

A Diabetic Retinopathy Classification Framework Based on Deep-Learning Analysis of OCT Angiography

Pengxiao Zang^{1,2}, Tristan T. Hormel¹, Xiaogang Wang³, Kotaro Tsuboi^{1,4}, David Huang¹, Thomas S. Hwang¹, and Yali Jia^{1,2}

¹ Casey Eye Institute, Oregon Health & Science University, Portland, OR, USA

² Department of Biomedical Engineering, Oregon Health & Science University, Portland, OR, USA

³ Shanxi Eye Hospital, Taiyuan, Shanxi, China

⁴ Department of Ophthalmology, Aichi Medical University, Nagakute, Japan

Correspondence: Yali Jia, Casey Eye Institute & Department of Biomedical Engineering, Oregon Health & Science University, 515 SW Campus Dr., Portland, OR 97239, USA.

e-mail: jiaya@ohsu.edu

Received: March 27, 2022

Accepted: June 20, 2022

Published: July 13, 2022

Keywords: diabetic retinopathy (DR); optical coherence tomography (OCT); deep learning; classification

Citation: Zang P, Hormel TT, Wang X, Tsuboi K, Huang D, Hwang TS, Jia Y. A diabetic retinopathy classification framework based on deep-learning analysis of OCT angiography. *Transl Vis Sci Technol.* 2022;11(7):10, <https://doi.org/10.1167/tvst.11.7.10>

Purpose: Reliable classification of referable and vision threatening diabetic retinopathy (DR) is essential for patients with diabetes to prevent blindness. Optical coherence tomography (OCT) and its angiography (OCTA) have several advantages over fundus photographs. We evaluated a deep-learning-aided DR classification framework using volumetric OCT and OCTA.

Methods: Four hundred fifty-six OCT and OCTA volumes were scanned from eyes of 50 healthy participants and 305 patients with diabetes. Retina specialists labeled the eyes as non-referable (nrDR), referable (rDR), or vision threatening DR (vtDR). Each eye underwent a 3 × 3-mm scan using a commercial 70 kHz spectral-domain OCT system. We developed a DR classification framework and trained it using volumetric OCT and OCTA to classify eyes into rDR and vtDR. For the scans identified as rDR or vtDR, 3D class activation maps were generated to highlight the subregions which were considered important by the framework for DR classification.

Results: For rDR classification, the framework achieved a 0.96 ± 0.01 area under the receiver operating characteristic curve (AUC) and 0.83 ± 0.04 quadratic-weighted kappa. For vtDR classification, the framework achieved a 0.92 ± 0.02 AUC and 0.73 ± 0.04 quadratic-weighted kappa. In addition, the multiple DR classification (non-rDR, rDR but non-vtDR, or vtDR) achieved a 0.83 ± 0.03 quadratic-weighted kappa.

Conclusions: A deep learning framework only based on OCT and OCTA can provide specialist-level DR classification using only a single imaging modality.

Translational Relevance: The proposed framework can be used to develop clinically valuable automated DR diagnosis system because of the specialist-level performance showed in this study.

Introduction

Diabetic retinopathy (DR) is a leading cause of preventable blindness globally.¹ Currently, DR classification uses fundus photographs or clinical examination to identify referable DR (rDR) and vision-threatening DR (vtDR). Eyes with worse than mild nonproliferative DR (NPDR) on the International Diabetic Retinopathy Severity Scale are considered rDR, and eyes with severe NPDR, proliferative DR

(PDR), or those with diabetic macular edema (DME) are considered vtDR.² An efficient and reliable classification system is essential in identifying patients who can benefit from treatment without an undue burden to the clinic. Eyes with rDR but without vtDR can be observed closely without referral to an ophthalmologist, helping preserve scarce resources for patients that require treatment. To do this safely requires an accurate stratification of patients into these categories.^{3,4}

Deep learning has enabled multiple reliable automated systems that classify DR from fundus

photographs.^{5–8} However, fundus photographs have a low sensitivity (60–73%) and specificity (67–79%) for detecting diabetic macular edema (DME), which accounts for the majority of vision loss in DR.^{9,10} This means that even when a network performs very well against a ground truth generated from fundus photographs, patients with DME may still frequently be misdiagnosed. Supplementing fundus photography with optical coherence tomography (OCT), which is the current gold standard for diagnosing macular edema, can avoid this problem.^{11–20} However, reliance on multiple imaging modalities is undesirable as it increases logistic challenges and cost.

Our group and others have demonstrated that OCT angiography (OCTA) can be used to stage DR according to fundus photography-derived DR severity scales using various biomarkers linked to capillary changes in DR.^{21–28} Because OCTA can be simultaneously acquired with structural OCT scans used for DME diagnosis, an automated system based on OCTA volume scans can potentially use a single imaging modality to accurately classify DR while avoiding low DME detection sensitivities and associated misdiagnoses that occur in systems based on just fundus photographs.

Despite this advantage, OCTA-based analyses require improvements. Previous methods for classifying DR using OCTA relied on accurate retinal layer segmentation and en face visualization of the 3D volume to visualize or measure biomarkers.^{29–34} However, with advanced pathology, retinal layer segmentation can become unreliable. This lowers OCTA yield rate, and may also lead to misclassification through segmentation errors. In addition, quantifying only specific biomarkers fails to make use of the information in the latent feature space of the OCT/OCTA volumes, which may be helpful for DR classification.³⁵

In this study, we propose an automated convolutional neural network (CNN)³⁶ that uses the volumetric OCT/OCTA to directly classify eyes as either non-rDR (nrDR) or rDR, and as either vtDR or eyes with referable but not vision-threatening DR (nvtDR). We also include a multiclass classification that classifies eyes as nrDR, rDR/nvtDR, or vtDR. To demonstrate which features the framework relies on to make the classification, the network also generates 3D class activation maps (CAMs).³⁷ Visualizations such as these are essential features of direct classification systems, because they allow graders to verify algorithm outputs. To the best of our knowledge, this is the first study to propose an automated multiclass DR severity-level classification framework based directly on OCT and OCTA volumes.

Methods

Data Acquisition

We recruited and examined 50 healthy participants and 305 patients with diabetes at the Casey Eye Institute, Oregon Health & Science University in the United States (50 healthy participants and 234 patients); Shanxi Eye Hospital in China (60 patients); and the Department of Ophthalmology, Aichi Medical University in Japan (11 patients). We included patients with diabetes with the full spectrum of disease from no clinically evident retinopathy to PDR. One or both eyes of each participant underwent 7-field color fundus photography and an OCTA scan using a commercial 70-kHz spectral-domain OCT (SD-OCT) system (RTVue-XR Avanti; Optovue Inc.) with 840-nm central wavelength. The scan depth was 1.6 mm in a 3.0 × 3.0 mm region (640 × 304 × 304 pixels) centered on the fovea. Two repeated B-frames were captured at each line-scan location. The structural images were obtained by averaging the two repeated and registered B-frames. Blood flow was detected using the split-spectrum amplitude-decorrelation angiography (SSADA) algorithm.^{23,38} For each volumetric OCT/OCTA, two continuously acquired volumetric raster scans (one x-fast scan and one y-fast scan) were registered and merged through an orthogonal registration algorithm to reduce motion artifacts.³⁹ In addition, the projection-resolved OCTA algorithm was applied to all OCTA scans to remove flow projection artifacts in the deeper layers.^{40,41} Scans with a signal strength index (SSI) lower than 50 were excluded. The data characteristics are shown below (Table 1). When the classes in our data set were not balanced, class weights were adjusted to prevent performance loss. Based on the data distribution showed in Table 1, the class weights for nrDR, r/nvtDR and vtDR were 0.76, 1.87, and 0.87, respectively.

