Ann Med Res

Current issue list available at Ann Med Res

Annals of Medical Research

journal page: annalsmedres.org

Explainable deep learning for the automated classification of macular diseases in OCT images

Murat Firat [a], Ilknur Tuncer Firat [b,*], Taner Tuncer [c]

[a] Malatya Turgut Özal University, Faculty of Medicine, Department of Ophthalmology, Malatya, Türkiye
[b] İnönü University, Faculty of Medicine, Department of Ophthalmology, Malatya, Türkiye
[c] Fırat University, Faculty of Engineering, Department of Computer Engineering, Elazığ, Türkiye
*Corresponding author: ilknurtuncer89@gmail.com (Ilknur Tuncer Firat)

MAIN POINTS

- A ResNet50-based deep learning model was developed to automatically classify 7 macular diseases using only OCT images, achieving a high overall accuracy of 95%.
- The model demonstrated strong interclass discrimination, especially distinguishing with high precision between morphologically similar conditions such as DME, RVO, and RAO.
- Grad-CAM heat maps were used to visualize the decision-making areas of the model, enhancing the transparency and clinical interpretability of the AI system.
- The system is highly effective in correctly identifying normal OCT scans (99.5%), which is crucial for community screening and reducing unnecessary clinical workload.
- The model showed promising performance in detecting surgically managed diseases such as vitreomacular interface disorders (VIDs), supporting its integration into triage and referral systems.

Cite this article as: Firat M, Tuncer Firat I, Tuncer T. Explainable deep learning for the automated classification of macular diseases in OCT images. *Ann Med Res*. 2026;33(5):205–214. doi: 10.5455/annalsmedres.2025.07.206.

ABSTRACT

Aim: Optical coherence tomography (OCT) is a widely used, noninvasive, rapid, and high-resolution imaging technique for diagnosing and monitoring macular diseases. Despite its clinical value, OCT image interpretation is time-consuming and requires expert knowledge, which may lead to inconsistencies in diagnosis. The objective of this study is to create an AI-based model that reliably and effectively categorizes macular diseases from OCT images, offering a workable solution in environments with restricted access to ophthalmology specialists.

Materials and Methods: A convolutional neural network model based on ResNet50 architecture was developed to classify OCT images into seven categories: age-related macular degeneration (AMD), diabetic macular edema (DME), epiretinal membrane (ERM), retinal artery occlusion (RAO), retinal vein occlusion (RVO), vitreomacular interface disease (VID), and normal (NO) controls. Grad-CAM was employed to enhance the interpretability of the model and support clinical usability.

Results: The model's macro-averaged precision, recall, and F1-score were 0.943 (95% confidence interval [CI]: 0.941–0.960), 0.940 (95% CI: 0.941–0.960), and 0.940 (95% CI: 0.941–0.960), respectively, with an overall accuracy of 0.950 (95% CI: 0.941–0.960). Grad-CAM visualizations confirmed the model's focus on relevant retinal regions, thus supporting diagnostic reliability and interpretability.

Conclusion: The explainable model demonstrated strong diagnostic performance and potential as a clinical decision-support tool, especially in environments with limited resources. The integration of explainable AI techniques, such as Grad-CAM, enhances trust in automated decision-making and offers significant potential in supporting non-expert users and early detection strategies.

Keywords: Retinal diseases, Artificial intelligence, Optical coherence tomography, Maculopathy

Received: Aug 05, 2025 **Accepted:** Dec 16, 2025 **Available Online:** May 22, 2026

INTRODUCTION

The diagnosis and follow-up of retinal diseases in ophthalmology practice requires a high level of expertise and time. During this process, specialists frequently utilize various imaging techniques, including visual acuity measurement, biomicroscopy, fundus photography, fundus fluorescein angiography (FFA), fundus autofluorescence (FAF), optical coherence tomography (OCT), and OCT angiography (OCTA) [1,2]. However, the clinical interpretation of these methods is time-consuming and creates a significant workload for specialists working under high patient volumes. Reducing the burden on specialists through preliminary assessment systems is crucial for the early and accurate diagnosis of RDs. To achieve this, we need reliable, easily trainable, and accessible diagnostic support systems that can be used by general physicians. Such systems, which offer high accuracy, do not re-

quire advanced expertise, and are low-cost, have become possible through AI-based approaches [3,4,5]. This has led to a growing interest in automated classification systems that aim to reduce specialist dependency and increase diagnostic reliability and speed. AI-based decision support systems stand out for their potential to standardize the process and alleviate the diagnostic burden, especially in regions with limited specialist access [6]. Today, optical coherence tomography (OCT) images have become one of the gold standard methods for assessing retinal diseases and play a critical role in macular disease diagnosis [7]. AI-based OCT data analysis improves clinical performance and offers large-scale applicability in community screening and remote monitoring scenarios. Recently, diagnostic support systems based on ML algorithms have been proposed for identifying retinal diseases [8, 9]. These studies employed classical algorithms, such as support vector machine (SVM), k-nearest neighbor (K-NN), decision tree (DT), and ensemble model (EM), by manually extracting features from fundus or OCT images. However, because manual feature extraction is time-consuming, labor-intensive, and has limited generalization capacity, deep learning-based decision support systems that perform automatic feature extraction and classification have become more popular.

Numerous studies have focused on the detection of diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration. Choudhary et al. used a VGG19-based classifier to distinguish choroidal neovascularization, drusen, and diabetic macular edema (DME) from normal OCT images. Their study achieved 99.17% accuracy with 84,568 OCT images [10]. Kabataş et al. used 1397 fundus images to distinguish DR and macular hole (MH) from normals. They achieved the highest accuracy of 93.97% with the ResNet50, InceptionV3, and Xception models [11]. Sunil et al developed a computer-aided diagnosis (CAD) system using fundus images for the diagnosis of DR stages. The proposed CNN model achieved an accuracy of 98.33% [12]. In a study involving 29,800 OCT images, Eghtedar et al. achieved 97.6% accuracy using a deep neural network (DNN) [13]. Rajan et al used 3 different models—OctDeepNet, OctDeepNet1, and OctDeepNet2—to classify normal, CNV, DME, and drusen. They achieved 98% accuracy by adjusting the model architecture for factors such as the number of layers, kernel size, and pooling [14]. Wang et al classified 8 retinal diseases using the ViT model and compared the results with those of existing methods, such as ResNet, VGG, DenseNet, and MobileNet. The ODIR-2019 dataset consists of fundus images and has a class imbalance. Therefore, the F1 score and AUC values were examined, and 0.932 ± 0.06 and 0.950 ± 0.03 were obtained, respectively [15]. Laouarem et al. combined the strengths of CNNs and visual transformers to classify 7 retinal diseases. They achieved accuracies of 99.40%, 97.00%, and 99.77% on the OCT-2017, OCT-C8, and OCT-2014 datasets, respectively [16]. To distinguish CNV, drusen, DME, and normal samples from OCT images, Akça et al. used the Vision Transformer (ViT), Token-to-Token Vision Transformer (T2T-ViT), and Mobile Vision Transformer (Mobile-ViT) algorithms. Accuracy rates of 95.14%, 96.07%, and 99.17% was achieved with the ViT, T2T-ViT, and Mobile-ViT models, respectively [17].

