

Etiology of Macular Edema Defined by Deep Learning in Optical Coherence Tomography Scans

Fabio Daniel Padilla-Pantoja¹, Yeison D. Sanchez², Bernardo Alfonso Quijano-Nieto¹, Oscar J. Perdomo³, and Fabio A. Gonzalez²

¹ Department of Ophthalmology, Faculty of Medicine, Universidad Nacional de Colombia, Bogotá, Colombia

² MindLab Research Group, Universidad Nacional de Colombia, Bogotá, Colombia

³ School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia

Correspondence: Fabio Daniel Padilla-Pantoja, Department of Ophthalmology, Universidad Nacional de Colombia, 111321 Bogotá, D.C., Colombia.
e-mail: fdpadillap@unal.edu.co

Received: May 15, 2022

Accepted: August 30, 2022

Published: September 28, 2022

Keywords: artificial intelligence; deep learning; computer vision; macular edema; optical coherence tomography

Citation: Padilla-Pantoja FD, Sanchez YD, Quijano-Nieto BA, Perdomo OJ, Gonzalez FA. Etiology of macular edema defined by deep learning in optical coherence tomography scans. *Transl Vis Sci Technol.* 2022;11(9):29.
<https://doi.org/10.1167/tvst.11.9.29>

Purpose: To develop an automated method based on deep learning (DL) to classify macular edema (ME) from the evaluation of optical coherence tomography (OCT) scans.

Methods: A total of 4230 images were obtained from data repositories of patients attended in an ophthalmology clinic in Colombia and two free open-access databases. They were annotated with four biomarkers (BMs) as intraretinal fluid, subretinal fluid, hyperreflective foci/tissue, and drusen. Then the scans were labeled as control or ocular disease among diabetic macular edema (DME), neovascular age-related macular degeneration (nAMD), and retinal vein occlusion (RVO) by two expert ophthalmologists. Our method was developed by following four consecutive phases: segmentation of BMs, the combination of BMs, feature extraction with convolutional neural networks to achieve binary classification for each disease, and, finally, multiclass classification of diseases and control images.

Results: The accuracy of our model for nAMD was 97%, and for DME, RVO, and control were 94%, 93%, and 93%, respectively. Area under curve values were 0.99, 0.98, 0.96, and 0.97, respectively. The mean Cohen's kappa coefficient for the multiclass classification task was 0.84.

Conclusions: The proposed DL model may identify OCT scans as normal and ME. In addition, it may classify its cause among three major exudative retinal diseases with high accuracy and reliability.

Translational Relevance: Our DL approach can optimize the efficiency and timeliness of appropriate etiological diagnosis of ME, thus improving patient access and clinical decision making. It could be useful in places with a shortage of specialists and for readers that evaluate OCT scans remotely.

Introduction

There is an upward trend of population aging and, consequently, a significant increase in the global burden of chronic diseases.¹ Diabetic macular edema (DME), neovascular age-related macular degeneration (nAMD), and retinal vein occlusion (RVO) are the three main exudative macular diseases and are also the most prevalent and potentially blinding diseases in older patients because of their structural and functional sequelae.^{2,3} They have a serious impact on patients and health care systems, and their

prompt recognition allows better therapeutic decision making.^{3,4} Macular edema (ME) can be associated with fluid accumulation that occurs in the extracellular space in the macula, which is an important area needed for sharp vision, fine detail, and color recognition.³

Optical coherence tomography (OCT) is the gold standard test for the appropriate evaluation, detection, and follow-up of ME, and nowadays, more than 30 million OCT scans are taken per year.⁵ Currently, the diagnosis relies on the manual evaluation of OCT scans, conducted by expert ophthalmologists, which is a subjective, challenging, and time-consuming process, because there is not such a large number of avail-

able specialists for a timely reading of the impressively huge and expansive number of examination results.^{6,7} OCT allows the best evaluation of ME by recognizing the proper location, extension, and significant disease patterns that are so helpful to determine the underlying pathology.⁸ The most representative findings are capable of being identified, processed, and quantified. They are known as biomarkers (BMs), which are certain characteristics that can be objectively measured and evaluated as indicators with diagnostic, predictive, and prognostic values.⁹

According to the literature, nAMD exhibits distinctive BMs such as drusen, hyperreflective foci/tissue (HRF/T), drusenoid pigment epithelial detachment (PED), and subretinal fluid (SRF).¹⁰ DME shows disorganization of retinal inner layers (DRIL), epiretinal membrane, intraretinal fluid (IRF), and HRF.¹¹ RVO associated with ME reveals typical BMs like retinal macrocystoid spaces, HRF, SRF, and external limiting membrane disruption.¹²

The growing incidence of these diseases increases the number of diagnostic tests with a very limited number of readers to analyze them. As a comprehensive strategy to overcome these needs, the application of artificial intelligence (AI) arises to support the timely reading of diagnostic images, which usually exhibit several patterns that can be challenging to recognize even by expert evaluators.^{13,14} A deep learning (DL) approach confers the inherent advantage of the optimized processing of a large amount of data in a very short time.¹³ Abramoff et al.¹⁵ conducted a pivotal study for automated diagnosis of diabetic retinopathy (DR), targeting superiority endpoints at sensitivity higher than 85% and specificity higher than 82.5%, becoming the first Food and Drug Administration–approved AI-based medical diagnostic algorithm. These values serve as a reference to compare AI-based models with the performance of an expert ophthalmologist.

Under the identification of specific BMs in OCT images, an etiological diagnosis of ME can be made, which consists of the proper recognition of the causative disease of ME, not only the recognition of ME as a single finding. However, a related automated approach has not yet been performed. So far, multi-class models have been developed for the recognition of diseases such as DR and AMD without discriminating the presence of ME as a key pathological finding.^{7,16–23} This article proposes a computational method based on DL to identify ME and supportive BMs that allow an automated approach to assess OCT scans and recognize the causative disease of ME among three major exudative retinal diseases.

Methods

Dataset

The OCT scans of this study were collected from two free open-access databases and one private dataset. The free open-access databases were the ZhangLab dataset,²⁴ which contains 207,130 OCT scans taken from patients with choroidal neovascularization, DME, drusen, and control, and the DUKE dataset,²⁵ which contains 269 SD-OCT volumes with 269,000 scans from people between 50 and 85 years of age with large drusen ($>125\ \mu\text{m}$) and AMD without any vitreoretinal surgery. The OCT volumes extracted from the ZhangLab and Duke datasets were acquired with 49 lines in a $6 \times 6\ \text{mm}$ cube. The complete set was exported as an E2E file into the free-open Labelbox digital platform, where annotations were performed. From these two open-access databases, 1343 images with the presence of ME and BMs (disease) and 1343 images with the absence of ME (control) were selected.