A masked trained retina specialist (author T.S.H.) graded seven-field color fundus photographs based on Early Treatment of Diabetic Retinopathy Study (ETDRS) scale.^{42,43} The presence of DME was determined using the central subfield thickness from structural OCT based on the Diabetic Retinopathy Clinical Research Network (DRCR.net) standard.³ We defined nrDR as ETDRS level better than 35 and without DME (we also included healthy eyes); referable DR as ETDRS level 35 or worse, or any DR with DME; r/nvtDR as ETDRS levels 35 to 47 without DME; and vtDR as ETDRS level 53 or worse, or any stage of DR with DME.² The participants were enrolled after signing an informed consent in accordance with

Table 1. Data for DR Classification

Characteristics	rDR Classification		vtDR Classification		Multiclass DR Classification		
Severity	nrDR	rDR	nvtDR	vtDR	nrDR	r/nvtDR	vtDR
Number of eyes/scans	199	257	280	176	199	81	176
Age, mean (SD), y	48.8 (14.6)	58.4 (12.1)	52.2 (14.7)	57.5 (12.3)	48.8 (14.6)	60.4 (14.7)	57.5 (12.3)
Female, %	50.8	49.0	50.0	49.4	50.8	48.2	49.4
No DR, %	83.4	1.6	59.3	2.2	83.4	0.0	2.2
Mild NPDR, %	16.6	0.0	11.8	0.0	16.6	0.0	0.0
Moderate NPDR, %	0.0	44.4	28.9	18.8	0.0	100.0	18.8
Severe NPDR, %	0.0	19.8	0.0	29.0	0.0	0.0	29.0
PDR, %	0.0	34.2	0.0	50.0	0.0	0.0	50.0
DME, %	0.0	32.3	0.0	47.2	0.0	0.0	47.2

DR, diabetic retinopathy; rDR, referable diabetic retinopathy; vtDR, = vision threatening diabetic retinopathy; r/nvtDR, referable but not vision threatening diabetic retinopathy; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy; DME, diabetic macular edema.

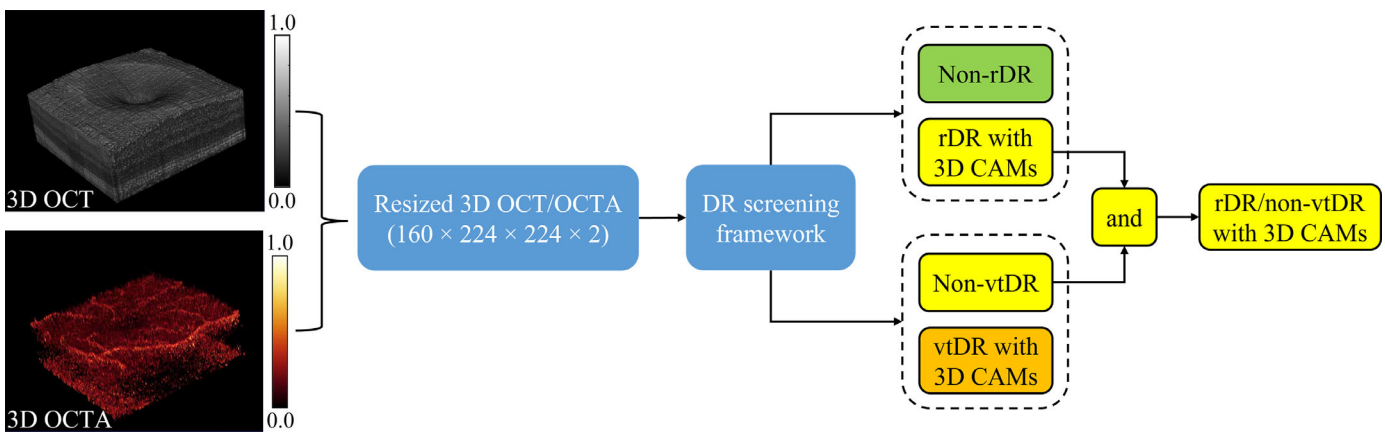


Figure 1. Automated DR classification framework using volumetric OCT and OCTA data as inputs. Inputs are first resized to $160 \times 224 \times 224 \times 2$ pixels (two channels 3D input with a $160 \times 224 \times 224$ structural and a $160 \times 224 \times 224$ angiographic volume). These inputs are fed into a DR screening framework based on a 3D CNN architecture. The network produces two outputs: a non-referable DR (nrDR) or referable DR (rDR) classification, and a non-vision-threatening (nvtDR) or vision threatening (vtDR) DR classification. The multiclass DR classification result is defined based on the rDR and vtDR classification results. Class activation maps (CAMs) are also output for each classification result.

an Institutional Review Board approved protocol. The study complied with the Declaration of Helsinki and the Health Insurance Portability and Accountability Act.

Data Inputs

OCT and OCTA generate detailed depth-resolved structural and microvascular information from the fundus (Fig. 1). Extracting DR-related features using neural networks can, however, be more challenging and time-consuming from 3D volumes such as those produced by OCTA than from 2D sources like fundus photography. To improve the computational and space efficiency of the framework, each volumetric OCT and OCTA were resized to $160 \times 224 \times 224$ voxels and

normalized to voxel values between 0 and 1. The input was the combination of each pair of resized volumes, giving final input dimensions of $160 \times 224 \times 224 \times 2$ pixels (see Fig. 1).

DR Classification Framework

A novel 3D CNN architecture (see Fig. 1) with 16 convolutional layers was designed and used as the core classifier in the DR classification framework (Supplementary Fig. S1). Five convolutional layers with stride two were used to downsample the input data. To avoid losing small but important DR-related features, diminishing convolutional kernel sizes were used in the five downsampling layers. We used batch normalization⁴⁴ after each 3D convolutional layer to increase

convergence speed. In order to improve the computational efficiency while ensuring the resolution of the features, most of the 3D convolutional layers were used with the middle size inputs (after the first downsampling, but before the last). A global average pooling layer was used after the last 3D convolutional layer to generate the 1D input for the output layers.

One subtlety in our approach for multiclass classification is the need to correctly identify rDR/nvDR eyes. Familiar frameworks for image classification like those used to diagnose medical conditions rely on the positive identification features associated with the malady. In our framework, rDR and vtDR classification works similarly by using rectified linear unit (ReLU) activations in the last convolutional layer and weight parameters of all the fully connected layers to guarantee positive-definite prediction values (Supplementary Fig. S2).^{45,46} However, the identification of r/nvDR does not depend on just the presence of rDR associated features, but also the absence of vtDR-associated features. To solve this issue, two parallel output layers were respectively used to detect rDR and vtDR at the same time (see Fig. 1). Each output layer was constructed by a fully connected layer with a softmax function (see Supplementary Fig. S2). The inputs data can be then classified as nrDR, r/nvDR, or vtDR based on rDR and vtDR classification outputs.