Maculopathies and OCT findings

Age-related macular degeneration (AMD) is one of the most common degenerative pathologies affecting the macula and is the primary cause of central vision loss, particularly in individuals aged >50 years. AMD develops because of the disruption of the interaction between the RPE, Bruch's membrane, and the choroid. In the dry form, drusen accumulate beneath the RPE, gradually weakening macular function. Abnormal vessels originating from the choroid (choroidal neovascularization) migrate into the subretinal space in the wet form, causing bleeding and fluid leakage [18]. This condition can be observed in OCT images as subretinal and intraretinal fluid accumulation, PED, RPE irregularities, and increased retinal thickness [19].

The macula is highly susceptible to edema, degeneration, and ischemic processes due to its high metabolic activity and structural characteristics. Macular edema develops due to increased capillary permeability, particularly in diabetic retinopathy and retinal vein occlusions. Edema affects the macula more severely than the peripheral retina due to its higher photoreceptor density and different vascular structure. DME is the most important complication of diabetic retinopathy, leading to vision loss. On OCT, it is characterized by increased macular thickness, cystic spaces, hyperreflective areas indicating exudate in the inner retinal layers, and sometimes subretinal fluid [20]. RVO, which is associated with hypertension and systemic vascular disease, is the second most common retinal vascular disease. Macular edema is the primary cause of vision loss. OCT reveals macular edema, increased retinal thickness, and intraretinal cystic spaces [21]. RAO is one of the most severe forms of retinal ischemia and is often associated with embolic events. OCT reveals marked hyperreflectivity and thickening of the inner retinal layers in the early stage and atrophy in the late stage [22].

Disruptions in the interactions between the vitreous cortex and the retinal surface cause vitreoretinal interface diseases (VIDs). Abnormally tight adhesions, incomplete vitreous separation, and the resulting fibroglial proliferation characterize these pathologies [23]. Examples of these conditions include VMT, VMA, macular hole, and epiretinal membrane (ERM). The most common complaints in these conditions include decreased visual quality, metamorphopsia, and blurred vision. OCT findings may reveal persistent adhesion of the vitreous cortex to the macula (vitreomacular adhesion) or traction (vitreomacular traction), as well as full-thickness (full-thickness macular hole) or partial (lamellar macular hole) defects in the macula [24]. ERM, a condition of this group of diseases, is characterized by the formation of a fibroglial mem-

https://doi.org/10.5455/annalsmedres.2025.07.206

brane on the inner surface of the retina. OCT imaging reveals a hyperreflective membrane, flattening of the inner retinal layers, and macular distortion [25].

Contributions of this study to the literature:

- Reduces inter-expert interpretation differences,
- It reduces the burden of manual evaluation,
- It enables the differentiation of multiple diseases,
- Provides explainable and interpretable disease information.

In this study, we developed an explainable ResNet50-based OCT classifier that simultaneously distinguished 7 macular conditions and provided Grad-CAM visualizations to support clinical interpretability.

MATERIALS AND METHODS

OCT is a valuable tool for creating detailed retinal microstructure images and diagnosing ocular conditions. This study used an open-access dataset consisting of 2064 OCT images labeled by experts according to retinal pathology, with predefined disease classes that eliminate the need for additional subjective clinical interpretation [26]. This dataset was published to provide researchers with access to a large set of labeled images that contribute to the development and improvement of algorithms for automated processing and analysis of OCT images for the early diagnosis and monitoring of eye diseases. The dataset consists of seven classes and includes OCT images of age-related macular degeneration, diabetic macular edema, epiretinal membrane, retinal artery occlusion, retinal vein occlusion, vitreomacular interface disease, and normal macula. The dataset exhibits an imbalanced class distribution. To address this issue, each image in the classes with fewer samples (DME, ERM, RAO, RVO, VID, and normal macula) was rotated 3° to the left and right, and symmetry operations were applied. As a result, in each minority class, the final sample size was simply determined by multiplying the original sample size by the fixed augmentation scheme, which reduced but did not completely eliminate the original class imbalance. This yielded an augmented dataset of 3 824 images (original n = 2 064). Table 1 shows the class distributions of the resulting dataset. After random augmentation, the dataset was divided into 2 parts: 80% training and 20% testing.

Table 1. Class distribution of the dataset.

Class	Original	Augmented
Age-related macular degeneration	1231	1231
Diabetic macular edema	147	441
Epiretinal Membrane	155	493
Normal	332	996
Retinal artery occlusion	22	132
Retinal venous occlusion	101	303
Vitreomacular Interface Disease (VMID)	76	228

Explainable ResNet50 model

The ResNet50 model was chosen for the classification of retinal diseases in this study because of its ability to recognize complex patterns and reduce overfitting, particularly in small and detailed medical datasets. Heat maps generated by the Grad-CAM method were used to better understand the obtained results, facilitate the interpretation of OCT images in the clinical setting, and increase the explainability of the model's decisions. The details of the model are provided below.

