

1 **Improved predictive diagnosis of diabetic macular edema based on hybrid models: an observational**
2 **study.**

3

4 Hughes-Cano JA¹, Quiroz-Mercado H², Hernández-Zimbrón LF^{2,3}, García-Franco R^{4,5}, Rubio Mijangos JF⁵,
5 López-Star E⁵, García-Roa M⁵, Lansingh VC^{5,6}, Olivares-Pinto U⁷, Thébault SC^{1*}.

6 ¹ Laboratorio de Investigación Traslacional en Salud Visual, Instituto de Neurobiología, Universidad
7 Nacional Autónoma de México (UNAM), Campus Juriquilla, Querétaro, Mexico.

8 ² Research Department, Asociación Para Evitar la Ceguera, Mexico City, Mexico

9 ³ Clínica de Salud Visual, Escuela Nacional de Estudios Superiores, Unidad León, Universidad Nacional
10 Autónoma de México (UNAM), León, Guanajuato, Mexico.

11 ⁴ Instituto de la Retina del Bajío (INDEREB), Prolongación Constituyentes 302 (Consultorios 410 y 411,
12 torre 3, Hospital San José), El jacal, 76187, Santiago de Querétaro, Querétaro, Mexico.

13 ⁵ Instituto Mexicano de Oftalmología (IMO), I.A.P., Circuito Exterior Estadio Corregidora Sn, Centro Sur,
14 76090 Santiago de Querétaro, Querétaro, Mexico.

15 ⁶ HelpMeSee, Inc., 20 West 36th Street, Floor 4, New York, NY, 10018-8005, USA

16 ⁷ Escuela Nacional de Estudios Superiores, Universidad Nacional Autónoma de México (UNAM), Campus
17 Juriquilla, Querétaro, Mexico.

18 * Corresponding author. Email: sthebault@comunidad.unam.mx

19 Abstract

20 Diabetic Macular Edema (DME) is the most common sight-threatening complication of type 2 diabetes.

21 Our goal was to develop an alternative method to optical coherence tomography (OCT) for DME diagnosis

22 by introducing spectral information derived from spontaneous electroretinogram (ERG) signals as a single

23 input or combined with eye fundus. To this end, an observational study was completed (n = 233

24 participants). Basal ERGs were used to generate scalograms and spectrograms via Wavelet and Fourier

25 transforms, respectively. Using transfer learning, distinct Convolutional Neural Networks (CNN) were

26 trained as classifiers for DME using OCT, scalogram, spectrogram, and fundus images. Input data were

27 randomly split into training and test sets with a proportion of 80 % to 20 %, respectively. The top

28 performers for each input type were selected, OpticNet-71 for OCT and DenseNet-201 for fundus and

29 non-evoked ERG-derived scalograms, to generate a combined model by assigning different weights for

30 each of the selected models. Model validation was performed using a dataset alien to the training phase

31 of the models. None of the models powered by non-evoked ERG-derived input performed well. Metrics

32 of the best hybrid models were all above 0.81 for fundus combined with non-evoked ERG-derived

33 information; and above 0.85 for OCT combined with non-evoked ERG-derived scalogram images. These

34 data show that the spontaneous ERG-based model improves all the performance metrics of the fundus

35 and OCT-based models, with the exception of sensitivity for the OCT model, to predict DME. Combining

36 non-evoked ERG with OCT represents an improvement to the existing OCT-based models, and combining

37 non-evoked ERG with fundus is a reliable and economical alternative for the diagnosis of DME in

38 underserved areas where OCT is unavailable.

39 Author summary

40 Providing an alternative diagnostic method to those that already exist for diabetic macular edema (DME)
41 that is reliable and physically and economically accessible is needed in places where optical coherence
42 tomography (OCT) is unavailable. In this work, we combined artificial intelligence (AI) classifying
43 techniques with information from a newly introduced signal that can be captured in a non-invasive
44 manner, the spontaneous oscillations of the electroretinogram (ERG). We found that if these signals alone
45 are ineffective in diagnosing DME cases, they improve the performance of AI models based on either eye
46 fundus or OCT in the prediction of DME. We therefore conclude that combining spontaneous ERG with
47 fundus, which is a basic optometric test even in underserved areas, represents a reliable alternative to
48 OCT for the diagnosis of DME. Also, combining OCT with spontaneous ERG signals will help ameliorate the
49 diagnosis of DME.

50 Introduction

51 Affecting 8.8 % of the world population and with an estimated number of cases reaching 783 million by
52 2045, type 2 diabetes is a modern pandemic (1,2). Diabetic macular edema (DME), characterized by the
53 accumulation of exudative fluid in the macula, is the most common form of retinopathy that threatens
54 vision in people with type 2 diabetes (3). DME is the leading cause of vision loss in diabetic individuals
55 from 20 to 74 years-old (4). In Mexico, its prevalence has been estimated at 6.6 %, but most importantly,
56 DME risk has been found to increase in the early stages of diabetes (5). As a major source of disability, it
57 is, therefore, necessary to find alternative diagnostic methods to those already existing that are effective,
58 accessible, and economical.

59 Artificial intelligence (AI) algorithms, particularly the start-of-the-art, continuously improving, machine
60 learning techniques are generating enormous interest in diagnosing various diseases (6). In this context,
61 AI algorithms known as convolutional neural networks (CNN) have been applied to the analysis of medical
62 images, showing robust performance in the diagnosis and detection of conditions such as pulmonary
63 tuberculosis from chest radiographs, malignant melanoma from skin photographs (6), and DME from
64 optical coherence tomography (OCT) and fundus images (7). Even though biomicroscopy examination of
65 the posterior pole of the eye remains the first step in DME diagnosis in many places, stereoscopic fundus
66 photographs lack stereopsis, rendering the diagnosis unreliable (8). In contrast, OCT provides quantitative
67 and qualitative biomarkers associated with visual and anatomical outcomes of DME (9), explaining why it
68 largely supplanted fundus examination in DME diagnosis. However, OCT devices are the prerogative of
69 rich countries.

70 Despite their good generalization features in various fields of medicine, CNNs come with a high
71 computational cost and require large amounts of medical data for efficient model training (6,7). Transfer
72 learning techniques address this issue by using pre-trained CNNs with millions of images of a different
73 nature (e.g., ImageNet dataset (10)), transferring the weights obtained from this process, and training the

74 network on a smaller dataset (6). Moreover, the development of hybrid models in AI (11), has enabled
75 the use of different types of algorithms or different data sources with similar algorithms to generate more
76 performant models (11,12).

77 In this study, we sought to improve the predictive power of fundus, as it is the most common DME test in
78 the world. To this end, we took advantage of non-evoked or spontaneous electroretinogram (ERG) signals
79 that have been recently shown to help predict risk factors of type 2 diabetes, including overweight,
80 obesity, and metabolic syndrome (13). This is in view that ERG alterations are detected in patients with
81 early DME (14,15), that portable devices can nowadays acquire ERG in a fully non-invasive way (16), and
82 those 5-minute ERGs, in the absence of any light flash, that is in mock conditions, and under photopic
83 conditions, which allows dispensing with mydriasis, is informative about early changes associated with
84 diabetes (17).

85 We generated hybrid CNN algorithms powered by different types of data, including images of OCT, fundus,
86 and non-evoked ERG spectrum obtained through either fast Fourier (spectrograms) or continuous wavelet
87 (scalograms) transforms (13). Our data show that, although it is impossible to distinguish patients with or
88 without DME using only non-evoked ERG spectral images, the performance of models using OCT and/or
89 fundus images as input data can be improved when combined with spectral images of spontaneous ERG
90 signals. Most notably, the hybrid model powered by fundus and mock ERG-derived wavelet scalogram
91 images performs as well as the OCT one.

92 Results

93 Mock ERG-based models for the predictive diagnosis of DME

94 We first asked whether the mock ERG signals *per se* help predict DME cases. To this end, two CNNs,
95 ResNet-50, and finely tuned DenseNet-201, powered by either the FFT or the wavelet-derived mock ERG
96 power spectrum, were tested. **Figure 1A, B** shows the ROC curves and confusion matrices for CNNs based
97 on FFT-derived information, while **Figure 1C, D** shows prediction results based on wavelet-derived
98 information. The performance metrics of externally validated predictions are summarized in **Figure 1E**.
99 For all models, the highest performance metric is specificity, and DenseNet-201 using scalograms is the
100 most sensitive one (**Figure 1E**), but none of them showed performance metrics above 0.80.

