British Journal of
Ophthalmology

Diagnostic Accuracy of Diabetic Retinopathy Grading by an Artificial Intelligence-enabled Algorithm compared with a human standard for Wide-field True-colour Confocal Scanner and Standard Digital Retinal Images

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1 **Diagnostic Accuracy of Diabetic Retinopathy Grading by an Artificial Intelligence-enabled**

2 **Algorithm compared with a human standard for Wide-field True-colour Confocal Scanner and**

3 **Standard Digital Retinal Images**

4 **Abraham Olvera-Barrios,[1-2] Tjebo Heeren,[1-2] Konstantinos Balaskas,[1] Louis Bolter,[3] Ryan**

5 **Chambers,[3] Catherine Egan,[1-2] Adnan Tufail,[1-2] John Anderson.[3]**

6   1.   Medical Retina, Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom

7   2.   University College London and Institute of Ophthalmology, London, United Kingdom

8   3.   Diabetes, Homerton University Hospital NHS Foundation Trust, London, United Kingdom

9

10

11

12 **Corresponding Author**

13 **Abraham Olvera-Barrios**

14 Moorfields Eye Hospital

15 162 City Road

16 London EC1V 2PD

17 United Kingdom

18 a.olvera@nhs.net

19

21

22 **Synopsis:**

23 The EyeArt software is able to classify true-colour, wide-field confocal images with comparable

24 accuracy and sensitivity to that of manual grading of standard digital photographs for diabetic

25 retinopathy.

27

Abstract

29 BACKGROUND: Photographic diabetic retinopathy screening requires labour-intensive grading of
30 retinal images by humans. Automated retinal image analysis software (ARIAS) could provide an
31 alternative to human grading. We compare the performance of an ARIAS using true-colour, wide-
32 field confocal scanning images and standard fundus images in the English National Diabetic Eye
33 Screening Programme (NDESP) against human grading.

34 METHODS: Cross-sectional study with consecutive recruitment of patients attending annual diabetic
35 eye screening. Imaging with mydriasis was performed (2-field protocol) with the EIDON platform
36 (CenterVue, Padua, Italy) and standard NDESP cameras. Human grading was carried out according to
37 NDESP Protocol. Images were processed by EyeArt v2.1.0 (Eyenuk Inc, Woodland Hills, CA). The
38 reference standard for analysis was the human grade of standard NDESP images.

39 RESULTS: We included 1257 patients. Sensitivity estimates for retinopathy grades were: EIDON
40 images 92.27%(95%CI:88.43-94.69%) for any retinopathy, 99%(95% CI.95.35-100%) for vision-
41 threatening retinopathy and 100%(95%CI:61-100%) for proliferative retinopathy. For NDESP images:
42 92.26%(95%CI:88.37-94.69%) for any retinopathy, 100%(95%CI:99.53-100%) for vision-threatening
43 retinopathy and 100%(95%CI:61-100%) for proliferative retinopathy. One case of vision-threatening
44 retinopathy(R1M1) was missed by the EyeArt when analysing the EIDON images, but identified by
45 the human graders. The EyeArt identified all cases of vision-threatening retinopathy in the standard
46 images.

47 CONCLUSION: EyeArt identified diabetic retinopathy in EIDON images with similar sensitivity to
48 standard images in a large-scale screening programme, exceeding the sensitivity threshold
49 recommended for a screening test. Further work to optimise the identification of "no retinopathy"
50 and to understand the differential lesion detection in the two imaging systems would enhance the
51 use of these two innovative technologies in a diabetic retinopathy screening setting.

52

54

INTRODUCTION

Diabetic retinopathy is one of the most common microvascular complications of diabetes.[1] There are 451 million people with diabetes worldwide, and this is projected to rise to 639 million in 2045.[2] An early diagnosis through regular clinical examination or grading of retinal photographs is essential to identify sight-threatening retinopathy and prevent diabetes-related visual impairment.[3,4] Annual screening of the retina is recommended for all patients with diabetes, but due to its prevalence, this represents a significant organisational and financial challenge that requires trained human graders.[5]

The automated grading of retinal images has improved due to advances in computational power, availability of big data sets, and due to the publicly available machine learning and neural network libraries. Artificial intelligence-enabled automated retinal image analysis software (ARIAS) now allow accurate and speedy detection of retinopathy without the need for human graders.[6–14] The diagnostic accuracy of ARIAS have been reported to be comparable to that of expert graders on 45 degree 2-field conventional digital photographs.[14,15]

In the English National Diabetic Eye Screening Programme (NDESP), screening is offered with two 45-degree fundus digital photographs per eye (macula- and disc-centred images) to every person with diabetes aged 12 years and older.[16] Progress in retinal imaging has led to a broader implementation of wide-field fundus imaging. Methods such as scanning laser ophthalmoscopy or digital confocal scanning pose possible advantages over standard imaging used in population screening programmes which may include, better or similar acquisition times, reduced rates of ungradable images in eyes with poor mydriasis[17] and more detailed visualisation of high-risk retinopathy features.[18] There is a well-recognised trade-off between wider field of view and the practicalities and costs involved in population screening. Nevertheless, some eyes present clinically significant diabetic retinopathy features outside the two 45-degree fields or the seven standard ETDRS fields.[18,19] Using ultra-wide field imaging, a subgroup of patients with predominantly peripheral lesions have shown increased risk of progression.[20] However, ultra-wide field imaging may miss posterior-pole and peripheral neovascular disease when compared with the ETDRS fields.[19] The EIDON platform (EIDON™; CenterVue, Padua, Italy) is a wide-field confocal scanner that obtains 60-degree (horizontal) true-colour fundus photographs per exposition by means of a white light-emitted diode illumination (440–650 nm).[21] Advantages of the EIDON over ultra-wide field imaging platforms are the true-colour nature of the images and absent distortion of posterior

86 pole, which may make this platform more sensitive to high-risk retinopathy features, including

87 proliferative disease.[18,19] However, no comparative study exists. The independent validity of the

88 performance results of ARIAS on 60 degree true-colour images of this nature and their clinical

89 applicability to high-volume screening programmes has not been evaluated.

90 This study aims to compare the screening performance of a commercially available ARIAS (EyeArt)

91 on images obtained with two platforms with different optical properties and field of view, the true-

92 colour wide-field confocal scanner (EIDON) and English NDESP-approved fundus cameras against

93 NDESP human grading in a large-scale, community-based diabetic retinopathy screening

94 programme.

95

97

MATERIALS AND METHODS

99 This cross-sectional, comparative study involved 1,257 adult patients (≥18 years) who had been

100 included in a previous cross-sectional, comparative service evaluation study of an imaging

101 platform.[18] The study protocols of both studies were registered and approved through the

102 research governance process at the Homerton University Hospital and adhered to the English NDESP

103 guidelines,[22,23] the tenets of the Declaration of Helsinki, and the UK Data Protection Act 2018.