ResNet50 is a residual architecture with a depth of 50 layers. Its primary layers are convolution, max pooling, sequential residual blocks, global average pooling, and a fully connected layer. Taking an $X \in R^{H \times W \times 3}$ image as input, the ResNet50 model uses a residual network structure to address the degradation problem (vanishing gradient). In this residual network block, the input information x is passed through the convolution → batch normalization → ReLU chain and combined with x again via the shortcut connection (Eq. 1).

$$y = F(x, \{W_i\}) + x \tag{1}$$

In the model, ($R^{2048} \rightarrow R^{c}$) was used instead of the original fully connected layer ($R^{2048} \rightarrow R^{1000}$) to represent the number of classes c. The difference between the model's output probabilities and the true labels was calculated using the cross-entropy loss function. The model output $z = f(x, \theta) \in R^c$ was used to calculate the softmax classification probabilities and the loss function (eq.2 and 3).

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{c} e^{z_j}} \tag{2}$$

$$L = -\sum_{i=1}^{c} y_i \log(\hat{y}_i) \tag{3}$$

Here, $y_i \in \{0, 1\}$ is one-hot coded.

Gradient-weighted Class Activation Mapping (Grad-CAM) was used to ensure the model's explainability of the classification results. In ResNet50, the activation maps in the final convolutional layer, which have the highest-level features for class distinction, were used for Grad-CAM. Activations indicate the pixel's importance in the classification.

For class c, weights are calculated as in Equation 4 using $y_c = f_c(x)$, the activation map $A^k \in R^{u \times v}$ obtained from the last convolution layer, and the gradients $\frac{(\partial y^c)}{(\partial A^k)}$ obtained by backpropagation.

$$\alpha_k^c = \frac{1}{Z} \sum_{i,j} \frac{\partial y^c}{\partial A_{i,j}^k} \tag{4}$$

https://doi.org/10.5455/annalsmedres.2025.07.206

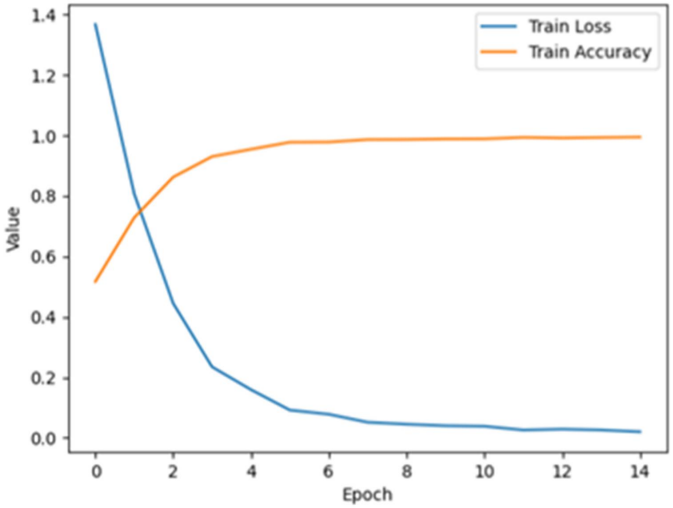

Figure 1. Training and validation of accuracy/loss curves across epochs for the ResNet50 model. These curves demonstrate the learning behavior and generalization performance of the model throughout the training process.

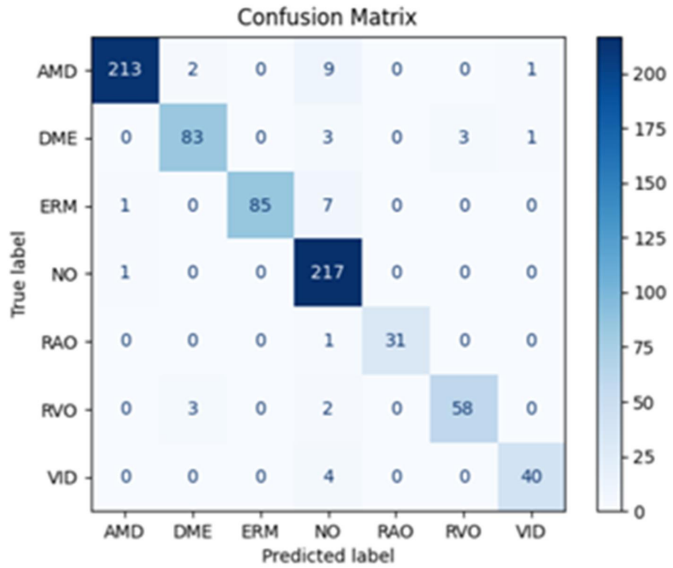

Figure 2. Confusion matrix summarizing the performance of the model across all 7 disease classes. Diagonal values indicate correct classifications, whereas off-diagonal values indicate misclassifications, enabling a visual analysis of class-wise accuracy.

Using the weights, the Grad-CAM heat map to be superimposed on the image is calculated as in eq.5.

$$L^{c}_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_{k} \alpha^{c}_{k} A^{k}\right) \tag{5}$$

RESULTS

A computer environment with an Intel(R) Core(TM) i7-9750H CPU @ 2.60 GHz (2.59 GHz), 8 GB RAM, and Python programming language was used to train and test the ResNet50 model. The random seed was fixed to 42 in all experiments; the same seed was used for Python/NumPy/PyTorch, and data splitting and shuffling were performed with

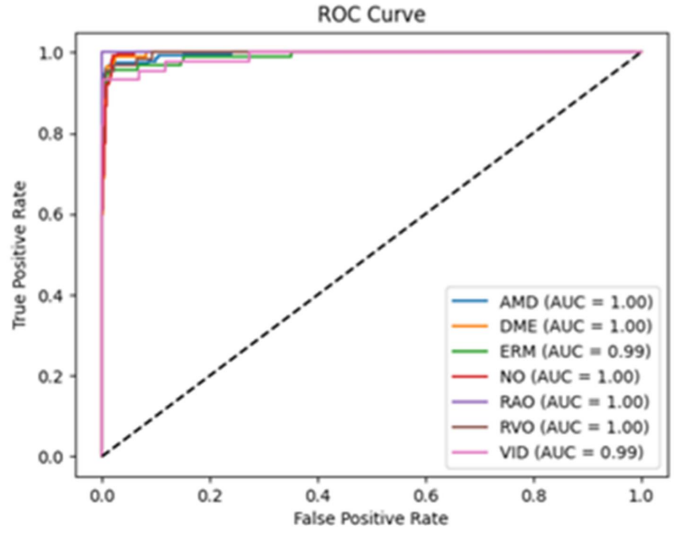

Figure 3. ROC curves for each class, illustrating the trade-off between sensitivity and specificity. AUC values provide a comparative metric for class-level performance.