101 In view of the poor performance of the models purely based on mock ERG-derived information to predict
102 DME, we next analyzed whether the mock ERG transform could help improve the predictive performance
103 of OCT and/or eye fundus images for DME.

104

105 Hybrid models for the predictive diagnosis of DME

106 To implement hybrid models that combine mock ERG transforms, OCT, and eye fundus images, we first
107 determined which models based on each of these entries independently were best, and then used them
108 as inputs for the hybrid models.

109 As previously shown, finely-tuned DenseNet-201 trained with mock ERG-derived scalograms was the least
110 bad in predicting DME (**Figure 1E**).

111 As for OCT-based models, external validation using 1,229 images from our database (**Table 1**) showed
112 better results for OpticNet-71. Compared to DenseNet-201, its precision (0.93 vs. 0.52), sensitivity (0.93
113 vs. 0.91), specificity (0.99 vs. 0.83), F1-score (0.93 vs. 0.72), accuracy (0.98 vs. 0.84), ROC AUC (0.96 vs.
114 0.87), and Cohen's Kappa (0.91 vs. 0.57) were largely higher (**S1 figure**, panels A, B, and E). Using the 164-
115 image set (**Table 1**) of Srinivasan *et al.* (23), we further confirmed that the best overall performance for
116 the predictive diagnosis of DME is for OpticNet-71, with metrics above 0.84 (**S1 figure**, panels C, D, and

117 E). The sensitivity for DenseNet-201 was greater than the one of OpticNet-71 (0.95 vs. 0.93); nonetheless,
 118 the rest of the metrics were not (Precision: 0.85 vs. 0.63, specificity: 0.94 vs. 0.8, F1-score: 0.89 vs. 0.75,
 119 accuracy: 0.94 vs. 0.84, ROC AUC: 0.94 vs. 0.88, and Cohen’s Kappa: 0.84 vs. 0.64, respectively) (**S1 figure**,
 120 panels C, D, and E).

	EYE FUNDUS			OCT			SCALOGRAMS AND SPECTROGRAMS		
	Dataset	DME	No DME	Dataset	DME	No DME	Dataset	DME	No DME
Training and Test	Present study	1,011	373	Kermany's	11,598	26,565	Present study	141	1,048
	Messidor	0	151	-	-	-	Present study with data augmentation	846	1,048
	Pachade's	0	180	-	-	-	-	-	-
	Giancardo's	0	115	-	-	-	-	-	-
Validation	Present study	42	122	Srinivasan's	36	53	Present study	42	122
	-	-	-	Present study	176	964	-	-	-

121 **Table 1.** Distribution and composition of the databases used. It is important to highlight that the images
 122 used for model training and testing were not used for validating the model performance. TO this end,
 123 independent images were used. Furthermore, the only datasets partially generated with data
 124 augmentation were those containing spectrograms and scalograms, but only for the DME group.

125
 126 Three different models were tested for eye fundus image-based DME predictive diagnosis and externally
 127 validated with n = 164 (DME: 42, No DME: 122, **Table 1**). **S2 figure** (panel A) shows the ROC curve,
 128 confusion matrix, and performance metrics for the Resnet-50 model. At the same time, similar data are
 129 reported in **S2 figure** (panels B and C) for the finely tuned MobileNet-V2 and DenseNet-201 models,
 130 respectively. If ResNet-50 had a high sensitivity (0.86) and MobileNet-V2 a very high specificity (0.98), the

131 DenseNet-201 model obtained overall more robust performance metrics, especially for F1-Score and
132 Cohen's Kappa that evaluate the performance of models trained with unbalanced classes (**S2 figure**, panel
133 C).

134 Next, based on the demonstration that OpcticNet-71 trained with OCT images, DenseNet-201 trained with
135 eye fundus images, and DenseNet-201 trained with scalograms were the best individual models, we
136 created hybrid models based on the linear combination of each individual model prediction matrices
137 (**Figure 2**). All possible combinations of weight factor values were systematically analyzed, considering
138 incremental steps of 1 %. Heat maps for performance metrics of all hybrid models are shown in **Figure 3**.
139 For the hybrid model that combined both OCT and eye fundus images, the highest metrics (all above >
140 0.87) were obtained when the weighting factor n1 was 0.60 and the n2 of 0.40 (**Figure 3A**), meaning that
141 the contribution of OCT images was of 60 %, while the one of fundus images was of 40 %. In these
142 conditions, precision reached 0.90, sensitivity 0.90, specificity 0.97, F1-score 0.90, accuracy 0.95, ROC AUC
143 0.94, and Cohen's Kappa 0.87.

144 For the hybrid model that combined both OCT and mock ERG-derived scalogram images, the model is
145 optimal when the n value for OCT is 0.60 and for mock ERG 0.40 (**Figure 3B**), since that, in these conditions,
146 precision reached 0.88, sensitivity 0.90, specificity 0.96, F1-score 0.89, accuracy 0.95, ROC AUC 0.93, and
147 Cohen's Kappa: 0.85. Most notably, these data show that including mock ERG-derived information
148 improved the specificity, precision, Cohen's Kappa, and accuracy of the OCT-based model (for reference,
149 see **S1 figure**, panel A).

150 For the hybrid model that combined eye fundus and mock ERG-derived scalogram images, optimization
151 occurred in a wide range of combinations from 50 % eye fundus and above and 50 % ERG and below,
152 respectively (**Figure 3C**). For example, when 70 % of the input comes from the eye fundus-based prediction
153 and 30 % from the mock ERG-based prediction, precision reached 0.84, sensitivity 0.88, specificity 0.94,
154 F1-score 0.86, accuracy 0.93, ROC AUC 0.91, and Cohen's Kappa: 0.81 (**Figure 3C**). All above 0.8

155 performance metrics of the DenseNet-201 model trained with eye fundus images (**S2 figure**, panel C)
156 increased when it is combined with the mock ERG model in the 70-30 % proportions, except for sensitivity.
157 We also created triple hybrid models that combined mock ERG, fundus, and OCT image inputs, and
158 observed that several combinations led to optimal performance metrics (**Figure 3D**). The most notable
159 optimization happened when $n_1 = 0.60$, $n_2 = 0.40$, and $n_3 = 0$, corresponding to the double hybrid model
160 that combined OCT and mock ERG-derived scalogram images (**Figure 3D**). Nonetheless, there are other
161 optimized combinations, like the $n_1 = 0.60$, $n_2 = 0.28$, and $n_3 = 0.12$, for OCT, eye fundus, and mock ERG-
162 derived scalogram images, respectively, that reached a precision of 0.88, sensitivity of 0.90, specificity of
163 0.96, F1-score of 0.89, accuracy of 0.95, ROC AUC of 0.93, and Cohen's Kappa of 0.85.
164 A single heat-map included all hybrid models' performances (**Figure 4**).

165 Discussion

166 DME continues to be the leading cause of preventable blindness in the working-age population in the
167 world (8). Therefore, early detection programs are extremely important in treating this DR complication.
168 Introducing new imaging modalities and technological advances has facilitated both early detection and
169 follow-up of patients with DME, particularly OCT angiography and AI. However, OCT exams are not
170 accessible to everyone in developing countries, due to their high cost and lack of equipment. In this
171 context, we showed that a CNN model powered by images of changes in the frequency spectrum of basal
172 ERG signals could not be used to diagnose DME, but that the performance of OCT and fundus image-based
173 classifiers in predicting DME can be improved by combining them with the mock ERG transform cues.

174
175 The poor performance of our model based on mock-ERG signals alone can be explained by the state of
176 the disease, the nature of the data, and the characteristics of the used models. At continuation, we will
177 discuss each one of these issues.

178 DME is a complication of DR, it, therefore, develops in association with different degrees of DR, ranging
179 from moderate non-proliferative to advanced proliferative stages (29). Furthermore, DR, classically
180 defined as microangiopathy affecting the retinal vessels (30), is known to affect retinal neurons (31).
181 Neuronal apoptosis begins with retinal ganglion cells but can also affect other retinal nerve cells, such as
182 bipolar cells, amacrine cells, and photoreceptors (31). This neuronal damage can be evidenced by latency
183 delays in multifocal ERG (32). Our training, test, and validation data contain ERG records from DME or
184 non-DME patients diagnosed with DR to some degree. Therefore, spectral images from mock ERGs likely
185 show similar changes in both DME and non-DME groups, confusing the models. We confirmed this
186 scenario, since the performance metrics of the models trained to distinguish DR and no DR classes
187 exceeded 0.60, unlike the values obtained for the validation of the DME class prediction, indicating that
188 the concomitant presence of DR with DME takes part into the poor performance of our models.

