AMERICAN ACADEMY OF OPHTHALMOLOGY®

Deep-Reticular Pseudodrusen-Net: A 3-Dimensional Deep Network for Detection of Reticular Pseudodrusen on OCT Scans

Amr Elsawy, PhD,[1] Tiarnan D.L. Keenan, PhD, MD,[2] Alisa T. Thavikulwat, MD,[2] Amy Lu, MD,[2] Sunil Bellur, MD,[2] Souvick Mukherjee, PhD,[2] Elvira Agron, MS,[2] Qingyu Chen, PhD,[1] Emily Y. Chew, MD,[2] Zhiyong Lu, PhD[1]

Objective: To propose Deep-RPD-Net, a 3-dimensional deep learning network with semisupervised learning (SSL) for the detection of reticular pseudodrusen (RPD) on spectral-domain OCT scans, explain its decision-making, and compare it with baseline methods.

Design: Deep learning model development.

Participants: Three hundred fifteen participants from the Age-Related Eye Disease Study 2 Ancillary OCT Study (AREDS2) and 161 participants from the Dark Adaptation in Age-related Macular Degeneration Study (DAAMD).

Methods: Two datasets comprising of 1304 (826 labeled) and 1479 (1366 labeled) OCT scans were used to develop and evaluate Deep-RPD-Net and baseline models. The AREDS2 RPD labels were transferred from fundus autofluorescence images, which were captured at the same visit for OCT scans, and DAAMD RPD labels were obtained from the Wisconsin reading center. The datasets were divided into 70%, 10%, and 20% at the participant level for training, validation, and test sets, respectively, for the baseline model. Then, SSL was used with the unlabeled OCT scans to improve the trained model. The performance of Deep-RPD-Net was compared to that of 3 retina specialists on a subset of 50 OCT scans for each dataset. En face and B-scan heatmaps of all networks were visualized and graded on 25 OCT scans with positive labels, using a scale of 1 to 4, to explore the models' decision-making.

Main Outcome Measures: Accuracy and area under the receiver-operating characteristic curve (AUROC).

Results: Deep-RPD-Net achieved the highest performance metrics, with accuracy and AUROC of 0.81 (95% confidence interval [CI]: 0.76–0.87) and 0.91 (95% CI: 0.86–0.95) on the AREDS2 dataset and 0.80 (95% CI: 0.75–0.84) and 0.86 (95% CI: 0.79–0.91) on the DAAMD dataset. On the subjective test, it achieved accuracy of 0.84 compared with 0.76 for the most accurate retina specialist on the AREDS2 dataset and 0.82 compared with 0.84 on the DAAMD dataset. It also achieved the highest visualization grades, of 3.26 and 3.32 for en face and B-scan heatmaps, respectively.

Conclusions: Deep-RPD-Net was able to detect RPD accurately from OCT scans. The visualizations of Deep-RPD-Net were the most explainable to the retina specialist with the highest accuracy. The code and pretrained models are publicly available at https://github.com/ncbi-nlp/Deep-RPD-Net.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science 2025;5:100655*

Supplemental material available at www.ophthalmologyscience.org.

Reticular pseudodrusen (RPD) are a form of drusen associated with age-related macular degeneration (AMD). In contrast to conventional drusen, which are located below the retinal pigment epithelium, RPD are located above the retinal pigment epithelium.[1,2] Hence, they are also known as subretinal drusenoid deposits. Reticular pseudodrusen represent an important AMD phenotype since their presence is strongly and independently associated with increased risk of progression to advanced AMD stages, i.e., neovascular AMD and geographic atrophy (GA).[3,4] Hence, the detection of RPD is highly useful for providing accurate prognostic information for individuals with AMD.

Reticular pseudodrusen are typically not visible on clinical examination or on the simple imaging modality of color fundus photography (CFP).[1,2,4] With the advent of more

https://doi.org/10.1016/j.xops.2024.100655
ISSN 2666-9145/24

advanced imaging methods, including fundus autofluorescence (FAF), near-infrared reflectance, and spectral-domain OCT,[5] RPD detection can be performed in specialized settings like dedicated image reading centers, but the task remains very challenging for ophthalmologists.[2,3] Typically, RPD appear as hypoautofluorescent spots on FAF. Owing to the more detailed 3-dimensional (3D) view of OCT, AMD features like RPD can often be characterized in more detail. Therefore, the use of artificial intelligence algorithms is of great interest for the difficult task of RPD detection.

In particular, deep learning is a subfield of artificial intelligence that has become the state-of-the-art in solving computer vision problems such as image classification. Deep convolutional neural networks (CNNs) can learn a complex hierarchy of nonlinear salient features that have a large discrimination power by being trained on labeled sets of images.[6] In the literature, Keenan et al[7] have used transfer learning-based models with FAF and CFP to detect RPD. Their FAF-based model achieved an area under the receiver operating characteristic curve (AUROC) of 0.939, and their CFP-based model achieved an AUROC of 0.832. Their models surpassed the performance of ophthalmologists.[7] Chen et al[8] have developed a multimodal, multitask deep network that uses both CFP and FAF in RPD detection. Their proposed network slightly enhanced the performance obtained using FAF images only,[8] with an AUROC of 0.933. Schwartz et al[9] have proposed both detection and segmentation models for RPD. They attempted to detect both RPD and conventional drusen from OCT scans using Inception-v1 architecture by replacing 2-dimensional (2D) convolutions with 3D convolutions. Their classification model achieved an AUROC of 0.99. However, they detected the presence of RPD or drusen as one class without distinguishing between them. Moreover, Schwartz et al quantified using the area of RPD on B-scans. They used a small, annotated dataset of 334 B-scans to develop and evaluate (60% training, 20% validation, and 20% testing) 2D segmentation U-Net. They reported a quite low agreement ($\sim$0.5x) in the area measurement between the graders and the area based on the segmentation model using $\sim$67 B-scans. They also reported a very low Dice score of their segmentation model, which suggests the difficulty of the RPD detection and segmentation on OCT scans.

In this work, we propose a classification network, Deep-RPD-Net, to detect RPD from OCT scans using scan-level binary labels. The proposed network consists of a 3D CNN, with every convolutional block having a squeeze-excitation block and a residual connection, and a classifier with 2 fully connected layers and SoftMax output. We obtained the RPD labels by transferring RPD labels from FAF images that were taken during the same visits as the OCT scans. We compared Deep-RPD-Net with other baseline networks that include a 2D CNN with average bagging and a 3D CNN without a squeeze excitation block.[10,11] We further enhanced Deep-RPD-Net training by using semi-supervised learning (SSL) on unlabeled data. The Gradient-weighted Class Activation Mapping (Grad-CAM)[12] heatmaps of all networks were visualized and graded for explainability. Finally, we compared the model performance to that of retina specialists.