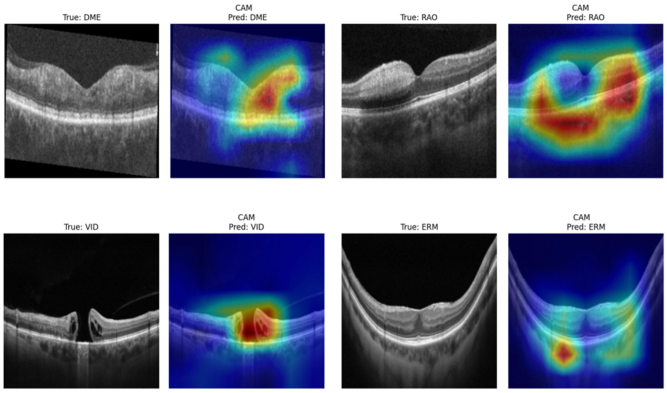

Figure 4. Examples of OCT images correctly classified with Grad-CAM–based attention heat maps. These visualizations demonstrate that the model accurately focuses on disease-relevant anatomical regions when making accurate predictions.

torch.Generator().manual_seed(42), CuDNN was set to deterministic and benchmark=False. Model performance parameters were analyzed using the accuracy, F1 score, precision, and sensitivity parameters obtained from the confusion matrix (eq.6-9). Additionally, ROC-AUC was used to determine model performance due to class imbalance in the dataset.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F_1 = \frac{2TP}{2TP + FN + FP} \tag{9}$$

https://doi.org/10.5455/annalsmedres.2025.07.206

Table 2. Training parameters.

Parameter	Value	Definition
Pretrained	True	ImageNet loaded pre-trained weights
Num Classes	7	Outputs equal the number of classes in the dataset
Loss Function	CrossEntropyLoss	Suitable for multiclass classification
Optimizer	Adam	The Adam optimization algorithm
Learning Rate	1e-4	Initial learning rate (%)
Epochs	15	Total number of training cycles
Batch Size	32	Training and validation
Train/Test Split	80%/20%	Splitting the ratio into training and test sets
Input Size	(224, 224)	All images were resized to this size.
Grad-CAM Layer	model.layer4[-1]	Final convolution layer used for the Grad-CAM

Grad-CAM: Gradient-weighted Class Activation Mapping.

Table 3. Precision, recall, and F1 score by class.

Class	Precision	Recall	F1 Score
AMD	0.995	0.947	0.970
DME	0.932	0.922	0.927
ERM	0.867	0.914	0.890
NO	0.943	0.995	0.969
RAO	0.911	0.969	0.940
RVO	1.000	0.921	0.959
VID	0.952	0.909	0.930

AMD: Age-related Macular Degeneration, DME: Diabetic Macular Edema, ERM: Epiretinal Membrane, NO: Normal, RAO: Retinal Artery Occlusion, RVO: Retinal Vein Occlusion, VID: Vitreomacular Interface Disease.

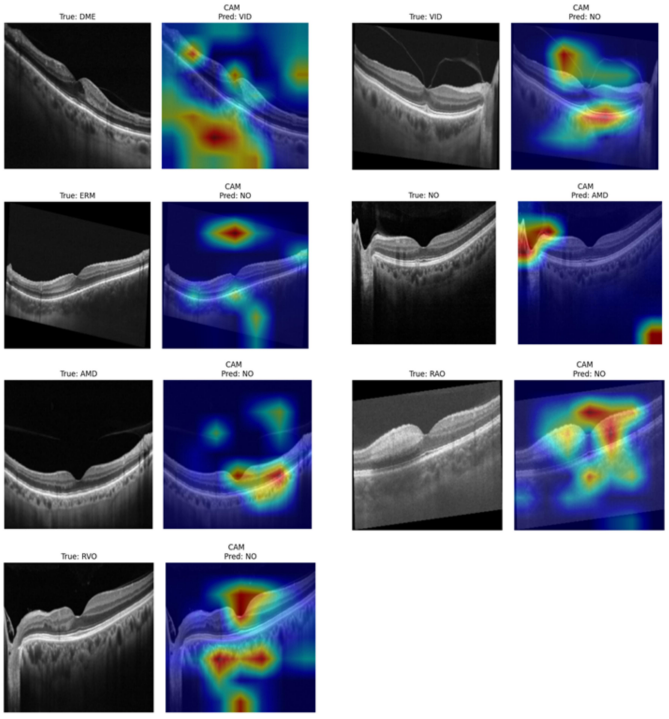

Figure 5. Examples of misclassified OCT images with Grad-CAM–based attention heat maps These cases illustrate the difficulty of the model in identifying subtle or ambiguous features and highlight the importance of explainability in evaluating model limitations.

Table 2 shows the parameters used for training the model. The dataset was first classified using the Vision Transformer, a widely used model in the literature, to demonstrate the superiority of the proposed model. Classification accuracy and f1-score values were 92.15% and 92%, respectively. Figure 1 shows the accuracy and loss function changes for training and validation as a result of the classification performed with ResNet50, Figure 2 shows the confusion matrix, and Figure 3 shows the ROC change.

The accuracy and loss graphs show that the loss value and accuracy are stable after the eighth epoch. This change indicates that the training process was successful. Furthermore, the receiver operating characteristic (ROC) curve was used to determine the discriminatory power of the model for each class. Since the change in all classes is close to the upper left corner, the model performed well and distinguished between classes. The AUC 95% confidence interval values were as follows: AMD, 0.996–1.000; DME, 0.994–1.000; ERM, 0.976–1.000; NO, 0.996–1.000; RAO, 0.991–1.000; RVO, 0.993–1.000; and VID, 0.969–1.000. The precision, recall, and F1 score values obtained from class after the elimination of the relative class imbalance are shown in the Table 3. The F1 score of 0.94, which is particularly relevant for class imbalance, indicates that the model successfully learned the dataset.

Heat maps (Grad-CAM) are created to visualize the regions to which the deep learning model pays attention during classification. This makes the decision-making process of the model more transparent and contributes significantly to clinical validation and explainability. Warm colors (red, orange, and yellow) represent diagnostically important regions during classification, whereas cool colors (blue tones) represent areas that the model ignores or deems unimportant for classification. Figures 4 and 5 present the correctly classified and misclassified images by the model, respectively, along with their corresponding heat maps of attention.