189 Even though the models trained for DR classification performed better than those for DME, they are not
190 robust enough to make a classification of sufficient quality for clinical use, which suggests that other
191 variables, like the nature of the data and the models themselves, are influencing the classification.

192 If, to the best of our knowledge, there is no record of the relevance of basal ERG transform images for
193 human disease prediction, previous works have reported the use of electrophysiological signals, such as
194 EEG, transformed into spectral images to train CNNs to recognize original data from synthetic ones
195 (33,34). As for this CNN model that performs relatively poorly (metrics of ~ 0.6 for EEG-derived scalograms)
196 (34), the time-frequency representations of our data may not express the necessary or sufficient
197 information (67). Our sampling frequency is 2 kHz, but some data may be lost when converting the original
198 signal to a spectral image. Spectrograms with too low spatial resolution have been shown to impair
199 predictive model performance (35). Additionally, our dataset consisting of just over 1,000 images is very
200 small compared to the 14 million image database usually used for CNN training (34). Data augmentation
201 helped us partly circumvent this limitation, but expanding our dataset is necessary for the near future.
202 Moreover, we dealt with the problem of missing data by using the transfer learning technique (10).
203 However, this may have initialized the weights for detecting object characteristics that are not present in
204 the spectrograms as proposed by Ruffini *et al.* (36), which may further explain why our methodology was
205 inefficient in generating the DME diagnosis through images of mock ERG transforms. In addition, given
206 the effectiveness of time series from spontaneous ERG oscillations in predicting risk factors for type 2
207 diabetes (17), the predictive value of time series data rather than images from mock ERG transforms in
208 classifying DME and non-DME cases remains to be studied.

209
210 Our data showed that, when combined with the best mock-ERG model, the performance of the best OCT
211 model to predict DME improved in terms of specificity, precision, Cohen's Kappa, and accuracy. However,
212 its sensitivity has decreased. Similar observations were found with the best fundus model, being more

213 performant when combined with the best mock-ERG model, except for its sensitivity that remained
214 unchanged.

215 It is recognized that mixed methods of prediction, which use multiple learning algorithms, improve the
216 performance of predictions obtained by individual learning processes (37). As observed for analyzing
217 sentiment in the text of low-resource languages (12) and for detecting esophageal cancer (38) by
218 combining CNNs and support vector machine, our hybrid models provided better results than individual
219 models to predict DME cases. It has been empirically verified that this improved performance relates to
220 the combination of distinct model characteristics and the variability in data of different origins (24). We
221 further believe that this improvement is partly due to the complementation of structural information with
222 functional data (26). Structural images, like OCT and fundus, are most commonly used for retinopathy
223 detection, including DME (39). These images are very useful when it comes to localize a lesion (40).
224 However, their very nature gives them the intrinsic spatial resolution limit and the caregiver's ability to
225 interpret the image. In addition, they offer a static view of the retina. In contrast, ERG provides
226 information on retinal activity, at a global or specific level (certain layers of the retina or cell type),
227 depending on the stimulation protocol used. ERG can inform about subtle changes long before any
228 structural alterations can be detected using images. Nevertheless, its application which consists of
229 exposing the subject to a series of specific light flash protocols can be time-consuming, often requires
230 dark adaptation and sometimes mydriasis, and is not recommended for DME detection (41), even though
231 patients with DME showed alterations in some ERG parameters (42). In this context, the predictive power
232 of 5-minute photopic ERGs in the absence of any light flash demonstrated by our data is a huge
233 improvement. More particularly, scalograms obtained from the mock-ERG transforms can be considered
234 functional images of the retina. Having both structural and functional information brings a more
235 comprehensive view of the retinal tissue to the model, thereby improving its capacity to predict DME.

236 Another issue concerns the performance metrics of the models during training and testing. Accuracy and
237 loss metrics were quite good for both the OCT and fundus models, but not for mock ERG-derived
238 scalogram models. The large number of OCT images used for training largely explains the training
239 performance of the OCT models (27,28). At the same time, the good quality and the transfer learning (43–
240 45) surely benefited the eye fundus models. In contrast, mock ERG-derived scalogram models were
241 subjected to overfitting since the accuracy of the testing exceeded that of the training (46). The small
242 amount of data available for training (141 spectral images) is likely responsible.

243 As previously mentioned, an interesting result is that, although the hybrid models generally performed
244 better, the single models that use OCT images as input were the most sensitive. Thanks to the vast OCT
245 image database (>30,000 images), the training was done, using 11,598 OCT images, thereby allowing the
246 model to extract information from a larger number of diseased subjects, facilitating the recognition of
247 true positives (28). In contrast, only 1,011 fundus images and 846 spectral images of mock-ERG after data
248 augmentation were available for these respective model training, likely accounting for the higher
249 sensitivity of the OCT models. We believe that when combining the models' probability matrices, the
250 mock-ERG and fundus models' low sensitivities tended to decrease the OCT models' high sensitivities. An
251 alternate explanation is that unlike fundus images, which only show the inner face of the posterior part
252 of the eyeball, allowing the observation of certain structures, such as the macula or the optic disc, and
253 provides with some qualitative information, OCT images provide quantitative information, like the retinal
254 layer thickness, as well as the presence or the absence of subretinal or intraretinal fluid (47). This
255 translates into a greater amount of information about the retina, increasing the sensitivity of the
256 prediction.

257 In our view, one of the most interesting results of this work is that the performance metrics of the
258 combined fundus plus mock ERG-derived scalogram model in the 70%:30% proportion are comparable to
259 that of the best OCT model. Both present Cohen's Kappa greater than 0.8 and F1-score greater than 0.85,

260 which, as previously mentioned, are the best metrics for evaluating models with unbalanced classes (28).
261 That the sensitivity of the OCT model is superior to the one of the hybrid model (0.93 vs. 0.88) has been
262 discussed above. These results become particularly interesting knowing that the cost of an OCT scanner
263 varies widely from US\$35,000 to \$100,000 (48). In comparison, the cost of a non-mydratic camera range
264 from US\$10,000 to \$20,000 (82) and portatile electroretinographs can be as cheap as approximately
265 US\$4,800 (49). On the eve of being certified, we also know about other ERG prototypes that could even
266 be cheaper. The purchase of the most expensive non-mydratic camera in conjunction with the
267 electroretinography gives an approximate cost of US\$25,000, which remains cheaper than purchasing the
268 most economical OCT device. In general terms, performing a fundus study plus ERG is cheaper than the
269 OCT study. Thanks to our model, this information is now helpful for DME predictive diagnosis. This is
270 relevant because eye care is inaccessible to many people (lack of specialists, insufficient infrastructure,
271 and transportation to clinics) (50). Our work contributes to the view that AI-based diagnostic methods can
272 solve the problem of the lack of specialists, particularly in the eye care area (50). Furthermore, our model
273 deals with already trained models, minimizing the computational cost. It is interesting to note that our
274 model can even be used on mobile devices (51), further reducing the use of resources.
275 In conclusion, using of the fundus-mock ERG hybrid model is viable and relevant for diagnosing of DME in
276 current medical practice.

277 [Materials and Methods](#)

278 [Ethics](#). The ethics committee for human participants of the Mexican Institute of Ophthalmology (IMO),
279 the National Committee of Ethics (reference: CONBIOÉTICA-09-CEI-006-20170306), and the Research
280 Committee at the Asociación Para Evitar la Ceguera (APEC, 17 CI 09 003 142) approved this study. Written
281 informed consent was provided by all subjects. All procedures were done according to the principles of
282 the Helsinki Declaration.

283 [Human data](#). Since we are introducing a completely new parameter, namely unevoked
284 electroretinogram (ERG) signals, for the predictive diagnosis of diabetic macular edema (DME), our study
285 serves as a pilot survey of the population and therefore we cannot determine the sample size (18). A total
286 of 321 adult subjects aged between 30 and 80 years (mean: 48.13 ± 0.71 years, 165 females) with or
287 without diabetes, were enrolled between February 26, 2015 and December 2019 and from September
288 2021 and December 12, 2023 in the IMO of Querétaro (mean age: 50.81 ± 1.57 years, 54 females) and
289 between August 10, 2021 and March 20, 2022 in the Asociación Para Evitar la Ceguera (APEC) in Mexico
290 City (mean age: 45.77 ± 1.20 years, 119 females). 233 (age mean: 44.31 ± 0.72 years, 118 females)
291 completed all tests required for the current study.