Methods

Datasets

The main dataset used in this study was curated from the Ancillary OCT Study of the Age-Related Eye Disease Study 2 (AREDS2), as described in previous studies.[11,13] The details of the AREDS2 design and protocol have been described previously.[14,15] In short, the AREDS2 was a multicenter phase III randomized controlled clinical trial designed to study the effects of nutritional supplements in participants at moderate to high risk of progression to late AMD. The participants were aged between 50 and 85 years at baseline (median age 74 years). At baseline, participants had either bilateral large drusen or large drusen in one eye and advanced AMD in the fellow eye. At each visit, the participants underwent eye examination and imaging by certified personnel using standardized protocols. The Ancillary OCT Study of AREDS2 enrolled AREDS2 participants from the Devers Eye Institute, Emory Eye Center, Duke Eye Center, and the National Eye Institute. The study was approved by the institutional review boards of the study sites, and written informed consent was obtained from all participants. It adhered to the tenets of the Declaration of Helsinki and complied with the Health Insurance Portability and Accountability Act. The study was registered at ClinicalTrials.gov (identifier NCT00734487). The participants underwent imaging using the Bioptigen Tabletop OCT system (Research Triangle Park) at the annual study visits.[16] The OCT scans had a surface area of 6.7 $\times$ 6.7 mm square centered at the fovea (with 100 B-scans, 1000 A-scans per B-scan, and 67 μm spacing between each B-scan).

The ground truth RPD labels were transferred from FAF images that were taken at the same visits as the OCT scans, as described in previous studies.[4,7] In short, the RPD labels assigned to each FAF image were from expert human graders at the Wisconsin Reading Center. The grading team was comprised of 6 graders: 4 primary graders and 2 senior graders as adjudicators. Reticular pseudodrusen were defined as clusters of discrete round or oval hypoautofluorescent lesions or confluent ribbon-like patterns with intervening areas of normal or increased autofluorescence; a minimum of 0.5-disc areas (approximately 5 lesions) was required. Two primary graders independently evaluated the FAF images for the presence of RPD. A senior grader adjudicated the final grade in the case of disagreement between the graders.[4] The level of agreement between the 2 primary graders for the presence or absence of RPD was 94%.[4]

The description of the AREDS2 OCT dataset is shown in Figure 1. In short, the dataset consisted of 1304 OCT scans from 315 participants, where participants contributed multiple OCT scans (i.e., from different study visits in consecutive years). For 826 of the 1304 OCT scans, the RPD labels were transferred from the FAF image to the corresponding OCT scan, as both FAF images and OCT scans were available at the same visit. Of these 826 OCT scans, 222 were labeled as positive for RPD, based on the corresponding FAF image.

A second dataset was used from the Study of Dark Adaptation in Age-Related Macular Degeneration (DAAMD). The details of the DAAMD design and protocol have been described previously.[17–19] In short, participants aged >50 years with no sign of AMD or early and intermediate stages of AMD were recruited from the study of DAAMD at the National Eye Institute, National

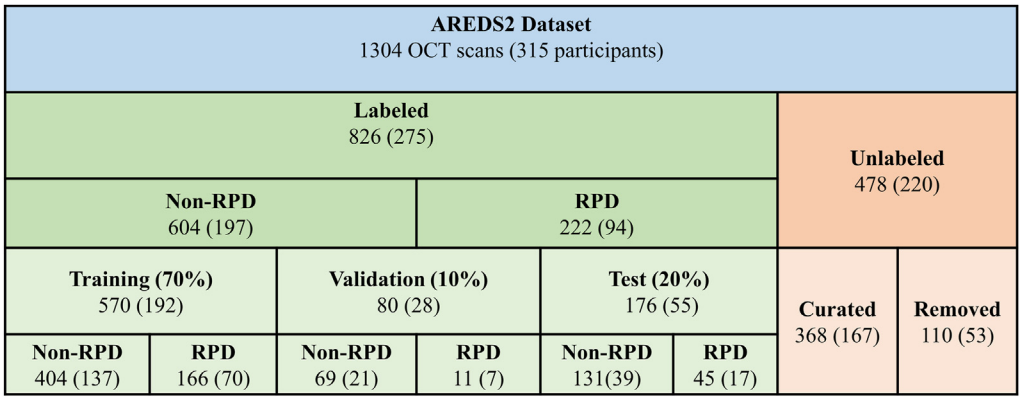

Figure 1. The description of the OCT dataset used in this study. AREDS2 = Age-Related Eye Disease Study 2; RPD = reticular pseudodrusen.

Institutes of Health. Only one eye per participant was included as the study eye. The participants underwent imaging using the Spectralis, Heidelberg Engineering OCT system at the annual study visits. The OCT scans had a surface area of 9 × 7.4 mm square centered at the fovea (with 121 B-scans and 768 A-scans per B-scan). The RPD labels were evaluated by the Wisconsin Reading Center using OCT and other modalities (FAF, near-infrared, or CFP).[17–20] This study was approved by the institutional review board of the National Institutes of Health and adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from all participants. It was registered at ClinicalTrials.gov (identifier NCT01352975). The description of the secondary OCT dataset is shown in Figure S2 (available at www.ophthalmologyscience.org). In short, the dataset consisted of 1479 OCT scans from 161 participants, where participants contributed multiple OCT scans. The RPD labels were available for 1366 OCT scans with 191 positive OCT scans.

Both datasets (i.e., labeled parts) were split at the participant level into independent 70% training, 10% validation, and 20% test sets. The remaining unlabeled OCT scans were curated by removing all scans of any participants existing in the validation or test sets to ensure that the network was not biased in validation or testing.

Image Preprocessing

The OCT B-scans were cropped to remove the bottom half that contained background noise only. Then, they were resized to 224 × 448 pixels, a large enough size to better capture RPD features since RPD are relatively small. The top 2 and bottom 2 OCT B-scans were excluded, i.e., 96 OCT B-scans were used (128 for the second dataset), to be to the power of 2, to have proper pooling. Thus, the final OCT scan size was set to 224 × 448 × 96 pixels. The OCT scan voxels were normalized between −1 and 1. Augmentation was applied to the OCT scans during the training using random flipping or erasing with a probability of 0.5.

