Deep Learning for Prediction of AMD Progression: A Pilot Study

Daniel B. Russakoff,[1] Ali Lamin,[2,3] Jonathan D. Oakley,[1] Adam M. Dubis,[2,3] and Sobha Sivaprasad[2,3]

[1]Voxeleron LLC, Pleasanton, California, United States
[2]NIHR Moorfields Biomedical Research Centre, London, United Kingdom
[3]UCL Institute of Ophthalmology, London, United Kingdom

Correspondence: Daniel B. Russakoff, Voxeleron LLC, 4695 Chabot Drive, Suite 200, Pleasanton, CA 94588, USA;
daniel@voxeleron.com.

PURPOSE. To develop and assess a method for predicting the likelihood of converting from early/intermediate to advanced wet age-related macular degeneration (AMD) using optical coherence tomography (OCT) imaging and methods of deep learning.

METHODS. Seventy-one eyes of 71 patients with confirmed early/intermediate AMD with contralateral wet AMD were imaged with OCT three times over 2 years (baseline, year 1, year 2). These eyes were divided into two groups: eyes that had not converted to wet AMD ($n = 40$) at year 2 and those that had ($n = 31$). Two deep convolutional neural networks (CNN) were evaluated using 5-fold cross validation on the OCT data at baseline to attempt to predict which eyes would convert to advanced AMD at year 2: (1) VGG16, a popular CNN for image recognition was fine-tuned, and (2) a novel, simplified CNN architecture was trained from scratch. Preprocessing was added in the form of a segmentation-based normalization to reduce variance in the data and improve performance.

RESULTS. Our new architecture, AMDnet, with preprocessing, achieved an area under the receiver operating characteristic (ROC) curve (AUC) of 0.89 at the B-scan level and 0.91 for volumes. Results for VGG16, an established CNN architecture, with preprocessing were 0.82 for B-scans/0.87 for volumes versus 0.66 for B-scans/0.69 for volumes without preprocessing.

CONCLUSIONS. A CNN with layer segmentation-based preprocessing shows strong predictive power for the progression of early/intermediate AMD to advanced AMD. Use of the preprocessing was shown to improve performance regardless of the network architecture.

Keywords: age-related macular degeneration, optical coherence tomography, image analysis, deep learning, choroidal neovascularization

Advanced age-related macular degeneration (AMD) is a leading cause of vision loss for people over 50 and accounts for 8.7% of all blindness worldwide.[1] AMD proceeds in distinct stages from early, to intermediate, to advanced. In advanced, wet (neovascular) AMD, blood vessel growth (choroidal neovascularization [CNV]) can lead to irreversible damage to the photoreceptors and rapid vision loss. Currently, patients can progress to wet AMD without symptoms or any measurable change. Thus, it is of the utmost importance to try and determine which patients are at the highest risk for conversion to wet AMD to allow intervention before permanent damage.

This ability to see subclinical neovascularization in the macula has been underutilized, in part due to the lack of therapeutics, but also because of its cost and discomfort.[2] While early studies using fluorescent dye imaging[3,4] have shown that subclinical irregularities (plaques, spots) are valuable biomarkers, new noninvasive techniques seek to build on this. The more recent findings correlating optical coherence tomography angiography (OCTA) with indocyanine green angiography imaging, for example, attempt to bridge early work to the newer technologies.[5–7] It is still an open area of research, however, and tremendous interest exists in utilizing more established imaging techniques, such as structural OCT and fundus photography, alongside more advanced algorithms to create clinical biomarkers stratifying a patient's level of risk of conversion to wet AMD. The motivation is pragmatic given that structural OCT imaging is the standard of care in the management of ocular diseases, is more affordable than OCTA, and has higher utilization and legacy data.

Structural OCT thus remains the most compelling modality to study for indications of subclinical CNV. And it, too, is an area of active research. de Sisternes et al.,[8] for example, used traditional, feature-based modeling techniques applied to a number of handcrafted features, or parameters, for prediction of conversion to wet AMD. The features used included volume, height, and reflectivity of drusen. In the case where the advanced AMD was geographic atrophy (GA) as opposed to neovascular, the same lab has developed similarly crafted features that were predictive of GA progression.[9] In this study, the best feature was thinning and loss of reflectivity of the inner/outer segment junction, a structural measure derived from the OCT data.

A similar combination of OCT-based structural features and visual acuity was used temporally across an initiation phase to characterize response to anti-VEGF treatment using a random forest classifier.[10] With areas under the curve (AUCs) between 0.7 and 0.8, the resulting model had comparable performance

TABLE 1. Demographics of the Study Subjects

Variable	Total	AMD Nonprogressors	AMD Progressors
Number of eyes	71	40	31
Age			
Mean (SD)	74 (8.5)	72 (8.7)	76 (7.5)
Median	76	72.5	77
Minimum–Maximum	57–91	57–89	62–91
Sex			
Female, *n* (%)	43 (60.6%)	20 (50%)	23 (74.2%)
Male, *n* (%)	28 (39.4%)	20 (50%)	8 (25.8%)
Laterality			
Right, *n* (%)	48 (67.6%)	26 (65%)	22 (71%)
Left, *n* (%)	23 (32.4%)	14 (35%)	9 (29%)
Follow-Up Exam/Conversion, months			
Mean (SD)	–	23.75 (1.33)	23.32 (2.06)
Median		24	24
Minimum–Maximum		20–26	17–27

to an expert human grader in predicting both low and high anti-VEGF treatment requirements. It is of interest that they found that temporally differential features were not shown to play an important, discriminatory role in their model's predictions and that a cross-sectional analysis, as is presented here, achieved the same performance. More recently, Schmidt-Erfurth et al.[11] used machine learning methods to assimilate various imaging, demographic, and genetic features to predict the likelihood of conversion from intermediate to advanced AMD. In a study of 495 eyes, they had separate models for conversion to wet AMD ($n = 114$) and GA ($n = 45$) and reported AUCs of 0.68 and 0.80, respectively, using 10-fold cross validation. The deep learning component used segmented hyperreflective foci in the OCT data producing an en face map of their location that generated nine separate numerical features based on location and distribution that were included in the final 71 features used. The predictive hallmarks for CNV were reported as "mostly drusen-centric."

In this work, we look to derive OCT-based biomarkers based on a deep learning classifier to help predict which patients will progress from early/intermediate AMD to wet AMD using OCT imaging data alone. For context, we also present the performance of more traditional machine learning classifiers using features akin to those of de Sisternes et al.[8] and Schmidt-Erfurth et al.[11]

METHODS

Patients with unilateral neovascular AMD and early or intermediate AMD in the fellow eye with approximately 2 years follow-up (17–27 months) were selected from the anti-VEGF in the AMD database of Moorfields Eye Hospital. Patients were imaged using two OCT devices, the 3D OCT-1000 (18 scans) and the 3D OCT-2000 (53 scans) in both eyes (Topcon, Tokyo, Japan). For each instrument, the 3D macular scan was used and produced 128 B-scans with 512 A-scans each over a 6 × 6-mm^2 area centered at the fovea.