292 Subjects underwent an anamnesis and an initial optometric examination to ensure that they were eligible
293 to participate. The exclusion criteria were ages outside 30 to 80 range, lens opacity, myopia greater than
294 6 diopters, glaucoma or other concomitant ophthalmologic disorders, ocular anomalies (e.g., surgery,
295 trauma), recent use of laser or anti-angiogenic intravitreal administration, and cornea problems that
296 disable ERG recordings. All tests were performed as described in (13). Patient diagnosis for DME, diabetic
297 retinopathy (DR), or other eye issues was established by experts at IMO (M.G.R., R.G.F., E.L.S.), APEC (H.Q.
298 and L.F.H.Z.).

299 [Electroretinogram \(ERG\) signals](#). Non-evoked ERGs were registered using customized protocols with
300 either RETIMAX (CSO), Moonpack (Metrovision), or RETeval (LKC Technologies) electroretinographs.

301 Under light conditions (~400 lux), the contour of the eye and the forehead of the subject were cleaned
302 before placing both recording and reference electrodes. ERGs consisted of 5-minute recordings in the
303 absence of any light flash under photopic conditions (400 lux). Recording conditions included a band-pass
304 filter of 0.3 Hz to 1 kHz and an acquisition frequency of 2 kHz. Subsequently, raw data were digitally
305 filtered between 0.3 and 40 Hz, as previously described (17). ERGs were then divided into one-minute
306 segments to maximize the number of samples. From the ERGs of 233 patients (151 from APEC and 82
307 from IMO), 1,353 one-minute ERG fragments were obtained, 183 from patients with DME and 1,170 from
308 patients without DME.

309 As part of this nascent research, we tested both Fourier and wavelet transforms. The Fourier transform
310 captures frequencies that persist over an entire signal, which may not serve for signals with short intervals
311 of characteristic oscillations (19). The wavelet transform is a good alternative because it decomposes a
312 function into a set of wavelets (20). Spectrograms were obtained from each one-minute ERG fragment
313 using Fast Fourier Transform (FFT) with the “Multitaper Spectrogram Code” tool (21). Similarly,
314 scalograms were generated using the Scipy 1.7.1 & Matplotlib 3.5.1 libraries for Python 3.8.8.

315 [Image datasets](#). We used OCT images from three public datasets. Kermany’s dataset contains 11,598
316 images with DME and 26,565 without DME (22); Srivasan’s is formed by macula-centered OCT images of
317 DME (n = 36) and non-DME (n = 53) patients (23), and our dataset ([https://github.com/Traslational-Visual-](https://github.com/Traslational-Visual-Health-Laboratory)
318 [Health-Laboratory](https://github.com/Traslational-Visual-Health-Laboratory)) contains 1,140 OCT images: 176 images with DME and 964 without DME.

319 We used eye fundus images from four open sources, including the MESSIDOR (24), Pachade’s (25),
320 Giancardo’s (26), and our own (<https://github.com/Traslational-Visual-Health-Laboratory>) datasets.
321 Fundus images with no pathology were pooled (n = 151 from MESSIDOR, n = 180 from Pachade’s, n = 115
322 from Giancarlo’s, and n = 495 from ours), as well as fundus images with DME (n = 1,053, only from our
323 dataset).

324 From the 1,353 non-evoked ERG-derived spectrogram and scalogram images, 164 were set apart for
325 model validation: 42 with DME and 122 without DME. The remaining 141 ERG images with DME were
326 subjected to data augmentation by modifying horizontal rotation, horizontal displacement, contrast,
327 brightness variation, and adding salt and pepper noise, as previously described (27). After data
328 augmentation, the non-evoked ERG-derived spectrogram and scalogram images dataset included 846
329 images with DME and 1,048 without DME.

330 Image dataset information is summarized in **Table 1**.

331 [Predictive models](#). Convolutional Neural Networks (CNN) were used, taking advantage of the transfer
332 learning technique (10). For ERG and eye fundus images, the following CNNs were used: ResNet-50,
333 MobileNet-V2, and DenseNet-201. For the last two models, fine-tuning was also used to add four dense
334 layers and two layers with LSTM networks at the end of the models. OpticNet-71 was used with OCT
335 images and a fine-tuned version of DenseNet-201.

336 The bundles of non-evoked ERG-derived spectrogram and scalogram, fundus, and OCT images from the
337 same patient were randomly assigned to training or test sets. A distribution percentage of 80 % training
338 and 20 % testing was used.

339 To verify the possibility that the ResNet-50 and DenseNet-201 models' performance is influenced by the
340 presence of DR, we reclassified the wavelet scalogram images used for training and testing into images
341 with ($n = 293$) or without some degree of DR ($n = 896$). To balance classes, scalograms from cases with DR
342 were subjected to data augmentation, as described in the Methods, leading to the generation of 879
343 scalograms with some degree of DR. Data were randomly divided into training (80 %) and test (20 %), and
344 the ResNet-50 and DenseNet-201 models were trained with the same parameters as described in the
345 Methods, to compare the results between the binary classification of patients with and without DME, with
346 the one of patients with and without DR. Validation was done using the previously used 164 images,
347 though they were recategorized into those with some degree of DR ($n = 77$) and those without DR ($n =$

348 87). **Table 2** summarizes the metric performance of the models trained to distinguish DR and no DR
 349 classes. All metrics exceeded 0.60, unlike the values obtained for the validation of the DME class
 350 prediction, indicating that the concomitant presence of DR with DME takes part into the poor
 351 performance of our models.

352

METRICS	MODELS			
	DME vs No DME		DR vs No DR	
	ResNet-50	DenseNet-201	ResNet-50	DenseNet-201
Precision	0.38	0.34	0.67	0.68
Accuracy	0.68	0.65	0.65	0.67
Sensitivity	0.36	0.40	0.66	0.70
Specificity	0.80	0.73	0.63	0.64
F1-Score	0.37	0.37	0.67	0.69
Cohen's Kappa	0.15	0.13	0.30	0.34

360

361 **Table 2.** Comparative table between the validations done for EMD classification and DR classification. In
 362 general, the metrics obtained in validation with the same models, trained during the same number of
 363 epochs, are better in validation than those of DR classification. All the metrics are above 0.60 and Cohen's
 364 Kappa is also improved.

365

366 **Hybrid model.** The best-performing models for each data type (OpticNet-71 for OCT, DenseNet-201
 367 fine-tuning model trained with scalogram images, and DenseNet201-fine-tuning model for fundus images)
 368 were chosen to create a hybrid model. Predictions were obtained for each model using the 164 validation
 369 images, and the resulting prediction matrices were then multiplied by a weight factor (**Figure 1**). Input

370 data (OCT, eye fundus, and scalogram images) came from the same eye of an individual and were taken
 371 the same day. To find the best weight values, all possible value combinations were tested. The weight
 372 distribution was selected based on the combination that yielded the best results for the hybrid model
 373 (Figure 2).

374 **Algorithm performance.** Models with a loss below 0.4 and accuracy above 0.8 during training and
 375 testing were selected (Table 3). To evaluate the performance of the models, 164 images from OCT, eye
 376 fundus, and ERG-derived scalograms and spectrograms alien to training and test phases were used: 42
 377 images with DME and 122 without DME in total, taken from 97 different subjects. The following metrics
 378 were calculated for each model: precision, sensitivity, specificity, accuracy, F1-score, the area under the
 379 curve (AUC) of the Receiver Operating Characteristic (ROC) curve, and Cohen’s Kappa (28).