The private dataset was provided from the repository of patients who attended an ophthalmology clinic in Colombia between 2015 and 2020. A total of 772 images with the presence of ME and BMs (disease) and 772 images with the absence of ME (control) were obtained. All images of all patients were supported by clinical records that included a full ophthalmologic examination, OCT, and fluorescein angiography (FA) assessment, and the respective confirmation of the proper diagnosis, performed by an experienced retinal specialist. These OCT scans were acquired using a Zeiss Cirrus HD-OCT 5000 device (Zeiss, Oberkochen, Germany) capturing the area of $6 \times 6 \times 2\ \text{mm}^3$ centered on the fovea and were also exported as an E2E file into the Labelbox platform. The two expert readers assessed the scan quality, ensuring it was suitable for determining the presence of pathological patterns. The poor-quality images were excluded. All scans were de-identified before being analyzed by expert readers to protect the safety and privacy of patients. The ethics approval for the research followed the Ethics Committee of the Faculty of Medicine of Universidad Nacional de Colombia (Ref. 018-182; November 12, 2020). Moreover, the study was conducted according to the tenets of Helsinki.

The ZhangLab dataset²⁴ originally included a representative cohort of patients with a distinct ethnicity, including Caucasian, Asian, Hispanic, African American, and mixed population. The DUKE dataset²⁵ included patients from the Age-Related Eye Disease Study 2 Ancillary SD-OCT Study, which was originally enrolled at clinical centers in the United States. Our patient characteristics for each diagnosis class are

Table 1. Images Per Dataset for Segmentation and Classification Of OCT Scans

Subsets	Biomarker	Training Set	Test Set	Validation Set	Total
Segmentation task					
1	Drusen	498	146	70	714
2	HRF/T	486	141	68	695
3	Subretinal fluid	242	70	34	346
4	Intraretinal fluid	250	74	34	358
Binary classification task					
5	nAMD	292	79	39	410
6	RVO	140	41	20	201
7	DME	107	32	15	154
Multiclass classification task					
8	nAMD	296	75	39	410
9	RVO	140	41	20	201
10	DME	107	32	15	154
11	Control	238	68	36	342

shown as supplemental information (Supplementary Table S1).

Finally, a total of 4230 images (half ME images and half control images) were collected from the three databases by two expert ophthalmologists according to the presence of ME with 2115 images and control (absence of ME) with 2115 images. The images were manually annotated with four key BMs and labeled with disease or control by two expert ophthalmologists. The segmented BMs included IRF, SRF, HRF/T, and drusen, taking into account retinal layers' delineation and the mutual agreement about the manual segmentations. IRF was considered as hyporeflective cystoid spaces within the surrounding retinal neuroepithelium, with a minimum size of 25 μm . SRF was taken as the hyporeflective space that separates the retina pigment epithelium from the photoreceptor layer. HRF were small reflective dots (<25 μm), and HRT was defined as larger areas than 25 μm of reflective material. Drusen were identified as small elevations between the pigment epithelium and Bruch membrane. Those elevations larger than 25 μm were considered as pigment epithelium detachment. In those cases of initial dissension or uncertainty, the most experienced retinal specialist, with more than 30 years of experience, decided on the proper segmentation and the most relevant findings.

ME recognition was based on the presence of fluid (IRF or SRF), and further classification of its causative disease was made between DME, nAMD, and RVO. The association of distinctive BMs for each disease was initially helpful to classify remotely the underlying pathology, according to the literature: the combination of drusen, HRF/T, PED, and SRF for nAMD¹⁰; IRF (cystoid spaces), diffuse HRF (more than 30 in number), DRIL (with loss of parallelism of retinal layers without the ability to distinguish them), in the

case of DME¹¹; macrocystoid spaces, SRF and perilesional HRF for RVO¹²; absence of fluid (SRF or IRF) and BMs in control images. Then, the initial classification was cross-checked and verified, by considering the true labels specified in the two open-access databases^{15,24} and the appropriate diagnosis performed by experienced retinal specialists in the case of images acquired from patients attended in the ophthalmology clinic (substantiated with evidence of clinical records, OCT, and FA evaluation). These last values were considered the gold standard or ground truth.

The 4230 images were randomly split into eleven different and independent subsets. Four subsets with BMs were generated for the segmentation task, and the remaining subsets were used for the classification tasks: three subsets for the binary classification task of each disease, and four subsets for multiclass classification between control images and the three different diseases. Finally, the balanced datasets were split into training, test, and validation sets containing 70%, 20%, and 10% of images, respectively. Moreover, the OCT scans from a single volume should belong to a single dataset for classification tasks to ensure heterogeneous data from the subsets. The distribution of images per set is presented in Table 1.

Deep Learning Proposed Method

The method comprises four consecutive phases: the segmentation and identification of BMs, the combination of BMs with mask predictions, the feature extraction performed by a convolutional neural network (CNN) for the binary classification task of each disease, and the multiclass classification of the three ocular diseases and control images.