104 The study protocol has been described elsewhere.[18] In brief, during the three-month study period

105 2,629 patients underwent routine photographic screening. All patients were asked if they were

106 willing to have an additional set of images taken with a second camera. A total of 1,257 patients had

107 this additional imaging. Written informed consent was obtained from all patients who accepted to

108 take part in the study.

109 Image acquisition and human grading

110 Image acquisition and grading has been described in detail in a previous publication.[18] The English

111 NDESP protocol was used in this study.[23,24] All patients underwent a 2-field imaging protocol with

112 mydriasis to capture two images per eye (one macula- and one disc-centred image) with the EIDON

113 confocal scanner and with English NDESP approved fundus cameras. A list of the approved fundus

114 cameras can be found in the NDESP guidance on camera approval.[25] In routine screening practice

115 and therefore in this study, additional images are often taken and stored on the screening software

116 to ensure that enough images of sufficient quality for retinal grading are obtained and to document

117 anterior segment pathology. No anterior segment images were captured with the EIDON platform.

118 The NDESP approved cameras image a field of 45-degrees horizontal and vertical with a resolution of

119 at least 12 megapixels for each capture.[25] A combination of the macula and disc-centred images

120 cover a field of 60 degrees horizontal x 45 degrees vertical (NDESP images). The wide-field true-

121 colour confocal scanner captures a field of 60-degrees horizontal x 50-degrees vertical with a

122 resolution of 14 megapixels per exposition.[21] A combination of the macula and disc-centred

123 images covers a field of 75 degrees horizontal x 50 degrees vertical (EIDON images).

124 The National Screening Committee UK classification for DR was utilised for grading. The grading

125 classification in order of increasing severity are no retinopathy (R0), background retinopathy (R1), no

126 maculopathy (M0), ungradable (U), maculopathy (M1), pre-proliferative retinopathy (R2) and

127 proliferative retinopathy (R3).[26] The NDESP images were managed and graded as protocol of the

128 current English NDESP pathway, where up to three levels of human graders in increasing order of

129 experience who meet the NDESP quality assurance standards assessed the images to determine a

130 disease severity grade and produce a "final grade" for each eye according to the highest level of

131 severity observed. Disagreements between level 1 and 2 human graders for episodes that are

132 potentially M1 or R2 are arbitrated by the level 3 human grader, whose assessment is final. A final

133 outcome human grade for the NDESP images was obtained after this. Referral to hospital eye service

134 is carried out for patients with grades M1, R2 and R3. Patients with an U grade are re-examined by

135 slit lamp biomicroscopy within the screening programme according to NDESP guidelines and

136 referred to the hospital for the above diabetic retinopathy grades or for other pathology.

137 The EIDON images were not introduced to the NDESP grading pathway because of their different

138 colour cast and higher pixel density when compared with the NDESP images. The EIDON images

139 were graded by a level 3 human grader with both wide experience grading in the NDESP and on the

140 manipulation techniques needed to grade EIDON images. The grader was masked to the outcome of

141 grading the standard images. The resultant EIDON images human grades were compared with the

142 final NDESP images human grades. All the patient encounters where there was a discrepancy

143 between the EIDON and NDESP images human grade were re-examined by a different experienced

144 level 3 grader within the screening programme and an ophthalmologist to obtain a consensus EIDON

145 images human grade.

146 Automated grading system

147 A commercially available ARIAS (EyeArt Software, v2.1.0, Eyenuk Inc, Woodland Hills, CA), was used

148 for this analysis. This automated system for diabetic retinopathy detection, with Conformité

149 Européen mark, offers a cloud-based platform for data analysis allowing scalable, elastic computing

150 for data processing. This version of the software combines core diabetic retinopathy analysis

151 algorithms of the version 1.2 with features derived from deep-learning multiple convolutional neural

152 networks.[27] The software is designed to identify cases of diabetic retinopathy that are R1 or above

153 and provides an output of "disease" or "no disease". An additional output of "refer" vs. "no refer" is

154 designed to identify cases which require referral (U grade or above) to eye hospital services. The

155 total of the images was uploaded into the cloud-based technology of the EyeArt as two separate

156 batches, one for the EIDON images and another one for the standard NDESP images. All processing

157 of the screening episodes was performed by the research team. The vendor was not allowed access

158 to the software or to the dataset during the study period. The batch process for both sets was

159 carried out overnight without any technical issues.

160   Reference standard

161   In a service evaluation study of the EIDON confocal scanner with human grading, it was evidenced

162   that the EIDON images were able to visualise high-risk retinopathy features which were missed by

163   the NDESP images.[18] Because of this, the selection of the reference standard can be debatable.

164   However, since this is a diabetic retinopathy screening study, the performance of the ARIAS was

165   assessed using the final NDESP image human grades as ground truth.

166   Statistical analysis

167   Statistical calculations were performed using R studio, version 1.1.463 (www.r-project.org). Accuracy

168   of sensitivity, false positive rate (specificity) and likelihood ratios were defined by 95% confidence

169   intervals (CIs) obtained using bootstrapping for the EIDON and NDESP image grades. In cases with

170   sensitivity estimates of 100%, the exact Clopper-Pearson method was used to obtain CI estimates.

171   Calculations were determined for any retinopathy (grades R1, U, M1, R2 or R3), vision-threatening

172   retinopathy (grades M1, R2 or R3) and for each grade of retinopathy separately.

173

174 **RESULTS**

175 A total of 2,508 eyes of 1,257 patients were included in the study. A total of 11,796 images were

176 obtained (5,061 and 6,735 images with the EIDON and standard fundus cameras, respectively). The

177 prevalence of retinopathy according to the final human NDESP grades for R0, R1, M1, R2 and R3 was

178 65.39%, 32.06%, 6.92%, 1.03%, and 0.48%, respectively. Prevalence of retinopathy according to final

179 human EIDON grades for R0, R1, M1, R2 and R3 was 57.68%, 39.14%, 7.08%, 1.67%, and 0.56%,

180 respectively. Table 1 shows the ARIAS sensitivity (detection rate) and false-positive rates for the

181 EIDON and NDESP images using the worst eye final human grade of the NDESP images as the

182 reference standard. The specificity for episodes graded as no retinopathy (R0M0) with the EIDON

183 images was 53%, and 74% with the NDESP images. The point estimates for sensitivity using the final

184 grade of the NDESP images as reference standard were: EIDON images 92.27% (95% CI 88.43-

185 94.69%) for any retinopathy, 99% (95% CI 95.35-100%) for vision-threatening retinopathy and 100%

186 (95% CI 61-100%) for proliferative retinopathy. Corresponding sensitivities for NDESP images were

187 92.26% (95% CI 88.37-94.69%) for any retinopathy, 100% (95% CI 99.53-100%) for vision-threatening

188 retinopathy and 100% (95% CI 61-100%) for proliferative retinopathy. The diagnostic accuracy for

189 the point estimates of Table 1 and likelihood ratios are shown in Table 2.

