AMERICAN ACADEMY OF OPHTHALMOLOGY®

Deep Learning to Predict the Future Growth of Geographic Atrophy from Fundus Autofluorescence

Anish Salvi, MS,[1] Julia Cluceru, PhD,[1] Simon S. Gao, PhD,[1] Christina Rabe, PhD,[1] Courtney Schiffman, PhD,[1] Qi Yang, PhD,[1] Aaron Y. Lee, MD, MSCI,[2] Pearse A. Keane, MD, FRCOphth,[3] Srinivas R. Sadda, MD,[4] Frank G. Holz, MD,[5] Daniela Ferrara, MD, PhD,[1] Neha Anegondi, MTech[1]

Purpose: The region of growth (ROG) of geographic atrophy (GA) throughout the macular area has an impact on visual outcomes. Here, we developed multiple deep learning models to predict the 1-year ROG of GA lesions using fundus autofluorescence (FAF) images.

Design: In this retrospective analysis, 3 types of models were developed using FAF images collected 6 months after baseline to predict the GA lesion area (segmented lesion mask) at 1.5 years, FAF images collected at baseline and 6 months to predict the GA lesion at 1.5 years, and FAF images collected 6 months after baseline to predict the GA lesion at 1 and 1.5 years. The 1-year ROG from the 6-month visit was derived by taking the difference between the GA lesion area (segmented lesion mask) at the 1.5-year and 6-month visits.

Participants: Patients enrolled in the following lampalizumab clinical trials and prospective observational studies: NCT02247479, NCT02247531, NCT02479386, and NCT02399072.

Methods: Datasets of study eyes from 597 patients were split into model training (310), validation (78), and test sets (209), stratified by baseline or initial lesion area, lesion growth rate, foveal involvement, and focality. Deep learning experiments were performed using the 2-dimensional U-Net; whole-lesion and multiclass models were developed.

Main Outcome Measures: The performance of the models was evaluated by calculating the Dice score, coefficient of determination (R^2), and the squared Pearson correlation coefficient (r^2) between the true and derived GA lesion 1-year ROG.

Results: The model using baseline and 6-month FAF images to predict GA lesion enlargement at 1.5 years had the best performance for the derived 1-year ROG. Mean Dice scores were 0.73, 0.68, and 0.70 in the training, validation, and test sets, respectively. The R^2 (0.77, 0.53, and 0.79) and r^2 (0.83, 0.61, and 0.79) showed similar trends across the 3 sets.

Conclusions: These findings show the potential of using baseline and/or 6-month visit FAF images to predict 1-year GA ROG using a deep learning approach. This work could potentially help support decision-making in clinical trials and more informed treatment decisions in clinical practice.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science 2025;5:100635*

Supplemental material available at www.ophthalmologyscience.org.

Geographic atrophy (GA) is an advanced form of age-related macular degeneration (AMD) that leads to loss of visual function.[1] It is defined by the presence of sharply demarcated atrophic lesions and is characterized by the loss of photoreceptors, retinal pigment epithelium (RPE), and choriocapillaris.[1] Approximately 40% of eyes with GA are considered legally blind,[2–4] with 29% losing at least 6 lines of vision on an Early Treatment Diabetic Retinopathy Study chart by 4 years.[5] Geographic atrophy lesions can be visualized by multiple imaging modalities, including color fundus photography, fluorescein angiography, fundus autofluorescence (FAF), near-infrared reflectance, and OCT.[6–10] Fundus autofluorescence is used in clinical trials to quantify GA lesion area and is based on topographic mapping of intrinsic fluorophores within lipofuscin granules in the postmitotic RPE.[8,9] The change in GA lesion area (mm^2) derived from FAF images over a defined period of time (i.e., GA growth rate) has been used as the primary anatomical endpoint for GA clinical trials.[10,11] As such, FAF remains an important noninvasive

https://doi.org/10.1016/j.xops.2024.100635
ISSN 2666-9145/24

tool for identifying and tracking GA progression over time and for assessing response to therapeutics.

Geographic atrophy is a slowly progressing disease, with GA growth rates varying among individuals; consequently, clinical trials in GA need to be relatively large to account for disease heterogeneity.[1,11] Accurate and personalized prediction of GA growth can be used for covariate adjustment, patient stratification, and potentially for patient enrichment to increase the confidence in interpreting the results of clinical trials.[11] Accurate prediction of the magnitude and direction of GA growth may also be relevant for patient counseling in clinical practice. In addition to predicting GA growth rate, predicting the location of the GA region of growth (ROG) may also be clinically relevant. This is because the topographic location of the GA lesion has an impact on visual function, especially when the central macular 1-mm subfield is affected by the lesion.[12,13] Although some studies have accurately predicted GA growth rates,[14–16] studies designed to predict the future GA ROG have been less successful.[17,18] In this study, we used data from lampalizumab phase 3 trials[19] and observational studies[20] to develop multiple state-of-the-art deep learning convolution neural network (CNN) models to accurately predict the future GA ROG using single or multiple visit FAF images. The objective of the data strategy and modeling approaches was to maximize the model's performance for 1-year ROG.