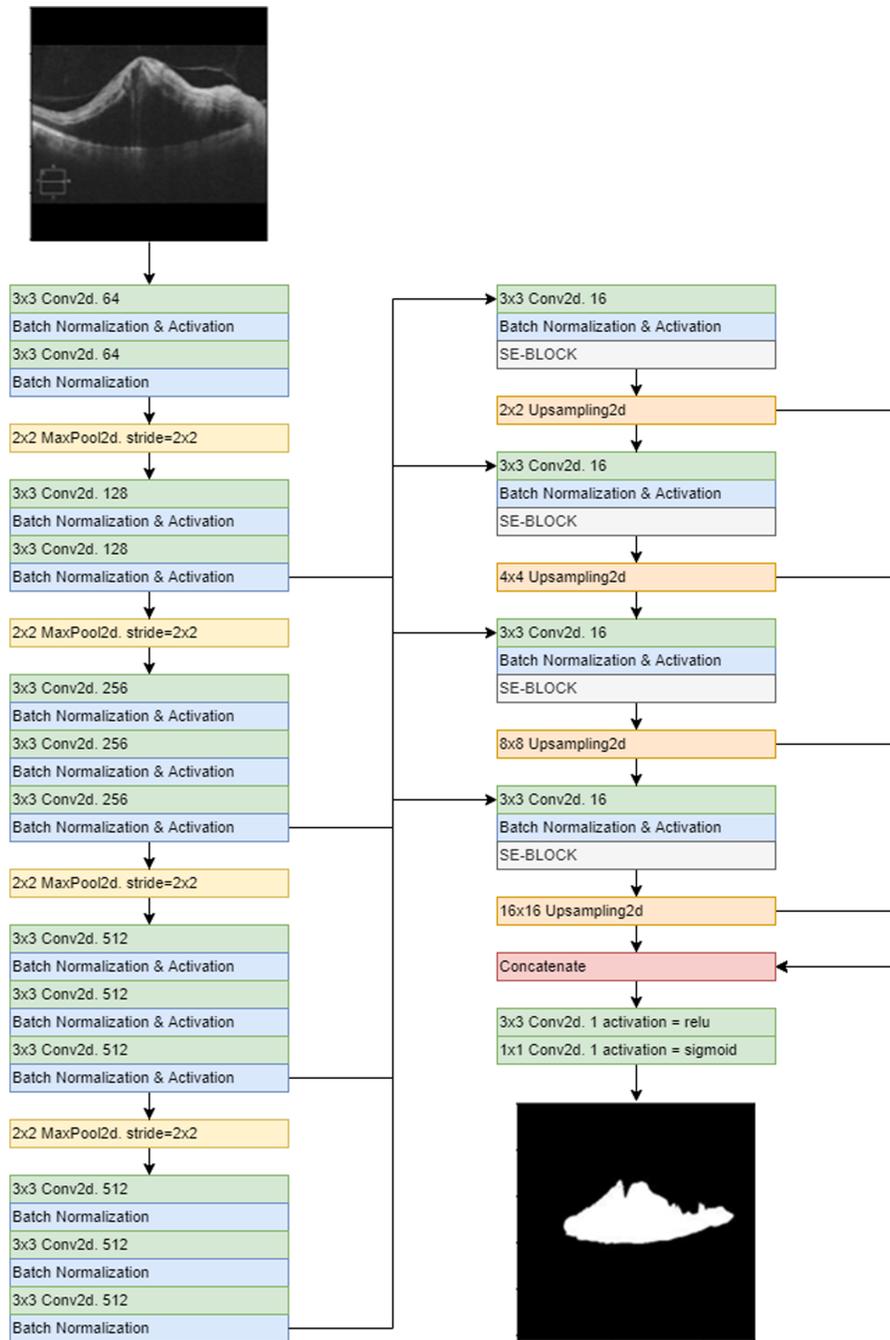


Figure 1. Proposed architecture for segmentation in optical coherence tomography scans. Each image was initially preprocessed and then explored with convolutional neural networks to achieve the proper segmentation of biomarkers, with the application of subsequent blocks and functions like SE blocks.

Segmentation and Identification of Biomarkers

Four CNNs architectures (Res-UNet++, SE-DRIU, SE-UNet, and DRIU) were independently trained to perform automatic recognition of four key BMs (IRF, SRF, Drusen, HRF/T) to diagnose exudative macular diseases. The training process of those CNNs was done with OCT scans according to the distribution reported in Table 1, and the manual segmentation of these BMs was done

manually by two expert ophthalmologists using the Labelbox digital platform as presented in the Methods section.

Figure 1 shows the CNN architecture proposed for the segmentation task. The images were first preprocessed: normalized, and data augmentation transformations such as flip and random crops over images were applied to raw scans for training segmentation models.

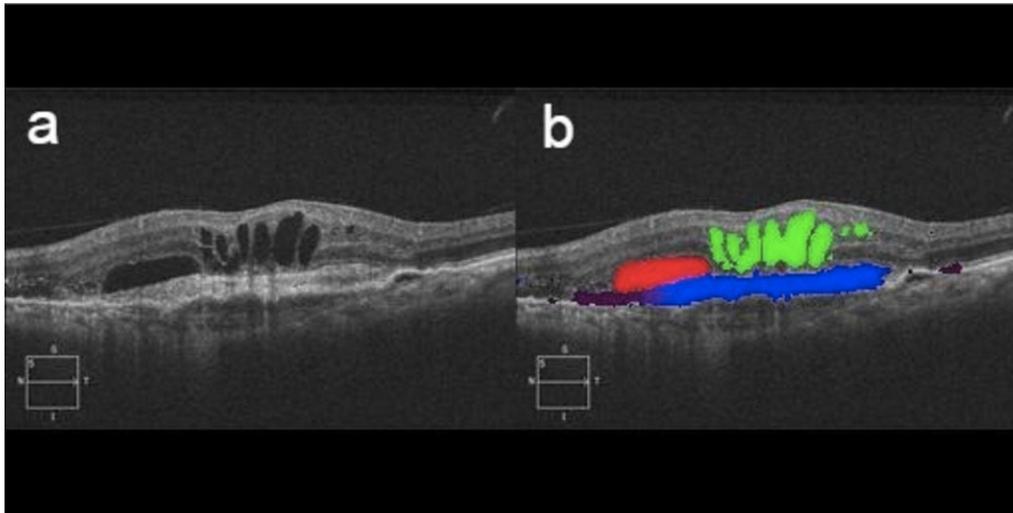


Figure 2. Example of combination of biomarkers in neovascular age-related macular degeneration. (a) Single scan. (b) Scan with subretinal fluid (red), intraretinal fluid (green), hyperreflective tissue (blue), and drusen (magenta) predicted by the model.

Specifically, the known CNN architectures UNET and DRIU (Deep Retinal Image Understanding) based on VGG-19 CNN were explored with some modifications.²⁶ Res-UNet++ is an architecture based on UNET architecture with residual blocks. The residual units allow learning the residual functions between inputs and outputs.²⁷ On Res-UNet++ squeeze-and-excitation (SE) blocks on the encoder path were added after each block.²⁸

Some modifications were explored and analyzed in DRIU architecture to ensure an outstanding performance in segmentation tasks: the first was adding batch normalization layers to improve the training speed and training convergence, and then the batch normalization layer also improved the network generalization ability; another change was adding SE blocks on the last layers of the CNN DRIU to create a new CNN termed as SE-DRIU.

Combination of BMs and Mask Predictions

The models used to predict BMs segmentation were combined and associated to recognize the etiology of ME. Three relevant studies^{10–12} supported the definition of diagnostic rules for the proper identification of diseases, through recognition and combination of certain biomarkers, as follows:

- nAMD: drusen, PED, SRF and HRT in the outer retinal layers (Keane et al.¹⁰) (Fig. 2)
- DME: IRF (Cystoid spaces), multiple HRF (> 30 in number), DRIL (Panozzo et al.¹¹)
- RVO: Macrocystoid spaces, SRF, perilesional HRF (Ozer et al.¹²)

- Control: absence of fluid (SRF and/or IRF) and BMs

Feature Extraction With CNNs to Perform the Binary Classification Task

The overall pipeline for segmentation and classification of OCT scans is shown in Figure 3. The first stage included automatic BMs segmentation (IRF, SRF, HRF/T, and drusen). Once a CNN model is trained for the recognition of each BM, a suitable prediction can be made per each scan, so they can be combined to identify all the BMs present in each scan. Then, the second step included the processing of the scan with combined BMs and certain patterns (as previously explained with the diagnostic rules), through a CNN, to achieve three different classifiers, one for each retinal disease. Every single classifier must be capable to determine the presence versus the absence of disease, as appropriate, for each of the three exudative diseases (nAMD, DME, and RVO).