190 The ARIAS correctly classified all vision-threatening retinopathy cases when using the NDESP image

191 grades. When using the EIDON images, the ARIAS correctly classified 98.8% of the vision-threatening

192 cases (1 case missed). The case missed by the ARIAS using the EIDON images was a case with a

193 human grade of R1M1 in both sets of images requiring routine referral. The proportion of cases of

194 vision-threatening diabetic retinopathy missed by the ARIAS was therefore 1.02% (1 case) with the

195 EIDON images. For the most severe retinopathy grade (R3, proliferative retinopathy), all cases were

196 correctly classified by the ARIAS in both imaging modalities (sensitivity of 100%).

197 The comparison of the human grading of EIDON and NDESP images of this sample,[18] evidenced

198 detection of 8 additional cases of R2 and 1 additional case of R3 with the EIDON images due to

199 visualisation of features outside the field of view of the NDESP images (2 cases), and by cause of

200 diabetic retinopathy feature visualisation within the 45-degree fields (8 cases) which were not

201 identified when grading the NDESP images. A figure illustrating these differences is found in a

202 previous publication comparing the human grading of EIDON and NDESP images.[18]

203 If the reference standard for comparison against the ARIAS outcome were the human grading of the

204 EIDON images, sensitivity for the detection of any retinopathy is 91.38% (95% CI 87.84-93.69), and

205 99.08% (95% CI 95.65-100) for vision-threatening retinopathy when processing the EIDON images.

206 When analysing the NDESP images, sensitivity for any retinopathy is 83.30% (95% CI 78.57-86.35),

207 and for vision-threatening retinopathy 100% (95% CI 99.53-100). The accuracy, sensitivity, false

208 positive rate and likelihood ratios defined by 95% confidence intervals when using the final EIDON

209 human grade as reference standard are available online as supplementary material (Supplementary

210 tables 1 to 4).

211 The ARIAS provides an alternative classification that attempts to identify cases which require referral

212 (grades U, M1, R2 and R3) as an alternative output ("refer" vs "no refer"). The findings of this

213 alternative output for the EIDON and NDESP images in terms of sensitivity (detection rate), false-

214 positive rates, and diagnostic accuracy of the point estimates are presented as supplementary

215 material (Supplementary tables 5 and 6). A marked effect for patients with R0M0 and R1M0 human

216 grades was found when comparing the ARIAS "disease" vs "no disease", and "refer" vs "no refer"

217 output in the EIDON images results (See tables 1 and 3, respectively). The ARIAS classified 53% of the

218 R0M0 episodes as "no disease", compared with 78% classified as "no refer". From the R1M0 cases,

219 9% were classified as "no disease", compared with 50% classified as "no refer". The NDESP images

220 automated grading showed a marked effect in patients graded R1M0 when comparing between

221 these two different ARIAS output. The ARIAS classified 9% of the R1M0 cases as "no disease",

222 compared with 63% classified as "no refer". The impact on the other retinopathy grades using this

223 comparison was less marked. Sensitivity for R1M1 reduced from 99% to 90% using the EIDON

224 images. A sensitivity reduction for R1M1 from 100% to 92% was found using the NDESP images. No

225 change was present for R3 grades in either image modality. The likelihood ratios for R1M1, R2 and

226 R3 detection were greater with the NDESP images and were even greater for both imaging

227 modalities in the output to identify referral (Tables 2 and 4).

228

229

230

231

232

233

234

235 Table 1. Screening performance of the EyeArt software with EIDON and NDESP (National Diabetic Eye Screening Programme) images compared with final

236 human grade of NDESP images using the classification of disease vs. no disease.

Human Grade (Worst Eye)	No. of Screening Episodes (Number, %[a])		EyeArt Outcome (Row %[b])							
			EIDON images				NDESP images			
			No disease		Disease		No disease		Disease	
Retinopathy grade	n	%	n	%	n	%	n	%	n	%
R0M0	822	65.39	438	53.28	384*	46.72	611	74.33	211*	25.67
R1M0	324	25.78	28	8.64	296	91.36	29	8.95	295	91.05
U	13	1.03	5	38.46	8	61.54	4	30.77	9	69.23
R1M1	79	6.28	1	1.27	78	98.73	0	0.00	79	100.00
R2	13	1.03	0	0.00	13	100.00	0	0.00	13	100.00
R2M0	9	0.72	0	0.00	9	100.00	0	0.00	9	100.00
R2M1	4	0.32	0	0.00	4	100.00	0	0.00	4	100.00
R3	6	0.48	0	0.00	6	100.00	0	0.00	6	100.00
R3M0	2	0.16	0	0.00	2	100.00	0	0.00	2	100.00
R3M1	4	0.32	0	0.00	4	100.00	0	0.00	4	100.00
Combination of grades										
R0M0, R1M0	1146	91.17	466	40.66	680	59.34	640	55.85	506	44.15
R1M1, R2, R3	98	7.80	1	1.02	97	98.98	0	0.00	98	100.00
U, R1M1, R2, R3	111	8.83	6	5.41	105	94.59	4	3.60	107	96.40

R1M0, U, R1M1, R2, R3	435	34.61	34	7.82	401	92.18	33	7.59	402	92.41
Total	1257	100.00	472	37.55	785	62.45	644	51.23	613	48.77

* Estimates refer to the proportion classified as disease present (i.e. false positives).

a Percentage of the total screened.

b Percentage within each human grade.

Grading classification in order of increasing severity are no retinopathy (R0), background retinopathy (R1), no maculopathy (M0), ungradable (U), maculopathy (M1), pre-proliferative retinopathy (R2) and proliferative retinopathy (R3).