As shown in Figure 4, the Grad-CAM attention maps in correctly classified cases focused on the pathological regions. In the correctly classified DME OCT image, which is characterized by small cystic spaces and increased thickness due to fluid accumulation in the inner layers of the retina, the attention map is concentrated in the central and thickened inner layers of the retina. In the OCT image of a correctly classified macular hole, which is a type of VID characterized by a full-

https://doi.org/10.5455/annalsmedres.2025.07.206

Table 4. Comparison of AI-based studies using OCT images to classify macular disease.

Reference	Model	Disease	Dataset	Performance
Wang et al, 2016 [27]	SMO	Dry AMD (n =15), DME (n =15), and NO (n = 15)	3 000+ OCT images	Accuracy: 0.993
Li et al, 2019 [28]	ResNet50	CNV, DME, Drusen, and NO (n = 2 796)	15 573 OCT images	Accuracy: 0.973 AUC: 0.995 Sensitivity: 0.963 Specificity: 0.985
Li et al, 2019 [29]	VGG-16	CNV, DME, Drusen, and NO	109 312 OCT images	Accuracy: 0.986 Sensitivity: 0.978 Specificity: 0.994 AUC: 1.000
Kuwayama et al, 2019 [30]	CNN (Caffe)	Wet AMD (n = 136), DR (n = 104), ERM (n = 90), NO (n = 570), and other diseases (n = 19)	1 200 OCT images	Precision/recall: Wet AMD 1.000/0.770; DR 0.780/1.000; ERM 0.750/0.750; NO 0.850/0.970
Bhatia et al, 2020 [31]	VGG-16	Dry AMD, wet AMD, DME, and NO	5 588 OCT volumes (162 721 B-scans)	AUC: ≥0.980 AUC of dry AMD: 0.850–0.990 Wet AMD area under the curve: 0.860–0.980 DME AUC: 0.930–0.990
Esfahani et al, 2023 [32]	CNN	Drusen, CNV, DME, and NO	3 133 OCT images	Accuracy: 0.987 Precision: 0.99 Recall: 0.962 F1: 0.971
Choudhary et al, 2023 [10]	VGG-19	CNV, DME, Drusen, and NO	84 568 OCT images	Accuracy: 0.992 Specificity: 0.995 Sensitivity: 0.990
Kayadibi et al, 2023 [33]	FD-CNN	CNV, DME, Drusen, NO (n = 4696)	84 495 OCT images (Duke + UCSD)	Accuracy: 0.996 (UCSD) Accuracy: 0.975 (Duke)
Proposed Method	ResNet50	AMD, DME, ERM, RAO, RVO, VID, and NO	3 824 OCT images (after augmentation; original n= 2 064)	Accuracy: 0.950; Precision: 0.943; Recall: 0.940;

SMO: Sequential Minimal Optimization, AMD: Age-related Macular Degeneration, DME: Diabetic Macular Edema, CNV: Choroidal Neovascularization, DR: Diabetic Retinopathy, ERM: Epiretinal Membrane, VID: Vitreomacular Interface Disease, RAO: Retinal Artery Occlusion, RVO: Retinal Vein Occlusion, NO: Normal, OCT: Optical Coherence Tomography, AUC: Area Under the Curve, CNN: Convolutional Neural Network, FD-CNN: Fully Dense Convolutional Neural Network, UCSD: University of California San Diego, F1: F1 Score, ROC: Receiver Operating Characteristic, VGG: Visual Geometry Group, ResNet50: 50-layer Residual Network.

thickness defect in the central fovea, the attention map focuses on VMT and the central defect. Significant hyperreflectivity and thickening of the inner retinal layers are observed in the OCT image of a correctly classified RAO. The attention map focuses on retinal thickening. In the correctly classified ERM example, the foveal contour is flattened. The attention map focuses on the flattened foveal contour.

Notably, there are mild cases where the features of the misclassified images are not evident. Mild cystic changes in the inner retinal layers typical of DME are observed at the center of the macula in the DME case that was misclassified as VID. It is noteworthy that there is slight retinal thickening. The attention map is shifted to the peripheral retinal regions. In the ERM OCT image that was misclassified as a normal macula, a thin hyperreflective line was observed adjacent to the inner retinal surface. The attention map is focused on the vitreous over the retina. In the OCT image of AMD that was misclassified as a normal macula, a slightly irregular RPE structure is observed, consistent with AMD. The attention map focuses primarily on the inner retinal layers. In the RVO OCT image that was misclassified as a normal macula, irregularity and mild retinal thickness increase are present in the inner retinal layers affecting the nasal fovea due to RVO. Cystoid macular edema was not evident. The attention map focuses on the peripheral lines rather than the central pathological area. In the VID OCT image that was misclassified as a normal macula, the vitreous cortex has a very small base, causing mild macular shrinkage, resulting in VID. The attention map focused on the adherent vitreous cortex but was misclassified. In the OCT image of a healthy macula that was misclassified as AMD, the optic disc was mistakenly evaluated as a sign of AMD. In the OCT image of an RAO case, hyperreflectivity and minimal thickening of the inner retinal layers are observed. The model successfully focused on retinal thickening but still misclassified it as normal.

https://doi.org/10.5455/annalsmedres.2025.07.206

DISCUSSION

The ResNet50 model demonstrated high performance in distinguishing 7 different macular diseases using only OCT images. The model achieved an overall accuracy of 0.95 (95% CI: 0.9408–0.9600), an F1-score of 0.94 (95% CI: 0.9405–0.9600), a precision of 0.943 (95% CI: 0.9408–0.9600), and a recall of 0.94 (95% CI: 0.9405–0.9600). These consistently high values across all metrics highlight the model's strong inter-class discrimination capability and its potential for clinical decision support system integration.