Methods

Datasets

Multiple models were developed and tested on retrospective data from study eyes of patients with GA enrolled in lampalizumab phase 3 clinical trials (Chroma [NCT02247479]; Spectri [NCT02247531]),[19] and observational studies (Proxima A [NCT02479386]; Proxima B [NCT02399072]).[20] Study eye inclusion criteria for these studies have been previously described.[19,20] Briefly, study eyes were required to have well-demarcated area(s) of GA secondary to AMD with no evidence of prior or active choroidal neovascularization and a total GA lesion size of 2.54 to 17.78 mm^2 (1–7 disc areas) residing completely within the blue-light FAF imaging field (field 2, 30 degrees, image centered on the fovea), with perilesional banded or diffuse hyperautofluorescence patterns on FAF images. If the GA was multifocal, ≥1 focal lesion must have been ≥1.27 mm^2 (≥0.5 disc areas). The trials adhered to the Declaration of Helsinki and were Health Insurance Portability and Accountability Act compliant. Protocols were approved by the institutional review board at each study site before the trials started. All patients provided written informed consent for future medical research and analyses.

The current study analyzed macular (field 2) 30 degree images (768 × 768 pixels or 1536 × 1536 pixels) from study eyes captured using the Spectralis HRA (Heidelberg Engineering, Inc). Because no treatment effect was observed on lesion growth rates in the phase 3 trials, data from all treatment arms were pooled for this analysis. Geographic atrophy lesion areas were graded at a central reading center by 2 trained graders (Grader 1 [G1] and Grader 2 [G2]), with an adjudicator in case of disagreements, on FAF images using RegionFinder software[a] (Heidelberg Engineering, Inc).[9] Before starting the grading process for follow-up visit images, the graders used automatic registration capability of RegionFinder software[a] to longitudinally register the follow-up visit FAF images to the screening visit FAF images.

Preprocessing

Training, Validation, and Test Sets. FAF images of study eyes were obtained at 6-month intervals from screening (SCR) to 2 years, with a standard deviation of approximately 5 days. Even though GA lesions were graded on all FAF images by 2 graders (Table 1), annotations or correspondent segmentation masks from both G1 (597 patients) and G2 (199 patients) (examples of which can be found in Fig 1) were not always available in a usable format. Consequently, this study used only G1 annotations for model development and testing. For each patient, we looked for the availability of 4 longitudinal visit FAF images and corresponding annotations for a 1.5-year period at 6-month intervals corresponding to trial visits. As a strategy to maximize the use of the limited expert annotations available, we looked at all possible combinations of time points available over the study duration. We encoded the time points (T) as T1 (SCR, week [W] 24 or W48 → baseline), T2 (W24, W48, or W72 → 6 months), T3 (W48, W72, or W96 → 1 year), and T4 (W72, W96, or W120 → 1.5 years) (Fig 2). In summary, T1 refers to the baseline visit, T2 to the 6-month visit, T3 to the 1-year visit, and T4 to the 1.5-year visit.

The performance of the CNN models is dependent on the quality of the longitudinal registration of FAF images and corresponding annotations. Consequently, additional curation steps were taken to avoid longitudinal registration errors. Briefly, we performed pairwise comparisons between annotations at 1) T1 and T2, 2) T2 and T3, 3) T3 and T4, and 4) T2 and T4. For each of these combinations, we obtained the Dice score coefficient (DSC), the change in lesion area (%), and the lesion growth rate (mm^2/year). In addition, we determined whether the lesion was extending beyond the FAF image field of view (true/false). To be included in the analysis, each of the pairwise comparisons must have achieved DSC > 0.7, −10% < change in lesion area (%) <100%, and all lesions must have been located within the FAF image field. Cases that achieved 0.7 < DSC < 0.9 were manually reviewed for registration errors and peripapillary lesions (lesions around the optic disc that may or may not be connected to the GA lesion and extend beyond the field of view); if observed, the patient was excluded from the study. The thresholds for the inclusion of annotations in the analysis were determined manually through visual checks. Although there were limited annotations in a usable format, some patients had annotations available at all visits from SCR to 2 years (SCR, W24, W48, W72, and W96). For these patients, there were 2 possible combinations of time points T1 → T2 → T3 → T4 over 1.5 years: SCR → W24 → W48 → W72 and W24 → W48 → W72 → W96. In such cases, only one combination was used so that no patients were duplicated across datasets. Also, in the cases where 2 or more combinations of timepoints were available, the "baseline" timepoint closest to SCR was chosen. All FAF images and corresponding annotations were resized to 768 × 768 pixels via nearest neighbor interpolation. Each FAF image was z-normalized[21] (a process used to normalize every pixel in an image so the mean of all values is 0 and the standard deviation is 1).