238   Table 2. Screening performance of the EyeArt software with EIDON and NDESP (National Diabetic Eye Screening Programme) images compared

239   with final human grade of NDESP images: 95% confidence limits and likelihood ratios using the classification disease vs. no disease

	EIDON		NDESP	
Human Grade (Worst Eye)	**Proportion classified as disease present**		**Proportion classified as disease present**	
	Estimate (95% Confidence Interval)	Likelihood ratio vs. R0 (95% Confidence Interval)	Estimate (95% Confidence Interval)	Likelihood ratio vs. R0 (95% Confidence Interval)
Retinopathy grade				
R0M0*	0.47 (0.42-0.50)	-	0.26 (0.22-0.29)	-
R1M0	0.91 (0.87-0.94)	1.74 (1.62-1.87)	0.91 (0.86-0.94)	2.67 (2.42-2.93)
U	0.63 (0.18-0.88)	0.87 (0.51-1.50)	0.70 (0.25-0.93)	1.50 (1.04-2.16)
R1M1	0.99 (0.94-1.00)	1.65 (1.57-1.74)	1 (0.96-1)	2.18 (2.04-2.34)
R2	1 (0.79-1.00)	1.61 (1.54-1.68)	1 (0.79-1)	1.92 (1.65-2.25)
R2M0	1 (0.72-1)	1.61 (1.54-1.68)	1 (0.72-1)	1.86 (1.50-2.30)
R2M1	1 (0.47-1)	1.60 (1.54-1.67)	1 (0.47-1)	2.06 (1.94-2.18)
R3	1 (0.61-1)	1.61 (1.54-1.68)	1 (0.61-1)	2.06 (1.95-2.18)
R3M0	1 (0.22-1)	1.60 (1.54-1.67)	1 (0.22-1)	2.05 (1.94-2.17)
R3M1	1 (0.47-1)	1.60 (1.54-1.67)	1 (0.47-1)	2.06 (1.94-2.18)
Combination of grades				

R0M0, R1M0	0.59 (0.55-0.62)	-	0.44 (0.40-0.47)	-
R1M1, R2, R3	0.99 (0.95-1)	1.67 (1.59-1.76)	1 (0.95-1)	2.21 (2.06-2.37)
U, R1M1, R2, R3	0.95 (0.88-0.98)	1.60 (1.50-1.71)	0.97 (0.91-0.99)	2.17 (2.01-2.34)
R1M0, U, R1M1, R2, R3	0.92 (0.88-0.95)	1.98 (1.83-2.14)	0.92 (0.88-0.95)	3.60 (3.19-4.06)

*Estimates refer to the proportion classified as disease present (i.e. false positives).

Grading classification in order of increasing severity are no retinopathy (R0), background retinopathy (R1), no maculopathy (M0), ungradable (U), maculopathy (M1), pre-proliferative retinopathy (R2) and proliferative retinopathy (R3).

240

241

242

243

244

245

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

246   **DISCUSSION**

247   The photographic screening of diabetic retinopathy is a complex multilevel task requiring

248   comprehensive image assessment by human graders. Our study demonstrates that the overall

249   screening performance and diagnostic accuracy of the EyeArt software using the EIDON images is

250   acceptable in terms of sensitivity for detection of any retinopathy and vision-threatening diabetic

251   retinopathy. The diagnostic accuracy of the EyeArt and EIDON images falls within the sensitivity

252   levels first set by the British Diabetic Association for a diabetic retinopathy screening test,[16]

253   achieving good sensitivity when compared with human graders. The use of diverse retinal images

254   with different quality and, in this case, from platforms with different optical properties in a large-

255   scale community screening programme, contributes to the literature and can be used to assess

256   future deployment of EyeArt with standard fundus photographs and/or true-colour wide-field

257   confocal scanning images in diabetic retinopathy screening programmes. Differences in performance

258   of ARIAS related to new imaging technology, new diagnostic algorithms, or new screening

259   populations (e.g. new ethnicity mix, higher prevalence of cataract) may affect the diagnostic

260   accuracy or the cost-effectiveness of this technology. When considering large scale, population-

261   based screening programmes, independent validation and rigorous standardisation will be needed

262   to ensure the potential of ARIAS to save vision is not squandered.

263   The sensitivity above 92% for any retinopathy and 100% for vision-threatening retinopathy found in

264   our analysis adds to the previous reports using standard fundus photographs.[14,15,27–29] A recent

265   study in 100,000 consecutive patients, using the EyePACS imaging protocol, from 404 primary care

266   clinics comparing the performance of the EyeArt v2.0 against grading from trained certified

267   ophthalmologists and optometrists, reported a sensitivity for any retinopathy and vision-threatening

268   disease of 91% and 98.5%,respectively.[27] Additionally, a report using images obtained from a

269   portable smartphone-based imaging device in 296 patients, found a sensitivity for any retinopathy

270   and vision-threatening diabetic retinopathy of 95.8% and 99.1%, respectively.[30]

271   The EyeArt picked up 99% of all the cases of vision-threatening retinopathy when analysing the

272   EIDON images, missing just 1 case of R1M1 requiring routine referral and not missing any of the

273   more severe R2 or R3 cases. Moreover, when analysing the NDESP images, the EyeArt correctly

274   classified all cases of vision-threatening disease, which was in agreement with previous work.[14,31]

275   Deployment of ARIAS within the NDESP before, or as a substitution, of the level 1 graders revealed

276   cost-effectiveness of the approach for the screening of diabetic retinopathy in both cases.[13,14,32]

277   Because of the referral pathway structure of the English NDESP, even if an ARIAS is overly sensitive,

278   the patient is likely to achieve the appropriate outcome at the end of the screening episode. Our

279 study evidenced a lower false-positive rate for R0M0 with both imaging modalities (47% and 26%
280 with the EIDON and NDESP images, respectively), hence higher specificity, when compared with a
281 previous version of the EyeArt software reported in the study of Tufail et al.[14] With this
282 performance, if the software were to be hypothetically deployed as a part of the English NDESP, the
283 EyeArt could reduce the need to grade R0M0 by half when using EIDON images and by almost two
284 thirds when using the NDESP images, a considerable workload reduction. It should be emphasized
285 that in the English NDESP, graders are mostly non-medical personnel and cost-effectiveness in
286 different health care settings would depend on the graders used, the salary band equivalents, and
287 their sensitivity and specificity for detection of vision-threatening retinopathy.

288 Previous studies have suggested that similar screening programmes using trained human graders
289 have a diagnostic accuracy comparable to that of ARIAS. Sensitivity for referable retinopathy for
290 human graders evaluating 1-field photographs with a reference standard defined as consensus grade
291 from expert graders, has been reported to be 91.9% on average (range 81.9% to 95.0%).[33] The
292 reference standard for comparison in our study was the human consensus grade of the NDESP
293 images and the design did not look at the accuracy of human graders. However, the diagnostic
294 accuracy obtained in our analysis is acceptable for a diagnostic test. Attention must be drawn to the
295 different imaging platforms. Following the two-field imaging protocol, the EIDON platform provides
296 images by means of confocal scanning with a total field of view after combination of the macula- and
297 disc-centred image of 75-degree horizontal x 50-degree vertical.[21] On the other hand the
298 combination of the macula- and disc-centred NDESP images provide a 60-degree horizontal x 45-
299 degree vertical field of view. The difference in the field of view and image quality due to the
300 different optics of each platform could play a role when deployed for analysis with the currently
301 available ARIAS. Furthermore, the EyeArt has not been optimised for EIDON images and algorithm
302 reference patterns might not be properly recognised by the current version of the software.
303 Although the performance is already acceptable, future iterations of EyeArt software may yield even
304 better performance if optimised for EIDON images.