Deep-RPD-Net. A high-level block diagram of the architecture of the proposed Deep-RPD-Net is shown in Figure 3. The proposed network comprises a 3D input layer, a 3D encoding network, a global average pooling layer, and a classifier. The encoder contains an initial block of two 3D convolutional layers with 64 filters followed by 4 encoding blocks, "E-Block;" each contains three 3D convolutional, one maximum pooling, and one squeeze excitation layer, as well as a residual connection. In all convolutions, we used the rectified linear unit activation function. The number of filters in each encoding block grows exponentially with a factor of 2. All maximum pooling operations are of size 2 × 2 × 2. The classifier consists of a dropout layer followed by a dense layer of length 256 with rectified linear unit activation function, a second dropout layer, and a dense layer with SoftMax activation. We used a dropout rate of 0.5.

For baseline comparisons, we developed a 2D network with average bagging, where the encoder is similar to that of Deep-RPD-Net but with 2D convolutions and without squeeze excitation blocks.[10] Also, we developed a 3D network with an encoder similar to that of Deep-RPD-Net but without squeeze excitation blocks.[11]

For all networks, the SoftMax activation function was used to generate the class likelihood probabilities; the binary predictions were made using the class with the highest probability. To train the networks, we used the Adam optimizer[21] with a learning rate of 0.0001, batch size of 2, and maximum epoch number of 50. We used oversampling to handle the imbalance in class size in our dataset. All experiments were performed using python 3.7 and TensorFlow 2.8 deep learning library running on a server with 48 Intel Xeon CPUs with 754 Gb RAM and an NVIDIA GeForce GTX 1080 Ti 32Gb GPU. To train the networks, we used categorical cross-entropy loss function.

To make use of the curated unlabeled OCT scans in improving performance, we used a SSL method, which is described in Figure 4. Briefly, the pretrained Deep-RPD-Net was used to generate pseudo-labels of the unlabeled OCT scans. The OCT scans with confidence (i.e., SoftMax probability) ≥0.90 were added to the training data. Then, the model was tuned using the extended data for 5 epochs. The process was repeated until there were no more OCT scans to add or if their number was ≤5% of the total unlabeled data size.

Performance Evaluation and Comparison

We used accuracy, precision, recall, F1 score, kappa (κ), and AUROC to evaluate and compare all networks. The 95% confidence intervals (CIs) were computed for all performance metrics. We evaluated the performance metrics on the 4 models: (1) a shared 2D network with average bagging, where the encoder is similar to that of Deep-RPD-Net, but with 2D convolutions and without squeeze excitation blocks (Base2D); (2) a 3D network with an encoder similar to that of Deep-RPD-Net, but without squeeze excitation blocks (Base3D); (3) the proposed network, Deep-RPD-Net; and (4) Deep-RPD-Net with SSL (Deep-RPD-Net+).

The AREDS2 and DAAMD datasets were used to train and validate the models. Due to significant heterogeneity (i.e., scan size in mm, images per scan, image size, image quality, image

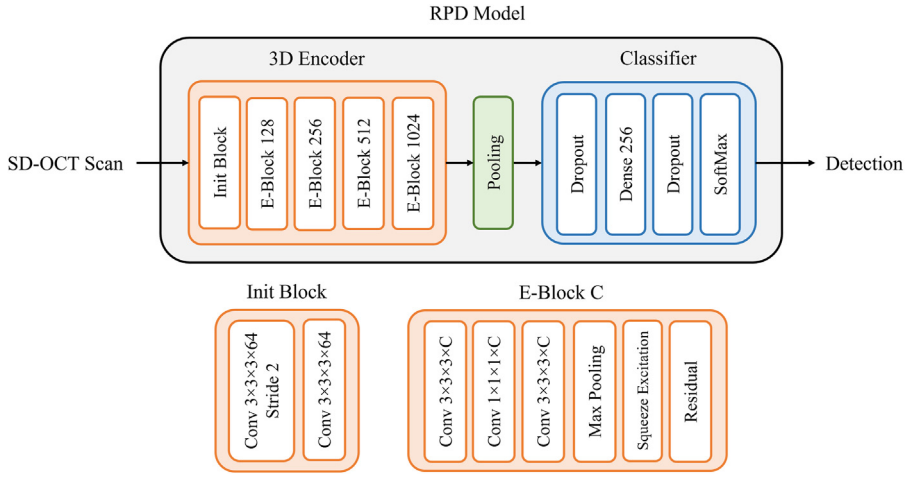

Fig. 3. The architecture of the proposed Deep-RPD-Net, where E-Block denotes the encoding blocking and C is the number of the channels in this block. The Deep-RPD-Net consists of a 3D encoder of 5 blocks and a global average pooling layer, and a classifier of 2 dropout and 2 dense layers. 3D = 3-dimensional; RPD = reticular pseudodrusen; SD-OCT = spectral-domain OCT.

averaging, image registration, etc.) between the 2 datasets, DAAMD could not be used in external validation.

Visualization of Deep-RPD-Net

The Grad-CAMs[22] were used to visualize the features learned by each network. To have a higher heatmap size, we used the Grad-CAM heatmaps of the last 2 convolutional layers and averaged them. The en face and OCT B-scan heatmaps were obtained as described by Elsawy et al.[11] In short, en face heatmaps were obtained by averaging the Grad-CAM heatmaps along the axial dimension. Heatmaps of OCT B-scans were obtained by slicing the Grad-CAM heatmap along the transverse dimension. All heatmaps were modulated by multiplying them by the en face or the OCT B-scan to emphasize the important location.

Subjective Grading by Retina Specialists

A masked subjective test was conducted by 3 retina specialists (T.K., A.T., and S.B.) on a subset of 50 OCT scans from the 2 test sets. The 3 retina specialists were at attending level with 6 years of experience for T.K., 3 years of experience for A.T., and 2 years of experience for S.B. The subjective dataset was randomly selected from the test dataset such that it contained 25 OCT scans with RPD and 25 OCT scans without RPD. The OCT scans were deidentified, and the distribution was not known to the retina specialists. The retina specialists were instructed to give a binary grade, i.e., 0 (absence of RPD) or 1 (presence of RPD), for each OCT scan.