Evaluation and Statistical Analysis

Overall accuracy, quadratic-weighted Cohen's kappa,⁴⁷ and area under the receiver operating characteristic curve (AUC) were used to evaluate the DR classification performance of our framework. Among these evaluation metrics, the AUCs were used as the primary metrics for rDR and vtDR classifications. For the multiclass DR classification, the quadratic-weighted kappa was used as the primary metric. Five-fold cross-validation was used in each case to explore robustness. From the whole data set, 60%, 20%, and 20% of the data were split for training, validation, and testing, respectively. Care was taken to ensure data from the same patients were only included in one of the training, validation, or testing data sets. The parameters and hyperparameters in our framework were trained and optimized only using the training and validation data set. In addition, adaptive label smoothing was used during training to reduce the overfitting.³⁴

Comparison With a 2D Input Approach

In contrast to the method proposed in this work, most OCT/OCTA-based DR classification algorithms

operate on 2D en face images.^{32–34,48} En face projections are popular input choices because (1) they correspond to the data representation most familiar to graders and (2) they typically reduce the size of the input data set relative to the full OCT/OCTA data volume by more than an order of magnitude, which simplifies network training. The trade-off with this data reduction is that networks analyzing en face images cannot learn all of the features latent in the full data volume, because many of these features will be removed during the step of slab projection for generating 2D maps. Furthermore, en face images are vulnerable to segmentation artifacts, which require a time-consuming review to correct.⁴⁹ For these reasons, models capable of analyzing OCT/OCTA volumes are desirable, but to be useful such models should reach performance parity with approaches using 2D inputs. To investigate, we compared our model with our previous 2D approach, which was a CNN designed around dense and continuous connection with adaptive rate dropout (*DcardNet*).³⁴ This 2D model was trained, validated, and evaluated based on the same data sets of our 3D model.

3D Class Activation Maps and Evaluation

For the detected rDR and vtDR cases, the 3D CAMs were generated by projecting the weight parameters from corresponding output layer back to the feature maps of the last 3D convolutional layer before global average pooling (see Supplementary Fig. S2). To assess whether or not the framework can correctly identify pathological regions, 3D CAMs were overlaid on en face or cross-sectional OCT and OCTA images. In order to generate the en face projections, an automated algorithm (commercial software provided by Optovue Inc.) segmented the following retinal layers (Supplementary Fig. S3): inner limiting membrane (ILM), nerve fiber layer (NFL), ganglion cell layer (GCL), inner plexiform layer (IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), ellipsoid zone (EZ), retinal pigment epithelium (RPE), and Bruch's membrane (BM). For the cases with severe pathologies, trained graders manually corrected the layer segmentation when necessary, using our custom designed COOL-ART grading software.⁵⁰ From OCT volumes, we generated the inner retinal (the slab between the Vitreous/ILM and OPL/ONL) thickness map, en face mean projection of OCT reflectance, and EZ en face mean projection (ONL/EZ to EZ/RPE). From OCTA volumes, we generated the superficial vascular complex (SVC), intermediate capillary plexus (ICP), and deep

capillary plexus (DCP) angiograms.^{26,51,52} The SVC was defined as the inner 80% of the ganglion cell complex (GCC), which included all structures between the ILM and IPL/INL border. The ICP was defined as the outer 20% of the GCC and the inner 50% of the INL. The DCP was defined as the remaining slab internal to the outer boundary of the OPL. The segmentation step and projection maps were just for evaluating the usefulness of 3D CAMs, not as input to the classification framework.

Results

Model performance was the best for rDR classification, followed by vtDR, then multi-level DR classification (Table 2, Fig. 2). For the multiclass DR classification, which classifies each case as nrDR, r/nvtDR, or

vtDR, we achieved a quadratic-weighted kappa 0.83, which is on par with the performance of ophthalmologists and retinal specialists (0.80 to 0.91).⁵³ The network was notably better at classifying rDR and vtDR compared to r/nvtDR (see Table 2). Most false positive r/nvtDR eyes were classified as vtDR (66.67%) instead of nrDR (33.33%).

The 3D model performed slightly better for rDR classification, and slightly worse for vtDR classification than our previous 2D model (Table 3). These mixed results indicate that our current model using volumetric data as input was able to train successfully enough to achieve parity with a 2D-image-based approach.

To demonstrate the deep-learning performance more explicitly, we compared the stratified ground truth with the network prediction with confusion matrices using the overall values from 5-fold cross-validation (Fig. 3). In the three confusion

Table 2. Automated DR Classification Performances

Metric	rDR Classification	vtDR Classification	Multiclass DR Classification
Overall accuracy	91.52% \pm 1.87%	87.39% \pm 2.02%	81.52% \pm 1.19%
Sensitivity	90.77% \pm 4.28%	82.22% \pm 2.83%	
Specificity	92.50% \pm 3.16%	90.71% \pm 3.46%	
AUC (mean \pm SD)	0.96 \pm 0.01	0.92 \pm 0.02	
Quadratic-weighted kappa	0.83 \pm 0.04	0.73 \pm 0.04	0.83 \pm 0.03

DR, diabetic retinopathy; rDR, referable diabetic retinopathy; vtDR, vision threatening diabetic retinopathy; AUC, area under the receiver operating characteristic curve.

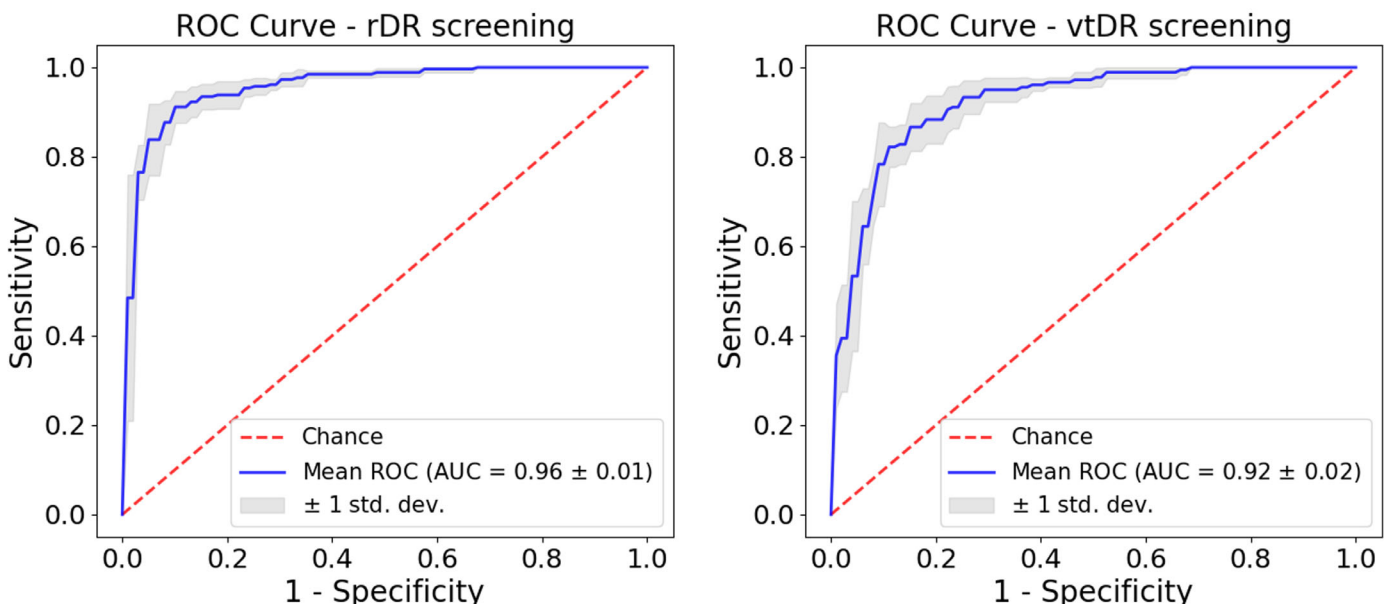


Figure 2. The mean receiver operating characteristic (ROC) curve derived from the five-fold cross-validation for rDR (left) and vtDR (right) classifications based on our DR classification framework. The models achieve an AUC of 0.96 ± 0.01 on rDR classification and AUC of 0.92 ± 0.02 on vtDR classification.