305 This study has certain limitations. Demography, duration of diabetes, ethnicity, time taken for
306 imaging with each imaging platform, and pupillary diameter of this dataset were not analysed. It has
307 been reported that the majority of the population who undergoes diabetic retinopathy screening are
308 older than 60 years.[34] Since the recruitment of the sample was carried out in a large-scale,
309 community-based screening programme, we have assumed that the demographic data of this
310 sample is likely to be representative of the population who undergoes screening on a yearly basis.
311 Ethnicity may influence the performance of an ARIAS due to different levels of fundus pigmentation

312 and fundus colour.[35] However, a study analysing a sample of 20,212 patients with White

313 European, Asian and Black African-Caribbean ethnicities found no strong evidence to suggest that

314 the sensitivity of a previous version of the ARIAS used in our study varies by ethnicity or sex.[14]

315 There might be a "black-box" issue with the EyeArt and the process of EIDON images because the

316 reference parameters or data points used by the software might not be the same as the ones used in

317 standard 45-degree colour fundus images, hence potential differences in grading. Further work is

318 needed to define if the wide-field true-colour images provide advantages in terms of diagnostic

319 accuracy with the EyeArt software. A health economic model may be warranted to evaluate the

320 deployment of these technologies in large-scale screening programmes.

321

322 **CONCLUSION**

323 The use of the EyeArt software for analysis of true-colour wide-field confocal scanning images

324 (EIDON) is shown to be accurate and sensitive enough for diabetic retinopathy screening in a large-

325 scale, community screening programme setting in comparison with standard retinal photographs per

326 English NDESP protocol. Proper implementation and use of the EyeArt, in conjunction with either

327 standard fundus photographs or true-colour wide-field confocal scanning images, and information

328 technology infrastructure, could address the evolving challenge of diabetic retinopathy screening

329 and improve the delivery of eye care.

331

339

COMPETING INTEREST

There are no competing interest to declare for any author.

342

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the North East London Diabetic Eye Screening Programme upon reasonable request.

351

CONTRIBUTORSHIP STATEMENT

353 As per ICMJE guidelines, all the authors agree to be accountable for all aspects of the work done on

354 this study. In addition, each individual author's contributions are:

355 A Olvera-Barrios. – Statistical analysis, interpretation of data, manuscript preparation and

356 manuscript approval.

357 T Heeren. – Statistical analysis, interpretation of data, manuscript preparation and manuscript

358 approval.

359 Konstantinos Balaskas. – Acquisition of data, manuscript preparation and manuscript approval.

360 Ryan Chambers. – Acquisition of data, manuscript preparation and manuscript approval.

361 Louis Bolter. – Acquisition of data, manuscript preparation and manuscript approval.

362 Adnan Tufail. – Study conception and design, interpretation of data, manuscript preparation and

363 manuscript approval.

364 Catherine Egan. – Study conception and design, interpretation of data, manuscript preparation and

365 manuscript approval.

366 John Anderson. – Study conception and design, interpretation of data, manuscript preparation and

367 manuscript approval.

368

370

REFERENCES

1   International Diabetes Federation, Federation ID. IDF Diabetes Atlas, 8th edn. Brussels,

     Belgium: International Diabetes Federation. *Int Diabetes Fed* Published Online First: 2017.

     doi:10.1289/image.ehp.v119.i03

2   Cho NH, Shaw JE, Karuranga S, *et al.* IDF Diabetes Atlas: Global estimates of diabetes

     prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract* 2018;**138**:271–81.

     doi:10.1016/j.diabres.2018.02.023

3   Mohamed Q, Gillies MC, Wong TY. Management of diabetic retinopathy: A systematic review.

     J. Am. Med. Assoc. 2007. doi:10.1001/jama.298.8.902

4   Ferris FL. How Effective Are Treatments for Diabetic Retinopathy? *JAMA J Am Med Assoc*

     Published Online First: 1993. doi:10.1111/j.1600-079X.1987.tb00846.x

5   Guariguata L, Whiting DR, Hambleton I, *et al.* Global estimates of diabetes prevalence for

     2013 and projections for 2035. *Diabetes Res Clin Pract* 2014;**103**:137–49.

     doi:10.1016/j.diabres.2013.11.002

6   Fleming AD, Goatman KA, Philip S, *et al.* Automated grading for diabetic retinopathy: A large-

     scale audit using arbitration by clinical experts. *Br J Ophthalmol* Published Online First: 2010.

     doi:10.1136/bjo.2009.176784

7   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 - J Am Med Assoc*

     Published Online First: 2016. doi:10.1001/jama.2016.17216

8   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;**94043**:1–7.

     doi:10.1001/jamaophthalmol.2019.2004

9   Sim DA, Keane PA, Tufail A, *et al.* Automated Retinal Image Analysis for Diabetic Retinopathy

     in Telemedicine. Curr. Diab. Rep. 2015. doi:10.1007/s11892-015-0577-6

10   Abràmoff MD, Niemeijer M, Suttorp-Schulten MSA, *et al.* Evaluation of a system for

     automatic detection of diabetic retinopathy from color fundus photographs in a large

     population of patients with diabetes. *Diabetes Care* Published Online First: 2008.

399   doi:10.2337/dc07-1312

400   11   Abràmoff MD, Reinhardt JM, Russell SR, *et al.* Automated Early Detection of Diabetic

401   Retinopathy. *Ophthalmology* Published Online First: 2010. doi:10.1016/j.ophtha.2010.03.046

402   12   Niemeijer M, Van Ginneken B, Cree MJ, *et al.* Retinopathy online challenge: Automatic

403   detection of microaneurysms in digital color fundus photographs. *IEEE Trans Med Imaging*

404   Published Online First: 2010. doi:10.1109/TMI.2009.2033909

405   13   Scotland GS, McNamee P, Fleming AD, *et al.* Costs and consequences of automated

406   algorithms versus manual grading for the detection of referable diabetic retinopathy. *Br J*

407   *Ophthalmol* Published Online First: 2010. doi:10.1136/bjo.2008.151126

408   14   Tufail A, Kapetanakis V V., Salas-Vega S, *et al.* An observational study to assess if automated

409   diabetic retinopathy image assessment software can replace one or more steps of manual

410   imaging grading and to determine their cost-effectiveness. *Health Technol Assess (Rockv)*

411   2016;**20**:1–72. doi:10.3310/hta20920

412   15   Tufail A, Rudisill C, Egan C, *et al.* Automated Diabetic Retinopathy Image Assessment

413   Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders.

414   *Ophthalmology* 2017;**124**:343–51. doi:10.1016/j.survophthal.2018.09.002

415   16   Scanlon PH. The English National Screening Programme for diabetic retinopathy 2003–2016.