Table 1. Distribution of Patients across Training, Validation, and Test Sets

Dataset Description	Training				Validation				Test			
Grader # (No. of Patients)	*G1 (310)*		*G2 (128)*		*G1 (78)*		*G2 (28)*		*G1 (209)*		*G2 (43)*	
Foveal involvement												
Nonsubfoveal	141	45.5%	55	43.0%	35	44.9%	17	60.7%	100	47.8%	17	39.5%
Subfoveal	169	54.5%	73	57.0%	43	55.1%	11	39.3%	109	52.2%	26	60.5%
Focal status												
Multifocal	238	76.8%	96	75.0%	61	78.2%	20	71.4%	165	78.9%	34	79.1%
Unifocal	72	23.2%	32	25.0%	17	21.8%	8	28.6%	44	21.1%	9	20.9%
Initial lesion area												
Large	155	50.0%	64	50.0%	39	50.0%	14	50.0%	104	49.8%	21	48.8%
Small	155	50.0%	64	50.0%	39	50.0%	14	50.0%	105	50.2%	22	51.2%
Lesion growth												
Fast	156	50.3%	65	50.8%	38	48.7%	13	46.4%	104	49.8%	21	48.8%
Slow	154	49.7%	63	49.2%	40	51.3%	15	53.6%	105	50.2%	22	51.2%
Visits												
SCR–W24–W48–W72	77	24.8%	33	25.8%	18	23.1%	8	28.6%	11	5.3%	2	4.7%
W24–W48–W72–W96	209	67.4%	84	65.6%	58	74.4%	18	64.3%	198	94.7%	41	95.3%
W48–W72–W96–W120	24	7.7%	11	8.6%	2	2.6%	2	7.1%	0	0.0%	0	0.0%
Eye												
OD	149	48.1%	57	44.5%	45	57.7%	18	64.3%	107	51.2%	22	51.2%
OS	161	51.9%	71	55.5%	33	42.3%	10	35.7%	102	48.8%	21	48.8%
Study												
Spectri	124	40.0%	37	28.9%	33	42.3%	11	39.3%	114	54.5%	18	41.9%
Chroma	93	30.0%	34	26.6%	25	32.1%	6	21.4%	95	45.5%	25	58.1%
Proxima A	42	13.5%	33	25.8%	9	11.5%	5	17.9%	0	0.0%	0	0.0%
Proxima B	51	16.5%	24	18.8%	11	14.1%	6	21.4%	0	0.0%	0	0.0%

G1 = Grader 1; G2 = Grader 2; OD = right eye; OS = left eye; SCR = screening; W = week.

Once we had the unique set of patients, we used the previously generated splits at the patient level[14] to divide the dataset (597 eyes) into development (388 eyes) and test sets (209 eyes). The development set was further divided into training (310 eyes) and validation (78 eyes). All the splits were stratified by baseline or initial lesion area (median: 9.03 mm^2), lesion growth rate (median: 1.61 mm^2), foveal involvement (subfoveal or nonsubfoveal lesions), and focality (multifocal or unifocal lesions), ensuring the distributions were similar across splits. The number of eyes in the development and test sets were different from those in our previous publication[14] due to the limited annotations available. The number of eyes included in the test set was higher because, by random chance, a higher percentage of the test set had annotations available. To increase the size of the development set, we also included Proxima B annotations. We made a conscious decision to use the previous splits[14] to have a fair comparison of the GA progression prediction approaches used in this study with our previous work. Fig 3 shows the data flow from the clinical trial data to the training, validation, and test sets. Table 1 shows the distribution of patients across each set.

Model Development

Deep learning experiments were performed using the 2-dimensional U-Net CNN.[22] The U-Net consists of an encoder (contracting path), which converts a medical image into feature maps, and a decoder (expanding path), which converts the feature maps into a probability map of equal size to the input image. In other words, the encoder extracts image features of different spatial resolutions, which are in turn used by the decoder to derive an accurate segmentation mask.

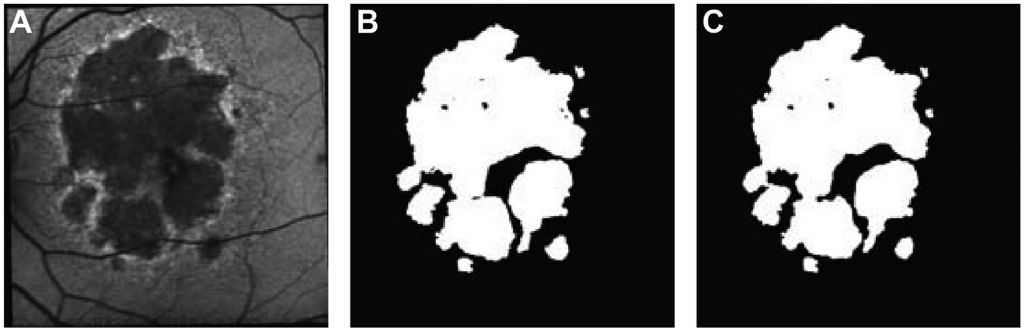

Figure 1. An example (**A**) fundus autofluorescence image; (**B**) Grader 1 annotation; and (**C**) Grader 2 annotation from a study eye with geographic atrophy.