In addition, the retina specialist with the highest accuracy on the previous task was instructed to grade the explainability of the visualizations for each network. A new subset of 25 OCT scans labeled with RPD was selected from the test set. The grading involved 2 tasks: (1) to grade the en face heatmaps generated by each network, and (2) to grade the B-scan heatmaps generated by each network. In each case, the task was to assess how closely the areas of high signal on the heatmap corresponded to retinal areas where RPD were indeed located. Grading was performed under masked conditions. A semiquantitative grading scale was used: poor (1), moderate (2), good (3), and very good (4) correspondence. Intermediate values were permitted, e.g., moderate/good (2.5). The final grade of each network was computed as the average of its individual grades.

Results

Performance of Deep-RPD-Net in Detecting RPD

The performance metrics of Deep-RPD-Net with SSL (Deep-RPD-Net+) and the other networks in detecting RPD

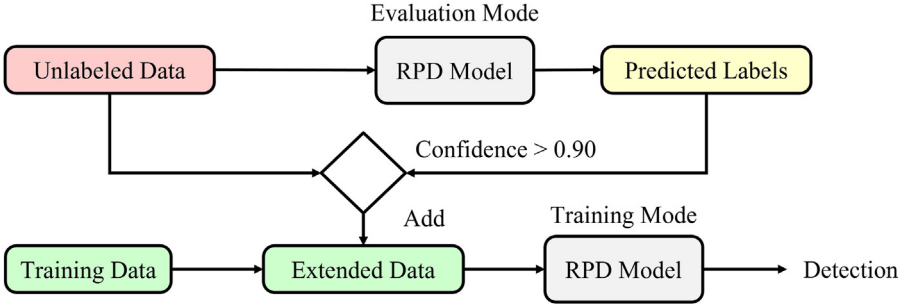

Fig. 4. One iteration of semisupervised learning method. The unlabeled data instances are labeled using the pretrained model, and the instances with confidence ≥0.90 are added to the training data. Then, the model is tuned using the extended data. The process repeats until there are no more or few instances to add. RPD = reticular pseudodrusen.

Table 1. The Average Performance Metrics Along With 95% Confidence Intervals for Deep-RPD-Net and Other Networks on the Two Test Datasets

Dataset	Network	Accuracy	Precision	Recall	F1 Score	κ	AUROC
AREDS2	Base2D	0.60 (0.53, 0.67)	0.37 (0.28, 0.47)	0.80 (0.68, 0.91)	0.50 (0.41, 0.60)	0.24 (0.13, 0.36)	0.68 (0.60, 0.77)
	Base3D	0.73 (0.66, 0.80)	0.48 (0.37, 0.60)	0.78 (0.66, 0.89)	0.59 (0.49, 0.69)	0.41 (0.28, 0.54)	0.81 (0.75, 0.88)
	Deep-RPD-Net	0.79 (0.73, 0.85)	0.57 (0.44, 0.71)	0.67 (0.53, 0.80)	0.62 (0.49, 0.72)	0.47 (0.33, 0.61)	0.88 (0.82, 0.93)
	Deep-RPD-Net+	**0.81 (0.76, 0.87)**	**0.59 (0.48, 0.71)**	**0.82 (0.71, 0.93)**	**0.69 (0.59, 0.79)**	**0.56 (0.43, 0.69)**	**0.91 (0.86, 0.95)**
DAAMD	Deep-RPD-Net	**0.84 (0.80, 0.88)**	0.42 (0.17, 0.75)	0.13 (0.04, 0.24)	0.20 (0.07, 0.34)	0.14 (0.01, 0.28)	0.76 (0.70, 0.82)
	Deep-RPD-Net-tuned	0.83 (0.79, 0.88)	**0.45 (0.31, 0.59)**	0.50 (0.35, 0.64)	0.47 (0.35, 0.59)	0.37 (0.23, 0.51)	0.79 (0.71, 0.86)
	Deep-RPD-Net+	0.80 (0.75, 0.84)	0.41 (0.31, 0.51)	**0.80 (0.68, 0.91)**	**0.54 (0.43, 0.64)**	**0.43 (0.31, 0.53)**	**0.86 (0.79, 0.91)**

AUROC = area under the receiver operating characteristic curve; AREDS2 = Age-Related Eye Disease Study 2; DAAMD = Dark Adaptation in Age-related Macular Degeneration Study.
The highest scores are in bold text.

from OCT scans are summarized in Table 1. On the AREDS2 dataset, Deep-RPD-Net+ achieved the highest overall performance scores with average metrics of 0.81 (95% CI: 0.76−0.87), 0.59 (95% CI: 0.48−0.71), 0.82 (95% CI: 0.71−0.93), 0.69 (95% CI: 0.59−0.79), 0.56 (95% CI: 0.43−0.69), and 0.91 (95% CI: 0.86−0.95) for accuracy, precision, recall, F1 score, κ, and AUROC, respectively. On the DAAMD dataset, Deep-RPD-Net+ achieved the highest performance scores with average metrics of 0.80 (95% CI: 0.68−0.91), 0.54 (95% CI: 0.43−0.64), 0.43 (95% CI: 0.31−0.53), and 0.86 (95% CI: 0.79−0.91) for recall, F1 score, κ, and AUROC, respectively.

The results of the subjective grading are summarized in Table 2. On the AREDS2 dataset, Deep-RPD-Net+ outperformed the 3 retina specialists, with performance metrics of 0.84, 0.84, 0.84, 0.84, 0.68, and 0.92 for accuracy, precision, recall, F1 score, κ, and AUROC, respectively. On the DAAMD dataset, Deep-RPD-Net+ outperformed one of the retina specialists with performance metrics of 0.82, 0.83, 0.80, 0.82, 0.64, and 0.88 for accuracy, precision, recall, F1 score, κ, and AUROC, respectively.

The receiver operating characteristic curves are shown in Figure 5, where the left figure shows the receiver operating characteristic curves of Deep-RPD-Net+ as well as other networks on the AREDS2 test set, the middle figure shows the results of Deep-RPD-Net+ against the graders on the AREDS2 subjective test, and the right figure shows the results of Deep-RPD-Net+ against the graders on the DAAMD subjective test. Deep-RPD-Net+ outperformed the other networks on the AREDS2 test set and outperformed the 3 retina specialists on the AREDS2 subjective test as well as one of them on the DAAMD subjective test.