Each eye was part of a three scan, 2-year protocol (baseline, year 1, year 2). The progressors ($n = 31$) were defined as fellow eyes with new-onset macular fluid on year 2 scans confirmed by fluorescein angiography to show the presence of CNV. The nonprogressors ($n = 41$) were the fellow eyes that had not converted on year 2 scans. The OCT scans used for prediction of conversion were the baseline scans ($n = 71$) with early or intermediate AMD corresponding to the progressor and nonprogressor groups. The average interval between baseline and year 2 scans was 2 years ± a standard deviation of 2 months (Table 1). In order to focus on CNV, cases of GA were excluded from the study. This work was approved by the Ethical Review Board of Moorfields Eye Hospital (ROAD 17/004) and adhered to the principles of the Declaration of Helsinki.

Of the 71 participants included in the study, 43 were female (60.6%) and 28 male (39.4%). The nonprogressors consisted of 20 females and 20 males. The progressors consisted of 23 females and 8 males. While age ranges were similar between the two cohorts, AMD progressors were on average older (unpaired Student's *t*-test; $P = 0.02$). Table 1 details these demographics.

Traditional Image Processing

Following earlier work,[8,11] we first assayed to perform the prediction using traditional image processing and machine learning techniques. All data sets were analyzed using patient data and layer-based biomarkers from OCT analysis software (Orion; Voxeleron LLC, Pleasanton, CA, USA). The software automatically segments the OCT volumes into seven retinal layers, allowing analysis of various metrics such as average thicknesses and volumes of the different layers (and of the drusen) within the Early Treatment Diabetic Retinopathy Study (ETDRS) zones[12] based on an automatic foveal centration. All segmentations were verified to be error free (AL and JDO), and then analyzed for separation using a state-of-the-art machine learning classifier. Example segmentations for both progressors and nonprogressors are shown in Figures 1 and 2, where we highlight more normal-looking retinas and also those with some obvious drusen. Multiple layer segmentation offers multiple parameters that can be analyzed in an effort to separate the two groups. An ETDRS grid has nine zones, and with seven average thicknesses being reported in each of these zones, we can use any combination of thicknesses or volumes over different regions to train a classifier to predict the class.

We used a 32-dimensional feature vector that comprised biomarkers from the segmentation as well as patient information (Tables 2, 3). We used a support vector machine (SVM), a well-defined, state-of-the-art machine learning classifier to perform the prediction.[13] The SVM was trained with radial basis functions for the kernel, and the free parameters (box constraint, kernel scale) were chosen empirically. We evaluated the SVM using 5-fold cross validation. 5-fold cross validation is performed by randomly partitioning the data into five equal subsets, or folds. One fold is used as a test set, whereas the other four folds are used to train a model. We created these folds taking care that, for a given run, no one patient's data ever appeared in both the training folds and the testing fold. By performing this procedure on each of the five folds in turn, we can generate a prediction for each data point. And, for the entire set of predictions, we generate a receiver operating characteristic (ROC) curve as well as its corresponding AUC (Fig. 3).

Deep Learning–Based Analysis

Our deep learning approach consists of a two-step process decoupling the image segmentation step from the classification step. This has the effect of allowing the classifier to focus specifically on the regions of interest. After the segmentation step, we tried two different convolutional neural networks (CNNs): (1) transfer learning using the popular VGG16 network[14] and (2) AMDnet, a novel, simplified architecture trained from scratch.
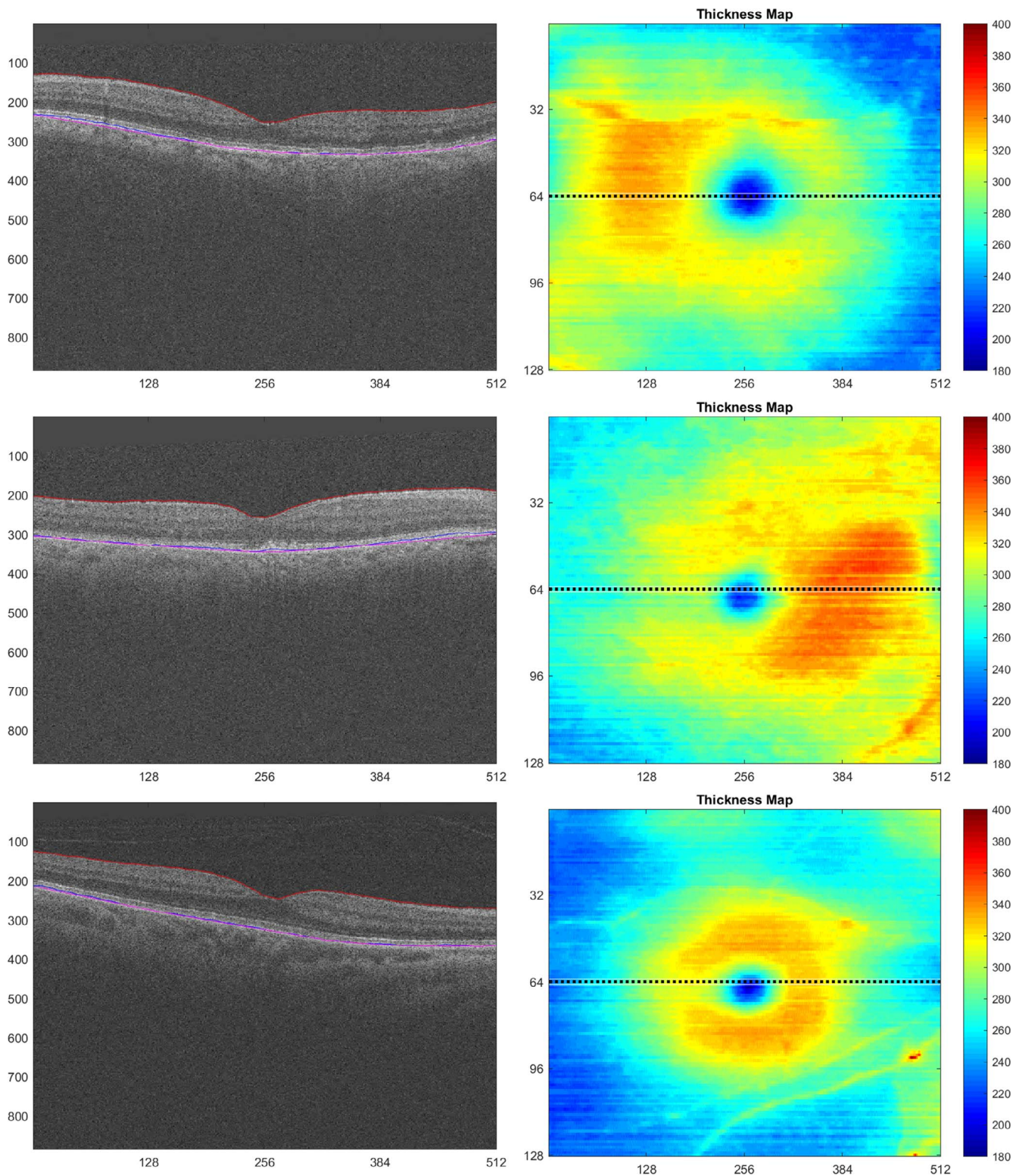
FIGURE 1. The left-hand side shows example segmentations in both the progressor (*top*, *bottom*) and nonprogressor (*middle*) groups. The *right-hand* side shows their corresponding total retinal thickness maps in micrometers.