OCT Scans Classification

The classification of the OCT scans was performed by combining the raw OCT scan and the predicted segmentation of SRF (red color), IRF (green color), and HRF/T (blue color) placing each prediction on an image color channel respectively. The drusen (magenta color) predicted image was generated with a linear combination resulting from stacking red and green channels, as presented in Figure 2. The new image obtained was superimposed on the original OCT scan and it was termed as a scan with predictions as shown in Figure 3.

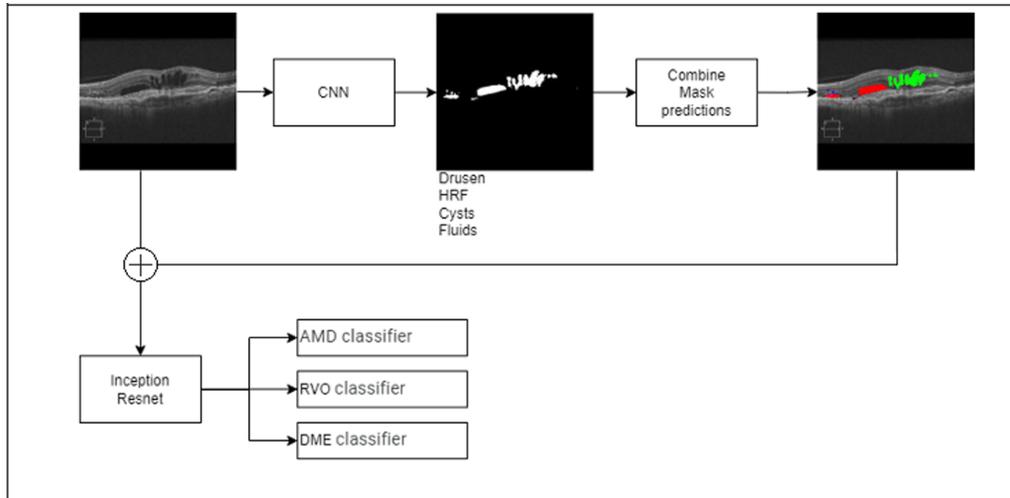


Figure 3. Block diagram of the proposed method to segment (upper) and classify (bottom) OCT scans. The predicted segmentation of each different biomarker was combined and superimposed on every single scan. Through the Inception-ResNet-v2 architecture, it was possible to achieve three different classifiers, one for each retinal disease.

Finally, an Inception-ResNet-v2 architecture was used for training the classification model using the new images obtained from data augmentation operations such as horizontal flip, zooming, and cropping images. Inception-ResNet-v2 architecture uses residual connections for Inception-v4 network.²⁹

Multiclass Classification of Diseases

The classification among the three macular exudative diseases or absence of disease (control) was achieved with a multiclass classification task. The multiclass approach allows joining the three separated and previously reported CNN architectures for each disease, through a multiple logistic regression operation with the Softmax activation function.³⁰ By integrating all of the architectures, a single scan is processed and analyzed with the entire CNNs, in such a way that every single scan is finally labeled with the corresponding class as control (absence or ME), DME, nAMD, or RVO associated with ME.

Performance Metrics

The Dice coefficient (DC) was used to evaluate the performance of the CNN architectures for the BMs segmentation tasks, which corresponds to a matrix that identifies the coincidence between the manual segmentation by the expert ophthalmologists and the system prediction. The range of DC is between 0 to 1, where 1 is a perfect agreement between original segmentation and predicted segmentation. In the classification task is intended that BMs were detected, so DC values >0.5

are required as shown in the next equation³¹:

$$DC = 2TP / (2TP + FP + FN) \quad (1)$$

where TP is the true-positive pixels, FP is the false-positive pixels, and FN is the false-negative pixels, resulting from comparing the original segmentation (ground-truth) and the predicted segmentation.

The performance of CNN in the classification task is done using the following metrics: accuracy, sensitivity, specificity, and the Cohen's Kappa coefficient (κ), which are shown below, where TN stands for true negative, ρ_{agree} is the probability of correct classification, and the ρ_{chance} is the probability of chance agreement.³¹

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (2)$$

$$Sensitivity = TP / TP + FN \quad (3)$$

$$Specificity = TN / TN + FP \quad (4)$$

Cohen's Kappa coefficient

$$= (\rho_{agree} - \rho_{chance}) / (1 - \rho_{chance}) \quad (5)$$

The value for κ varies from 0 to 1, according to the level of agreement, where 0 indicates no agreement, and 0–0.20 means slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement,

0.61–0.80 substantial agreement, 0.81–0.99 almost perfect agreement and 1 states for perfect agreement.³¹

Results

Biomarker Segmentation

Res-UNet++, SE-DRIU, SE-UNet, and DRIU architectures were explored for the segmentation task. The CNNs were trained for 200 epochs. Table 2 shows the results of DC for BMs segmentation.

Figure 4 shows the segmentation process of BMs in OCT scans, which includes the result of manual segmentation performed by expert readers, the automatic segmentations of HRF/T, IRF, and drusen performed by the Res-UNet++, and the SRF segmentation performed by SE-DRIU.

Binary Classification

The classification of OCT was done using an Inception-ResNet-v2. The performance obtained with the Res-UNet++ segmentations was systematically analyzed. Table 3 shows the accuracy, sensitivity, specificity, and κ of the binary classification task performed by the model. Figure 5 presents the ROC curves and confusion matrices for retinal classification tasks provided by our trained model.

Multiclass Classification

The Softmax activation function in the ResNet-v2 architecture produces a probability of belonging to one of the three diseases and control groups for each image.²¹ Then, OCT class is the highest probability value provided by ResNet-v2 architecture and it was used to calculate all performance metrics. Figure 6 shows the confusion matrix for multiclass classification tasks and the ROC curves for individual prediction of each disease. The values of diagnostic accuracy,

sensitivity, specificity, and κ of our proposed model are reported in Table 4. The area under the curve values obtained for DME, nAMD, RVO, and control images were 0.99, 0.98, 0.96, and 0.97, respectively.

The calculation of the Gradient Weighted Class Activation Map (Grad-CAM) was implemented to provide a visual explanation of the more relevant areas to classify into a disease.³² Figure 7 shows model interpretability with feature maps as heat maps, where red zones mean activated parts of our proposed method, to distinguish the main characteristics that the model considers when it differentiates diseases. For example, drusen, SRF, and HRF/T were the most significant BMs for detecting nAMD (Fig. 7a). DME was activated in regions with IRF, HRF, and DRIL (Fig. 7b). RVO associated with ME exhibited macrocystoid spaces, HRF, and SRF (Fig. 7c). The control scans had neither fluid nor biomarker findings. (Fig. 7d).