Table 3. Comparison Between Our 3D Model and Previous 2D Model

Models	rDR Classification		vtDR Classification	
	Overall Accuracy	AUC	Overall Accuracy	AUC
2D model	89.67% \pm 2.50%	0.95 \pm 0.02	88.99% \pm 0.84%	0.94 \pm 0.02
3D model	91.52% \pm 1.87%	0.96 \pm 0.01	87.39% \pm 2.02%	0.92 \pm 0.02

Stratified Ground Truth	rDR screening		vtDR screening		Multiple DR screening		
	nrDR						
	r/nvtDR						
	vtDR/nDME						
	vtDR/DME						
	nrDR	rDR	nvtDR	vtDR	nrDR	r/nvtDR	vtDR
	Predicted		Predicted		Predicted		
	185	15	198	2	185	13	2
	12	68	56	24	12	44	24
	8	87	24	71	8	16	71
	4	81	8	77	4	6	77

Figure 3. Three confusion matrices for referable DR (rDR) classification, vision threatening DR (vtDR) classification, and multiclass DR classification based on the overall five-fold cross-validation results. The vtDR was split as non-DME (nDME) and DME in the matrices. The correctly and incorrectly classified cases are shaded blue and orange, respectively.

matrices, the vtDR cases were separated into non-DME (nDME) and DME to investigate whether the presence of DME can affect rDR and vtDR classification accuracy. In the rDR classification task, we found the classification accuracies of vtDR/nDME and vtDR/DME to be similar (87/95 and 81/85). For vtDR classification, the network identified cases with DME (77/85) with a greater accuracy than nDME cases (71/95), which may imply DME features were likely influential for decision making. In the multi-level classification, the network misclassified 16 of 95 vtDR/nDME cases as r/nvtDR. In addition, most of the r/nvtDR cases with false-positive results were classified as vtDR. Only two nrDR cases were misidentified as vtDR.

To better understand network decision making, we produced CAMs for some example cases. The CAM output of an r/nvtDR case points to dilated vessels in the DCP and a perifoveal area of decreased vessel

density (Fig. 4). Meanwhile, in a vtDR case without DME, the CAMs have a larger area of high attention (Fig. 5), indicating that the DR pathology is more pervasive throughout the volume. In addition to pointing to areas of decreased vessel density, the CAM overlaid on a structural OCT B-scan points to an area with abnormal curvature of the retinal layers. Finally, for a vtDR case with DME, the CAM pointed to areas with intraretinal cysts and abnormal curvature of the retinal layers on structural OCT, as well as decreased vessel density and abnormally dilated vessels on OCTA (Fig. 6). This is clearly an improvement over our previous 2D CAM output (Fig. 7),³⁴ which identified changes in the perifoveal region, but missed other pathologies, such as intraretinal cysts and abnormally dilated vessels. Based on the distribution of the highlighted regions from all the 3D CAMs, we found the non-perfusion areas near fovea and most fluids were preferentially selected by our framework for

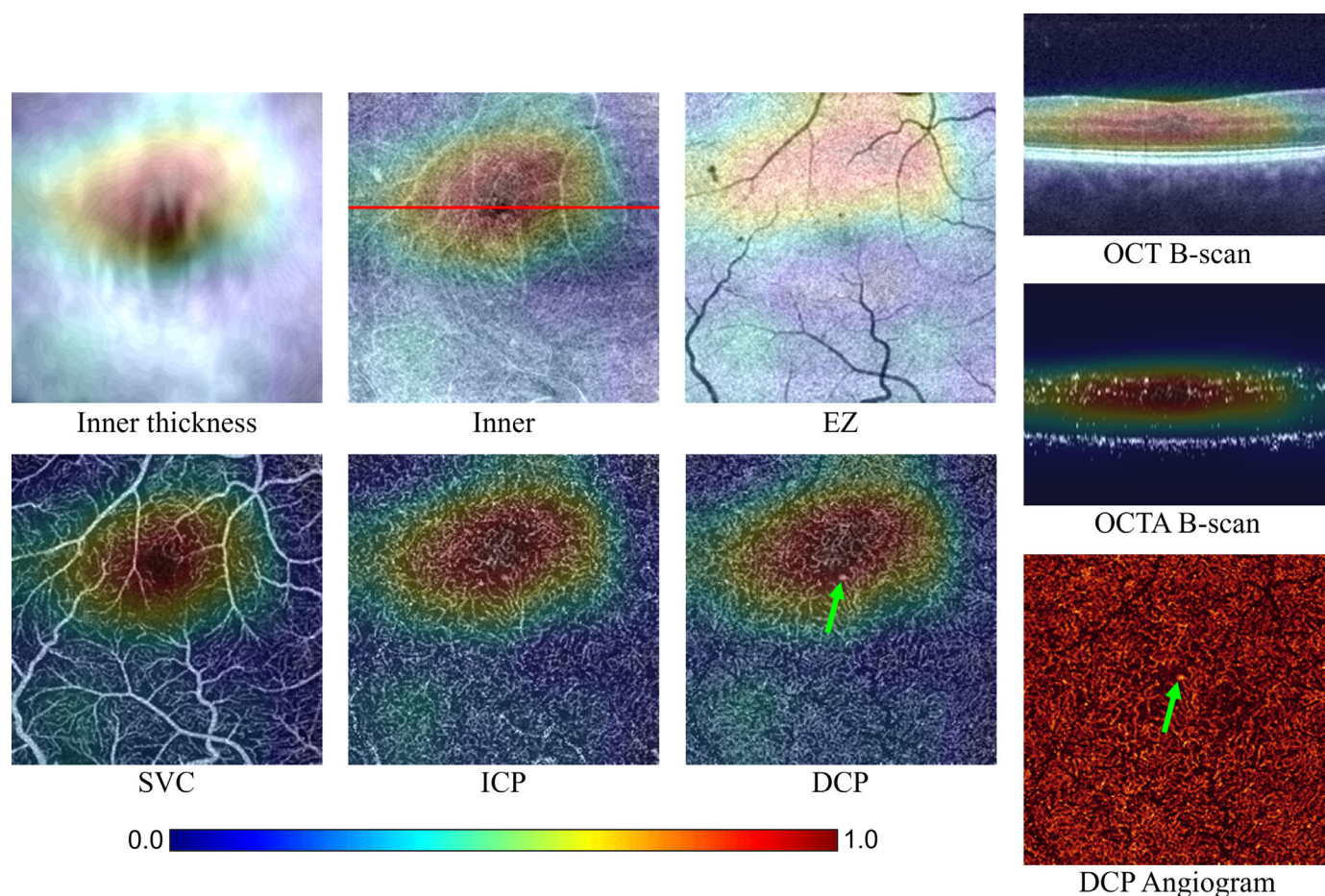


Figure 4. Class activation maps (CAMs) based on the referable DR (rDR) output layer of our framework for data from an eye with rDR without vision threatening DR (vtDR). Six en face projections covered with the corresponding projections of the 3D CAMs are shown. Extracted CAMs for an OCT and OCTA B-scans (red line in the inner retina en face projection) are also shown. The deep capillary plexus (DCP) angiogram without a CAM is shown so that the pathology highlighted by the corresponding CAM can be more easily identified. The green arrows indicate an abnormal vessel in the DCP. For descriptions of the regions projected over to produce the en face images, see the caption for Supplementary Figure S3.

decision making. In addition, the nonperfusion areas at the boundary of the inputs were barely selected by our framework.