			Training		Test	
Model	Input	# Epochs	Accuracy	Loss	Accuracy	Loss
OpticNet-71	OCT	20	0.98	0.05	0.99	0.01
Finely-tuned DenseNet-201	OCT	100	0.97	0.06	0.94	0.18
ResNet-50	Eye fundus	50	1.00	0.01	0.99	0.03
Finely-tuned MobileNet-V2	Eye fundus	100	0.96	0.10	0.99	0.03
Finely-tuned DenseNet-201	Eye fundus	100	0.96	0.10	0.96	0.05
ResNet-50 (FFT)	FFT	50	0.94	0.18	0.93	0.16
Finely-tuned DenseNet-201	FFT	200	0.84	0.37	0.94	0.18
ResNet-50	Wavelet	50	0.98	0.10	0.96	0.16
Finely-tuned DenseNet-201	Wavelet	200	0.88	0.27	0.95	0.15

392 **Table 3.** Models' characteristics during training and testing, including the number of epochs for training
393 and both the accuracy and loss after testing. OCT, optical coherence tomography images; FFT, FFT-derived
394 spectrograms from basal ERGs; Wavelet, Wavelet transform-derived spectrograms from basal ERGs.

395

396 [Code availability](#). To facilitate the reproducibility of our data analyses, the Python code and
397 documentation for the analysis are available online ([https://github.com/Traslational-Visual-Health-](https://github.com/Traslational-Visual-Health-Laboratory)
398 [Laboratory](#)).

399 Acknowledgements

400 We thank all volunteers under the care of the APEC, IMO, ENES León and INDEREB, as well as the many
401 physicians and technicians who worked on data collection.

402 References

- 403 1. Association AD. 11. Microvascular Complications and Foot Care: Standards of Medical
404 Care in Diabetes—2021. *Diabetes Care* [Internet]. 2021 Jan 1 [cited 2021 Aug 24];44(Supplement
405 1):S151–67. Available from: https://care.diabetesjournals.org/content/44/Supplement_1/S151
- 406 2. Singer ME, Dorrance KA, Oxenreiter MM, Yan KR, Close KL. The type 2 diabetes ‘modern
407 preventable pandemic’ and replicable lessons from the COVID-19 crisis. *Prev Med Rep*. 2022
408 Feb;25:101636.
- 409 3. Tan GS, Cheung N, Simó R, Cheung GCM, Wong TY. Diabetic macular oedema. *Lancet*
410 *Diabetes Endocrinol*. 2017 Feb;5(2):143–55.
- 411 4. Abcouwer SF, Gardner TW. Diabetic retinopathy: loss of neuroretinal adaptation to the
412 diabetic metabolic environment. *Ann N Y Acad Sci* [Internet]. 2014 [cited 2021 Aug
413 24];1311(1):174. Available from: [/pmc/articles/PMC4154702/](https://pubmed.ncbi.nlm.nih.gov/254154702/)
- 414 5. Graue-Hernandez EO, Rivera-De-La-Parra D, Hernandez-Jimenez S, Aguilar-Salinas CA,
415 Kershenobich-Stalnikowitz D, Jimenez-Corona A. Prevalence and associated risk factors of diabetic
416 retinopathy and macular oedema in patients recently diagnosed with type 2 diabetes. *BMJ Open*
417 *Ophthalmol*. 2020 Mar 10;5(1):e000304.
- 418 6. Pramanik R, Sarkar S, Sarkar R. An adaptive and altruistic PSO-based deep feature
419 selection method for Pneumonia detection from Chest X-rays. *Appl Soft Comput*. 2022
420 Oct;128:109464.
- 421 7. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence
422 and deep learning in ophthalmology. *British Journal of Ophthalmology*. 2019 Feb;103(2):167–75.
- 423 8. Browning D, Stewart M, Lee C. Diabetic macular edema: Evidence-based management.
424 *Indian J Ophthalmol*. 2018;66(12):1736.

- 425 9. Szeto SKH, Hui VWK, Siu V, Mohamed S, Chan CKM, Cheung CYL, et al. Recent Advances
426 in Clinical Applications of Imaging in Retinal Diseases. *Asia-Pacific Journal of Ophthalmology*. 2022
427 Dec 13;20.
- 428 10. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A Comprehensive Survey on Transfer
429 Learning. 2019 Nov 6;
- 430 11. Dang CN, Moreno-García MN, De la Prieta F. Hybrid Deep Learning Models for Sentiment
431 Analysis. *Complexity*. 2021 Aug 12;2021:1–16.
- 432 12. Md Shad Akhtar, Ayush Kumar. A Hybrid Deep Learning Architecture for Sentiment
433 Analysis.
- 434 13. Noguez Imm R, Muñoz-Benitez J, Medina D, Barcenas E, Molero-Castillo G, Reyes-Ortega
435 P, et al. Preventable risk factors for type 2 diabetes can be detected using noninvasive spontaneous
436 electroretinogram signals. *PLoS One*. 2023 Jan 12;18(1):e0278388.
- 437 14. Luu CD, Szental JA, Lee SY, Lavanya R, Wong TY. Correlation between Retinal Oscillatory
438 Potentials and Retinal Vascular Caliber in Type 2 Diabetes. *Invest Ophthalmol Vis Sci*. 2010 Jan
439 1;51(1):482–6.
- 440 15. Tehrani NM, Riazi-Esfahani H, Jafarzadehpur E, Mirzajani A, Talebi H, Amini A, et al.
441 Multifocal Electroretinogram in Diabetic Macular Edema; Correlation with Visual Acuity and
442 Optical Coherence Tomography. *J Ophthalmic Vis Res [Internet]*. 2015 Apr 1 [cited 2023 Mar
443 17];10(2):165. Available from: [/pmc/articles/PMC4568615/](#)
- 444 16. Kato K, Kondo M, Sugimoto M, Ikesugi K, Matsubara H. Effect of Pupil Size on Flicker ERGs
445 Recorded With RET *eval* System: New Mydriasis-Free Full-Field ERG System. *Investigative
446 Ophthalmology & Visual Science*. 2015 Jun 5;56(6):3684.

- 447 17. Ramsés Noguez Imm, Julio Muñoz-Benitez. Spontaneous electroretinogram signals to
448 screen people with early risk factors for diabetic retinopathy. [cited 2022 Nov 11]; Available from:
449 <https://www.medrxiv.org/content/10.1101/2022.06.26.22276881v1>
- 450 18. O'Neill B. Sample size determination with a pilot study. PLoS One. 2022 Feb
451 15;17(2):e0262804.
- 452 19. J. V. Candy. MULTITAPER SPECTRAL ESTIMATION: An Alternative to the Welch
453 Periodogram Approach [Internet]. Lawrence Livermore National Laboratory. 2019 [cited 2021 Sep
454 23]. Available from: <https://www.osti.gov/servlets/purl/1560107>
- 455 20. Santa María F, Kouro Rodrigo Musalem M E SR. Técnicas Modernas en Automática Tutorial
456 introductorio a la Teoría de Wavelet.
- 457 21. Prerau MJ, Brown RE, Bianchi MT, Ellenbogen JM, Purdon PL. Sleep Neurophysiological
458 Dynamics Through the Lens of Multitaper Spectral Analysis. Physiology. 2017 Jan;32(1):60–92.
- 459 22. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying
460 Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018
461 Feb;172(5):1122-1131.e9.
- 462 23. Srinivasan PP, Kim LA, Mettu PS, Cousins SW, Comer GM, Izatt JA, et al. Fully automated
463 detection of diabetic macular edema and dry age-related macular degeneration from optical
464 coherence tomography images. Biomed Opt Express. 2014 Oct 1;5(10):3568.
- 465 24. Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, et al. FEEDBACK ON A
466 PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE. Image Analysis &
467 Stereology. 2014 Aug 26;33(3):231.
- 468 25. Pachade S, Porwal P, Thulkar D, Kokare M, Deshmukh G, Sahasrabuddhe V, et al. Retinal
469 Fundus Multi-Disease Image Dataset (RFMiD): A Dataset for Multi-Disease Detection Research.