Discussion

ME is a leading cause of decreased visual acuity and potential irreversible visual loss, and the main causes correspond to three particularly frequent diseases in older patients: AMD, DR, and RVO.^{2,3} They may even coincide in the same patient and usually exhibit similar characteristics that challenge the proper diagnosis of the underlying disease, even by experienced retinal specialists.³ The prognostic value and therapeutic orientation depend on their adequate recognition, including the selection of the most suitable molecule in cases of intravitreal management with antiangiogenic and steroid agents.^{3,4} The clinical evaluation is supported with diagnostic tools such as FA and OCT, which is considered the gold standard for the diagnosis and monitoring of ME.⁵ No study has designed an automated algorithm for the recognition of the cause of ME among the three main retinal exudative diseases from the exclusive evaluation of OCT images to the

Table 2. Dice Coefficient for Segmentation of Biomarkers Over Tested Architectures

CNNs	Subretinal Fluid		HRF/T		Intraretinal Fluid		Drusen	
	Validation	Test	Validation	Test	Validation	Test	Validation	Test
Res-UNet++	0.6650	0.6213	0.5728	0.5686	0.8269	0.8211	0.6346	0.6037
SE-DRIU	0.6968	0.6657	0.5834	0.5432	0.8183	0.8000	0.6806	0.5973
SE-UNet	0.6639	0.6594	0.5344	0.4537	0.8321	0.7200	0.5967	0.5269
DRIU	0.6508	0.6213	0.5369	0.5209	0.8151	0.7877	0.6646	0.5291

The best values in the test set are in bold.
CNNs, convolutional neural networks.

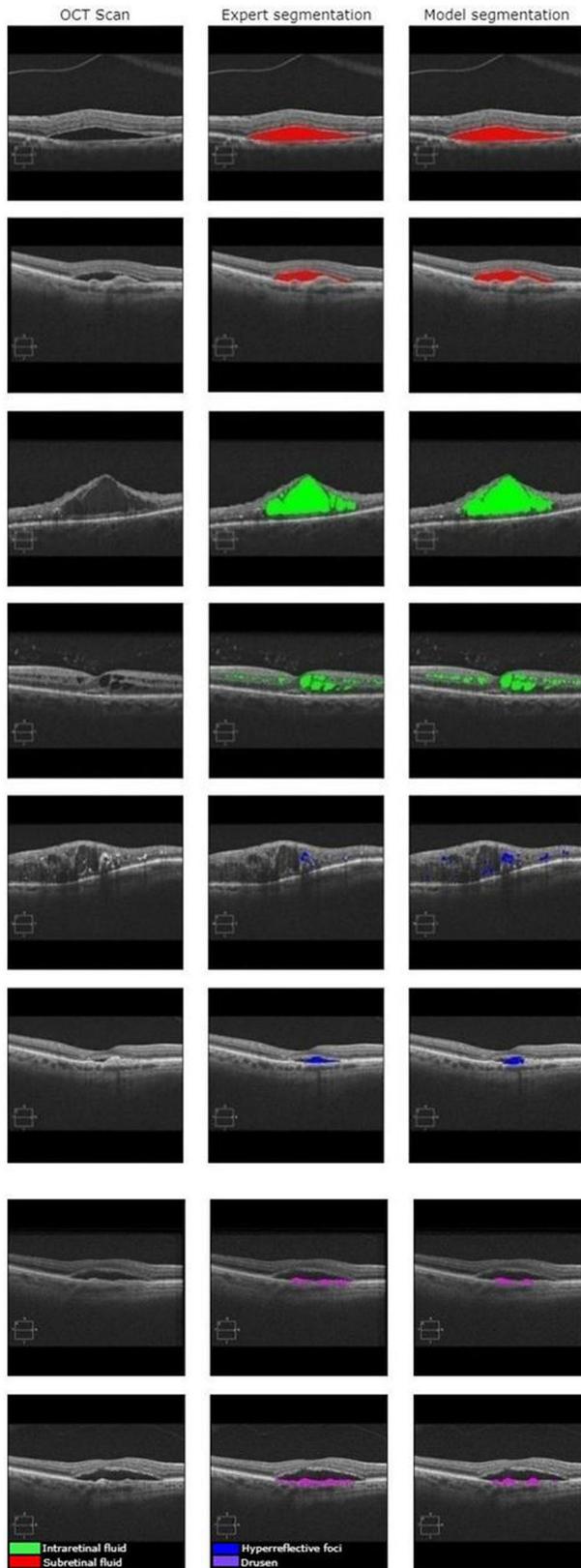


Figure 4. Examples of segmentation of the four key biomarkers: subretinal fluid (red), intraretinal fluid (green), hyperreflective foci and tissue (blue), and drusen (magenta). From left to right: original optical coherence tomography scan, biomarker segmentation performed by expert ophthalmologists, and model prediction.

Table 3. Accuracy, Sensitivity, Specificity, and Cohen's Kappa Coefficient for Binary Classification Task on the Test Set for the Three Major Exudative Retinal Diseases

Disease	Test			Cohen's Kappa Coefficient
	Accuracy	Sensitivity	Specificity	
nAMD	0.96	0.93	1	0.924
RVO	0.99	0.98	1	0.99
DME	1	1	1	1

best of our knowledge. Previous works are limited to the recognition of retinal fluid and its location, as well as the isolated identification of certain BMs,^{16–23} which are not grouped for the classification of the causative disease of ME. Additionally, the specific case of RVO associated with ME has not been explored with an automated approach.

This study proposes a DL method applied to OCT scans for the automatic segmentation of BMs and the classification of macular diseases. Our proposed approach method achieved a state-of-the-art performance, showing an improvement over the original architecture (DRIU) for OCT scans segmentation. Res-UNet++ and SE-DRIU CNNs presented better results for fluid segmentation compared to DRIU and UNet models. However, our proposed SE-DRIU method got better performance because it reduced the number of trainable parameters, which means less time for training and predicting.²⁷

Our method is based on the recognition of key BMs and their appropriate combination for the diagnostic approximation of the causal disease of ME, according to the literature,^{10–12} which is in specific the analytical exercise that an expert reader should normally do, with the challenge of differentiating similar findings between the main underlying conditions and sometimes without the support of clinical information or additional diagnostic tools.