Segmentation-Based Preprocessing. The 71 volumes were decomposed into 9088 B-scans that were preprocessed using the aforementioned layer segmentation software to identify the inner limiting membrane (ILM) and Bruch's membrane (Fig. 4). Each B-scan was then cropped from the ILM to a fixed offset (390 μm) below Bruch's membrane and

resampled to a uniform size (Fig. 5). The offset used was designed to capture choroidal information over a fixed area beneath the choriocapillaris. It was chosen based on work from Manjunath et al.[15] to represent 2 SD above the mean subfoveal choroidal thickness in a population with AMD. This
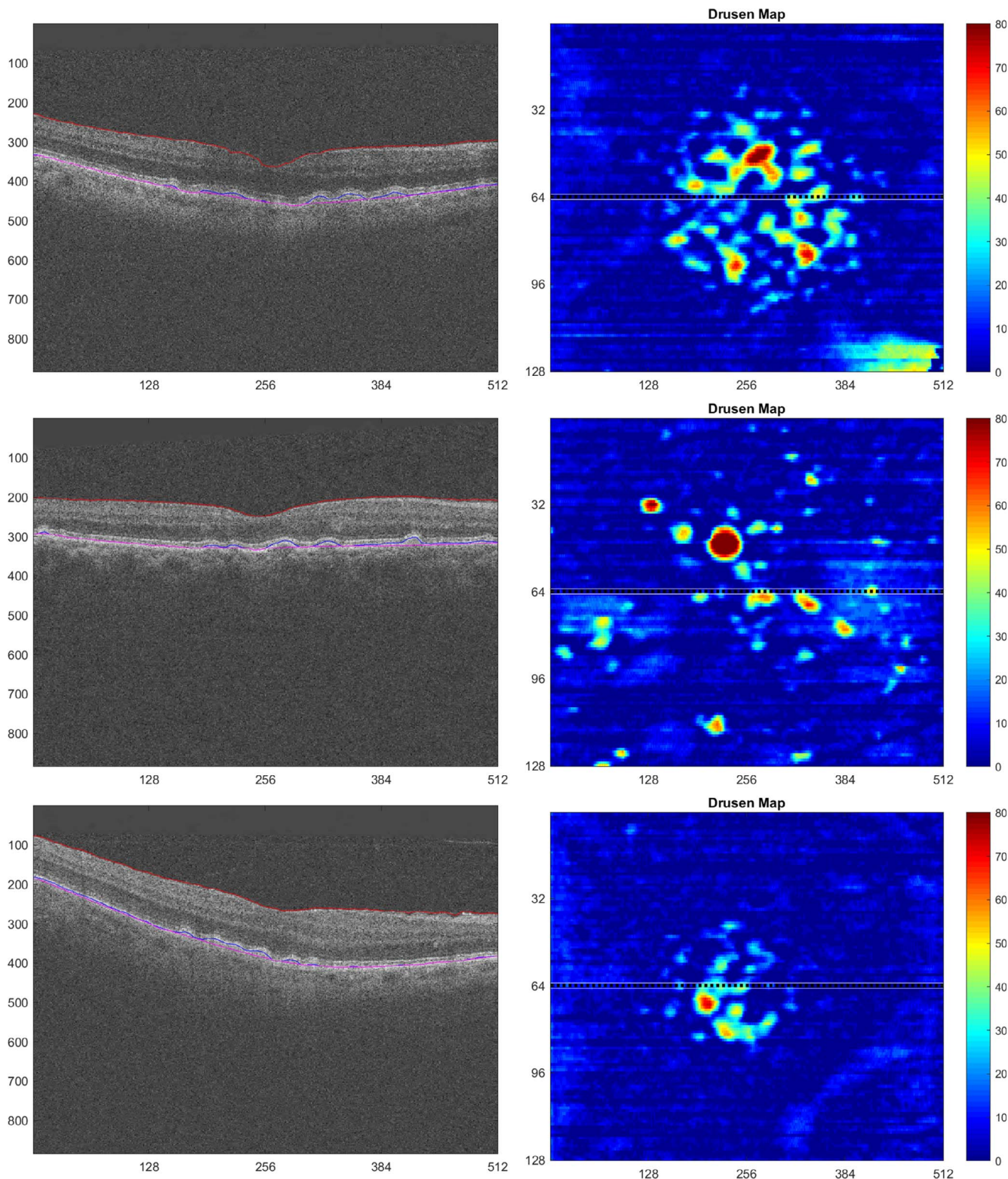
FIGURE 2. The *left-hand* side shows example segmentations in both the progressor (*top*, *bottom*) and nonprogressor (*middle*) groups. The *right-hand* side shows their corresponding drusen thickness maps in micrometers.

preprocessing was performed to reduce the variance of the training set and create some invariance to scale.

A Transfer Learning Model. To evaluate the preprocessing, an existing, well-established deep CNN (VGG16)[14] was fine-tuned using transfer learning based on the standard strategy of retraining only the fully connected layers of the model.[16] We used the original paper's fully connected layer sizes (4096 neurons each), changing only the final layer from 1000 neurons to 2 neurons to fit our problem. Similar to Rattani,[16] we experimented with simpler versions with a

TABLE 2. The 32 Features Used to Train the SVM Classifier

Layer	RNFL	GC-IPL	INL	OPL	ONL	PR	RPE to Bruch's	TRT	Patient-Based
Zones	IA,OA,TA	IA,OA,TA	IA,OA,TA,D	IA,OA,TA,D	IA,OA,TA,D	IA,OA,TA,D	IA,OA,TA,D	IA,OA,TA,D	Age, Sex
Features	3	3	4	4	4	4	4	4	2