- 470 Data 2021, Vol 6, Page 14 [Internet]. 2021 Feb 3 [cited 2022 May 1];6(2):14. Available from:
471 <https://www.mdpi.com/2306-5729/6/2/14/htm>
- 472 26. Giancardo L, Meriaudeau F, Karnowski TP, Li Y, Garg S, Tobin KW, et al. Exudate-based
473 diabetic macular edema detection in fundus images using publicly available datasets. *Med Image*
474 *Anal.* 2012 Jan;16(1):216–26.
- 475 27. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *J*
476 *Big Data.* 2019 Dec 6;6(1):60.
- 477 28. Sokolova M, Japkowicz N, Szpakowicz S. Beyond Accuracy, F-Score and ROC: A Family of
478 Discriminant Measures for Performance Evaluation. In 2006. p. 1015–21.
- 479 29. J. Andonegui, L. Jiménez. Edema macular diabético [Internet]. 2008 [cited 2021 Sep 9].
480 Available from: [https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-](https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272008000600004)
481 [66272008000600004](https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272008000600004)
- 482 30. Muñoz de Escalona-Rojas JE, Quereda-Castañeda A, García-García O. Actualización de la
483 retinopatía diabética para médicos de atención primaria: hacia una mejora de la medicina
484 telemática. *SEMERGEN - Medicina de Familia.* 2016 Apr;42(3):172–6.
- 485 31. Ciprés M, Satue M, Melchor I, Gil-Arribas L, Vilades E, Garcia-Martin E. Neurodegeneración
486 retiniana en pacientes diabéticos tipo 2 sin retinopatía diabética. *Arch Soc Esp Oftalmol.* 2022
487 Apr;97(4):205–18.
- 488 32. McAnany JJ, Persidina OS, Park JC. Clinical electroretinography in diabetic retinopathy: a
489 review. *Surv Ophthalmol.* 2022 May;67(3):712–22.
- 490 33. Türk Ö, Sirac M. Epilepsy Detection by Using Scalogram Based Convolutional Neural
491 Network from EEG Signals. 2017;
- 492 34. Audun Eltvik. Deep Learning for the Classification of EEG Time-Frequency
493 Representations. [Norwegian]: Norwegian University of Science and Technology; 2018.

- 494 35. Ramos R, Olvera A, Olmos I. Analysis of EEG Signal Processing Techniques based on
495 Spectrograms. Benemérita Universidad Autónoma de Puebla. 2017;
- 496 36. Ruffini G, Ibañez D, Castellano M, Dubreuil-Vall L, Soria-Frisch A, Postuma R, et al. Deep
497 Learning With EEG Spectrograms in Rapid Eye Movement Behavior Disorder. *Front Neurol*. 2019
498 Jul 30;10.
- 499 37. Sancho Fernando. Métodos combinados de aprendizaje [Internet]. 2018 [cited 2022 Nov
500 5]. Available from: <http://www.cs.us.es/~fsancho/?e=106>
- 501 38. Xue DX, Zhang R, Feng H, Wang YL. CNN-SVM for Microvascular Morphological Type
502 Recognition with Data Augmentation. *J Med Biol Eng*. 2016 Dec 10;36(6):755–64.
- 503 39. Cheung C, Tang F, Ting D, Tan G, Wong T. Artificial Intelligence in Diabetic Eye Disease
504 Screening. *Asia-Pacific Journal of Ophthalmology*. 2019;
- 505 40. Las técnicas de neuroimagen [Internet]. [cited 2022 Nov 5]. Available from:
506 <https://www.centroacadia.es/tecnicas-neuroimagen/>
- 507 41. Robson AG, Nilsson J, Li S, Jalali S, Fulton AB, Tormene AP, et al. ISCEV guide to visual
508 electrodiagnostic procedures. *Documenta Ophthalmologica*. 2018 Feb 3;136(1):1–26.
- 509 42. NM T, H RE, E J, A M, H T, A A, et al. Multifocal Electroretinogram in Diabetic Macular
510 Edema; Correlation with Visual Acuity and Optical Coherence Tomography. *J Ophthalmic Vis Res*
511 [Internet]. 2015 Apr 1 [cited 2021 Sep 9];10(2):165–71. Available from:
512 <https://pubmed.ncbi.nlm.nih.gov/26425320/>
- 513 43. VAIBHAV KUMAR. MobileNet vs ResNet50 – Two CNN Transfer Learning Light
514 Frameworks. 2020.
- 515 44. Pin K, Ho Chang J, Nam Y. Comparative Study of Transfer Learning Models for Retinal
516 Disease Diagnosis from Fundus Images. *Computers, Materials & Continua*. 2022;70(3):5821–34.

- 517 45. Patel R, Chaware A. Transfer Learning with Fine-Tuned MobileNetV2 for Diabetic
518 Retinopathy. In: 2020 International Conference for Emerging Technology (INCET). IEEE; 2020. p. 1–
519 4.
- 520 46. Rebecca Roelofs*. A Meta-Analysis of Overfitting in Machine Learning. 33rd Conference
521 on Neural Information Processing Systems. 2019;
- 522 47. T S, SM W, H B, F E, A S, AM P, et al. Fully Automated Detection and Quantification of
523 Macular Fluid in OCT Using Deep Learning. *Ophthalmology* [Internet]. 2018 Apr 1 [cited 2021 Sep
524 23];125(4):549–58. Available from: <https://pubmed.ncbi.nlm.nih.gov/29224926/>
- 525 48. OCT machine prices, comparisons and general must-knows - Medilex [Internet]. [cited
526 2022 Nov 11]. Available from: <https://www.medilexonline.com/oct-machine-prices/>
- 527 49. Al-Otaibi H, Al-Otaibi MD, Khandekar R, Souru C, Al-Abdullah AA, Al-Dhibi H, et al. Validity,
528 Usefulness and Cost of RET *eval* System for Diabetic Retinopathy Screening. *Transl Vis Sci Technol*.
529 2017 Jul 17;6(3):3.
- 530 50. Cheng D, Babij R, Cabrera D, Yuan M, Port A, Mckenney AS, et al. Effective Low-Cost
531 Ophthalmological Screening With a Novel iPhone Fundus Camera at Community Centers. *Cureus*.
532 2022 Aug 17;
- 533 51. Superior EP, Fragua Baeza Á, Ángel M, García G. Ejecución de redes neuronales en móviles
534 Android con aceleración hardware mediante Keras y Tensorflow Lite. 2021 [cited 2022 Nov 11];
535 Available from: <https://repositorio.uam.es/handle/10486/698211>
- 536
537

538 Figure legends

539 **Figure 1.** Diagram showing hybrid model operation. The best validated performing models were selected
540 for each separate type of input (basal electroretinogram (ERG) scalogram, optical coherence tomography
541 (OCT), and eye fundus images): OpticNet-71 for OCT and finely tuned DenseNet-201 for eye fundus and
542 basal ERG scalograms. Predictions with the aforementioned models were done for each type of input
543 (OCT, eye fundus, or scalogram images, $n = 164$ cases in total) and the result was multiplied by a scalar
544 quantity n_i that represents the adjustment percentage for each model. The three resulting values were
545 then summed up to obtain a new prediction matrix, using the predictions made by each model. The
546 maximum value (Max value) is obtained for each matrix row, the output being 1 with diabetic macular
547 edema (DME) and 2 without DME.

548
549 **Figure 2.** Performance of the spontaneous ERG-based models to predict DME. Receiver-Operating
550 Characteristic (ROC) curves and confusion matrixes corresponding to the **(A)** ResNet-50 model fed with
551 fast Fourier Transform (FFT)-derived spectrograms, **(B)** finely-tuned DenseNet-201 model fed with FFT-
552 derived spectrograms, **(C)** ResNet-50 model fed with Wavelet Transform-derived scalograms, and **(D)**
553 finely-tuned DenseNet-201 fed with Wavelet Transform-derived scalograms of spontaneous ERGs. The
554 validation dataset included $n = 42$ cases with DME and $n = 122$ cases without DME. **(E)** Summary of all
555 above-models' performance metrics.

556
557 **Figure 3.** Optimization of hybrid models. Heatmaps showing the performance metrics of double and triple
558 hybrid models, according to the proportion assigned to each used models, i.e. **(A)** OCT and eye fundus
559 images, **(B)** OCT and basal ERG-derived scalogram images, **(C)** eye fundus and ERG-derived images, and
560 **(D)** all three different types of data, as indicated by the percentages in the X-axis. All combinations
561 summed up to 100 %. For the triple hybrid model (D), the remanent percentage of each 10 %-subdivision

562 of the OCT-axis (e.g. 60 % in the case of 40 % assigned to OCT) is divided between eye fundus and ERG-
563 derived images, as indicated in the two lower X-axes. Optimizers looked for the optimal n_1, n_2, n_3 values,
564 as introduced in **Figure 2**.

565
566 **Figure 4.** DME-predicting models' comparison. Heatmap showing the performance metrics obtained for
567 each tested models, including the ones based on one type of input (OCT, eye fundus or ERG-derived
568 images), as well as double and triple hybrid models. External validation dataset: n = 42 for DME and n =
569 122 without DME.

