DR	1 (3.1%)	102 (59.3%)	
Mild NPDR	3 (9.4%)	31 (18.0%)	
Moderate NPDR	20 (62.5%)	35 (20.3%)	
Severe NPDR	7 (21.9%)	3 (1.7%)	
PDR	1 (3.1%)	1 (0.6%)	
History of DME			< 0.001
No	21 (72.4%)	153 (98.1%)	
Yes	8 (27.6%)	3 (1.9%)	
Previous DME treatments			< 0.001
No	22 (75.9%)	153 (98.1%)	
Yes	7 (24.1%)	3 (1.9%)	
Hard exudates[a]			< 0.001
Absent	18 (56.3%)	165 (95.4%)	
Present	14 (43.8%)	8 (4.6%)	

Table 5 continued

	False positives ($N = 32$)	True negatives ($N = 174$)	p value
Microaneurysms[a]			< 0.001
Absent	3 (9.4%)	121 (69.9%)	
Present	29 (90.6%)	52 (30.1%)	
ERM[a]			0.06
Absent	28 (87.5%)	166 (96.0%)	
Present	4 (12.5%)	7 (4.0%)	
CMT ≥ 291 μm			0.005
No	15 (46.9%)	126 (72.4%)	
Yes	17 (53.1%)	48 (27.6%)	

DME diabetic macular edema, *SD* standard deviation, *DM* diabetes mellitus, *IOL* intraocular lens, *DR* diabetic retinopathy, *NPDR* non-proliferative diabetic retinopathy, *PDR* proliferative diabetic retinopathy, *ERM* epiretinal membrane, *CMT* central macular thickness

[a]This variable contains missing data. Percentages are calculated on the available sample

in population selection may not fully represent the population in a screening setting, as the exclusion of age-related macular degeneration patients from the study. The study did not differentiate between different treatments for DME, so the effect of specific molecules on the misclassification rate could not be determined. Lastly, the study focused on white patients, and the findings may not apply to other ethnicities.

CONCLUSIONS

In conclusion, although ARIAS demonstrated high sensitivity for DME, its diagnostic performance varied in specific subgroups. Older patients and those without classic DME-related funduscopic lesions had lower sensitivity. On the other hand, eyes with inactivated DME had lower specificity. Therefore, ARIAS systems cannot be solely relied upon for DME surveillance, and additional imaging with SD-OCT is still necessary. Consideration of multi-center studies or external validation would strengthen the study's impact and support the broader implementation of ARIAS in clinical practice.

REFERENCES

1. Baena-Díez JM, Peñafiel J, Subirana I, et al. Risk of cause-specific death in individuals with diabetes: a competing risks analysis. Diabetes Care. 2016;39(11):1987–95. https://doi.org/10.2337/DC16-0614.

2. Wilkinson CP, Ferris FL, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology. 2003;110(9):1677–82. https://doi.org/10.1016/S0161-6420(03)00475-5.

3. Ixcamey M, Palma C. Diabetic macular edema. Dis Mon. 2021;67(5):101138. https://doi.org/10.1016/J.DISAMONTH.2021.101138.

4. Bandello F, Cicinelli M, Parodi M. Anti-VEGF molecules for the management of diabetic macular edema. Curr Pharm Des. 2015;21(32):4731–7. https://doi.org/10.2174/1381612821666150909095756.

5. Lattanzio R, Cicinelli MV, Bandello F. Intravitreal steroids in diabetic macular edema. Dev Ophthalmol. 2017;60:78–90. https://doi.org/10.1159/000459691.

6. Cicinelli MV, Rabiolo A, Zollet P, Capone L, Lattanzio R, Bandello F. Persistent or recurrent diabetic macular edema after fluocinolone acetonide 0.19 mg implant: risk factors and management. Am J Ophthalmol. 2020;215:14–24. https://doi.org/10.1016/J.AJO.2020.03.016.

7. Bandello F, Battaglia Parodi M, Lanzetta P, et al. Diabetic macular edema. Dev Ophthalmol. 2017;58:102–38. https://doi.org/10.1159/000455277.

8. Li JPO, Liu H, Ting DSJ, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. Prog Retin Eye Res. 2021;82:100900. https://doi.org/10.1016/J.PRETEYERES.2020.100900.

9. Fleming AD, Goatman KA, Philip S, Prescott GJ, Sharp PF, Olson JA. Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts. Br J Ophthalmol. 2010;94(12):1606–10. https://doi.org/10.1136/BJO.2009.176784.

10. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–10. https://doi.org/10.1001/JAMA.2016.17216.

11. Gulshan V, Rajan RP, Widner K, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. JAMA Ophthalmol. 2019;137(9):987–93. https://doi.org/10.1001/JAMAOPHTHALMOL.2019.2004.