Discussion

In this study, we proposed a CNN-based automated DR classification framework that operates directly on volumetric OCT/OCTA data without requiring retinal layer segmentation. This framework classified cases into clinically actionable categories (nrDR, r/nvtDR, and vtDR) using a single imaging modality. For multi-class DR classification, the framework achieved a quadratic-weighted kappa of 0.83 ± 0.03 , which is on par with the performance of human ophthalmologists and retinal specialists (0.80 to 0.91).⁵³ The network also

demonstrated robust performance on both rDR and vtDR classification ($AUC = 0.96 \pm 0.01$ and 0.92 ± 0.02 , respectively).

The framework used feature-rich structural OCT and OCTA volumes as inputs and a deep-learning model as the core classifier to achieve a high level of performance. The majority of DR classification algorithms to date have been based on fundus photographs.^{5–8} However, fundus photographs detect DME with only about a 70% accuracy relative to structural OCT, whereas DME accounts for the majority of vision loss in DR.^{9,10} Our method, on the other hand, actually performs better in the presence of DME (see Fig. 3).

Our image labels appealed to structural OCT to detect DME, and so did not adhere exactly to the ETDRS scale (the current gold standard for DR grading), which uses only seven field fundus

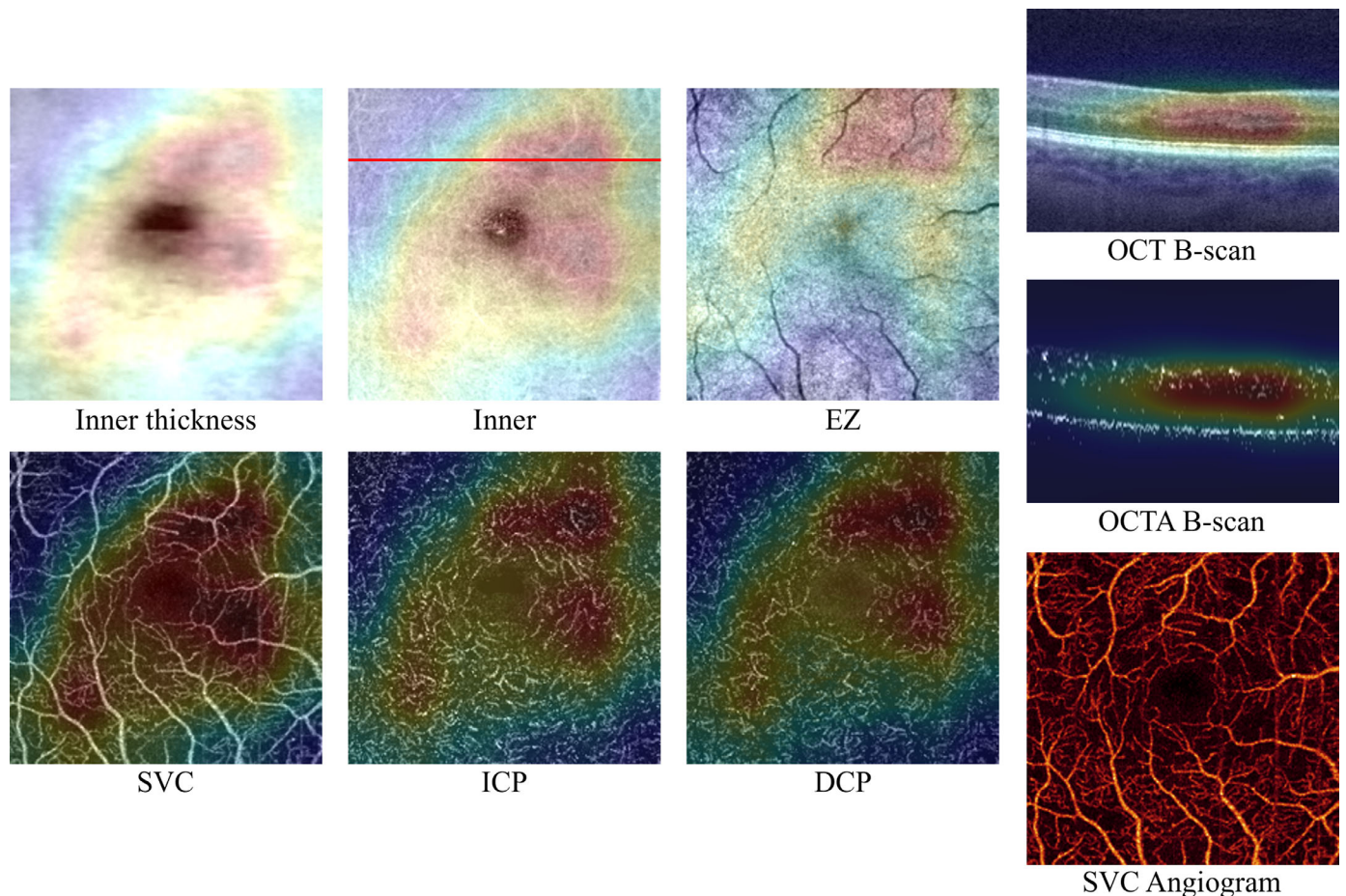


Figure 5. Class activation maps (CAMs) based on the vision threatening DR (vtDR) output layer of our framework for data from an eye with vtDR but without DME. Six en face projections covered with corresponding projections of the 3D CAMs are shown. Extracted CAMs for an OCT and OCTA B-scan (red line in the inner retina en face projection) are also shown. An SVC angiogram without a CAM is also shown to help identify pathological features for DR diagnosis. The SVC CAM indicates that the framework learned to identify non-perfusion areas, which are known biomarkers for DR diagnosis. For descriptions of the regions projected over to produce the en face images, see the caption for Supplementary Figure S3.

photographs. This prevented our model from learning to misdiagnose eyes based on the presence of DME not detected by fundus photography. However, at the same time, OCTA may not recapitulate every feature in fundus photography used for staging DR on the ETDRS scale. For example, OCTA does not detect intraretinal hemorrhages and may not detect all microaneurysms.²³ Achieving comparable performance to fundus photograph-based automated classification frameworks indicates that these disadvantages were surmounted by our approach.