416   Acta Diabetol. 2017. doi:10.1007/s00592-017-0974-1

417   17   Silva PS, Horton MB, Clary D, *et al.* Identification of Diabetic Retinopathy and Ungradable

418   Image Rate with Ultrawide Field Imaging in a National Teleophthalmology Program.

419   *Ophthalmology* 2016;**123**:1360–7. doi:10.1016/j.ophtha.2016.01.043

420   18   Olvera-Barrios A, Tufail A, Egan C, *et al.* Comparison of True-colour Wide-field Confocal

421   Scanner Imaging (EIDON) with Standard 2-field Fundus Photography for Diabetic Retinopathy

422   Screening. *Br J Ophthalmol* Published Online First: 2020. doi:10.1136/bjophthalmol-2019-

423   315269

424   19   Silva PS, Cavallerano JD, Sun JK, *et al.* Peripheral lesions identified by mydriatic ultrawide field

425   imaging: Distribution and potential impact on diabetic retinopathy severity. *Ophthalmology*

426   2013;**120**:2587–95. doi:10.1016/j.ophtha.2013.05.004

427   20   Silva PS, Cavallerano JD, Haddad NMN, *et al.* Peripheral lesions identified on ultrawide field

428   imaging predict increased risk of diabetic retinopathy progression over 4 years.

429   *Ophthalmology* 2015;**122**:949–56. doi:10.1016/j.ophtha.2015.01.008

430   21   EIDON-Brochure. The First True-Color Wide-Field Confocal Scanner.

431   https://www.centervue.com/wp-content/uploads/2016/05/EIDON-Brochure_REV02-

432   160307_US.pdf (accessed 8 May 2019).

433   22   Core National Diabetic Eye Screening Programme team. Diabetic Eye Screening Feature Based

434   Grading Forms, Version 1.4. 2012.

435   https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_d

436   ata/file/402295/Feature_Based_Grading_Forms_V1_4_1Nov12_SSG.pdf (accessed 21 May

437   2019).

438   23   Taylor D. Diabetic Eye Screening Programme Grading definitions for referable disease Public

439   Health England leads the NHS Screening Programmes. 2017.

440   https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_d

441   ata/file/582710/Grading_definitions_for_referrable_disease_2017_new_110117.pdf

442   (accessed 21 May 2019).

443   24   NHS Diabetic Screening Programme. Operational Guidance.

444   2015.https://www.gov.uk/government/collections/diabetic-eye-screening-commission-and-

445   provide (accessed 21 May 2019).

446   25   NHS Diabetic Screening Programme. Diabetic eye screening: guidance on camera approval -

447   GOV.UK. https://www.gov.uk/government/publications/diabetic-eye-screening-approved-

448   cameras-and-settings/diabetic-eye-screening-guidance-on-camera-approval (accessed 4 Nov

449   2019).

450   26   Harding S, Greenwood R, Aldington S, *et al.* Grading and disease management in national

451   screening for diabetic retinopathy in England and Wales. *Diabet Med* 2003;**20**:965–71.

452   doi:10.1111/j.1464-5491.2003.01077.x

453   27   Bhaskaranand M, Ramachandra C, Bhat S, *et al.* The Value of Automated Diabetic

454   Retinopathy Screening with the EyeArt System: A Study of More Than 100,000 Consecutive

455   Encounters from People with Diabetes. *Diabetes Technol Ther* 2019;:dia.2019.0164.

456   doi:10.1089/dia.2019.0164

457   28   Kapetanakis V V., Rudnicka AR, Liew G, *et al.* A study of whether automated diabetic

458   retinopathy image assessment could replace manual grading steps in the English national

459   screening Programme. *J Med Screen* 2015;**22**:112–8. doi:10.1177/0969141315571953

460 29 Bhaskaranand M, Cuadros J, Ramachandra C, *et al.* EyeArt + EyePACS: Automated Retinal
461   Image Analysis For Diabetic Retinopathy Screening in a Telemedicine System. 2017.
462   doi:10.17077/omia.1033

463 30 R. R, R. S, R.M. A, *et al.* Automated diabetic retinopathy detection in smartphone-based
464   fundus photography using artificial intelligence. *Eye* Published Online First: 2018.
465   doi:10.1038/s41433-018-0064-9

466 31 Tufail A, Rudisill C, Egan C, *et al.* Automated Diabetic Retinopathy Image Assessment
467   Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders.
468   *Ophthalmology* 2017;**124**:343–51. doi:10.1016/j.ophtha.2016.11.014

469 32 Scotland GS, McNamee P, Philip S, *et al.* Cost-effectiveness of implementing automated
470   grading within the national screening programme for diabetic retinopathy in Scotland. *Br J
471   Ophthalmol* Published Online First: 2007. doi:10.1136/bjo.2007.120972

472 33 Goatman KA, Philip S, Fleming AD, *et al.* External quality assurance for image grading in the
473   Scottish Diabetic Retinopathy Screening Programme. *Diabet Med* Published Online First:
474   2012. doi:10.1111/j.1464-5491.2011.03504.x

475 34 Scanlon PH, Malhotra R, Greenwood RH, *et al.* Comparison of two reference standards in
476   validating two field mydriatic digital photography as a method of screening for diabetic
477   retinopathy. *Br J Ophthalmol* 2003;**87**:1258–63. doi:10.1136/bjo.87.10.1258

478 35 Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning.
479   *Ophthalmology* Published Online First: 2017. doi:10.1016/j.ophtha.2017.02.008

480

Supplementary Table 1. Screening performance of EyeArt with EIDON and National Diabetes Eye Screening Programme (NDESP) images compared with the human grade of the EIDON images as ground truth using the classification of "disease" vs. "no disease"

Manual Grade (Worse Eye)	No. of Screening Episodes (Number, %)		EyeArt Outcome (Row %) EIDON images				NDESP images			
			No disease		Disease		No disease		Disease	
Retinopathy grade	n	%	n	%	n	%	n	%	n	%
R0M0	725	57.68	426	58.76	299	41.24	555	76.55	170	23.45
R1M0	414	32.94	43	10.39	371	89.61	85	20.53	329	79.47
U	12	0.95	2	16.67	10	83.33	4	33.33	8	66.67
R1M1	78	6.21	1	1.28	77	98.72	0	0.00	78	100.00
R2	21	1.67	0	0.00	21	100.00	0	0.00	21	100.00
R2M0	15	1.19	0	0.00	15	100.00	0	0.00	15	100.00
R2M1	6	0.48	0	0.00	6	100.00	0	0.00	6	100.00
R3	7	0.56	0	0.00	7	100.00	0	0.00	7	100.00
R3M0	2	0.16	0	0.00	2	100.00	0	0.00	2	100.00
R3M1	5	0.40	0	0.00	5	100.00	0	0.00	5	100.00
Combination of grades										
R0M0, R1M0	1139	90.61	469	41.18	670	58.82	640	56.19	499	43.81
R1M1, R2, R3	106	8.43	1	0.94	105	99.06	0	0.00	106	100.00
U, R1M1, R2, R3	118	9.39	3	2.54	115	97.46	4	3.39	114	96.61
R1M0, U, R1M1, R2, R3	532	42.32	46	8.65	486	91.35	89	16.73	443	83.27
Total	1257	100.00	472	37.55	785	62.45	644	51.23	613	48.77