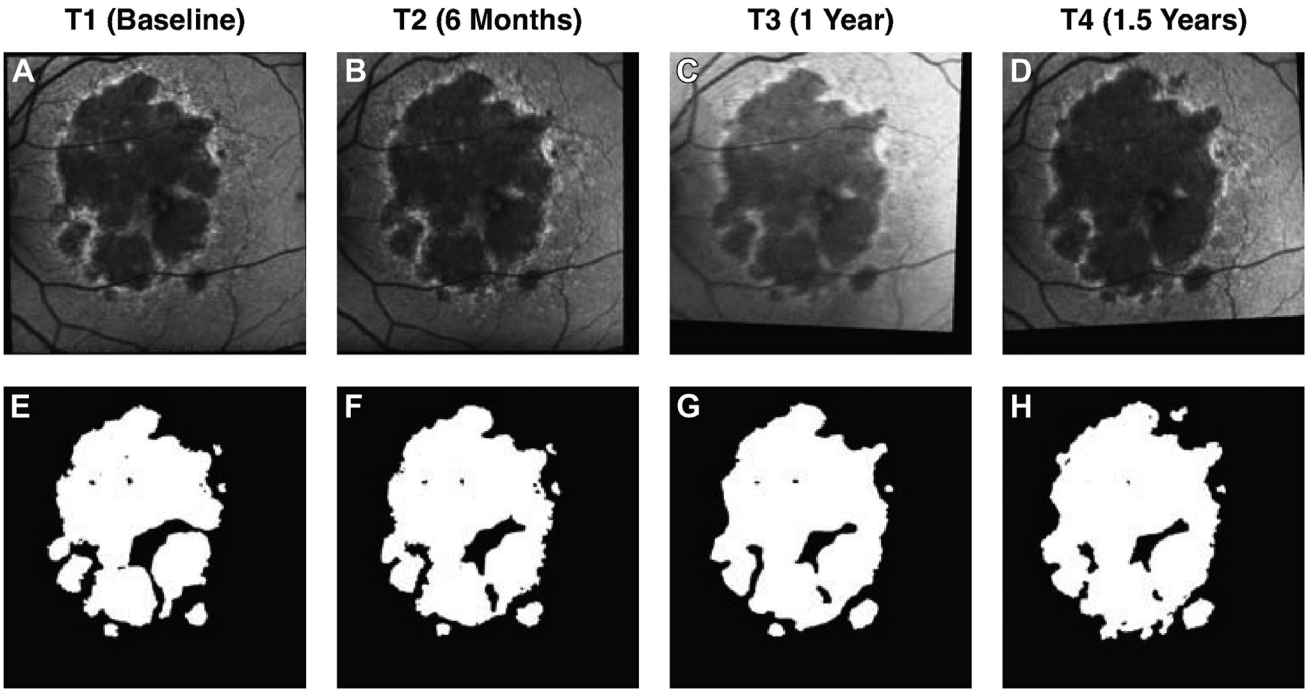

Figure 2. Fundus autofluorescence images showing geographic atrophy (GA) lesion progression at (**A**) baseline (time point [T] 1); (**B**) 6 months (T2); (**C**) 1 year (T3); and (**D**) 1.5 years (T4). (**E**–**H**) Corresponding GA lesion annotations.

The same encoder and decoder specifications were used for all the models developed in this study.[23] The encoder used a 34-layer residual neural network (ResNet)[24] backbone to extract features at different resolutions, an encoder depth of 4, and pretrained ImageNet weights, where the encoder depth signifies the number of stages and the feature size decreases with each additional stage.[25] The decoder uses batch normalization between the convolutional and activation layers and has a depth of 4 with decoder channels of (128, 64, 32, and 16). Training parameters (data augmentation, scheduled learning rate reductions on loss plateaus, batch size 4, maximum epochs 200 [with early stopping], optimizer AdamW[26]) were trained on G1 annotations exclusively, and similar parameters were used across experiments. Data augmentation consisted of randomized horizontal and vertical flipping, addition of noise, and minor translation and scaling. Three different whole-lesion models and 3 multiclass models were developed using PyTorch.[27] Hyperparameter optimization was conducted across learning rate (1e-3, 1e-4, 5e-4, 1e-5), weight decay (1e-2, 1e-4), and loss functions[28] via a grid search, with 48 separate models being developed for each model type (6 total types). The search was designed to maximize the GA ROG DSC, which was calculated during postprocessing.