Unlike many previous studies in the literature that focused primarily on common retinal pathologies, our study includes a broader classification spectrum by incorporating less frequently addressed conditions, such as RAO, RVO, and VID. Although early studies, such as those by Wang et al. [27] and Li et al. [28, 29], achieved high accuracy with limited disease classes, their clinical applicability remains restricted. Kuwayama et al. [30] and Bhatia et al. [31] expanded class diversity but faced limitations in sample size or reliance on commercial systems, which may reduce generalizability. Similarly, Choudhary [10], Esfahani [32] and Kayadibi [33] reported high accuracy but were limited to more balanced and prevalent disease classes. In contrast, our study offers a more comprehensive approach by successfully classifying seven retinal conditions using real-world data, achieving robust performance (accuracy: 0.95, F1 score: 0.94), and demonstrating compatibility with clinical scenarios—highlighting its originality and practical relevance. We extend the typical four-class OCT taxonomy by adding RAO, RVO, and VID, resulting in a seven-class scheme that is aligned with clinical referral and optimized for triage. We also implemented a ViT baseline; although the ViT study (17) reported 99.17% accuracy on the four-class OCT2017 benchmark, our seven-class, referral-oriented task—including RAO/RVO/VID—is inherently harder, and our model achieves Accuracy 0.95 and F1 0.94. Furthermore, considering the prevalence of these diseases [34-36], providing early diagnosis opportunities with a low-cost and accessible AI-based screening system could offer significant gains for public health [37]. Previous studies similar to ours are summarized and presented in Table 4 for comparison.

In terms of model architecture, while many existing works employed VGG-16/19 or custom-built architectures (e.g., FD-CNN) [10, 29, 33], we utilized the ResNet50 model, which balances network depth, generalizability, and computational efficiency through residual connections. Furthermore, the integration of Grad-CAM–based visual explanations enhances interpretability and supports clinical trust in the model's outputs—an aspect often lacking in other studies [27,32]. Compared with related work, our model performs reliably even with imbalanced datasets, offers transparent decision-making through explainable AI, and uses an open-source architecture that promotes practical implementation. Collectively, these features position our model as a technically robust and clinically applicable contribution to the current literature.

The 99.5% accuracy of normal samples is a critical advantage, particularly for community screening and preliminary assessment systems. Low false-positive rates in normal individuals increase patient confidence and prevent unnecessary burden on the health care system. This supports the utility of AI-based systems as triage tools. Additionally, high precision in detecting normal cases can improve system efficiency in large-scale screening programs by reducing the number of cases referred for unnecessary specialist evaluation.

AMD detection with 94.67% accuracy is highly valuable for early diagnosis and timely intervention. Early-stage vitamin supplementation, risk factor management, and timely initiation of anti-vascular endothelial growth factor therapies, which can be applied in wet cases, are key to reducing vision loss [38]. This aligns with global efforts to reduce preventable vision impairment due to AMD through earlier therapeutic interventions.

The ability of the model to distinguish morphologically similar vascular pathologies such as DME (92.22%), RVO (96.67%), and RAO (96.88%) is crucial for triage systems. Despite the structural similarities between DME and RVO on macular OCT, the model's misdiagnosis of these 2 diseases is only 2.0%. This is a particularly valuable distinction in clinical decision-making processes, such as anti-VEGF treatment planning and systemic assessment differentiation [39]. The model may aid in prompt therapeutic direction by providing automated differentiation of these pathologies, especially in settings with limited retinal specialists.

The ability of the model to distinguish VID with 90.91% accuracy is a significant achievement from both technical and clinical perspectives. The VID class includes pathologies, such as VMT, VMA, ERM, and macular hole, that may require surgical management. The correct identification of these cases is essential for timely surgical referral, potentially preventing irreversible visual deterioration in progressive cases.

The ERM, a subtype of the VID class, was classified by the model with 91.4% accuracy. This result demonstrates that ERM is highly successful in identifying cases with isolated ERM, as it frequently accompanies other groups [24].

The most significant difference between VID and other macular diseases is that many cases are managed surgically rather than medically. In particular, macular holes and advanced ERM require vitreoretinal surgery, such as vitrectomy [40]. Accurate identification of VID can reduce time lost not only in terms of disease diagnosis but also by directing patients to the appropriate specialist (vitreoretinal surgeon). This can minimize unnecessary examinations, re-evaluations, and surgical interventions. Therefore, the model's VID performance may improve referral efficiency and reducing time to interven-

https://doi.org/10.5455/annalsmedres.2025.07.206

tion.

Attention heat maps visually represent the OCT image regions that the model focuses on when making classification decisions. The model was found to effectively identify anatomical regions with disease-specific structural abnormalities in correctly classified examples. In misclassified cases, the pathologies that reveal class characteristics in the OCT images were subtle, and the attention maps were generally concentrated in scattered or irrelevant regions. Beyond low disease severity, attenuation of acute signs over time (e.g., subacute/chronic phases) may also hinder feature extraction from a single B-scan, and weak cues can divert attention toward artifacts, extramacular structures (such as vitreous hyperreflectivity), or disk-containing slices, as illustrated by the Grad-CAM examples. Methodologically, these errors could be mitigated by incorporating phase metadata (acute vs. chronic), applying lesion-centric attention regularization with penalties for off-target regions, re-weighting borderline cases via focal loss, and enforcing slice quality control (excluding peripapillary, tilted, or low signal-to-noise scans). These observations that ambiguous or borderline findings may challenge the generalizability of the model, emphasizing the potential role of human-AI collaboration to improve diagnostic safety in such cases. Consistent with these observations, misdiagnoses in our cohort mainly arose from confusion between diseased classes with overlapping OCT appearances (particularly DME, ERM, VID, and RVO), rather than between diseased and normal eyes, and were exacerbated by the limited sample size of some categories.

Limitations

However, certain limitations must be explicitly acknowledged. First, this study used a single-center dataset, which may limit the model's generalizability across different populations and imaging devices. Second, the retrospective nature of the dataset could introduce selection bias, as the data may not reflect the full clinical spectrum of MDs. Third, class imbalance in the dataset may still impact model sensitivity in underrepresented disease groups, although partially addressed through augmentation. Accordingly, prospective validation using multicenter, balanced datasets is necessary to confirm performance across diverse settings. In addition, we did not perform calibration analysis (e.g., ECE, Brier score, and reliability diagrams) or ablation analyses, which limited our assessment of probability calibration and component effects. Finally, the study did not incorporate multimodal data or clinical parameters, which could further enhance classification accuracy and clinical relevance in future models. Moreover, because data augmentation was performed before the train–test split, augmented variants of the same original image may have been present in both sets, potentially leading to a slight overestimation of test performance. Therefore, future work with larger, more balanced multicenter datasets, multimodal inputs, and class-balanced or cost-sensitive training strategies may help reduce such errors.