Supplementary Table 2. Screening performance of EyeArt software using EIDON and National Diabetes Eye Screening Programme (NDESP) images compared with the human grade of the EIDON images as ground truth: 95% confidence limits and likelihood ratios using the EyeArt classification "disease" vs. "no disease"

Manual Grade (Worst Eye)	EIDON		NDESP	
	Proportion classified as disease present		**Proportion classified as disease present**	
	Estimate (95% Confidence Interval)	Likelihood ratio vs. R0 (95% Confidence Interval)	Estimate (95% Confidence Interval)	Likelihood ratio vs. R0 (95% Confidence Interval)
Retinopathy grade				
R0M0*	0.41 (0.36-0.45)	-	0.23 (0.19-0.26)	-
R1M0	0.99 (0.94-1)	1.82 (1.69-1.97)	0.79 (0.74-0.83)	2.36 (2.12-2.62)
U	0.84 (0.42-1)	1.34 (1.04-1.73)	0.67 (0.22-0.92)	1.37 (0.92-2.06)
R1M1	0.99 (0.94-1)	1.64 (1.56-1.73)	1 (0.96-1)	2.2 (2.07-2.35)
R2	1 (0.87-1)	1.62 (1.55-1.69)	1 (0.87-1)	2.09 (1.97-2.21)
R2M0	1 (0.82-1)	1.61 (1.54-1.68)	1 (0.82-1)	2.08 (1.96-2.2)
R2M1	1 (0.61-1)	1.61 (1.54-1.68)	1 (0.61-1)	2.06 (1.95-2.18)
R3	1 (0.65-1)	1.61 (1.54-1.68)	1 (0.65-1)	2.06 (1.95-2.18)
R3M0	1 (0.22-1)	1.6 (1.54-1.67)	1 (0.22-1)	2.05 (1.94-2.17)
R3M1	1 (0.55-1)	1.61 (1.54-1.68)	1 (0.55-1)	2.06 (1.95-2.18)
Combination of grades				
R0M0, R1M0	0.59 (0.55-0.62)	-	0.44 (0.39-0.47)	-
R1M1, R2, R3	0.99 (0.95-1)	1.68 (1.59-1.77)	1 (0.97-1)	2.27 (2.13-2.42)
U, R1M1, R2, R3	0.98 (0.93-1)	1.66 (1.57-1.75)	0.97 (0.91-0.99)	2.21 (2.05-2.37)
R1M0, U, R1M1, R2, R3	0.91 (0.88-0.94)	2.22 (2.02-2.43)	0.83 (0.79-0.86)	3.55 (3.1-4.07)

*Estimates refer to the proportion classified as disease present (i.e. false positives)

Grading classification in order of increasing severity are no retinopathy (R0), background retinopathy (R1), no maculopathy (M0), ungradable (U), maculopathy (M1), pre-proliferative retinopathy (R2) and proliferative retinopathy (R3).

Supplementary Table 3. Screening performance of EyeArt with EIDON and National Diabetes Eye Screening Programme (NDESP) images compared with the human grade of the EIDON images as ground truth using the classification of "refer" vs. "no refer"

			EyeArt Outcome (Row %)							
			EIDON images				NDESP images			
Manual Grade (Worse Eye)	No. of Screening Episodes (Number, %)		No refer		refer		No refer		refer	
Retinopathy grade	n	%	n	%	n	%	n	%	n	%
R0M0	725	57.68	578	79.72	147	20.28	675	93.10	50	6.90
R1M0	414	32.94	223	53.86	191	46.14	277	66.91	137	33.09
U	12	0.95	4	33.33	8	66.67	10	83.33	2	16.67
R1M1	78	6.21	10	12.82	68	87.18	8	10.26	70	89.74
R2	21	1.67	1	4.76	20	95.24	2	9.52	19	90.48
R2M0	15	1.19	1	6.67	14	93.33	2	13.33	13	86.67
R2M1	6	0.48	0	0.00	6	100.00	0	0.00	6	100.00
R3	7	0.56	0	0.00	7	100.00	0	0.00	7	100.00
R3M0	2	0.16	0	0.00	2	100.00	0	0.00	2	100.00
R3M1	5	0.40	0	0.00	5	100.00	0	0.00	5	100.00
Combination of grades		0.00								
R0M0, R1M0	1146	91.17	801	69.90	338	29.49	952	83.07	187	16.32
R1M1, R2, R3	98	7.80	11	11.22	95	96.94	10	10.20	96	97.96
U, R1M1, R2, R3	111	8.83	15	13.51	103	92.79	20	18.02	98	88.29
R1M0, U, R1M1, R2, R3	435	34.61	238	54.71	294	67.59	297	68.28	235	54.02
Total	1257	100.00	816	64.92	441	35.08	972	77.33	285	22.67

Grading classification in order of increasing severity are no retinopathy (R0), background retinopathy (R1), no maculopathy (M0), ungradable (U), maculopathy (M1), pre-proliferative retinopathy (R2) and proliferative retinopathy (R3).

Supplementary Table 4. Screening performance of EyeArt software using EIDON and National Diabetes Eye Screening Programme (NDESP) images compared with the human grade of the EIDON images as ground truth: 95% confidence limits and likelihood ratios using the EyeArt classification "referable" vs. "non-referable retinopathy"