Whole-Lesion Model

The T used for whole-lesion model development were: T1 (baseline), T2 (6 months), T3 (1 year), and T4 (1.5 years). The models were trained on the T4 whole-lesion ground truth. The 3 whole-lesion models are described as: 1) Model 1 or Simple U-Net, which infers the T4 whole lesion from the T2 FAF; 2) Model 2 or multichannel U-Net, which infers the T4 whole lesion from the combination of T1 and T2 FAF images (T1 and T2 images are incorporated as 2 channels of a single image); and 3) Model 3 or Sequential Label U-Net, which infers the T3 and T4 whole lesions, respectively, from the T2 FAF image (Table 2; Fig 4). T3 and T4 images are predicted as 2 channels of a single image. The loss functions included the Dice, Dice cross-entropy, Tversky (a = 0.6, b = 0.4), Tversky (a = 0.4, b = 0.6), and focal losses (Supplementary Table S3, available at https://www.ophthalmologyscience.org/).

Whole-Lesion Postprocessing

For model inference, we applied the sigmoid function to each prediction and set the threshold at >0.5 to define the final resultant prediction. We also experimented with Otsu threshold but did not see any performance improvement.[29] The 1-year ROG of GA was derived by taking the difference between the predicted T4 lesion and grader T2 lesion annotation (Supplementary Fig S5; Supplementary Table S3, available at https://www.ophthalmologyscience.org/). Thus, the 3 whole-lesion models are Models 1, 2, and 3.

Multiclass Model

The Ts used for the multiclass models were: T1 (baseline), T2 (6 months), T3 (1 year), and T4 (1.5 years). Thereby, T4−T2 would be the 1-year ROG of GA, and T3−T2 would be the 6-month ROG. The multiclass models were trained on a multiclass ground truth. The 3 multiclass segmentation models are described as (1) Model 4 or Simple U-Net, which infers the classes T2 whole lesion and T4−T2 ROG from the T2 FAF; (2) Model 5 or Multichannel U-Net, which infers the classes T2 whole lesion and T4−T2 ROG from the combination of T1 and T2 FAF images (T1 and T2 images are incorporated as 2 channels of a single image); and (3) Model 6 or Sequential U-Net, which infers the classes T2 whole lesion, T4−T3 ROG, and T3−T2 ROG from the T2 FAF (Table 2; Fig 6). The labels T4−T2 and T2 in the case of Model 4 and Model 5 and T4−T3, T3−T2, and T2 in the case of Model 6 are different classes in the one channel of a single image. The loss functions included the Dice, generalized Dice, generalized Dice focal, Dice