Explainability of the Visualizations of Deep-RPD-Net

The visualization grading is summarized in Table 3. Deep-RPD-Net+ outperformed the other networks, with an average grade (range 1−4) of 3.26 on the en-face heatmaps and of 3.32 on the OCT B-scans.

Representative examples of the en-face heatmaps for each network are shown in Figure 6 for the AREDS2 dataset and in Figure S7 (available at www.ophthalmologyscience.org) for the DAAMD dataset. For each en-face OCT B-scan, the corresponding FAF image is also shown for easier interpretation of the visualizations. Deep-RPD-Net+ tended to have high levels of correspondence between

Table 2. The Subjective Grading of Retina Specialists Against Deep-RPD-Net on Subsets of the Two Test Subsets (i.e., Each Has 50 OCT Scans)

Dataset	Network	Accuracy	Precision	Recall	F1 Score	κ	AUROC
AREDS2	Grader1	0.76	0.78	0.72	0.75	0.52	-
	Grader2	0.62	0.67	0.48	0.56	0.24	-
	Grader3	0.56	0.54	0.88	0.67	0.12	-
	Deep-RPD-Net+	**0.84**	**0.84**	**0.84**	**0.84**	**0.68**	**0.92**
DAAMD	Grader1	0.84	**0.90**	0.76	0.83	0.68	-
	Grader2	0.80	0.83	0.76	0.79	0.60	-
	Grader3	**0.88**	0.81	**1.00**	**0.89**	**0.76**	-
	Deep-RPD-Net+	0.82	0.83	0.80	0.82	0.64	**0.88**

AUROC = area under the receiver operating characteristic curve; AREDS2 = Age-Related Eye Disease Study 2; DAAMD = Dark Adaptation in Age-related Macular Degeneration Study.
The highest scores are in bold text.

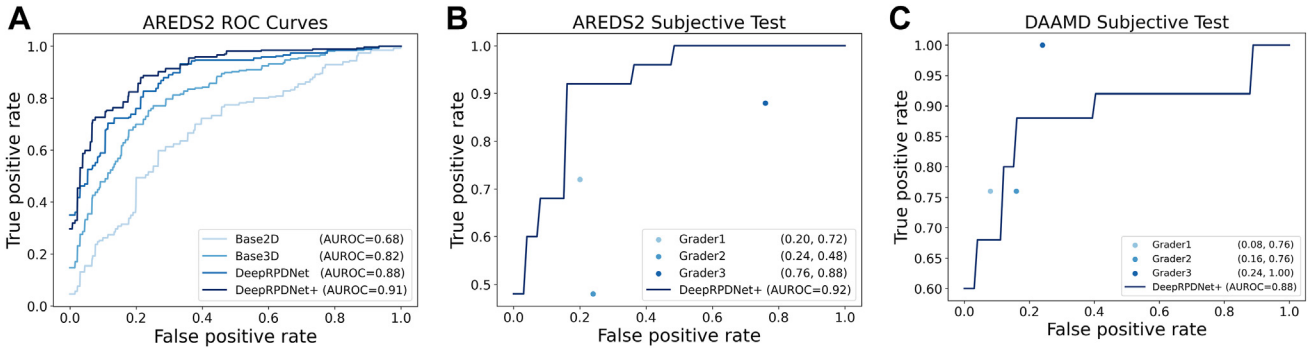

Fig. 5. The ROC curves for all networks over the AREDS2 test set (left) and the ROC curve of the Deep-RPD-Net versus the graders on the subjective test for AREDS2 (middle) and DAAMD (right). AREDS2 = Age-Related Eye Disease Study 2; AUROC = area under the receiver operating characteristic curve; DAAMD = Dark Adaptation in Age-related Macular Degeneration Study; ROC = receiver operating characteristic.

areas of high signal and retinal areas where RPD were located. It was also able to distinguish between RPD and other features, such as GA. By contrast, some of the lower-performing networks often had areas of high signal in retinal areas with GA (and without RPD).

Representative examples of the B-scan heatmaps of Deep-RPD-Net+ are shown in Figure 8 for the AREDS2 dataset. Again, Deep-RPD-Net+ tended to have high levels of correspondence between areas of high signal and retinal areas where RPD were located. Areas of high signal were also often present in areas of retinal thinning directly overlying RPD; this is consistent with the features known to accompany RPD.[23,24] Unlike some of the other lower-performing networks, it was also able to distinguish between RPD and other features, such as GA and conventional drusen.

Discussion

Deep-RPD-Net achieved the highest performance metrics, compared with those of the other networks, in the detection of RPD on OCT scans (see Table 1 and Fig 4). In general, the 3D networks achieved superior performance compared with that of the 2D network; this is consistent with previous observations reported by Elsawy et al.[10,11] However, the precision in detecting RPD was relatively low, which might relate to RPD not being well represented (i.e., distribution of easy and difficult cases is not similar across sets) in the training, validation, and test sets, as well as a degree of data imbalance. The possibility of model overfitting seems less likely since the 3D models outperformed the baseline 2D model. Also, this performance level (i.e., AUROC) in detecting RPD on OCT scans was very similar to the performance level of previous deep learning models on FAF images and superior to that of previous deep learning models on CFP images, as reported by Keenan et al.[7] This suggests a high level of performance by Deep-RPD-Net for RPD detection on OCT scans.

Table 3. The Subjective Grading of the Most Accurate Retina Specialists on a 25 Positive OCT Scans, Where the Grades Are Poor (1), Moderate (2), Good (3), and Very Good (4)

Network	En Face Grade	OCT B-Scan Grade
Base2D	0.98	0.52
Base3D	1.82	1.78
Deep-RPD-Net	3.00	3.02
Deep-RPD-Net+	**3.26**	**3.32**

The highest scores are marked in bold text.

Semisupervised learning lies between supervised learning, where the data are fully labeled, and unsupervised learning, where the data have no labels. The effect of SSL comes when the data are partially labeled. Semisupervised learning uses the weak supervision of a model trained on the labeled part of the data to provide pseudo-labels of the unlabeled part. The pseudo-labeled data are then added to the training data. The pseudo-labels can be seen as noisy labels. Deep-learning models have been shown to be robust to moderate levels of noisy labels.[25] Thus, this may help make the model more stable due to having more training data. In our work, SSL could significantly improve the model performance and explainability as suggested by the results. In particular, SSL helped to improve the precision on the AREDS2 dataset. While there was a drop in the precision on the DAAMD dataset, SSL helped to improve the recall and F1 score significantly.