Another important feature in our framework design is the use of a deep-learning model for the classifier. Compared to previously published OCT/OCTA-based DR classification algorithms, the proposed framework has several innovations. One advantage is the use of the volumetric OCT/OCTA, instead of preselected features from segmented en face images. This means

that correlations or structures within the data volume that may be difficult for a human to identify can still be incorporated into the decision making in our framework. Two dimensional approaches may miss important features without access to cross-sectional information, as happens with color fundus photography and DME.³³ As a corollary, our framework may then also have a greater capacity to improve with more training data because no data are removed by projection. Moreover, accurate retinal layer segmentation is required to generate the en face images. In severely diseased eyes, automated layer segmentations often fail. Mis-segmented layers can introduce artifacts into en face images unless they are manually corrected, a labor-intensive task that may not be clinically practical. By using volumetric data, our framework avoids this issue entirely. Another advantage built into our framework is the ability to detect both rDR and vtDR. This

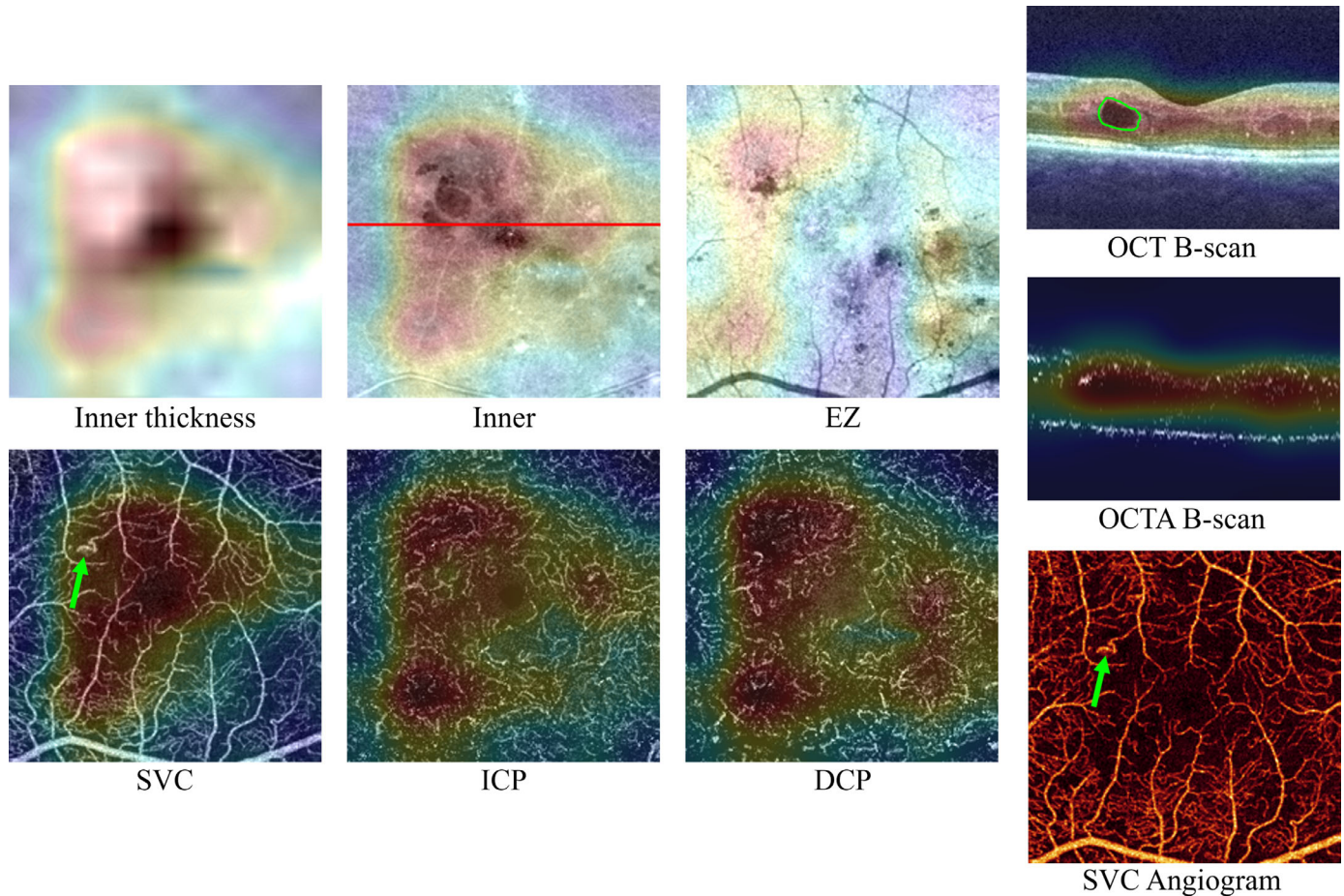


Figure 6. Class activation maps (CAMs) based on vision threatening DR (vtDR) output layer of our framework for data from an eye with vtDR and DME. Six en face projections covered with the corresponding projections of 3D CAMs are shown. Extracted CAMs for an OCT and OCTA B-scan (red line in the inner retina en face projection) are also shown. The SVC angiogram without a CAM is shown to more readily observe pathology. The green arrow in the SVC CAM shows an abnormal vessel, which can also be seen in the angiogram. Central macular fluid is marked by green circle on the OCT B-scan. The CAM allocated high weights to both of these regions. For descriptions of the regions projected over to produce the en face images, see the caption for Supplementary Figure S3.

higher level of granularity makes a more efficient use of resources possible compared to solutions that only identify rDR.^{5,7,29–33}

A final significant advantage in our framework is the inclusion of 3D CAMs. While independent of model performance, generating CAMs allow clinicians to interpret the classification results and ensure model outputs are correct. This is important because, outside of visualizations such as CAMs, users cannot in general ascertain how deep learning algorithms arrive at a classification decision. However, in medical imaging, it is essential to be able to verify and understand these classification decisions because doing so could prevent misdiagnosis. Black-box algorithms, such as deep learning algorithms, may hide important biases that could prove to be disadvantageous for certain groups. This risk can be lowered when the results are interpretable. With our framework this is possi-

ble. The CAMs in this work were generated volumetrically. Compared to 2D CAMs, the current framework using 3D OCT/OCTA as inputs can identify and learn relevant features (see Fig. 6, Fig. 7). The resulting CAMs consistently highlighted macular fluid (see Fig. 6), demonstrating that the model did indeed learn relevant features because central macular fluid is the most important biomarker for detecting DME.³⁵ We also found our 3D CAMs pointed to other key features, such as lower vessel density and dilated capillaries (see Fig. 4, Fig. 5). Although the 3D CAM did not identify all DR features (e.g. certain regions with lower vessel density were ignored), it found many key features, indicating that our framework has successfully learned relevant features and that 3D CAMs could be useful in clinical review. In addition, the purpose of generating 3D CAMs is not necessarily to find all DR biomarkers, but simply to highlight