Manual Grade (Worst Eye)	EIDON		NDESP	
	Proportion classified as refer		Proportion classified as disease present	
	Estimate (95% Confidence Interval)	Likelihood ratio vs. R0 (95% Confidence Interval)	Estimate (95% Confidence Interval)	Likelihood ratio vs. R0 + R1M0 (95% Confidence Interval)
Retinopathy grade				
R0M0*	0.2 (0.16-0.23)	-	0.07 (0.04-0.09)	-
R1M0	0.46 (0.39-0.51)	-	0.33 (0.26-0.38)	-
U	0.67 (0.22-0.92)	1.92 (1.28-2.89)	0.16 (0-0.42)	0.73 (0.21-2.61)
R1M1	0.87 (0.76-0.94)	2.76 (2.45-3.11)	0.9 (0.79-0.96)	4.92 (4.27-5.67)
R2	0.96 (0.77-1)	2.8 (2.47-3.16)	0.91 (0.68-1)	4.2 (3.53-5.01)
R2M0	0.94 (0.67-1)	2.71 (2.32-3.17)	0.88 (0.56-1)	3.96 (3.16-4.95)
R2M1	1 (0.61-1)	2.88 (2.67-3.1)	1 (0.55-1)	4.48 (4.04-4.97)
R3	1 (0.65-1)	2.88 (2.67-3.11)	1 (0.65-1)	4.5 (4.05-4.99)
R3M0	1 (0.22-1)	2.86 (2.65-3.08)	1 (0.22-1)	4.43 (4-4.91)
R3M1	1 (0.55-1)	2.87 (2.66-3.1)	1 (0.55-1)	4.47 (4.03-4.96)
Combination of grades				
R0M0, R1M0	0.3 (0.26-0.32)	-	0.16 (0.13-0.19)	-
R1M1, R2, R3	0.9 (0.81-0.95)	2.98 (2.67-3.33)	0.91 (0.82-0.96)	5.52 (4.78-6.37)
U, R1M1, R2, R3	0.87 (0.78-0.93)	2.94 (2.63-3.29)	0.83 (0.73-0.9)	5.06 (4.34-5.9)
R1M0, U, R1M1, R2, R3	0.55 (0.49-0.59)	2.73 (2.31-3.21)	0.44 (0.38-0.48)	6.41 (4.82-8.51)

*Estimates refer to the proportion classified as refer (i.e. false positives)

Grading classification in order of increasing severity are no retinopathy (R0), background retinopathy (R1), no maculopathy (M0), ungradable (U), maculopathy (M1), pre-proliferative retinopathy (R2) and proliferative retinopathy (R3).

Supplementary Table 5. Screening performance of the automated retinal imaging analysis software with EIDON and National Diabetes Eye Screening Programme (NDESP) images compared with final human grade of NDESP images using the classification of referable vs. non-referable retinopathy

			EyeArt Outcome (Row %[b])							
			EIDON images				NDESP images			
Human Grade (Worst Eye)	No. of Screening Episodes (Number, %[a])		No refer		refer		No refer		refer	
Retinopathy grade	n	%	n	%	n	%	n	%	n	%
R0M0	822	65.39	640	77.86	182[*]	22.14	758	92.21	63	7.66
R1M0	324	25.78	162	50.00	162	50.00	203	62.65	121	37.35
U	13	1.03	6	46.15	7	53.85	4	30.77	9	69.23
R1M1	79	6.28	8	10.13	71	89.87	6	7.59	73	92.41
R2	13	1.03	0	0.00	13	100.00	0	0.00	13	100.00
R2M0	9	0.72	0	0.00	9	100.00	0	0.00	9	100.00
R2M1	4	0.32	0	0.00	4	100.00	0	0.00	4	100.00
R3	6	0.48	0	0.00	6	100.00	0	0.00	6	100.00
R3M0	2	0.16	0	0.00	2	100.00	0	0.00	2	100.00
R3M1	4	0.32	0	0.00	4	100.00	0	0.00	4	100.00
Combination of grades										
R0M0, R1M0	1146	91.17	802	69.98	344	30.02	961	83.86	184	16.06
R1M1, R2, R3	98	7.80	8	8.16	90	91.84	6	6.12	92	93.88
U, R1M1, R2, R3	111	8.83	14	12.61	97	87.39	10	9.01	101	90.99
R1M0, U, R1M1, R2, R3	435	34.61	176	40.46	259	59.54	213	48.97	222	51.03
Total	1257	100.00	816	64.92	441	35.08	971	77.25	285	22.67

[*] Estimates refer to the proportion classified as disease present (i.e. false positives).

[a] Percentage of the total screened.

[b] Percentage within each human grade.

Grading classification in order of increasing severity are no retinopathy (R0), background retinopathy (R1), no maculopathy (M0), ungradable (U), maculopathy (M1), pre-proliferative retinopathy (R2) and proliferative retinopathy (R3).

Supplementary Table 6. Screening performance of the EyeArt software with EIDON and National Diabetes Eye Screening Programme (NDESP) images compared with final human grade of NDESP images: 95% confidence limits and likelihood ratios using the classification referable vs. non-referable retinopathy.

Human Grade (Worst Eye)	EIDON		NDESP	
	Proportion classified as refer		Proportion classified as refer	
	Estimate (95% Confidence Interval)	Likelihood ratio vs. R0 (95% Confidence Interval)	Estimate (95% Confidence Interval)	Likelihood ratio vs. R0 (95% Confidence Interval)
Retinopathy grade				
R0M0*	0.22 (0.18-0.25)	-	0.08 (0.05-0.09)	-
R1M0	0.50 (0.42-0.55)	-	0.37 (0.30-0.43)	-
U	0.54 (0.11-0.82)	1.30 (0.68-2.49)	0.70 (0.25-0.93)	3.27 (2.25-4.77)
R1M1	0.90 (0.79-0.96)	2.85 (2.55-3.19)	0.93 (0.83-0.98)	5.05 (4.39-5.82)
R2	1 (0.79-1)	2.91 (2.70-3.14)	1 (0.79-1)	4.24 (3.55-5.08)
R2M0	1 (0.72-1)	2.89 (2.68-3.12)	1 (0.72-1)	4.07 (3.23-5.12)
R2M1	1 (0.47-1)	2.87 (2.66-3.09)	1 (0.47-1)	4.46 (4.02-4.94)
R3	1 (0.61-1)	2.88 (2.67-3.10)	1 (0.61-1)	4.48 (4.04-4.97)
R3M0	1 (0.22-1)	2.86 (2.65-3.08)	1 (0.22-1)	4.43 (4.00-4.91)
R3M1	1 (0.47-1)	2.87 (2.66-3.09)	1 (0.47-1)	4.46 (4.02-4.94)
Combination of grades				
R0M0, R1M0	0.30 (0.26-0.33)	-	0.16 (0.13-0.18)	-
R1M1, R2, R3	0.91 (0.82-0.96)	3.04 (2.73-3.38)	0.92 (0.84-0.97)	5.51 (4.78-6.36)
U, R1M1, R2, R3	0.87 (0.77-0.93)	2.90 (2.59-3.25)	0.90 (0.81-0.96)	5.59 (4.82-6.47)
R1M0, U, R1M1, R2, R3	0.60 (0.53-0.64)	2.72 (2.34-3.16)	0.51 (0.44-0.56)	6.75 (5.22-8.72)

*Estimates refer to the proportion classified as refer (i.e. false positives).

Grading classification in order of increasing severity are no retinopathy (R0), background retinopathy (R1), no maculopathy (M0), ungradable (U), maculopathy (M1), pre-proliferative retinopathy (R2) and proliferative retinopathy (R3).