The closest work to our study has been done by Schwartz et al.[9] They have attempted to detect and quantify RPD on OCT scans. Their detection model detected both RPD and conventional drusen as one class without distinguishing between them. Thus, the performance of RPD detection is masked because conventional drusen have more salient features that are easier to detect compared to those of RPD. Differentiating between RPD and conventional drusen is of critical importance because they carry different risk profiles for progression to late AMD. Moreover, Schwartz et al quantified RPD using the area of RPD on B-scans using 2D segmentation models trained on a small dataset of annotated B-scans. They reported low agreement in area measurement between the graders and the model as well as low intergrader agreement. They also reported a very low

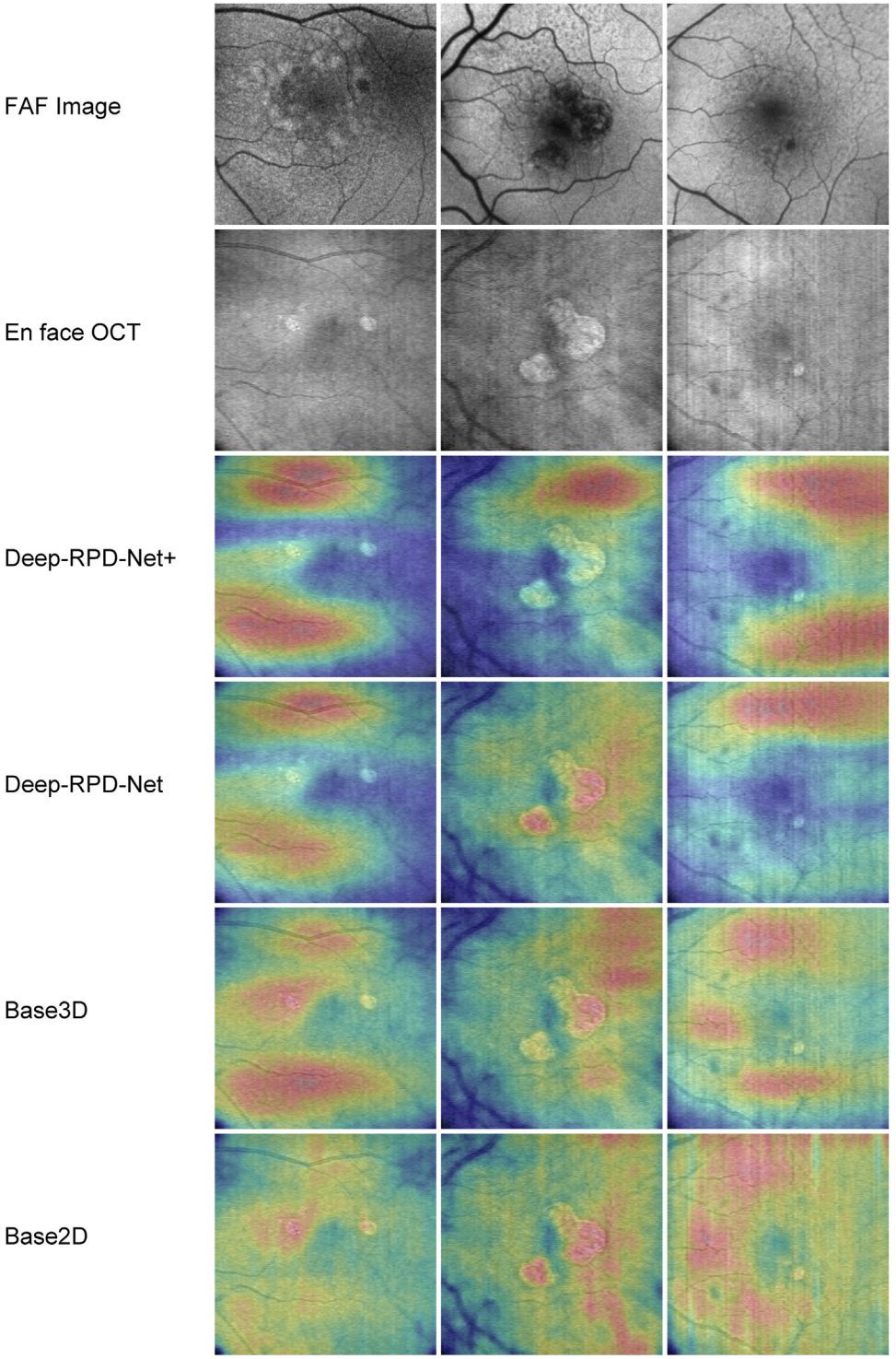

Fig. 6. Visualization of all networks on OCT scans labeled with RPD presence from the AREDS2 dataset. AREDS2 = Age-Related Eye Disease Study 2; FAF = fundus autofluorescence; RPD = reticular pseudodrusen.

Dice score for their segmentation model performance. Their results show that obtaining accurate RPD annotations is poorly reliable due to the low agreement between graders and difficult due to the low performance of segmentation model. In our work, we focused on RPD detection alone, i.e., using a single label per OCT scan, which is a novel approach; RPD quantification was outside the scope of this study.

The results of the subjective test demonstrate that grading RPD is a difficult task for retina specialists on the AREDS2