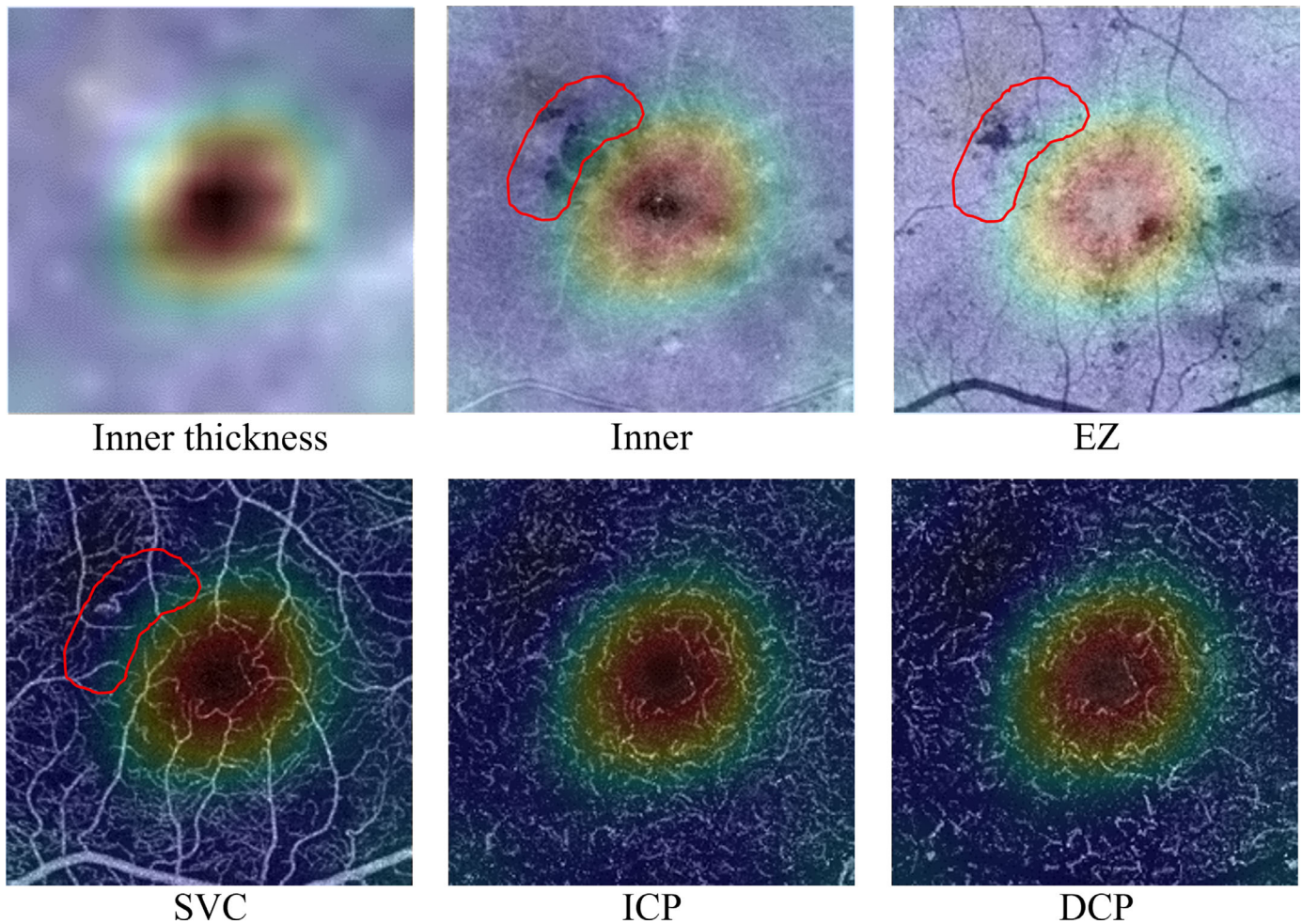


Figure 7. Two-dimensional class activation maps (CAMs) generated by our previous study for data from an eye with vtDR and DME. Six en face projections (see Supplementary Fig. S3 for details) covered with the same 2D CAMs are shown. The abnormal vessels and central macular fluid, which were highlighted regions in the 3D CAMs in Figure 6, were not weighted highly by the 2D CAM algorithm (red circles in the inner, EZ, and SVC CAMs).

the features used by the network to make decisions. That the network ignored some known DR-associated features is interesting, because it implies that these features were not critical for diagnosing DR at a given severity.

There are aspects of our framework that could be improved in future work. The sensitivity for r/nvtDR classification ($55.00\% \pm 15.51\%$) was lower than the other two grades ($92.50\% \pm 3.16\%$ for nrDR and $81.11\% \pm 2.08\%$ for vtDR). Larger data sets with more r/nvtDR cases could help mitigate this performance gap, and it is worth noting that most r/nvtDR misclassifications resulted in vtDR classifications. Although this is obviously not optimal, this outcome at least spares patients with referable DR from failing to receive needed clinical attention. In addition, the r/nvtDR is a middle DR severity which also makes the classification more challenge than the other two grades. The classi-

fication performance for rDR ($AUC = 0.96 \pm 0.01$) also outperforms vtDR ($AUC = 0.92 \pm 0.02$). Our model relied on a small scan region (3.0×3.0 mm) at the central macula.^{25,27,54} However, a larger scan area with appropriate sampling density (e.g. not lower than 10 $\mu\text{m}/\text{pixel}$) could still improve the DR classification performance, as there are key DR features, such as neovascularization and venous beading, that are typically outside the 3×3 mm region. Because these features are associated with more advanced stages of DR, exploring models that use larger fields of view may preferentially improve vtDR diagnosis. Therefore, in the future, we hope to improve DR classification performance with larger data sets and scans with larger field of view. In addition, to improve the reliability of our evaluation results, we also hope to test our framework on an external data set based on federated learning.

Conclusion

We proposed a fully automated DR classification framework using 3D OCT and OCTA as inputs. Our framework achieved reliable performance on multiclass DR classification (nrDR, rDR/nvtDR, and vtDR), and produces 3D CAMs that can be used to interpret the model's decision making. By using our framework, the number of imaging modalities required for DR classification was reduced from fundus photographs and OCT to an OCTA procedure alone. This accuracy of the model output in this study also suggests the combination of OCT/OCTA and deep learning could perform well in a clinical setting.

Acknowledgments

Supported by National Institutes of Health (R01 EY027833, R01 EY024544, P30 EY010572, T32 EY023211, and UL1TR002369); Unrestricted Departmental Funding Grant and William & Mary Greve Special Scholar Award from Research to Prevent Blindness (New York, NY).

Disclosure: **P. Zang**, None; **T.T. Hormel**, None; **X. Wang**, None; **K. Tsuboi**, None; **D. Huang**, Optovue, Inc. (F, I, C, P); **T.S. Hwang**, None; **Y. Jia**, Optovue, Inc. (F, P), Optos (P)

References

1. Wilkinson CP, Ferris FL, III, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110(9):1677–1682.
2. Wong TY, Sun J, Kawasaki R, et al. Guidelines on diabetic eye care: the international council of ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings. *Ophthalmology*. 2018;125(10):1608–1622.
3. Flaxel CJ, Adelman RA, Bailey ST, et al. Diabetic retinopathy preferred practice pattern. *Ophthalmology*. 2020;127(1):66–145.
4. Antonetti DA, Klein R, Gardner TW. Diabetic retinopathy. *N Engl J Med*. 2012;366:1227–1239.
5. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124(7):962–969.
6. Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57(13):5200–5206.
7. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–2410.
8. Ghosh R, Ghosh K, Maitra S. Automatic detection and classification of diabetic retinopathy stages using CNN. *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*. 2017;pp. 550–554.
9. Lee R, Wong TY, Sabanayagam C. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. *Eye Vis*. 2015;2(1):1–25.
10. Prescott G, Sharp P, Goatman K, et al. Improving the cost-effectiveness of photographic screening for diabetic macular oedema: a prospective, multi-centre, UK study. *Br J Ophthalmol*. 2014;98(8):1042–1049.
11. Huang D, Swanson EA, Lin C P, et al. Optical coherence tomography. *Science*. 1991;254(5035):1178–1181.
12. Virgili G, Menchini F, Casazza G, et al. Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy. *Cochrane Database Syst Rev*. 2015;1:CD008081.
13. Kinyoun J, Barton F, Fisher M, Hubbard L, Aiello L, Ferris F, III; The ETDRS Research Group. Detection of diabetic macular edema: ophthalmoscopy versus photography—Early Treatment Diabetic Retinopathy Study Report Number 5. *Ophthalmology*. 1989;96(6):746–750.
14. Bhavsar KV, Subramanian ML. Risk factors for progression of subclinical diabetic macular oedema. *Br J Ophthalmol*. 2011;95(5):671–674.
15. Bressler NM, Miller KM, Beck RW, et al.; Diabetic Retinopathy Clinical Research Network. Observational study of subclinical diabetic macular edema. *Eye (Lond)*. 2012;26(6):833–840.
16. Browning DJ, Fraser CM. The predictive value of patient and eye characteristics on the course of subclinical diabetic macular edema. *Am J Ophthalmol*. 2008;145(1):149–154.
17. Browning DJ, Fraser CM, Clark S. The relationship of macular thickness to clinically graded diabetic retinopathy severity in eyes without clinically detected diabetic macular edema. *Ophthalmology*. 2008;115(3):533–539.