The segmentation of BMs and the performance of the model for their detection revealed interesting findings that are evident in clinical practice when recognizing and interpreting OCT images. Thus, for example, IRF exhibited the best recognition performance, while HRF had lower values of DC. This is explained by the greater facility for the right identification and demarcation of the retinal cystoid spaces, because of their size and the convenient contrast with the surrounding neuroepithelium, both at the time of reading by an expert and when determining their presence by the automated model. On the other hand, smaller and often multiple findings with difficult

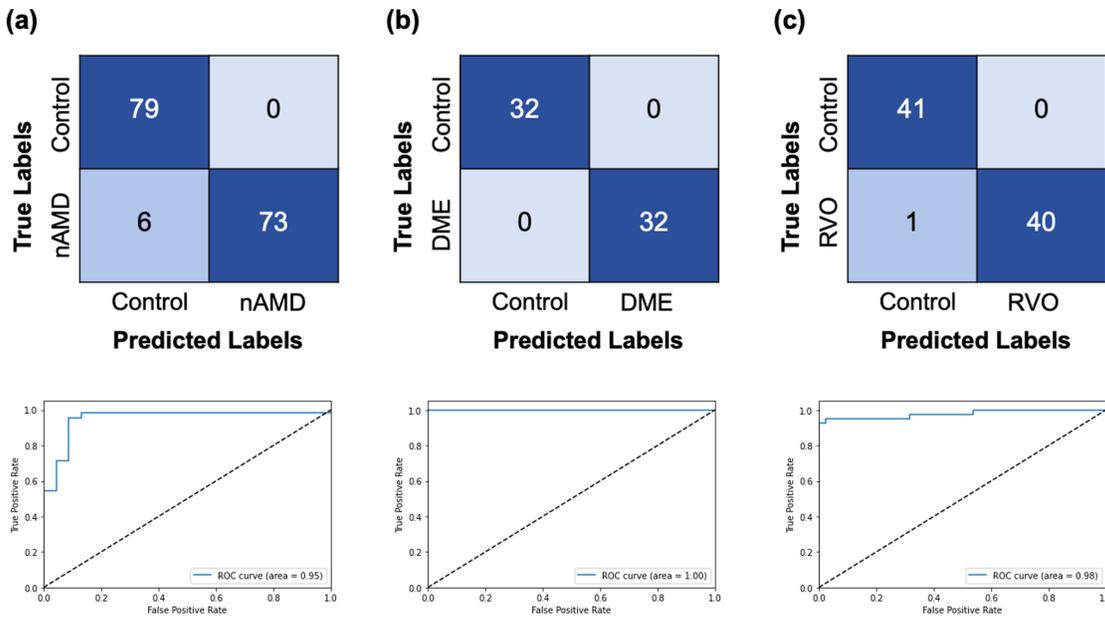


Figure 5. Confusion matrixes and ROC curves for binary classification task of each disease. (a) nAMD. (b) DME. (c) RVO.

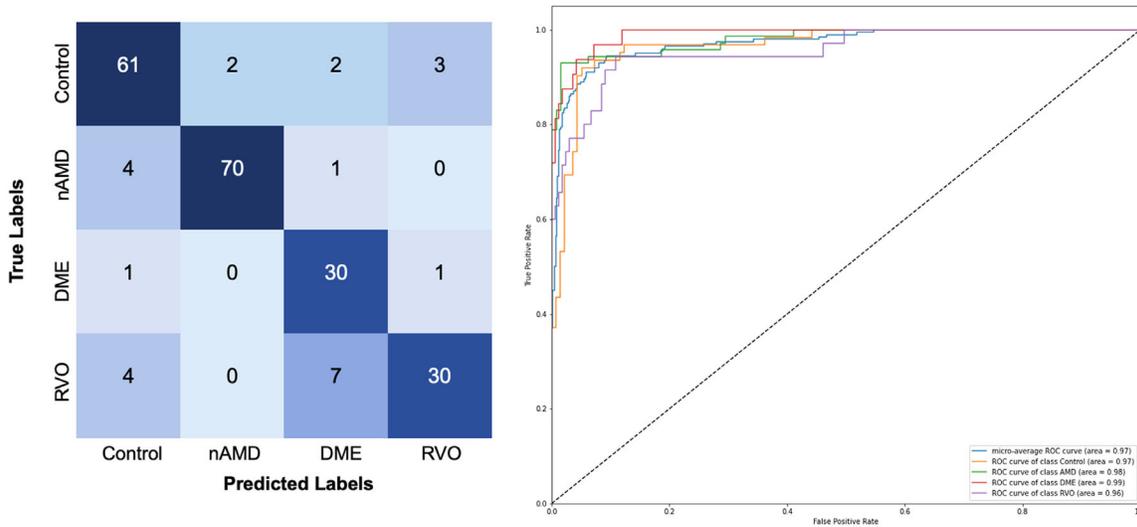


Figure 6. ROC curves, area under the curve (AUC) and global confusion matrix for the multiclass classification task. The model was trained for the appropriate recognition of four classes: nAMD, DME, RVO, and control images.

Table 4. Accuracy, Sensitivity, Specificity, and Cohen’s Kappa Coefficient for the Multiclass Classification of the Three Major Exudative Retinal Diseases and Control Images for the Test Set

Disease	Multiclass Classification			Mean Cohen’s Kappa Coefficient
	Accuracy	Sensitivity	Specificity	
Control	0.93	0.96	0.85	0.84
nAMD	0.97	0.98	0.93	
RVO	0.93	0.97	0.73	
DME	0.94	0.94	0.93	

differentiation of contrast from the surrounding retinal layers, such as drusen and specifically in the case of HRF, make manual segmentation difficult for precise demarcation and it becomes a great challenge to guarantee coverage of the entire of these findings, which can also be confused if they are located close to the retinal hyperreflective layers. These same difficulties that the clinicians undergo, even the most experienced, are also present at the time of being evaluated and performed by an automated method.

The accuracy, sensitivity, and specificity achieved by our model were comparable to the performance of an expert specialist for the classification of the three