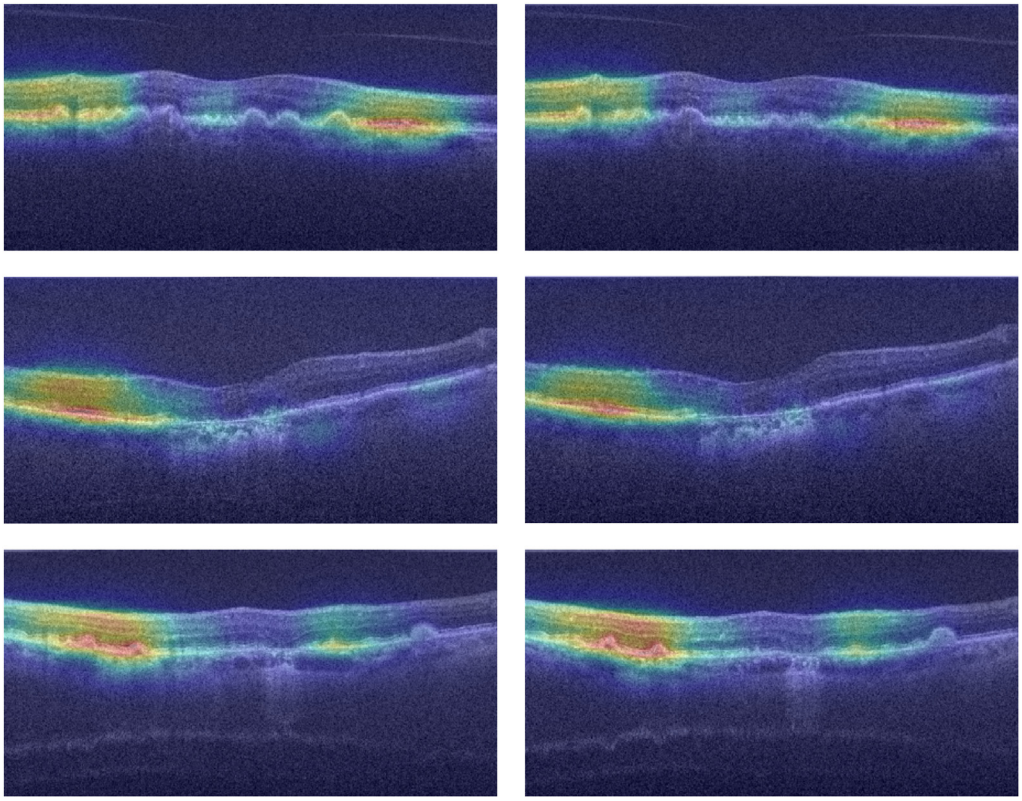

Fig. 8. Examples of the Grad-CAM heatmaps for Deep-RPD-Net+ on OCT B-scans from the AREDS2 dataset. Deep-RPD-Net+ tended to highlight areas with reticular pseudodrusen (RPD). It could also distinguish between RPD and other features, such as geographic atrophy and conventional drusen. AREDS2 = Age-Related Eye Disease Study 2; Grad-CAM = Gradient-weighted Class Activation Mapping.

dataset and quite less difficult on the DAAMD dataset due to the better OCT scan quality. Deep-RPD-Net outperformed the retina specialists on the AREDS2 dataset and outperformed one of them on the DAAMD dataset. The DAAMD dataset is highly imbalanced as compared with the AREDS2 dataset, so this could be a reason for the performance drop. Moreover, grading by humans is time-consuming and requires detailed examination of each of the ∼100 (or 121) B-scans in the full OCT scan. This makes the prospect of an automated algorithm for RPD detection highly attractive.

The visualization results suggest that Deep-RPD-Net is not only reliable in detecting RPD on OCT scans but also has a high level of explainability and interpretability in its decision-making.

The strengths of this work include: (1) using a large multicenter OCT dataset curated from the AREDS2, which makes the findings more representative and generalizable; (2) verifying the findings using another large OCT dataset from a different machine and curated from the study of dark adaptation in AMD by the National Eye Institute; (3) detecting RPD on OCT scans with very high performance metrics, comparable to those obtained on FAF images; (4) performing subjective testing to show the performance of the proposed network against that of retina specialists; and 5) generating heatmaps that can provide interpretable decisions.

The limitations of this work include: (1) detecting RPD presence/absence at the image level without segmentation or quantification, which is a very difficult task even for expert human graders; (2) using only OCT data for detection without using other modalities simultaneously (i.e., the aim of this work was to develop deep-learning models able to work using OCT data alone); and (3) limited generalizability of the models trained on one OCT machine to data from other OCT machines due to heterogeneity of the OCT scans (i.e., scan size, number of images per scan, image size, image quality, etc.), which is outside the scope of this study.

In conclusion, we propose Deep-RPD-Net, a 3D CNN for the detection of RPD on OCT scans. Deep-RPD-Net was able to detect RPD from OCT scans with high accuracy, compared with that of other networks. It outperformed 3 retina specialists on a subjective grading test. The visualizations of Deep-RPD-Net were more interpretable, compared with those of the other networks, and could highlight RPD and their associated features. The code and pretrained models will be publicly available at https://github.com/ncbi-nlp/Deep-RPD-Net for the transparency and reproducibility of this study and to provide a benchmark for further studies.

Footnotes and Disclosures

Originally received: April 22, 2024.
Final revision: September 12, 2024.
Accepted: November 12, 2024.
Available online: November 19, 2024. Manuscript no. XOPS-D-24-00122.

[1] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland.

[2] Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, Maryland.

Emily Y. Chew, the Editor-in-Chief of this journal, was recused from the peer-review process of this article and had no access to information regarding its peer-review.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The authors made the following disclosures:

Q.C.: Grants − National Library of Medicine, 1K99LM014024.

S.B.: Grants and consultant − Sanofi Pharmaceuticals (consultant for R&D, work not connected to submitted manuscript).

T.K.: Patents − Methods and Systems for Predicting Rates of Progression of Age-related Macular Degeneration (co-inventor on a patent application); Participation on a Data Safety Monitoring Board or Advisory Board − Voting member of the Safety Monitoring Committee for the APL2-103 interventional study (Apellis Pharmaceuticals), 2020−2021.

E.Y.C.: Participation on a Data Safety Monitoring Board or Advisory Board − NBM Bio DSMC (no travel and no funding for participation [by zoom]), Genentech Advisory Board (no funds involved), 4DMT (no travel and no funding for participation [by zoom]); Receipt of medical writing − A manuscript was prepared with a medical writer in 2018.

This work was supported by the Intramural Research Program of the National Library of Medicine and the National Eye Institute, National Institutes of Health.

HUMAN SUBJECTS: No human subjects were included in this study. The Ancillary OCT Study of AREDS2 enrolled AREDS2 participants from the Devers Eye Institute, Emory Eye Center, Duke Eye Center, and the National Eye Institute. The study was approved by the institutional review boards of the study sites, and written informed consent was obtained from all participants. It adhered to the tenets of the Declaration of Helsinki and complied with the Health Insurance Portability and Accountability Act. The study was registered at ClinicalTrials.gov (identifier NCT00734487). For the DAAMD dataset, the study was approved by the institutional review board of the National Institutes of Health and adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from all participants. It was registered at ClinicalTrials.gov (identifier NCT01352975).

No animal subjects were used in this study.