The thicknesses were taken as averages with the fovea-centered ETDRS grid: IA, inner annulus (circle of 3-mm diameter, less the central 1-mm subfield); OA, outer annulus (circle of 6-mm diameter, less the central 1-mm subfield and the IA); TA, total annulus = IA + OA; D, entire 6-mm-diameter circle. The layer names follow the APOSTEL recommendations[27]: RNFL, retinal nerve fiber layer; GC-IPL, ganglion-cell + inner plexiform layer; INL, inner nuclear layer; OPL, outer plexiform layer; ONL, outer nuclear layer; PR, photoreceptor complex; as well as RPE to Bruch's and the TRT, total retinal thickness. Foveal thicknesses of the RNFL and GC-IPL were excluded.

smaller number of neurons, settling on 512 and 128 neurons for the first two fully connected layers, respectively. This process was applied to both the raw and preprocessed B-scans. The raw and preprocessed B-scans were resized to 224 × 224 to match VGG16's expected input. The training was run for 2500 epochs using stochastic gradient descent with Nesterov momentum and a learning rate of 5e-5. To avoid overtraining, we used early stopping. This procedure stops the training when the loss on a held-out validation set fails to improve for a prespecified number of epochs, termed the *patience*. For this experiment, we set the patience at 20 epochs. The resulting classifiers were evaluated using the exact same 5-fold cross validation folds from the prior, traditional image processing analysis.

The AMDnet Model. Alternate architectures were explored in an effort to further improve the results. We tried both deeper, more complex networks as well as shallower, simpler ones and eventually settled on the latter. AMDnet (Figs. 6, 7) consists of just three convolutional layers with varying amounts of pooling. The number of parameters for this model is just over 2 million versus more than 27 million (12 million trainable) for VGG16. Given the relatively small size of the data set, we took care to regularize this model in three specific ways:

1. We used dropout regularization with a percentage of 45% at the end of all but one of the convolutional and fully connected layers. Dropout essentially acts during training on each batch to randomly remove a percentage of the previous layer's neurons. Dropout has the effect of averaging an ensemble of classifiers, which produces more robust results and resists overtraining.[17]

2. We used L2 regularization for each of the convolutional layers, which penalizes very large weights and has the effect of simplifying the model. Simpler models generalize better, which also works to prevent overtraining.

3. We used *maxnorm* regularization for the dense layers, which also works to simplify the model by requiring the norm of a given layer's weights to be less than a

prespecified value. As above, simpler models are harder to overtrain.

Figure 7 has a detailed breakdown of the architecture of AMDnet. We evaluated AMDnet using the exact same 5-fold cross validation and folds as described above.

Feature Analysis

In an effort to tease out what latent features the classifier is relying on, and perhaps learn something about the disease process itself, we also performed an occlusion sensitivity analysis[18] of the outputs of the neural network. The occlusion analysis shows the regions of the image that are most discriminative with respect to a specific class. Such visualizations help interpret the overall results, especially in asking whether the method makes basic sense and whether artifacts or irrelevant features are driving the performance. This we revisit more thoroughly in the discussion.

RESULTS

Traditional Image Processing

The summary results of the SVM analysis are shown in Figure 3 on the left. With AUCs in the range of 0.74 to 0.82 (mean = 0.78), these results are consistent with what has been previously reported for this type of approach.[8,11]

Another consideration was given to potential bias introduced based on machine type as the two Topcon devices use different spectrometers, resulting in different axial resolutions. To investigate this, we added the dimensionality of the scan's axial resolution (either 480 or 885 pixels) as a feature, acting as an instrument flag. All features were scaled to zero mean and unit variance as part of the training process (test features being scaled based on the learned ranges). We reran the best-performing SVM from the previous experiment using this new feature set and report the results in Figure 3 on the right. The mean AUC of 0.79 for this experiment suggests that having

TABLE 3. Quantitative Report of 8 of the 32 Features From Table 2

	Nonprogressors		Progressors		
Layer (annulus)	Mean	SD	Mean	SD	*P* Value
TRT	8.23789	0.3796	8.14838	0.501817	0.39494
RNFL	1.139194	0.1038	1.161975	0.170421	0.488885
GC-IPL	1.916145	0.1717	1.759685	0.18552	*0.000464*
INL	0.980921	0.0669	0.97634	0.078946	0.792206
OPL	0.681369	0.1018	0.624325	0.136326	*0.04734*
ONL	2.14774	0.1556	2.230325	0.273673	0.113082
PR	1.236373	0.0818	1.247537	0.094527	0.59592
RPE to Bruch's	0.136155	0.0888	0.148176	0.065638	0.529916

The total annulus thickness features for each layer. Values are in mm^3. Statistically significant values with $P < 0.05$ are in bold italics.
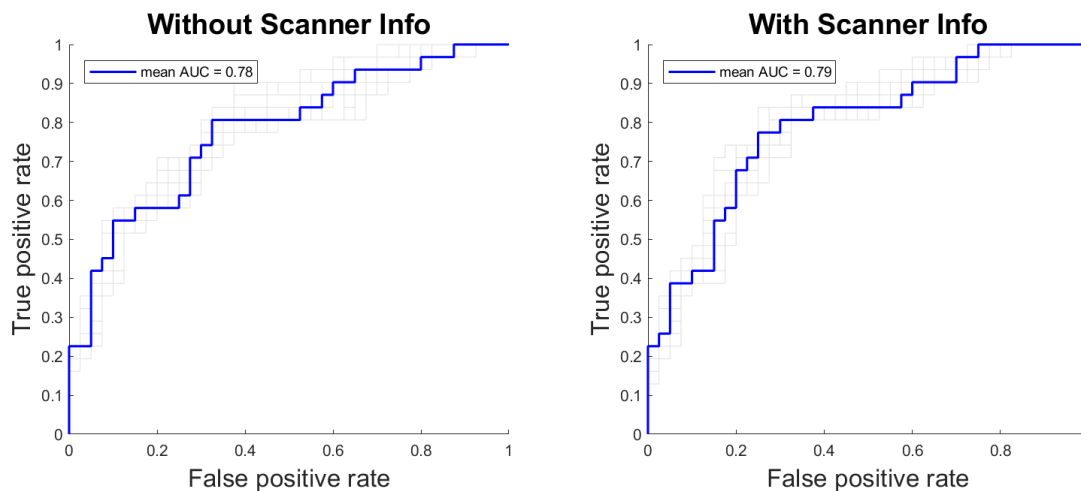
FIGURE 3. The above compares the performance of the SVM classifier (*left*) with the same data plus instrument type added as a feature (*right*). This information appears to offer very little improvement in the classification.

knowledge of the instrument appears to add little to no additional information to the classifier's performance.

Finally, we explore the potential bias of the follow-up interval on the performance of the classifier. Following the experiments above, we chose an operating point with a false-positive rate of 0.25 and looked at the true positives, true negatives, false positives, and false negatives with respect to follow-up interval. We conclude, based on Figure 8, that small variations in the follow-up interval do not introduce a large bias into the results.