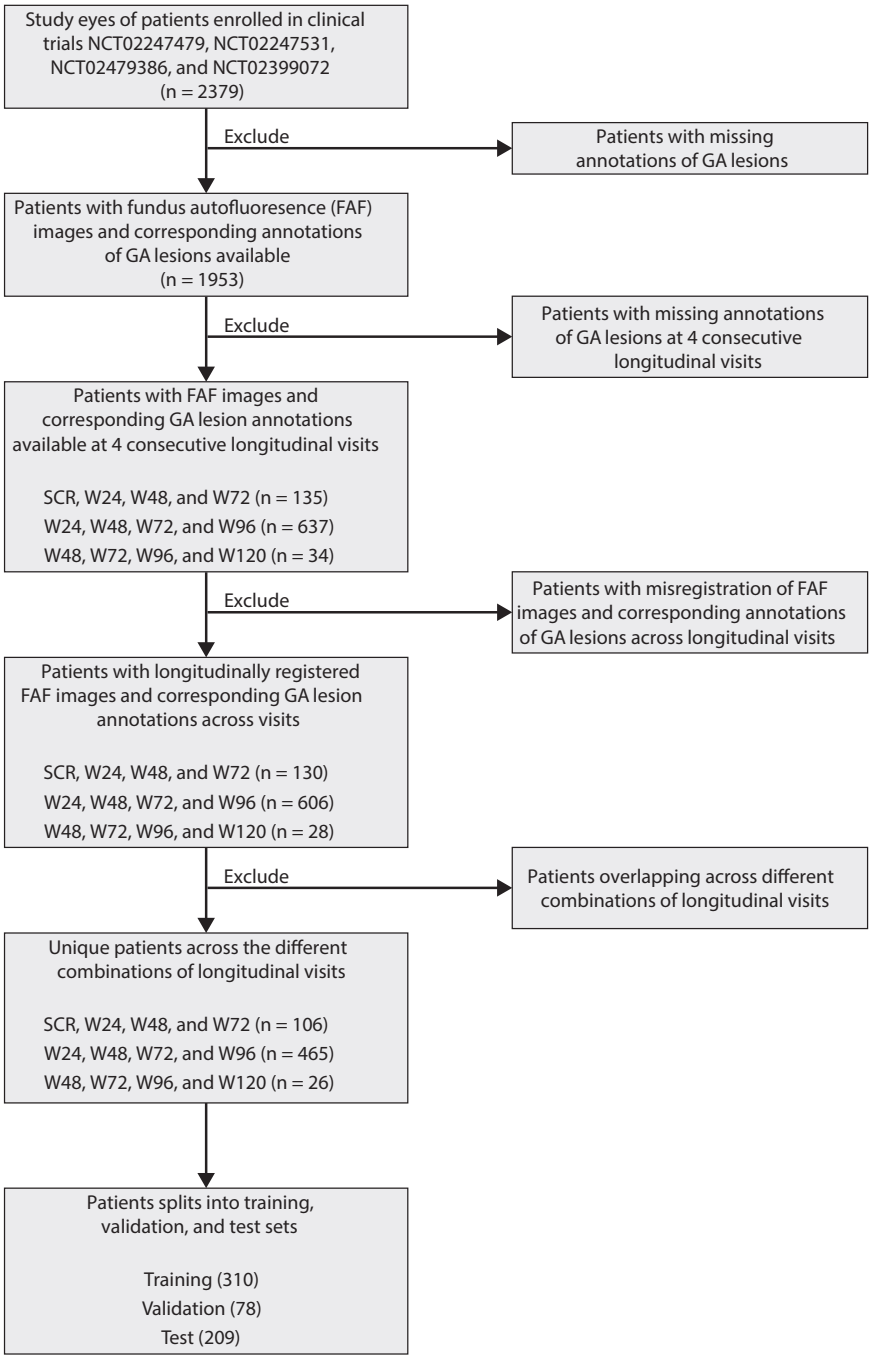

Figure 3. Data flow from clinical trials to training, validation, and test sets. FAF = fundus autofluorescence; GA = geographic atrophy; SCR = screening; W = week.

cross-entropy, Dice focal, and focal losses (Supplementary Table S3, available at https://www.ophthalmologyscience.org/). The rationale for using generalized losses as opposed to Tversky was to explore whether class imbalance impacted the model's calibration because the ROGs were smaller than the whole lesions.

Multiclass Postprocessing

Method 1 (Multiclass). For model inference, we applied the softmax function channel-wise to each prediction with the highest probability defining the final prediction. The background was considered a separate class as during training. For the first 2 multiclass models, the T4—T2 ROG of GA was predicted by the models, and no further postprocessing was required (Supplementary Fig S7, available at https://www.ophthalmologyscience.org/). For the third multiclass model, the T4—T2 ROG was generated by adding the T3—T2 ROG and the T4—T3 ROG (Supplementary Fig S7, available at https://www.ophthalmologyscience.org/). Thus, the 3 multiclass models with this method of postprocessing were Model 4 multiclass, Model 5 multiclass, and Model 6 multiclass.

Table 2. Summary of the Models Developed

Model #	Type	Description	FAF	Segmentation	Motivation
1	Whole lesion	Simple U-Net	T2	T4	Standard approach
2		Multichannel U-Net	T1 and T2	T4	Incorporating temporal information
3		Sequential growth U-Net	T2	T3 + T4	Introducing more targets
4	Multiclass	Simple U-Net	T2	T2 + T4−T2	Standard approach
5		Multichannel U-Net	T1 and T2	T2 + T4−T2	Incorporating temporal information
6		Sequential growth U-Net	T2	T2, T4−T3 + T3−T2	Introducing more targets

FAF = fundus autofluorescence; T = time point; T1 = baseline; T2 = 6 months; T3 = 1 year; T4 = 1.5 years.

Method 2 (Multiclass Whole Lesion). T4 whole lesion was the sum of the T2 lesion prediction and the T4−T2 ROG of GA prediction ([T3−T2] + [T4−T3] for the third multiclass model) (Supplementary Fig S8, available at https://www.ophthalmologyscience.org/; Table 4). The ROG was determined based on the difference between the derived T4 whole lesion and the G2 T2 lesion annotation. This method was implemented to reduce the variability in the ROG prediction coming from the predicted T2 lesion. Thus, the 3 multiclass models with this method of postprocessing were Model 4 multiclass whole lesion, Model 5 multiclass whole lesion, and Model 6 multiclass whole lesion.

The goal of the data strategy and modeling approaches used in this work was to maximize the model's performance for 1-year ROG. Therefore, we report our best-performing models with the corresponding hyperparameters across all postprocessing methods according to the T4−T2 ROG DSC with respect to the annotations of G1 (Supplementary Table S5, available at https://www.ophthalmologyscience.org). We also provide the squared Pearson correlation coefficient (r^2) of the observed versus predicted ROG (Table 4). Some of our models were miscalibrated, resulting in lower coefficient of determination (R^2) values. To fix this, we estimated a linear calibration function (using population least squares linear regression of observed versus predicted 1-year ROG) from the validation set. We then transformed the predictions in the test set with this calibration function to obtain recalibrated predictions. It is worth noting that the linear calibration function was very similar to estimated calibration curves (fit either with a generalized additive model or with the Pool-Adjacent-Violators Algorithm) in the validation set, suggesting that linear calibration was acceptable.[30] When predictions are transformed via the calibration line, R^2 is equal to the squared Pearson correlation between the predictions and the observed outcomes. The calibration was performed only on the GA lesion area (mm^2) and quantified from the predicted ROG and not on the pixel intensities.