12. Sim DA, Keane PA, Tufail A, Egan CA, Aiello LP, Silva PS. Automated retinal image analysis for diabetic retinopathy in telemedicine. Curr Diab Rep. 2015. https://doi.org/10.1007/S11892-015-0577-6.

13. Niemeijer M, Van Ginneken B, Cree MJ, et al. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. IEEE Trans Med Imaging. 2010;29(1):185–95. https://doi.org/10.1109/TMI.2009.2033909.

14. Scotland GS, McNamee P, Fleming AD, et al. Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. Br J Ophthalmol. 2010;94(6):712–9. https://doi.org/10.1136/BJO.2008.151126.

15. 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. https://doi.org/10.3310/HTA20920.

16. 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):343–51. https://doi.org/10.1016/J.OPHTHA.2016.11.014.

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):723–8. https://doi.org/10.1136/BJOPHTHALMOL-2020-316594.

18. Lois N, Cook JA, Wang A, et al. Evaluation of a new model of care for people with complications of diabetic retinopathy: the EMERALD study. Ophthalmology. 2021;128(4):561–73. https://doi.org/10.1016/J.OPHTHA.2020.10.030.

19. Ipp E, Liljenquist D, Bode B, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy. JAMA Netw Open. 2021;4(11):e2134254. https://doi.org/10.1001/JAMANETWORKOPEN.2021.34254.

20. Im JHB, Jin YP, Chow R, Yan P. Prevalence of diabetic macular edema based on optical coherence tomography in people with diabetes: a systematic review and meta-analysis. Surv Ophthalmol. 2022;67(4):1244–51. https://doi.org/10.1016/J.SURVOPHTHAL.2022.01.009.

21. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform. 2014;48:193–204. https://doi.org/10.1016/J.JBI.2014.02.013.

22. Maguire MG. Assessing intereye symmetry and its implications for study design. Investig Ophthalmol Vis Sci. 2020;61(6):27. https://doi.org/10.1167/IOVS.61.6.27.

23. Ying GS, Maguire MG, Glynn RJ, Rosner B. Calculating sensitivity, specificity, and predictive values for correlated eye data. Investig Ophthalmol Vis Sci. 2020;61(11):29. https://doi.org/10.1167/IOVS.61.11.29.

24. Shi L, Wu H, Dong J, Jiang K, Lu X, Shi J. Telemedicine for detecting diabetic retinopathy: a systematic review and meta-analysis. Br J Ophthalmol. 2015;99(6):823–31. https://doi.org/10.1136/BJOPHTHALMOL-2014-305631.

25. Kortuem K, Fasler K, Charnley A, et al. Implementation of medical retina virtual clinics in a tertiary eye care referral centre. Br J Ophthalmol. 2018;102(10):1391–5. https://doi.org/10.1136/BJOPHTHALMOL-2017-311494.

26. Abràmoff MD, Folk JC, Han DP, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol. 2013;131(3):351–7. https://doi.org/10.1001/JAMAOPHTHALMOL.2013.1743.

27. Olvera-Barrios A, Heeren TFC, 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):265–70. https://doi.org/10.1136/BJOPHTHALMOL-2019-315394.

28. 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):635–43. https://doi.org/10.1089/DIA.2019.0164.

29. 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):71–7. https://doi.org/10.1016/J.ORET.2020.06.016.

30. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye (Lond). 2018;32(6):1138–44. https://doi.org/10.1038/S41433-018-0064-9.

31. Wang YT, Tadarati M, Wolfson Y, Bressler SB, Bressler NM. Comparison of prevalence of diabetic macular edema based on monocular fundus photography vs optical coherence tomography. JAMA Ophthalmol. 2016;134(2):222–8. https://doi.org/10.1001/JAMAOPHTHALMOL.2015.5332.

32. Domalpally A, Ip MS, Ehrlich JS. Effects of intravitreal ranibizumab on retinal hard exudate in diabetic macular edema: findings from the RIDE and RISE phase III clinical trials. Ophthalmology. 2015;122(4):779–86. https://doi.org/10.1016/J.OPHTHA.2014.10.028.

33. Tsai WS, Thottarath S, Gurudas S, et al. Characterization of the structural and functional alteration in eyes with diabetic macular ischemia. Ophthalmol Retina. 2023;7(2):142–52. https://doi.org/10.1016/J.ORET.2022.07.010.

34. Varadarajan AV, Bavishi P, Ruamviboonsuk P, et al. Predicting optical coherence tomography-derived diabetic macular edema grades from fundus photographs using deep learning. Nat Commun. 2020. https://doi.org/10.1038/S41467-019-13922-8.

35. Panozzo G, Cicinelli MV, Augustin AJ, et al. An optical coherence tomography-based grading of diabetic maculopathy proposed by an international expert panel: the European School for Advanced Studies in Ophthalmology classification. Eur J Ophthalmol. 2020;30(1):8–18. https://doi.org/10.1177/1120672119880394.