Deep Learning

The results comparing the effect of preprocessing (Fig. 9) are presented at both the B-scan and volume levels. The prediction value for the volume level analysis was calculated by taking the mean of each volume's individual B-scan predictions. For VGG16, with preprocessing, the AUC was 0.82 at the B-scan level and 0.87 at the volume level, whereas the same run without preprocessing (only scaling to match the VGG16 input) had AUCs of 0.67 and 0.69, respectively. The results for the same 5-fold validation for AMDnet are shown in Figure 10. We achieve a marked improvement with AMDnet at the B-scan level (0.89) and at the volume level (0.91). Of interest, we also performed simple augmentation of the data (adding small rotations plus noise) but were unable to improve the algorithm's performance. This very clearly demonstrates the benefits of preprocessing as, regardless of network and evaluation metric, the performance improves each time.

The results of the feature analysis, shown in Figure 11, illustrate that the areas around the retinal pigment epithelium
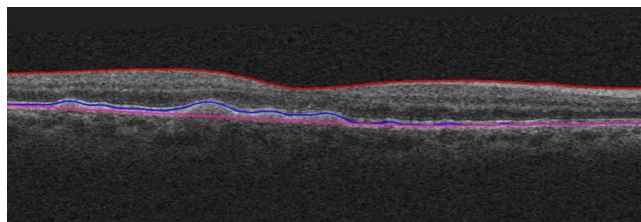
(RPE) and choroid seem to be the most useful to the classifier in making its predictions. In particular, this analysis shows that pixels around the RPE have the largest impact on the final score of the classifier in the case of nonprogressors while progressors seem to have more sub-RPE choroidal involvement. In addition, we stacked the occlusion sensitivity maps into volumes looking for a pattern in the en face direction (Fig. 12). These results suggest a stronger response nasally for nonprogressors, while the progressors rely more on the temporal region.

DISCUSSION

We have reported on a use of deep learning to predict conversion to wet AMD using OCT imaging. The results show clear separation between the progressors and the nonprogressors, and the occlusion sensitivity analysis indicates that relevant features are brought to bear by the technique. We considered, as a comparison, a clinical study where retinal specialists would try, using their clinical experience, to divine which patients would convert to neovascular AMD. We put this question to three retinal specialists who indicated that, given the lack of validated biomarkers for this problem, they would not feel comfortable making this decision, even for a research study. In the following, we add context to the findings, discuss their clinical relevance, present some limitations of the study, and close with some conclusions.

One of the major challenges in the clinical management of patients with early/intermediate AMD is the assessment of risk of conversion, and any metrics supportive of this assessment are welcome. Structural OCT data have been used to create anatomical biomarkers such as thickness and volumetric measures, but despite being researched for several years, compelling indicators of conversion have yet to emerge. Instead, interest has turned to OCTA, where subclinical neovascularization is being observed and studies are being carried out on how to quantify these observations such that they can be deployed clinically. OCTA instrumentation is, however, less widely used, and longitudinal data are less readily available. In addition, OCTA data have greater dependence on variations in signal strength across different systems and are vulnerable to projection artifacts that make it difficult to assess flow as a reliable biomarker, especially in the case of neovascularization underneath the RPE (type 1).[19] With the advent of more advanced feature extractors and classifiers

FIGURE 4. Example B-scan showing the automated segmentation (ILM in *red*, RPE in *blue*, and Bruch's membrane in *magenta*) used for the preprocessing. In this example, we clearly see a signal in the choroid, albeit diminished below the drusen.
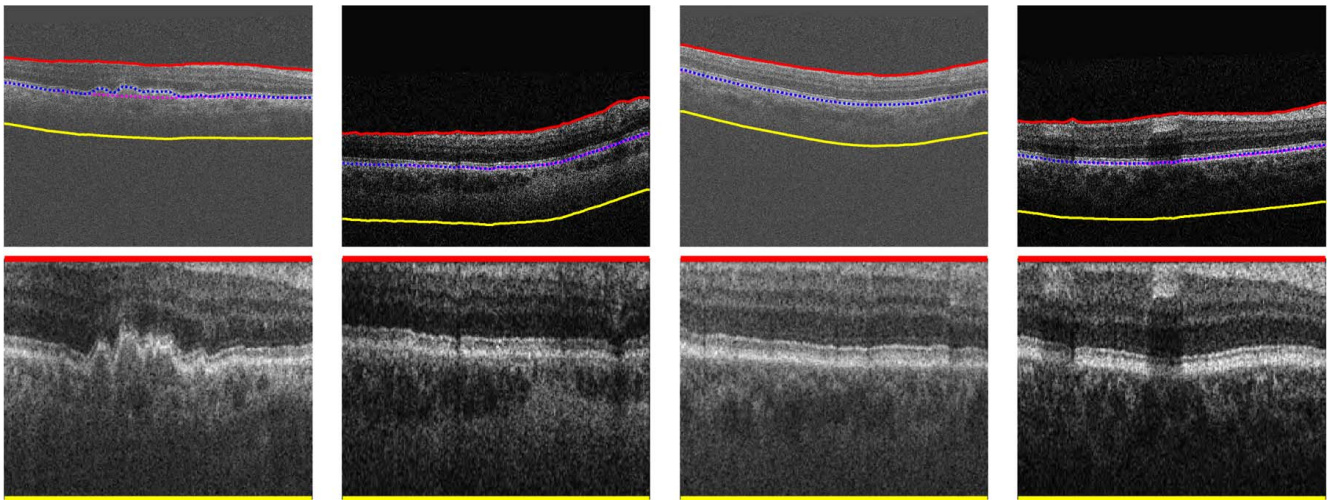
FIGURE 5. An example of the preprocessing used to normalize the B-scans. The *top row* shows B-scans from a Topcon OCT scanner, and the *bottom row* shows the corresponding images with normalization applied. The data are cropped between the ILM (*red*) and a fixed offset (390 μm) (*yellow*) from Bruch's membrane (*magenta-dashed*), which is itself estimated as a baseline to the RPE (*blue-dashed*). Normalization in this way greatly reduces the variance in the training set and allows for robust training of smaller data sets as well as better generalizability. Note that, despite this being a spectral-domain OCT (SD-OCT) device, the signal in the choroid is apparent and strong in each case.

facilitated through deep learning, we have revisited and further mined the OCT data sets for signals that, akin to OCTA, might be supportive of the subclinical assessment of nonexudative neovascularization.