Author Contributions:

Conception and design: Elsawy, Keenan, Chew, Zhiyong Lu

Data collection: Elsawy, Keenan, Thavikulwat, Amy Lu, Bellur, Mukherjee, Agron, Chen, Chew, Zhiyong Lu

Analysis and interpretation: Elsawy, Keenan

Obtained funding: N/A

Overall responsibility: Elsawy, Keenan, Thavikulwat, Amy Lu, Bellur, Mukherjee, Agron, Chen, Chew, Zhiyong Lu

Abbreviations and Acronyms:
2D = 2-dimensional; **3D** = 3-dimensional; **AMD** = age-related macular degeneration; **AREDS2** = Age-Related Eye Disease Study 2; **AUROC** = area under the receiver-operating characteristic curve; **CFP** = color fundus photography; **CI** = confidence interval; **CNN** = convolutional neural network; **DAAMD** = Dark Adaptation in Age-related Macular Degeneration Study; **FAF** = fundus autofluorescence; **GA** = geographic atrophy; **Grad-CAM** = Gradient-weighted Class Activation Mapping; **RPD** = reticular pseudodrusen; **SSL** = semisupervised learning.

Keywords:
Reticular pseudodrusen, OCT, Deep learning, Detection, Age-related macular degeneration.

Correspondences:
Zhiyong Lu, PhD, National Center for Biotechnology Information (NCBI), National Libary of Medicine (NLM), BG 38A RM 10N1003A MSC 3825 8600 Rockville Pike, Bethesda, MD 20894-3825. E-mail: luzh@ncbi.nlm.nih.gov; and Emily Y. Chew, MD, Clinical Trials Branch, National Eye Institute (NEI), BG 10-CRC RM 3-2531 MSC 120410 Center Dr, Bethesda, MD 20892-1204. E-mail: echew@nei.nih.gov.

References

1. Zweifel SA, Spaide RF, Curcio CA, et al. Reticular pseudodrusen are subretinal drusenoid deposits. *Ophthalmology*. 2010;117:303−312.e1.
2. Wightman AJ, Guymer RH. Reticular pseudodrusen: current understanding. *Clin Exp Optom*. 2019;102:455−462.
3. Spaide RF, Ooto S, Curcio CA. Subretinal drusenoid deposits AKA pseudodrusen. *Surv Ophthalmol*. 2018;63:782−815.
4. Domalpally A, Agrón E, Pak JW, et al. Prevalence, risk, and genetic association of reticular pseudodrusen in age-related macular degeneration: age-related eye disease study 2 report 21. *Ophthalmology*. 2019;126:1659−1666.
5. de Boer JF, Cense B, Park BH, et al. Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Opt Lett*. 2003;28:2067−2069.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436−444.
7. Keenan TD, Chen Q, Peng Y, et al. Deep learning automated detection of reticular pseudodrusen from fundus autofluorescence images or color fundus photographs in AREDS2. *Ophthalmology*. 2020;127:1674−1687.
8. Chen Q, Keenan TD, Allot A, et al. Multimodal, multitask, multiattention (M3) deep learning detection of reticular pseudodrusen: toward automated and accessible classification of age-related macular degeneration. *J Am Med Inf Assoc*. 2021;28:1135−1148.
9. Schwartz R, Khalid H, Liakopoulos S, et al. A deep learning framework for the detection and quantification of reticular pseudodrusen and drusen on optical coherence tomography. *Trans Vis Sci Technol*. 2022;11:3.
10. Elsawy A, Kenan TD, Chen Q, et al. *Attention-based 3D convolutional networks for detection of geographic atrophy from optical coherence tomography scans*. In: *Medical Imaging 2023: Image Processing*. 12464. SPIE; 2023:764−768.
11. Elsawy A, Keenan TD, Chen Q, et al. Deep-GA-net for accurate and explainable detection of geographic atrophy on OCT scans. *Ophthalmol Sci*. 2023;3:100311.

12. Selvaraju RR, Cogswell M, Das A, et al. *Grad-cam: Visual explanations from deep networks via gradient-based localization*. *Proceedings of the IEEE international conference on computer vision*. 2017:618−626.
13. Shi X, Keenan TDL, Chen Q, et al. Improving interpretability in machine diagnosis: detection of geographic atrophy in OCT scans. *Ophthalmol Sci*. 2021;1:100038.
14. Arslan J, Samarasinghe G, Benke KK, et al. Artificial intelligence algorithms for analysis of geographic atrophy: a review and evaluation. *Trans Vis Sci Technol*. 2020;9:57.
15. Christenbury JG, Folgar FA, O'Connell RV, et al. Progression of intermediate age-related macular degeneration with proliferation and inner retinal migration of hyperreflective foci. *Ophthalmology*. 2013;120:1038−1045.
16. Leuschen JN, Schuman SG, Winter KP, et al. Spectral-domain optical coherence tomography characteristics of intermediate age-related macular degeneration. *Ophthalmology*. 2013;120:140−150.
17. Flynn OJ, Cukras CA, Jeffrey BG. Characterization of rod function phenotypes across a range of age-related macular degeneration severities and subretinal drusenoid deposits. *Invest Ophthalmol Vis Sci*. 2018;59:2411−2421.
18. Flamendorf J, Agrón E, Wong WT, et al. Impairments in Dark adaptation are associated with age-related macular degeneration severity and reticular pseudodrusen. *Ophthalmology*. 2015;122:2053−2062.
19. Chen KG, Alvarez JA, Yazdanie M, et al. Longitudinal study of Dark adaptation as a functional Outcome measure for age-related macular degeneration. *Ophthalmology*. 2019;126:856−865.
20. Davis MD, Gangnon RE, Lee LY, et al. The Age-Related Eye Disease Study severity scale for age-related macular degeneration: AREDS Report No. 17. *Arch Ophthalmol*. 2005;123:1484−1498.
21. Kingma DP, JJapa Ba. *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980. 2014.
22. Zhou B, Khosla A, Lapedriza A, et al. *Learning deep features for discriminative localization*. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:2921−2929.
23. Chiang TT, Keenan TD, Agrón E, et al. Macular thickness in intermediate age-related macular degeneration is influenced by disease severity and subretinal drusenoid deposit presence. *Invest Ophthalmol Vis Sci*. 2020;61:59.
24. Duic C, Mukherjee S, Pfau K, et al. Local and global associations of reticular pseudodrusen in age-related macular degeneration. *Ophthalmol Retina*. 2024;8:646−656.
25. Rolnick D, Veit A, Belongie S, Shavit N. Deep learning is robust to massive label noise. *arXiv*. 2017. https://doi.org/10.48550/arXiv.1705.10694.