18. Ruia S, Saxena S, Gemmy Cheung CM, Gilhota JS, Lai TY. Spectral domain optical coherence tomography features and classification systems for diabetic macular edema: a review. *Asia Pac J Ophthalmol (Phila)*. 2016;5(5):360–367.
19. Olson J, Sharp P, Goatman K, et al. Improving the economic value of photographic screening for optical coherence tomography–detectable macular oedema: a prospective, multicentre, UK study. *Health Technol Assess*. 2013;17(51):1–142.
20. Schmidt-Erfurth U, Garcia-Arumi J, Bandello F, et al. Guidelines for the management of diabetic macular edema by the European Society of Retina Specialists (EURETINA). *Ophthalmologica*. 2017;237(4):185–222.
21. Makita S, Hong Y, Yamanari M, et al. Optical coherence angiography. *Optics Express*. 2006;14(17):7821–7840.
22. An L, Wang RK. In vivo volumetric imaging of vascular perfusion within human retina and choroids with optical micro-angiography. *Optics Express*. 2008;16(15):11438–11452.
23. Jia Y, Tan O, Tokayer J, et al. Split-spectrum amplitude-decorrelation angiography with optical coherence tomography. *Optics Express*. 2012;20(4):4710–4725.
24. Jia Y, Bailey ST, Hwang TS, et al. Quantitative optical coherence tomography angiography of vascular abnormalities in the living human eye. *Proc Natl Acad Sci* 2015;112(18):E2395–E2402.
25. Hwang TS, Zhang M, Bhavsar K, et al. Visualization of 3 Distinct Retinal Plexuses by Projection-Resolved Optical Coherence Tomography Angiography in Diabetic Retinopathy. *JAMA Ophthalmol*. 2016;134(12):1411–1419.
26. Zhang M, Hwang TS, Dongye C, et al. Automated Quantification of Nonperfusion in Three Retinal Plexuses Using Projection-Resolved Optical Coherence Tomography Angiography in Diabetic Retinopathy. *Investig Ophthalmol Vis Sci*. 2016;57(13):5101–5106.
27. Hwang TS, Gao SS, Liu L, et al. Automated quantification of capillary nonperfusion using optical coherence tomography angiography in diabetic retinopathy. *JAMA Ophthalmol*. 2016;134(4):367–373.
28. Hwang TS, Jia Y, Gao SS, et al. Optical coherence tomography angiography features of diabetic retinopathy. *Retina*. 2015;35(11):2371.
29. Sandhu HS, Eltanboly A, Shalaby A, et al. Automated diagnosis and grading of diabetic retinopathy using optical coherence tomography. *Investig Ophthalmol Vis Sci*. 2018;59(7):3155–3160.
30. Sandhu HS, Eladawi N, Elmogy M, et al. Automated diabetic retinopathy detection using optical coherence tomography angiography: a pilot study. *Br J Ophthalmol*. 2018;102(11):1564–1569.
31. Alam M, Zhang Y, Lim JI, et al. Quantitative optical coherence tomography angiography features for objective classification and staging of diabetic retinopathy. *Retina*. 2020;40(2):322–332.
32. Heisler M, Karst S, Lo J, et al. Ensemble Deep Learning for Diabetic Retinopathy Detection Using Optical Coherence Tomography Angiography. *Transl Vis Sci Technol*. 2020;9(2):20.
33. Le D, Alam M, Yao CK, et al. Transfer learning for automated OCTA detection of diabetic retinopathy. *Transl Vis Sci Technol*. 2020;9(2):35.
34. Zang P, Gao L, Hormel TT, et al. DcardNet: Diabetic Retinopathy Classification at Multiple Levels Based on Structural and Angiographic Optical Coherence Tomography. *IEEE Trans Biomed Eng*. 2021;68(6):1859–1870.
35. You Q S, Tsuboi K, Guo Y, et al. Comparison of central macular fluid volume with central subfield thickness in patients with diabetic macular edema using optical coherence tomography angiography. *JAMA Ophthalmol*. 2021;139(7):734–741.
36. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444.
37. Zhou B, Khosla A, Lapedriza A, et al. Learning Deep Features for Discriminative Localization. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2016:2921–2929. Available at: <https://arxiv.org/abs/1512.04150v1#:~:text=Learning%20Deep%20Features%20for%20Discriminative%20Localization%20%20,%20%20%20%20%20%20https%3A%2F%2Fdoi.org%2F10.48550%2F%2F1512.04150%20...%20>.
38. Gao SS, Liu G, Huang D, et al. Optimization of the split-spectrum amplitude-decorrelation angiography algorithm on a spectral optical coherence tomography system. *Opt Lett*. 2015;40(10):2305–2308.
39. Kraus MF, Liu JJ, Schottenhamml J, et al. Quantitative 3D-OCT motion correction with tilt and illumination correction, robust similarity measure and regularization. *Biomed Opt Express*. 2014;5(8):2591–2613.
40. Zhang M, Hwang TS, Campbell JP, et al. Projection-resolved optical coherence tomographic angiography. *Biomed Opt Express*. 2016;7(3):816–828.

41. Wang J, Zhang M, Hwang TS, et al. Reflectance-based projection resolved optical coherence tomography. *Biomed Opt Express*. 2017;8(3):1536–1548.
42. Early Treatment Diabetic Retinopathy Study Research Group. Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS report number 12. *Ophthalmology*. 1991;98(5):823–833.
43. Haneda S, Yamashita H. International clinical diabetic retinopathy disease severity scale detailed table [Article in Japanese]. *Nihon Rinsho*. 2002;68(Suppl 9):228–235.
44. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*. 2015, <https://doi.org/10.48550/arXiv.1502.03167>.
45. Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. *Conference: Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. June 21–24, 2010, Haifa, Israel. 2010:807–814.
46. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. *AISTATS 11: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2011;15:315–323.
47. Cohen J. A coefficient of agreement for nominal scales. *Educational Psychological Measurement*. 1960;20(1):37–46.
48. Hormel TT, Hwang TS, Bailey ST, et al. Artificial intelligence in OCT angiography. *Prog Retin Eye Res*. 2021;85:100965.
49. Hormel TT, Huang D, Jia Y. Artifacts and artifact removal in optical coherence tomographic angiography. *Quantitative Imaging Med Surg*. 2021;11(3):1120.
50. Zhang M, Wang J, Pechauer AD, et al. Advanced image processing for optical coherence tomographic angiography of macular diseases. *Biomed Opt Express*. 2015;6(12):4661–4675.
51. Campbell JP, Zhang M, Hwang TS, et al. Detailed vascular anatomy of the human retina by projection-resolved optical coherence tomography angiography. *Sci Rep*. 2017;7:42201.
52. Hormel TT, Wang J, Bailey ST, et al. Maximum value projection produces better en face OCT angiograms than mean value projection. *Biomed Opt Express*. 2018;9(12):6412–6424.
53. Krause J, Gulshan V, Rahimy E, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology*. 2018;125(8):1264–1272.
54. Hwang TS, Hagag AM, Wang J, et al. Automated quantification of nonperfusion areas in 3 vascular plexuses with optical coherence tomography angiography in eyes of patients with diabetes. *JAMA Ophthalmol*. 2018;136(8):929–936.