CONCLUSION

In this study, the ResNet50 model was evaluated as an effective AI approach for the high-accuracy classification of macular diseases from OCT images. In particular, the accurate identification of normal individuals, successful differentiation of common diseases such as AMD, and effective classification of VID requiring surgical intervention hold great potential for clinical decision-making. The high interclass discrimination power of the model enabled selective results even among diseases with similar OCT morphology. To the best of our knowledge, our seven-class design—adding RAO, RVO, and VID—goes beyond four-class OCT benchmarks and better mirrors referral-driven workflows, enabling triage and timely surgical referral. These findings support the utility of AI-supported systems in population-based screening, primary care, and clinical triage. Future studies should test the model's generalizability and clinical applicability with larger and more heterogeneous datasets.

Ethics Committee Approval: Our study did not involve invasive procedures on humans or animals and used a publicly available open-access optical coherence tomography dataset. Therefore, approval from the ethics committee was not necessary.

Informed Consent: Not necessary for this manuscript.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors have no conflicts of interest to declare.

Author Contributions: M.F: Conseption, Design, Supervision, Data Collection and/or Processing, Analysis and/or Interpretation, Writing, Critical Review; İ.T.F: Conseption, Design, Supervision, Analysis and/or Interpretation, Writing, Critical Review; T.T: Design, Materials, Data Collection and/or Processing, Analysis and/or Interpretation, Literature Review, Writing, Critical Review.

Financial Disclosure: This research did not receive any specific grant from public, commercial, or not-for-profit funding agencies.

Artificial Intelligence Disclosure: The authors declare that an artificial intelligence–based tool (ChatGPT 5) was used solely for language editing and improving the clarity and readability of the manuscript. The AI tool did not contribute to the study design, data analysis, interpretation of results, or scientific conclusions. All content was critically reviewed and approved by the authors, who take full responsibility for the manuscript.

REFERENCES

1. Wong TY, Sabanayagam C. Strategies to Tackle the Global Burden of Diabetic Retinopathy: From Epidemiology to Artificial Intelligence. *Ophthalmologica*. 2020;243(1):9-20. doi: 10.1159/000502387.