Results

Table 4 shows uncalibrated and calibrated R^2 values for all models. The inter-grader agreement between G1 and G2 annotations are shown in Table 6. This can serve as an upper limit of performance of the prediction models (i.e., a prediction of the 1-year ROG of GA will not exceed the agreement in the annotated 1-year ROG between graders). According to Table 4, uncalibrated R^2, calibrated R^2, r^2, and Dice scores for validation (0.53, 0.61, 0.61, and 0.68) and

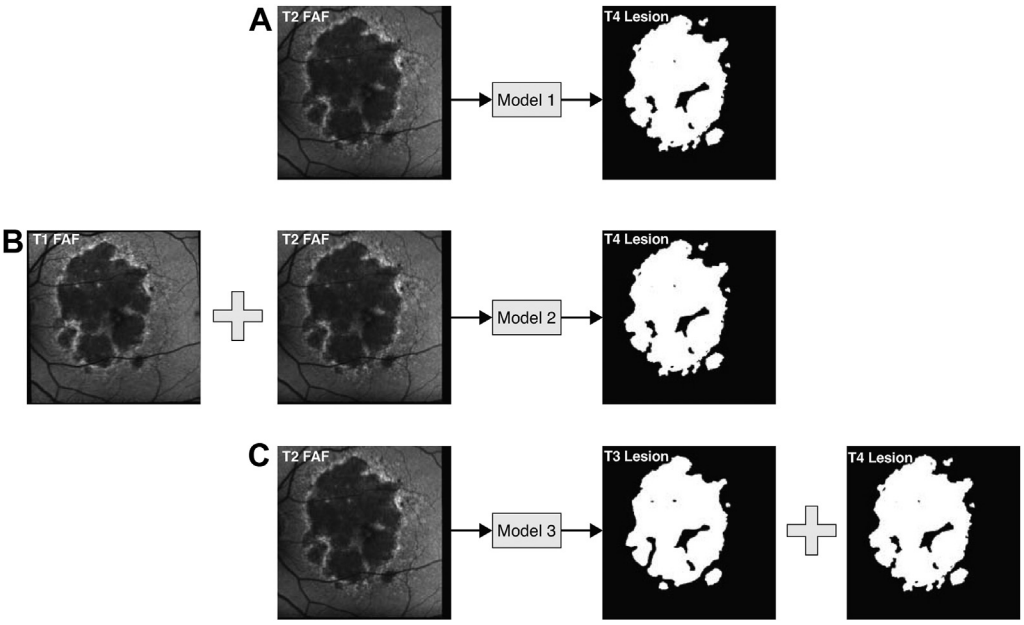

Figure 4. (**A**) Model 1 for time point (T) 4 (1.5 years) whole lesion segmentation from 6 months (T2) fundus autofluorescence (FAF); (**B**) Model 2 for T4 whole lesion segmentation from baseline (T1) and T2 FAF; (**C**) Model 3 for 1 year (T3) and T4 whole lesion segmentation from T2 FAF.