An immediate interpretation of the findings is that the neural network has discovered specific patterns indicative of pathologic change. OCT-based features identified in early CNV have been previously reported.[20–23] The analysis we report on, however, looks at data before any clinically observable signs of conversion, so consideration must be given to more subtle features, including textural changes that are perhaps occurring as a direct result of early physiological changes. Pathology detection using OCT texture analysis has itself been previously researched.[24] Such approaches failed to gain traction, but in the advent of better computational resources and the more sophisticated learning approaches, we envisage a resurgence in such work. The texture descriptors were examples of handcrafted features, a technique that has been superseded by the ability to instead learn the features through deep learning. Similarly, in the work from de Sisternes et al.,[8] Niu et al.,[9] and Schmidt-Erfurth et al.,[11] the features were manually crafted and, through extensive use of regression, applied to temporal data in their final models. By learning the features in a

systematic way afforded by deep neural networks, more powerful and better regularized solutions are now possible. Very important to the method, however, is the preprocessing of the input data via a segmentation step that (1) gives us some invariance to instrumentation and (2) allows the network to concentrate on tissue of interest. This is somewhat akin to the recent work by De Fauw et al.[25] in which their classification scheme uses a separate segmentation step, here using a U-Net deep learning architecture,[26] and then classifying the homogenous tissue regions into referral classes using a second deep learning architecture, one that is very similar in composition to that used in this study. In our work, however, we do not disregard the image intensities and distributions as they are critical to our method in differentiating the classes.

Another interesting finding is the difference in the en face occlusion sensitivity maps between progressors and non-progressors (Fig. 12). Further investigation is needed, but this difference could potentially be due to the presence of more photoreceptors nasally or large arterioles nasally skewing the choroidal density.

This study is not without some limitations. Although this is a large and balanced data set, more data would help better support our conclusions. To address this, unbiased estimates of
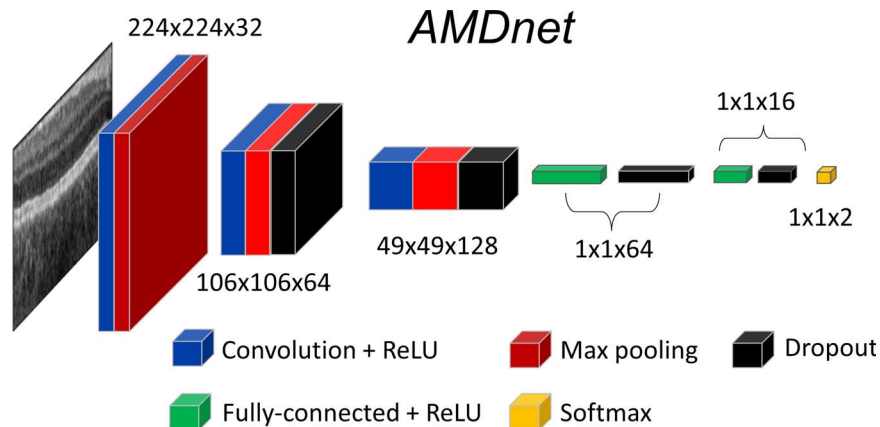

FIGURE 6. A schematic of the architecture of AMDnet.

Layer	Input	Size	Stride	Padding	Regularization	Activation	Output	No. parameters
Conv1	224x224	5x5x32	1	Valid	L2 (l=1E-3)	ReLU	220x220x32	832
MaxPool1	220x220x32	2x2	1	-	-	-	110x110x32	0
Conv2	110x110x32	5x5x64	1	Valid	L2 (l=1E-3)	ReLU	106x106x64	51,264
MaxPool2	106x106x64	2x2	1	-	-	-	53x53x64	0
Dropout1 (45%)	-	-	-	-	-	-	-	0
Conv3	53x53x64	5x5x128	1	Valid	L2 (l=1E-3)	ReLU	49x49x128	204,928
MaxPool3	49x49x128	3x3	1		--	-	16x16x128	0
Dropout2 (45%)	-	-	-	-	-	-	-	0
Flatten1	16x16x128	-	-	-	-	-	32768x1	0
Dense1	32768x1	-	-	-	maxnorm (m=2)	ReLU	64x1	2,097,216
BatchNorm1	64x1	-	-	-	-	-	64x1	256
Dropout3 (45%)	-	-	-	-	-	-	-	0
Dense2	64x1	-	-	-	maxnorm (m=2)	ReLU	16x1	1,040
BatchNorm2	16x1	-	-	-	-	-	16x1	64
Dropout4 (45%)	-	-	-	-	-	-	-	0
Dense3	16x1	-	-	-	-	Softmax	2x1	34

FIGURE 7. A detailed breakdown of AMDnet.

performance are reported, including the cross validation approach given in the method section, where care was taken to evenly balance the cohorts in the test and training sets, ensuring same subject data were not used across data sets. As a pilot study, however, the findings are compelling.

In addition, the current deep learning model is applied only on B-scans, and those results are aggregated to make a final prediction. This is different from the traditional machine learning approaches that use features derived from consideration of the entire volume. Future work will need to investigate the application of the deep learning approach directly on the full volumes as, potentially, a more natural way of finding patterns of subclinical CNV.
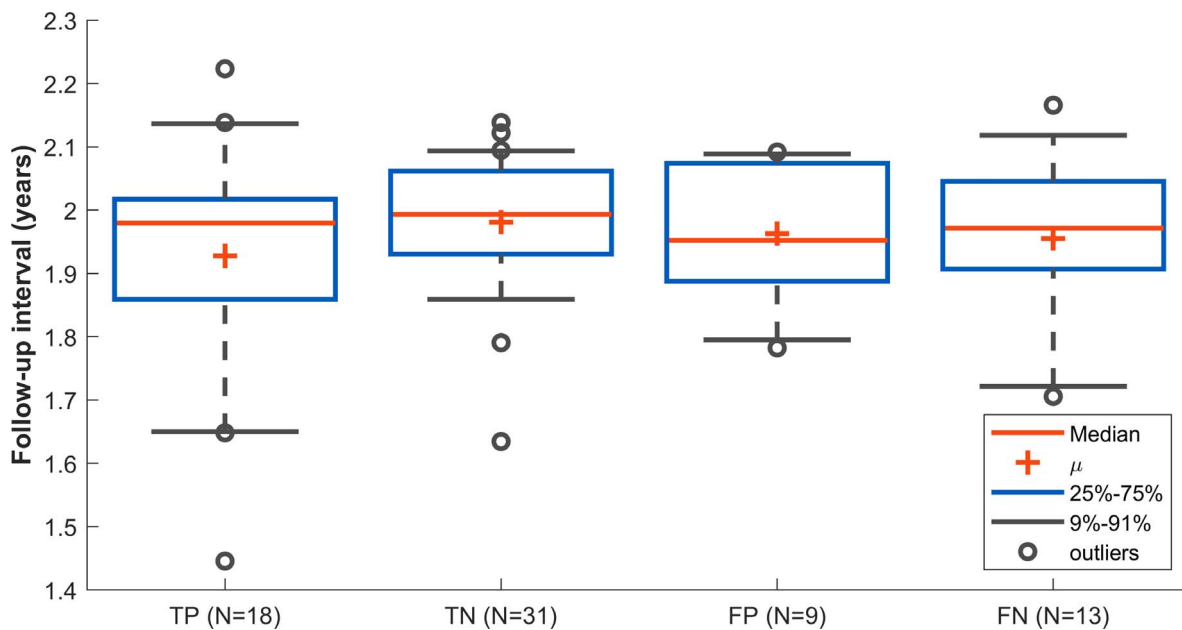