https://doi.org/10.5455/annalsmedres.2025.07.206

2. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for the detection of DR in primary care offices. *NPJ Digit Med.* 2018;1:39. doi: 10.1038/s41746-018-0040-6.
3. Stankiewicz A, Marciniak T, Budna N, Chwałek R, Dziedzic M. Classification of OCT Images of the Human Eye Using Mobile Devices. *Appl Sci.* 2025;15(6):2937. doi: 10.3390/app15062937.
4. Akinniyi O, Rahman MM, Sandhu HS, El-Baz A, Khalifa F. Multi-Stage Classification of Retinal OCT Using Multi-Scale Ensemble Deep Architecture. *Bioengineering (Basel).* 2023;10(7):823. doi: 10.3390/bioengineering10070823.
5. Khan A, Pin K, Aziz A, Han JW, Nam Y. Optical Coherence Tomography Image Classification Using Hybrid Deep Learning and Ant Colony Optimization. *Sensors (Basel).* 2023;23(15):6706. doi: 10.3390/s23156706.
6. Heydon P, Egan C, Bolter L et al. Prospective evaluation of an AI-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. *Br J Ophthalmol.* 2021;105(5):723-728. doi: 10.1136/bjophthalmol-2020-316594.
7. Song Y, Wang Q, Du Y, et al. Prognostic biomarkers of treatment-naïve central retinal vein occlusion with macular edema. *Eur J Med Res.* 2025;30(1):594. doi: 10.1186/s40001-025-02884-x.
8. Özdaş MB, Uysal F, Hardalaç F. Classification of Retinal Diseases in Optical Coherence Tomography Images Using Artificial Intelligence and Firefly Algorithm. *Diagnostics (Basel).* 2023;13(3):433. doi: 10.3390/diagnostics13030433.
9. Abd El-Khalek AA, Balaha HM, Alghamdi NS, et al. A concentrated machine learning-based classification system for age-related macular degeneration (AMD) diagnosis using fundus images. *Sci Rep.* 2024;14(1):2434. doi: 10.1038/s41598-024-52131-2.
10. Choudhary A, Ahlawat S, Urooj S, Pathak N, Lay-Ekuakille A, Sharma N. A Deep Learning-Based Framework for Retinal Disease Classification. *Healthcare (Basel).* 2023;11(2):212. doi: 10.3390/healthcare11020212.
11. Kabataş B, Ölmez E. Deep learning-based decision support system for the classification of retinal diseases: diabetic retinopathy and macular hole. *Artif Intell Theory Appl.* 2025;5(1):51-62.
12. Sunil SS, Shrivindhya A. Computer-aided detection of diabetic retinopathy on fundus images via statistical classification. 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS). doi: 10.1109/IC3SIS54991.2022.9885359.
13. Alizadeh Eghtedar R, Vard A, Malekahmadi M, Peyman A. A new computer-aided diagnosis tool based on deep learning methods for automatic detection of retinal disorders from OCT images. *Int Ophthalmol.* 2024;44(1):110. doi: 10.1007/s10792-024-03033-9.
14. Rajan R, Kumar SN. Deep learning architectures for OCT images retinal disease classification. *SN Comput Sci.* 2025;6(2):155. doi: 10.1007/s42979-025-03715-w.
15. Wang D, Lian J, Jiao W. Multilabel classification of retinal disease via a novel vision transformer model. *Front Neurosci.* 2024;17:1290803. doi: 10.3389/fnins.2023.1290803.
16. Laouarem A, Kara-Mohamed C, Bourennane EB, Hamdi-Cherif A. HTC-retina: A hybrid retinal diseases classification model using transformer-Convolutional Neural Network from optical coherence tomography images. *Comput Biol Med.* 2024;178:108726. doi: 10.1016/j.compbiomed.2024.108726.
17. Akça S, Garip Z, Ekinci E, Atban F. Automated classification of choroidal neovascularization, diabetic macular edema, and drusen from retinal OCT images using vision transformers: a comparative study. *Lasers Med Sci.* 2024;39(1):140. doi: 10.1007/s10103-024-04089-w.
18. Sheth JU, Stewart MW, Narayanan R et al. Macular neovascularization. *Surv Ophthalmol.* 2025;70(4):653-675. doi: 10.1016/j.survophthal.2024.08.003.
19. Spaide RF, Jaffe GJ, Sarraf D, et al. Consensus Nomenclature for Reporting Neovascular Age-Related Macular Degeneration Data: Consensus on Neovascular Age-Related Macular Degeneration Nomenclature Study Group. *Ophthalmology.* 2020;127(5):616-636. doi:10.1016/j.ophtha.2019.11.004.
20. Das R, Spence G, Hogg RE, Stevenson M, Chakravarthy U. Disorganization of Inner Retina and Outer Retinal Morphology in Diabetic Macular Edema. *JAMA Ophthalmol.* 2018;136(2):202-208. doi: 10.1001/jamaophthalmol.2017.6256.
21. Song Y, Wang Q, Du Y, et al. Prognostic biomarkers of treatment-naïve central retinal vein occlusion with macular edema. *Eur J Med Res.* 2025;30(1):594. doi: 10.1186/s40001-025-02884-x.
22. Ahn SJ, Woo SJ, Park KH, Jung C, Hong JH, Han MK. Retinal and choroidal changes and visual outcome in central retinal artery occlusion: an optical coherence tomography study. *Am J Ophthalmol.* 2015;159(4):667-676. doi: 10.1016/j.ajo.2015.01.001.
23. Phillips JD, Hwang ES, Morgan DJ, Creveling CJ, Coats B. Structure and mechanics of the vitreoretinal interface. *J Mech Behav Biomed Mater.* 2022;134:105399. doi: 10.1016/j.jmbbm.2022.105399.
24. Duker JS, Kaiser PK, Binder S, et al. The International Vitreomacular Traction Study Group classification of vitreomacular adhesion, traction, and macular hole. *Ophthalmology.* 2013;120(12):2611-2619. doi: 10.1016/j.ophtha.2013.07.042.
25. Besagar S, Chennupati S, Ji X, Chen Q, Thomas AS, Finn AP. Optical coherence tomography biomarkers in the epiretinal membrane. *Ophthalmic Surg Lasers Imaging Retina.* 2025;56(7):422-428. doi: 10.3928/23258160-20250416-01.
26. Kulyabin M, Zhdanov A, Nikiforova, et al. OCTDL: Optical coherence tomography dataset for image-based deep learning methods. *Sci Data.* 2024;11(1):365. doi: 10.1038/s41597-024-03182-7.
27. Wang Y, Zhang Y, Yao Z, Zhao R, Zhou F. Machine learning-based detection of age-related macular degeneration and diabetic macular edema from optical coherence tomography images. *Biomed Opt Express.* 2016;7(12):4928-4940. doi: 10.1364/BOE.7.004928.
28. Li F, Chen H, Liu Z, et al. Deep learning-based automated detection of retinal diseases using OCT images. *Biomed Opt Express.* 2019;10(12):6204-6226. doi: 10.1364/BOE.10.006204.
29. Li F, Chen H, Liu Z, Zhang X, Wu Z. Fully automated detection of retinal disorders using image-based deep learning. *Graefes Arch Clin Exp Ophthalmol.* 2019;257(3):495-505. doi: 10.1007/s00417-018-04224-8.
30. Kuwayama S, Ayatsuka Y, Yanagisono D, et al. Automated detection of macular diseases using optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images. *J Ophthalmol.* 2019;2019:6319581. doi: 10.1155/2019/6319581.
31. 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(8):1549-1557. doi: 10.1097/IAE.0000000000002640.
32. Riazi Esfahani P, Reddy AJ, Nawathey N, et al. Deep learning classification of drusen, choroidal neovascularization, and diabetic macular edema in OCT images. *Cureus.* 2023;15(7):e41615. doi: 10.7759/cureus.41615.
33. Kayadibi İ, Güraksın GE. An explainable fully dense fusion neural network with deep support vector machine for retinal disease determination. *Int J Comput Intell Syst.* 2023;16(1):28. doi: 10.1007/s44196-023-00210-z.
34. GBD 2021 Global AMD Collaborators. Global burden of vision impairment due to age-related macular degeneration, 1990–2021, with forecasts to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Glob Health.* 2025;13(7):e1175-e1190. doi: 10.1016/S2214-109X(25)00143-3.
35. Candan Ö, Orman G, Ünlü N, Üney G, Burcu A. Prevalence of retinal vascular diseases in a Turkish tertiary care hospital in Türkiye: a hospital-based epidemiologic study. *Turk J Ophthalmol.* 2025;55(1):16-23. doi: 10.4274/tjo.galenos.2025.88262.

https://doi.org/10.5455/annalsmedres.2025.07.206

36. McKibbin M, Farragher T, Shickle D. Vitreoretinal interface abnormalities in middle-aged adults with visual impairment in the UK Biobank study: prevalence, impact on visual acuity and associations. *BMJ Open Ophthalmol.* 2017;1(1):e000057. doi: 10.1136/bmjophth-2016-000057.
37. Heloterä H, Viita AM, Laine J. The evolution of workload associated with anti-VEGF treatments for AMD, DME, RVO, and mCNV in the Hospital District of Southwest Finland. *Clin Ophthalmol.* 2024;18:3645-3655. doi: 10.2147/OPTH.S479816.
38. Guymer RH, Rosenfeld PJ, Ehlers JP, et al. Designing the next generation of clinical trials in intermediate AMD—a consensus driven, pragmatic, proof-of-concept early intervention study. *Exp Eye Res.* 2025;255:110340. doi: 10.1016/j.exer.2025.110340.
39. 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. doi: 10.1167/tvst.11.9.29.
40. Er-Reguyeg Y, Doukkali S, Hébert M et al. Anatomical and functional outcomes of lamellar macular hole and epiretinal membrane foveoschisis surgery: predictive factors and associated complications: a retrospective interventional study. *Clin Ophthalmol.* 2025;19:1365-1376. doi: 10.2147/OPTH.S499493.

https://doi.org/10.5455/annalsmedres.2025.07.206