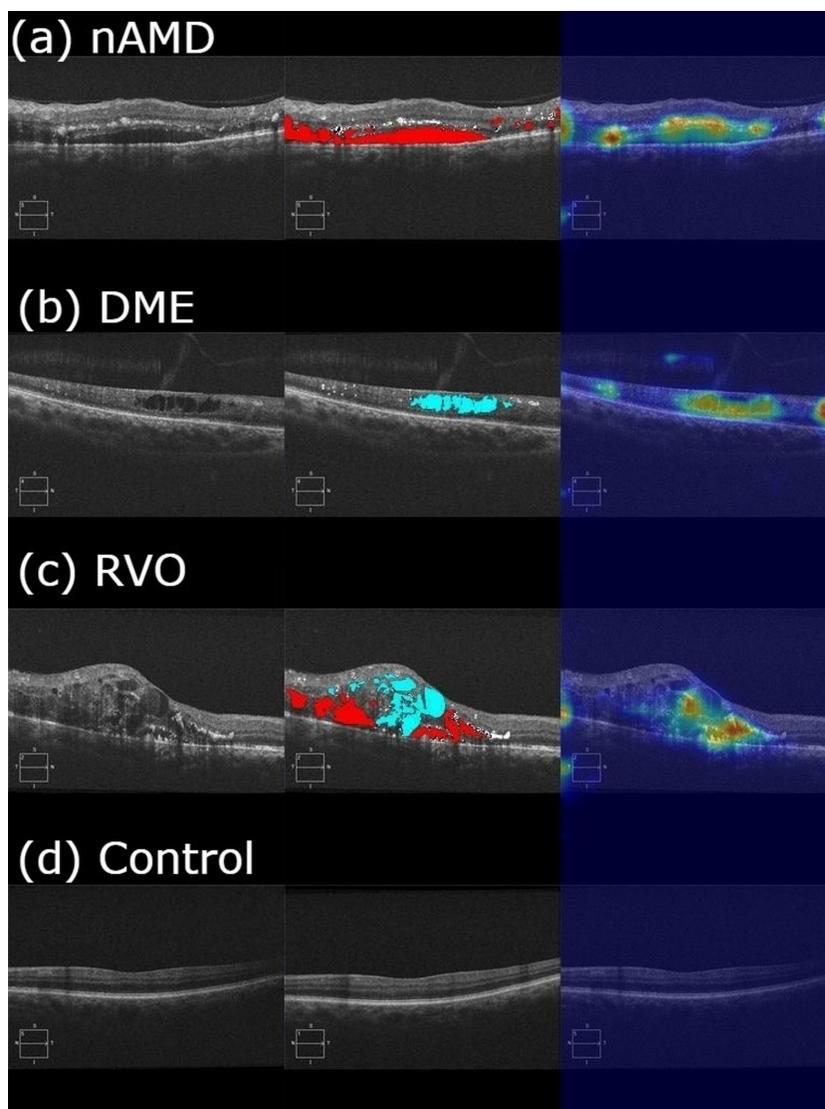


Figure 7. Heatmap visualization of regions considered by the model to perform disease classification. The original optical coherence tomography scan, manual annotation performed by expert ophthalmologists, and model prediction are shown from left to right to provide a visual comparison. (a) Detection of subretinal fluid, hyperreflective foci/tissue and drusen was highlighted for recognition of nAMD. (b) Detection of intraretinal fluid and diffuse hyperreflective tissue conducts to recognition of DME. (c) Macrocystoid spaces, subretinal fluid and hyperreflective foci are highlighted on RVO. (d) No fluid or biomarker was highlighted by the deep learning model in the control scan.

diseases, as verified by the good results obtained with the κ , which also confirms a remarkable interobserver concordance. The best results were obtained for the DME and nAMD classification. Although the specificity for RVO was lower compared with the other diseases, the manual detection and segmentation is not an easy task due to the great similarity in ME patterns between DME and associated with RVO, which can be confused even by highly experienced readers.

Li et al.³³ developed a classification algorithm for the automatic detection of choroidal neovascularization, DME, drusen, and normal images on OCT scans using the ResNet50 neural network. They achieved

an outstanding classification performance with an accuracy of 0.973, a sensitivity of 0.963, and a specificity of 0.98533. Tsuji et al.²² proposed a method to improve classification accuracy by replacing convolution neural networks with a capsule network and achieved an accuracy of 0.996. Taking these models as a baseline to compare our results, a method was proposed with excellent accuracy, sensitivity, and specificity for the classification of ME caused by the three major macular exudative diseases and normal images, in a way comparable to the reading of an expert specialist. The application of our model could be especially useful in the support of the diagnostic

process at different moments of the overall process of patient care. Thus, for example, it can be supportive at the primary care level for optometrists and general practitioners as a screening tool. It can also be very useful for general ophthalmologists in the diagnostic and referral process, as well as for retina specialists in making clinical decisions, collecting information for the evaluation of local epidemiology, and the predictive study of these conditions.³⁴ It can also be incorporated into digital health strategies such as telemedicine, in light of the additional challenges posed by public health contingencies, such as the recent SARS-CoV-2 pandemic, to try to overcome these added barriers to prompt care.³⁵

Moreover, there was an interobserver comparison. The scans were properly labeled and classified by two experienced ophthalmologists, who also had access to medical records and other diagnosis tools such as FA, to be consistent with the right classification in the case of the private dataset, and to the true labels specified in the two open-access databases (ZhangLab²⁴ and Duke²⁵ datasets). It is well known that in some locations, very few ophthalmologists read the scans remotely, and the demand outstrips supply. Then, our method could be especially useful to optimize the efficiency and timeliness of appropriate diagnosis, as well as clinical decision making, thus improving patient access and care, particularly in places with few readers who must issue their medical opinions without other supportive tools.

The limitations of this study include the limited number of OCT scans, expert readers, and the restraint to one ophthalmology center and two free open-access databases. Moreover, the use of retrospective data restricts the opportunity to include clinical information and imaging follow-up, which may enhance the performance of the model. Although the focus of our method was on the three major exudative macular diseases, other retinal conditions could be associated with the presence of retinal fluid, as is already the case of vitreomacular traction syndrome (VMT). The VMT disorders are often clearer and more consistent for their remote identification with the exclusive evaluation of OCT images, considering the evident finding of epiretinal membranes and the tractional pull on the macula, with the consequent alteration of retinal architecture. Because of the relative ease in recognizing cases of VMT without the need for supportive automated tools, the classification of these conditions was not included in this study. However, it is recognized that it is an important differential diagnosis that should be explored together with the three main exudative macular diseases in future studies.

In an attempt to include ME images from patients of different ethnicity, our study included: OCT scans from the Latin American population attended at the ophthalmology clinic, random images acquired from the ZhangLab dataset (which originally included a representative cohort of Caucasian, Asian, Hispanic, African American, and mixed population)²⁴ and random images obtained from the DUKE dataset that included the United States Population.²⁵ However, the inclusion of multiple ophthalmological centers and the proper demographic characterization of their patients in different locations will allow assessing the generalizability in future studies. Then this study offers a basic architecture that can be enriched by multiple ophthalmological centers, feeding the system with a greater number of images and expert readers, to improve the diagnostic behavior.

The Grad-CAM exploration opened the black box of the model by conferring interpretability and a visual explanation of the performance of the CNNs that highlights key BMs and their combination. This issue should be validated in future studies by performing proper training of CNNs on raw OCT scans. In this regard, the model provides qualitative information that is of great importance for clinical practice. Because of that, it could be a teaching tool for the reading and interpretation of OCT scans by comparing the performance of students with different levels of training and the model lecture, which would be theoretically comparable to the diagnosis made by a retina professor. In future studies, the integration of complementary models for the diagnosis, treatment, and prognosis of diseases could provide a valuable strategy for the comprehensive clinical analysis of patients, promoting the opportunity for timely attention and clinical decision making in a fast, efficient, and reliable way.

In this article, we also release our entire image dataset for the segmentation of biomarkers and the classification of ME diseases, including the codes for their corresponding models. This allows future researchers to compare their performance with our method, as well as to enrich the architecture of the artificial neural network (<https://github.com/yeisonlegarda/EtiologicClassificationOfMacularEdemaUsing-A-DeepLearningApproachOpticalCoherenceTomographyScans>).