FIGURE 8. A box and whiskers analysis of the SVM results for a specific operating point (FP rate = 0.25). The box represents the 25th and 75th percentiles, while the whiskers are the 9th and 91st percentiles, respectively. The follow-up interval does not seem to have a marked effect on the results.
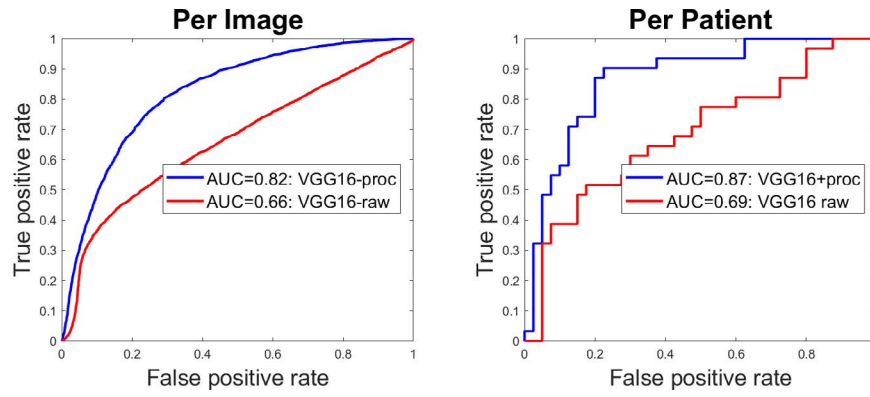
FIGURE 9. Per B-scan (*left*) and per patient (*right*) ROC and AUC results for the fine-tuned VGG16 CNN using segmentation-based preprocessing (*blue*) and just simple resizing (*red*). As expected, preprocessing to reduce the variance of the input data dramatically improves the results.
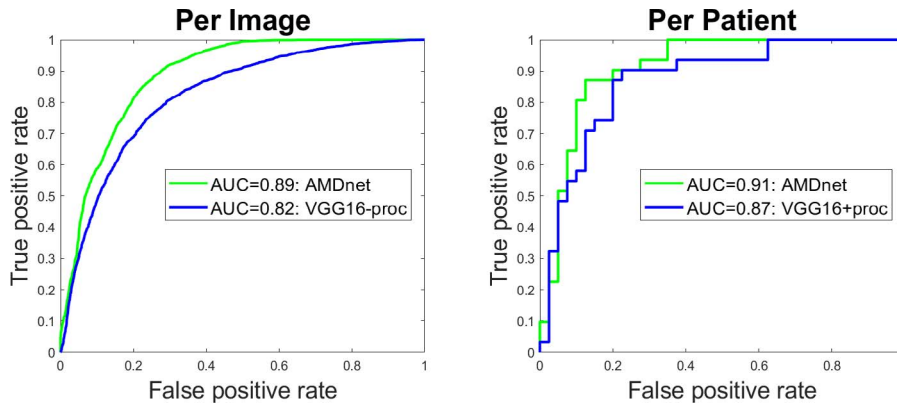

FIGURE 10. Per B-scan (*left*) and per patient (*right*) ROC and AUC results for AMDnet (*green*) and VGG16 with preprocessing (*blue*). The simplified AMDnet architecture shows improvements across both sets.
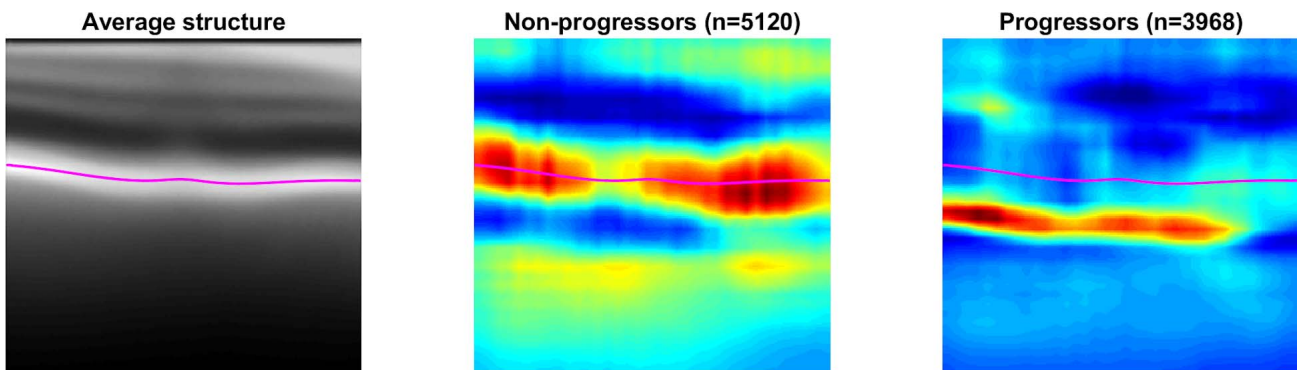
FIGURE 11. Occlusion sensitivity analysis for progressors (*right*) and nonprogressors (*middle*). These images were derived by averaging the occlusion analysis outputs for all B-scans in their respective groups. The average structure for all B-scans is shown on the *left*, and the mean location of Bruch's membrane in all scans is plotted in *magenta*. This analysis shows that, in particular, pixels around the RPE have the largest impact on the final score of the classifier for nonprogressors. It also suggests more sub-RPE or choroidal involvement for progressors.
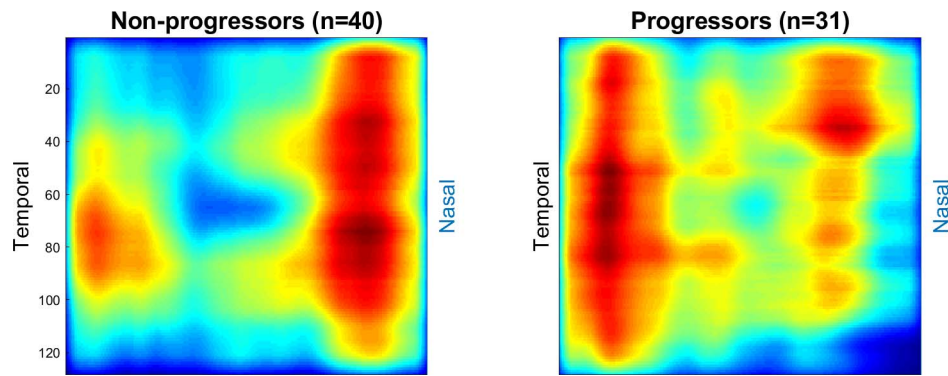
FIGURE 12. En face visualization of occlusion sensitivity analysis. The results from each B-scan's occlusion analyses were stacked into a volume, and all of these volumes were averaged. An en face image of the average volume is displayed for nonprogressors (*left*) and progressors (*right*). The nonprogressors seem to have more relevant features in the nasal side of the volumes while the progressors show the opposite effect.