570

571 Supporting information

572 **S1 figure.** Performance of all OCT models. ROC curves and confusion matrixes for the **(A)** OpticNet-71
573 model validated with n = 1,229 cases (n = 212 with DME and n = 1,017 without DME), **(B)** the Finely-tuned
574 DenseNet-201 model validated with n = 1229 (n = 212 with DME and n = 1,017 without DME), **(C)** the
575 OpticNet-71 model validated with n = 164 (n = 42 with DME and n = 122 without DME), and **(D)** the finely-
576 tuned DenseNet-201 model validated with n = 1,229 (n = 212 with DME and n = 1017 without DME). **(D)**
577 Summary of all above-models' performance metrics.

578
579 **S2 figure.** Performance of the Eye Fundus-based model. ROC curves, confusion matrixes, and performance
580 metrics for the **(A)** ResNet-50, **(B)** finely-tuned MobileNet-V2, and **(C)** finely-tuned DenseNet-201 models.
581 The validation dataset included a total of 164 eye fundus images (n = 42 with DME and n = 122 without
582 DME).

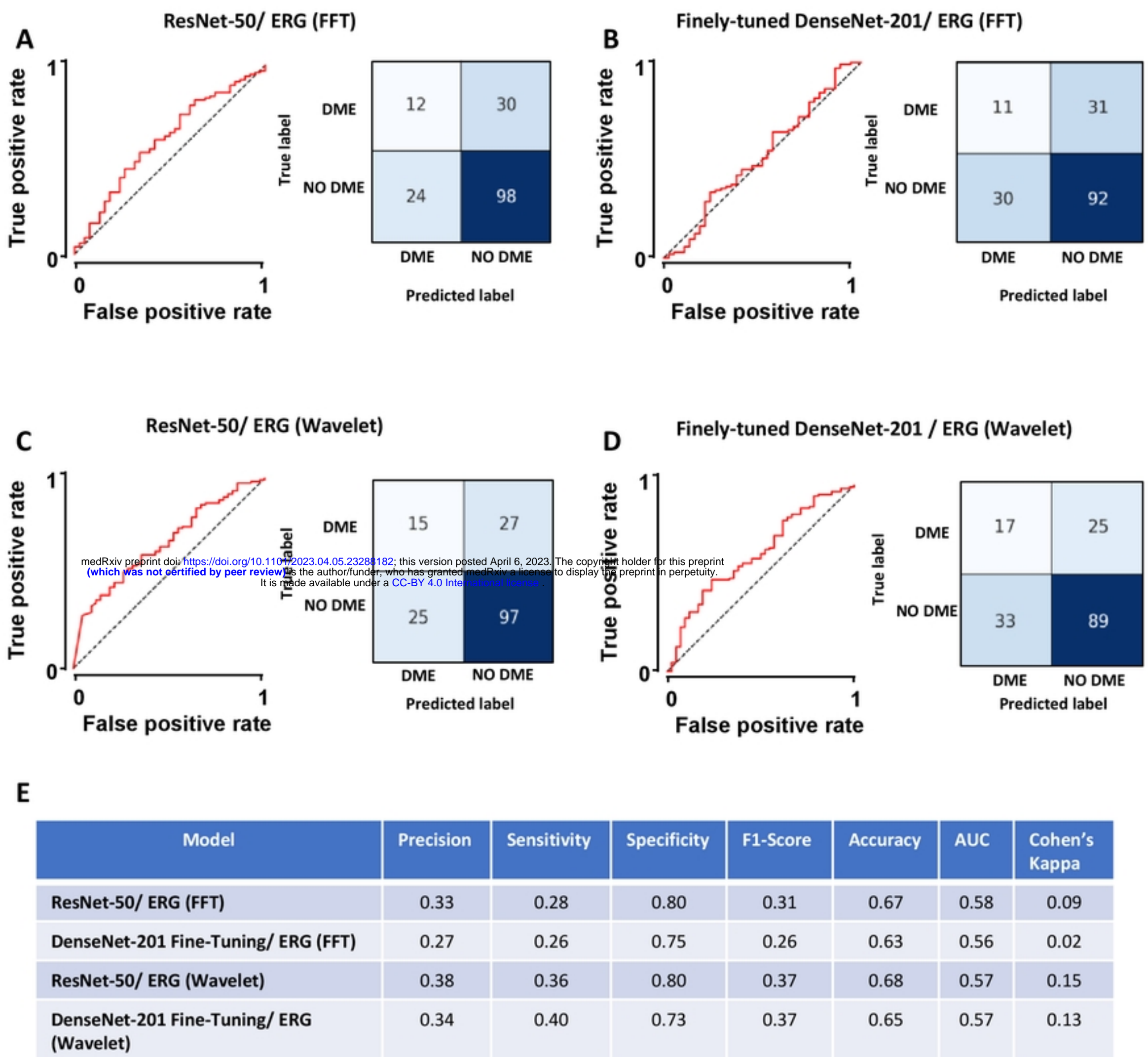
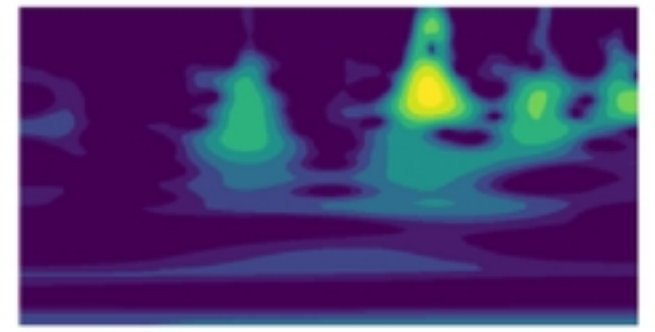
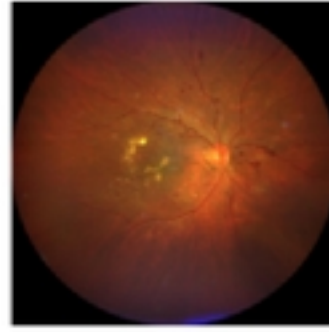
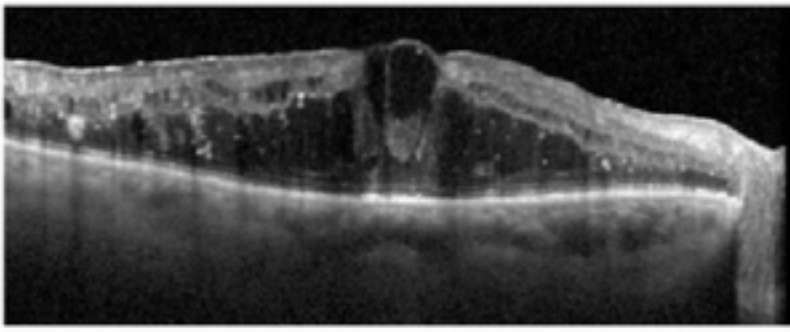


Figure 1



OPTICNET-71

DENSENET 201

DENSENET 201



PREDICTIONS



PREDICTIONS



PREDICTIONS

$$\begin{array}{c}
 \begin{matrix} \text{[CLASS 1 | CLASS 2]} \\ (x_{1,1} \quad x_{1,2}) \\ (x_{2,1} \quad x_{2,2}) \\ (x_{3,1} \quad x_{3,2}) \end{matrix} \times n_1 \quad + \quad \begin{matrix} \text{[CLASS 1 | CLASS 2]} \\ (y_{1,1} \quad y_{1,2}) \\ (y_{2,1} \quad y_{2,2}) \\ (y_{3,1} \quad y_{3,2}) \end{matrix} \times n_2 \quad + \quad \begin{matrix} \text{[CLASS 1 | CLASS 2]} \\ (z_{1,1} \quad z_{1,2}) \\ (z_{2,1} \quad z_{2,2}) \\ (z_{3,1} \quad z_{3,2}) \end{matrix} \times n_3
 \end{array}$$

medRxiv preprint doi: <https://doi.org/10.1101/2023.04.05.23288182>; this version posted April 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

$$\begin{array}{c}
 \begin{matrix} (x_{1,1}, x_{1,2}) * n_1 + (y_{1,1}, y_{1,2}) * n_2 + (z_{1,1}, z_{1,2}) * n_3 \\ (x_{2,1}, x_{2,2}) * n_1 + (y_{2,1}, y_{2,2}) * n_2 + (z_{2,1}, z_{2,2}) * n_3 \\ (x_{3,1}, x_{3,2}) * n_1 + (y_{3,1}, y_{3,2}) * n_2 + (z_{3,1}, z_{3,2}) * n_3 \end{matrix} = \begin{matrix} \text{[CLASS 1 | CLASS 2]} \\ (w_{1,1} \quad w_{1,2}) \\ (w_{2,1} \quad w_{2,2}) \\ (w_{3,1} \quad w_{3,2}) \end{matrix} \xrightarrow{\text{MAX VALUE}} \begin{matrix} \text{[CLASS 1]} \\ \text{[CLASS 2]} \\ \text{[CLASS 2]} \end{matrix}
 \end{array}$$