Conclusion

Our method not only recognizes macular edema, as has been explored in previous studies through fluid identification, but it also may classify its cause among

three different maculopathies, which are the major exudative retinal diseases. The proper recognition of key BMs and their specific combination, including location and quantity of the findings, allowed pattern determination of ME that in turn achieved the right identification of the underlying disease. The Grad-CAM exploration opened the black box of the developed mathematical model, conferring interpretability to our automated method and a corresponding visual explanation of the performance of the CNNs, revealing the importance of the automated identification of BMs and their association. Furthermore, this model may help to deal with the high demand for several tests, lifting the burden for ophthalmologists, particularly in places where the availability of experts is scarce. Likewise, this approach may become especially useful considering that it is not uncommon that many readers must issue their medical opinions exclusively from the evaluation of OCT scans, with no access to medical records and other diagnostics and guidance tools, for which an automated etiological approach could provide greater accuracy in diagnosis, allowing them to make the timeliest and appropriate medical decisions.

Acknowledgments

The authors thank the management staff of the Ophthalmology Clinic “Oftalmocenter Ltda.” for providing the private data used in this study.

Disclosure: **F.D. Padilla-Pantoja**, None; **Y.D. Sanchez**, None; **B.A. Quijano-Nieto**, None; **O.J. Perdomo**, None; **F.A. Gonzalez**, None

References

1. Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Heal*. 2017;5(12):e1221–e1234.
2. Tranos PG, Wickremasinghe SS, Stangos NT, et al. Macular edema. *Surv Ophthalmol*. 2004;49:470–490.
3. Daruich A, Matet A, Moulin A, et al. Mechanisms of macular edema: Beyond the surface. *Prog Retin Eye Res*. 2018;63:20–68.
4. Das UN. Diabetic macular edema, retinopathy and age-related macular degeneration as inflammatory conditions. *Arch Med Sci*. 2016;12:1142–1157.
5. Zhong P, Wang J, Guo Y, et al. Multiclass retinal disease classification and lesion segmentation in OCT B-scan images using cascaded convolutional networks. *Appl Opt*. 2020;59:10312–10320.
6. Kapoor R, Walters SP, Al-Aswad LA. The current state of artificial intelligence in ophthalmology. *Surv Ophthalmol*. 2019;64:233–240.
7. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol*. 2019;20(5):e262–e273.
8. Munk M, Sacu S, Huf W, et al. Differential diagnosis of macular edema of different pathophysiologic origins by spectral domain optical coherence tomography. *Retina*. 2014;34:2218–2232.
9. Califf RM. Biomarker definitions and their applications. *Exp Biol Med*. 2018;243:213–221.
10. Keane PA, Patel PJ, Liakopoulos S, et al. Evaluation of age-related macular degeneration with optical coherence tomography. *Surv Ophthalmol*. 2012;57:389–414.
11. Panozzo G, Cicinelli MV, Augustin AJ, et al. An optical coherence tomography-based grading of diabetic maculopathy proposed by an international expert panel: The European School for Advanced Studies in Ophthalmology classification. *Eur J Ophthalmol*. 2020;30:8–18.
12. Ozer MD, Batur M, Mesen S, et al. Evaluation of the Initial optical coherence tomography parameters in anticipating the final visual outcome of central retinal vein occlusion. *J Curr Ophthalmol*. 2020;32(1):46–52.
13. Yang LWY, Ng WY, Foo LL, et al. Deep learning-based natural language processing in ophthalmology: applications, challenges and future directions. *Curr Opin Ophthalmol*. 2021;32(5):397–405.
14. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172:1122–1131.
15. Abramoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1:39.
16. Chen Z, Li D, Shen H, et al. Automated segmentation of fluid regions in optical coherence tomography B-scan images of age-related macular degeneration. *Opt Laser Technol*. 2020;122:105830.
17. Fang L, Wang C, Li S, et al. Attention to Lesion: Lesion-Aware Convolutional Neural Network for Retinal Optical Coherence Tomography Image Classification. *IEEE Trans Med Imaging*. 2019;38:1959–1970.

18. Kuwayama S, Ayatsuka Y, Yanagisono D, et al. Automated detection of macular diseases by optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images. *J Ophthalmol.* 2019;2019:6319581.
19. Bhatia KK, Graham MS, Terry L, et al. Disease classification of macular optical coherence tomography scans using deep learning software: validation on independent, multicenter data. *Retina.* 2020;40:1549–1557.
20. Hassan B, Qin S, Ahmed R, et al. Deep learning based joint segmentation and characterization of multi-class retinal fluid lesions on OCT scans for clinical use in anti-VEGF therapy. *Comput Biol Med.* 2021;136:104727.
21. Liu X, Bai Y, Cao J, et al. Joint disease classification and lesion segmentation via one-stage attention-based convolutional neural network in OCT images. *Biomed Signal Process Control.* 2022;71:103087.
22. Tsuji T, Hirose Y, Fujimori K, et al. Classification of optical coherence tomography images using a capsule network. *BMC Ophthalmol.* 2020;20:114.
23. Karri SP, Chakraborty D, Chatterjee J. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. *Biomed Opt Express.* 2017;8:579–592.
24. Kermany D, Zhang K, Goldbaum M. Large dataset of labeled optical coherence tomography (OCT) and chest X-ray images. Mendeley Data, v3; 2018, <http://dx.doi.org/10.17632/rscbjbr9sj.3>.
25. Farsiu S, Chiu SJ, O'Connell RV, et al. Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography. *Ophthalmology.* 2014;121:162–172.
26. Li MX, Yu SQ, Zhang W, et al. Segmentation of retinal fluid based on deep learning: application of three-dimensional fully convolutional neural networks in optical coherence tomography images. *Int J Ophthalmol.* 2019;12:1012.
27. Jha D, Smedsrud PH, Riegler MA, et al. ResUNet++: an advanced architecture for medical image segmentation. *2019 IEEE Int Symp Multimed.* 2019:225–2255.
28. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.* 2018:7132–7141.
29. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the impact of residual connections on learning. *31st AAAI Conf Artif Intell AAAI 2017.* 2016:4278–4284.
30. Zhou L, Li Q, Huo G, et al. Image classification using biomimetic pattern recognition with convolutional neural networks features. *Comput Intell Neurosci.* 2017;2017:3792805.
31. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol.* 2004;11:178–189.
32. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis.* 2016;128:336–359.
33. Li F, Chen H, Liu Z, et al. Deep learning-based automated detection of retinal diseases using optical coherence tomography images. *Biomed Opt Express.* 2019;10:6204.
34. O'Byrne C, Abbas A, Korot E, et al. Automated deep learning in ophthalmology: AI that can build AI. *Curr Opin Ophthalmol.* 2021;32:406–412.
35. Sundararajan M, Schallhorn JM, Doan T, et al. Changes to ophthalmic clinical care during the coronavirus disease 2019 pandemic. *Curr Opin Ophthalmol.* 2021;32:561–566.