Another limitation could perhaps also be considered a strength of the method given the positive results and the indication that information in the choroid is of importance to the performance. This is namely the SD-OCT scanner used (Topcon 3d OCT); it has a light source of 840 nm, which offers limited depth penetration given its relatively short wavelength. Longer wavelengths are preferred for resolving detail in the choroid even if these lose some axial resolution. However, through simple review of the B-scans (see Figs. 1, 2, for example), one can see clear choroidal signal in the OCT data. And conversely, this speaks to the strength of the method as even with this limited penetration, there is clearly information in the choroidal regions of the data that is being used to discriminate progressors from nonprogressors (Fig. 11). We are currently collecting data to test the method using other devices, including swept-source OCT as well as depth-enhanced imaging, a spectral-domain approach that puts the focal plane (point of greatest signal) lower in the image.

This study is on a population of unilateral neovascular AMD eyes that have a high risk of conversion. Therefore, studying the nonprogressors and progressors in this enriched cohort allowed us to better target the pathologic area. As this is the case, however, it is not known how the models and results would generalize to patients with bilateral early/intermediate AMD, who constitute the majority of the at-risk population. Again, this is an interesting avenue of research that we would also like to look at in more detail.

To conclude, we report that a deep learning CNN with layer segmentation-based preprocessing shows strong predictive power with respect to the progression of early/intermediate AMD to advanced AMD. Such adjunct analysis could be useful in, for example, setting the frequency of patient visits and guiding interventions.

References

1. Wong WL, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health.* 2014;2:106–116.

2. de Oliveira Dias JR, Zhang Q, Garcia JMB, et al. Natural history of subclinical neovascularization in nonexudative age-related macular degeneration using swept-source OCT angiography. *Ophthalmology.* 2018;125:255–266.

3. Hanutsaha P, Guyer DR, Yannuzzi LA, et al. Indocyanine-green videoangiography of drusen as a possible predictive indicator of exudative maculopathy. *Ophthalmology.* 1998;105:1632–1636.

4. Schneider U, Gelisken F, Inhoffen W, Kreissig I. Indocyanine green angiographic findings in fellow eyes of patients with unilateral occult neovascular age-related macular degeneration. *Int Ophthalmol.* 1997;21:79–85.

5. Chung CY, Tang HHY, Li SH, Li KKW. Differential microvascular assessment of retinal vein occlusion with coherence tomography angiography and fluorescein angiography: a blinded comparative study. *Int Ophthalmol.* 2018;38:1119–1128.

6. Hirano T, Kakihara S, Toriyama Y, Nittala MG, Murata T, Sadda S. Wide-field en face swept-source optical coherence tomography angiography using extended field imaging in diabetic retinopathy. *Br J Ophthalmol.* 2018;102:1199–1203.

7. Roisman L, Zhang Q, Wang RK, et al. Optical coherence tomography angiography of asymptomatic neovascularization in intermediate age-related macular degeneration. *Ophthalmology.* 2016;123:1309–1319.

8. de Sisternes L, Simon N, Tibshirani R, Leng T, Rubin DL. Quantitative SD-OCT imaging biomarkers as indicators of age-related macular degeneration progression. *Invest Ophthalmol Vis Sci.* 2014;55:7093–7103.

9. Niu S, de Sisternes L, Chen Q, Rubin DL, Leng T. Fully automated prediction of geographic atrophy growth using quantitative spectral-domain optical coherence tomography biomarkers. *Ophthalmology.* 2016;123:1737–1750.

10. Bogunovic H, Waldstein SM, Schlegl T, et al. Prediction of anti-VEGF treatment requirements in neovascular AMD using a machine learning approach. *Invest Ophthalmol Vis Sci.* 2017; 58:3240–3248.

11. Schmidt-Erfurth U, Waldstein SM, Klimscha S, et al. Prediction of individual disease conversion in early AMD using artificial intelligence. *Invest Ophthalmol Vis Sci.* 2018;59:3199–3208.

12. Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics. ETDRS report number 7. *Ophthalmology*. 1991;98:741–756.

13. Boser B, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational Learning Theory*. Pittsburgh, PA: ACM; 1992:144–152.

14. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*. 2014:1409–1556.

15. Manjunath V, Goren J, Fujimoto JG, Duker JS. Analysis of choroidal thickness in age-related macular degeneration using spectral-domain optical coherence tomography. *Am J Ophthalmol*. 2011;152:663–668.

16. Rattani A, Derakhshania R. On fine-tuning convolutional neural networks for smartphone based ocular recognition. *IJCB*; 2017:762–767.

17. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv*. 2012:1207.0580.

18. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. *arXiv*. 2013:1311.2901.

19. Novais EA, Adhi M, Moult EM, et al. Choroidal neovascularization analyzed on ultrahigh-speed swept-source optical coherence tomography angiography compared to spectral-domain optical coherence tomography angiography. *Am J Ophthalmol*. 2016;164:80–88.

20. Mukkamala SK, Costa RA, Fung A, Sarraf D, Gallego-Pinazo R, Freund KB. Optical coherence tomographic imaging of sub-retinal pigment epithelium lipid. *Arch Ophthalmol*. 2012; 130:1547–1553.

21. Querques G, Georges A, Ben Moussa N, Sterkers M, Souied EH. Appearance of regressing drusen on optical coherence tomography in age-related macular degeneration. *Ophthalmology*. 2014;121:173–179.

22. Sato T, Kishi S, Watanabe G, Matsumoto H, Mukai R. Tomographic features of branching vascular networks in polypoidal choroidal vasculopathy. *Retina*. 2007;27:589–594.

23. Spaide RF. Enhanced depth imaging optical coherence tomography of retinal pigment epithelial detachment in age-related macular degeneration. *Am J Ophthalmol*. 2009;147: 644–652.

24. Gossage KW, Tkaczyk TS, Rodriguez JJ, Barton JK. Texture analysis of optical coherence tomography images: feasibility for tissue classification. *J Biomed Opt*. 2003;8:570–575.

25. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24:1342–1350.

26. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. MICCAI: Springer; 2015:234–241.

27. Cruz-Herranz A, Balk LJ, Oberwahrenbrock T, et al. The APOSTEL recommendations for reporting quantitative optical coherence tomography studies. *Neurology*. 2016;86:2303–2309.