Figure 2

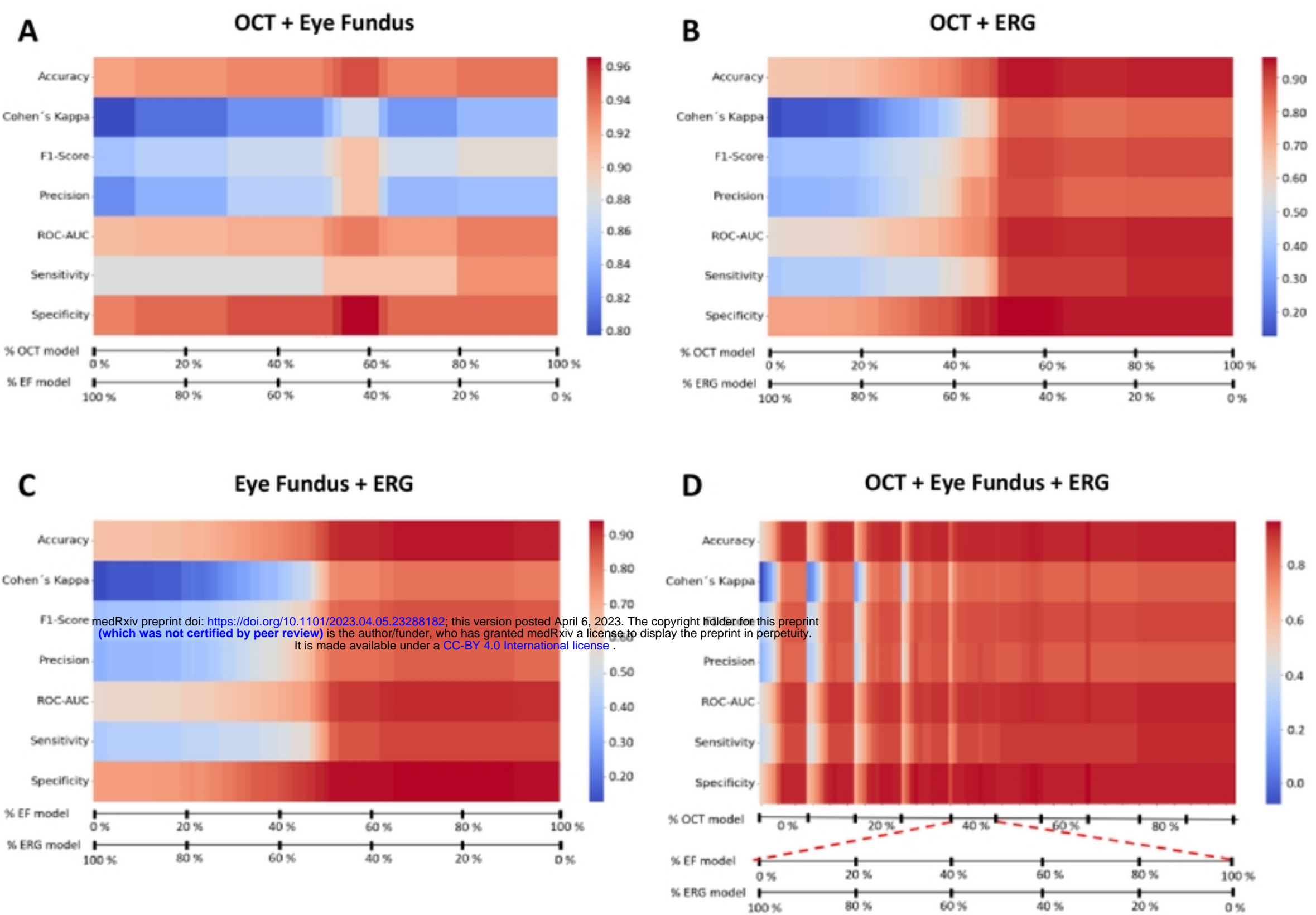


Figure 3

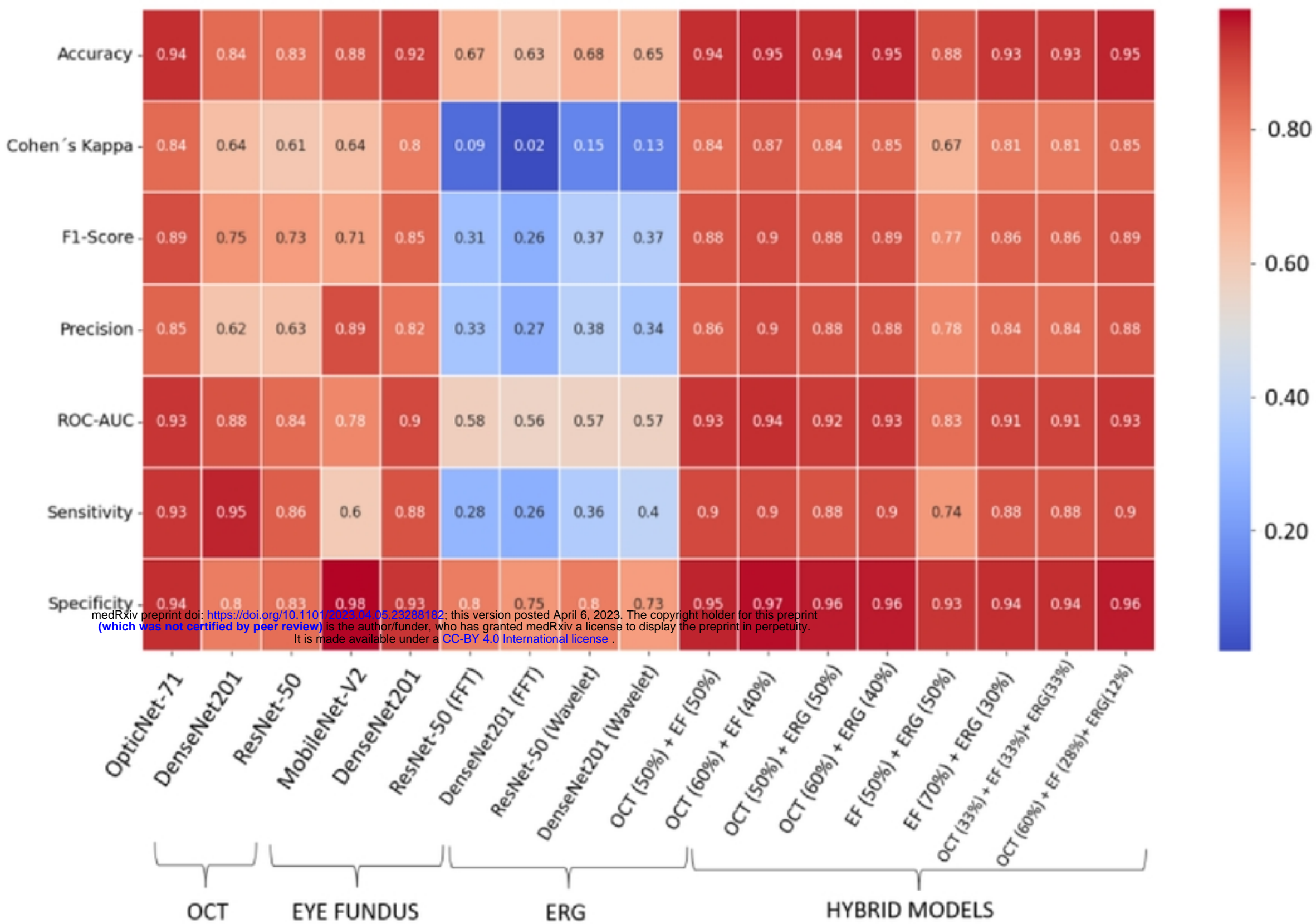


Figure 4