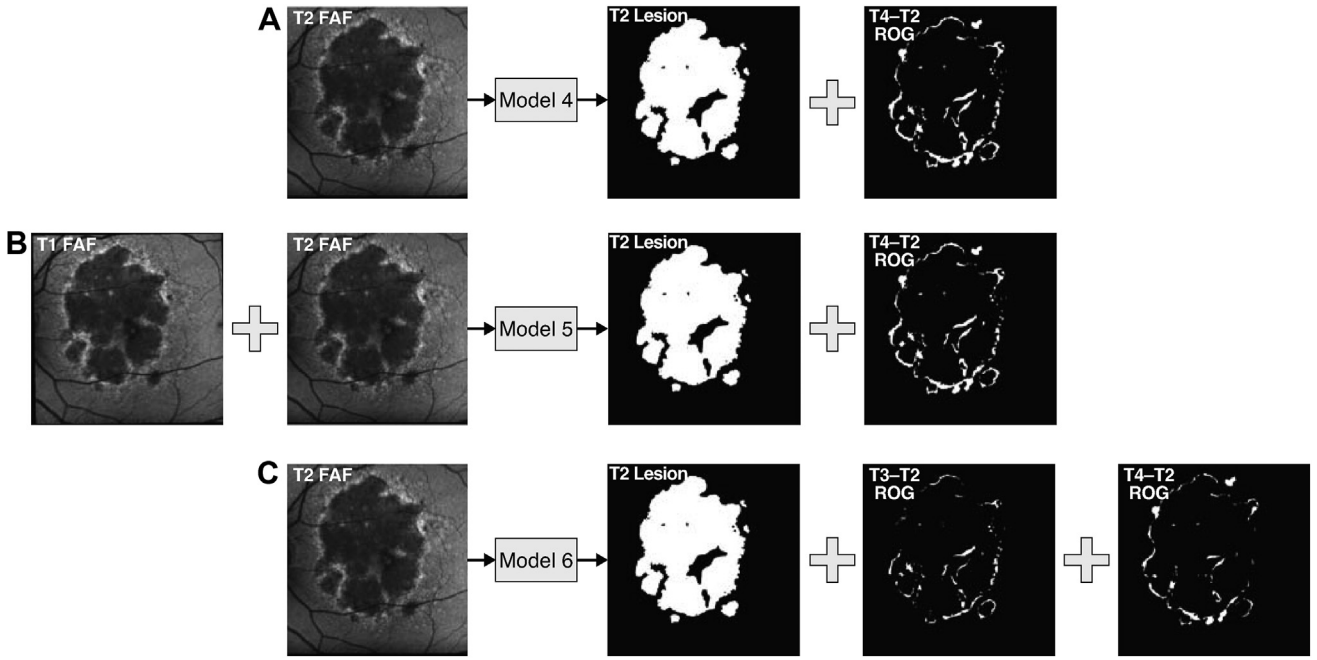

Figure 6. **A**) Model 4 for T2 whole lesion and T4–T2 ROG predictions from T2 FAF; (**B**) Model 5 for T2 whole lesion and T4–T2 ROG predictions from T1 and T2 FAF; (**C**) Model 6 for T2 whole lesion, T3–T2 ROG, and T4–T3 ROG predictions from T2 FAF. FAF = fundus autofluorescence; ROG = region of growth; T = time point; T1 = baseline; T2 = 6 months; T3 = 1 year; T4 = 1.5 years; T4–T2 ROG = 1 year ROG; T3–T2 ROG = 6-month ROG.

test sets (0.79, 0.73, 0.79, and 0.70) for Model 2 using T1 and T2 FAF images to predict GA lesion at T4 had the best performance for the derived 1-year ROG. Supplementary Table S7 (available at www.ophthalmologyscience.org) shows the precision and recall scores for all the models. Supplementary Table S8 (available at www.ophthalmologyscience.org) shows the r^2 of the square root transformed ROG between G1 and the models. The performance of the models was consistent between the treatment arms.

Figure 9 shows overlays of ground truth and predicted 1-year ROG over T2 FAF images for the best-performing model of each type. For all models, the 1-year ROG ground truth was 2.38 mm^2, and 1-year predicted ROG was 2.18 mm^2 for Model 2 (Dice 0.72), 3.24 mm^2 for Model 5 multiclass (Dice 0.58), and 3.09 mm^2 for Model 5 multiclass whole lesion (Dice 0.69). The scatter plots of the predicted 1-year ROG area versus ground truth 1-year ROG area of the best-performing model of each type (Model 2, Model 5 multiclass, Model 5 multiclass whole

Table 4. Summary of Results for 1-Year Lesion Growth (T4–T2 ROG) with Respect to G1 Annotations

1-Year Lesion Growth (G1 ROG, T4–T2) (No. of Patients)		Training (310)			Validation (78)				Test (209)			
Model #	*Method*	*DSC*	R^2	*Pearson* r^2	*DSC*	$^*R^2$	$^{\&}R^2$	*Pearson* r^2	*DSC*	$^*R^2$	$^{\&}R^2$	*Pearson* r^2
1	Whole lesion	0.68 ± 0.10	0.27	0.61	0.65 ± 0.10	0.22	0.56	0.56	0.66 ± 0.10	0.32	0.43	0.46
2		0.73 ± 0.08	0.77	0.83	0.68 ± 0.09	0.53	0.61	0.61	0.70 ± 0.09	0.79	0.73	0.79
3		0.68 ± 0.09	0.03	0.53	0.65 ± 0.11	<0	0.49	0.49	0.66 ± 0.10	<0	0.35	0.37
4	Multiclass	0.54 ± 0.10	<0	0.46	0.52 ± 0.11	<0	0.54	0.54	0.53 ± 0.11	0.12	0.39	0.42
5		0.57 ± 0.10	0.03	0.62	0.55 ± 0.12	<0	0.58	0.58	0.57 ± 0.11	0.37	0.54	0.61
6		0.56 ± 0.11	<0	0.52	0.52 ± 0.12	<0	0.54	0.54	0.54 ± 0.11	0.001	0.42	0.46
4	Multiclass whole lesion	0.69 ± 0.09	0.44	0.59	0.65 ± 0.10	0.37	0.56	0.56	0.67 ± 0.10	0.38	0.43	0.45
5		0.70 ± 0.10	<0	0.62	0.68 ± 0.10	0.08	0.62	0.62	0.70 ± 0.10	0.34	0.62	0.65
6		0.67 ± 0.11	0.08	0.55	0.64 ± 0.11	0.08	0.56	0.56	0.66 ± 0.11	0.11	0.37	0.39

DSC = Dice score coefficient; G1 = Grader 1; r^2 = correlation coefficient; $^{\&}R^2$ = calibrated coefficient of determination; $^*R^2$ = uncalibrated coefficient of determination; ROG = region of growth; T = time point; T1 = baseline; T2 = 6 months; T3 = 1 year; T4 = 1.5 years.

Table 6. Summary of Grader Agreement

1-Year Lesion Growth (T4–T2 ROG) (No. of Patients)	Training (128)			Validation (28)			Test (43)		
G1 and G2	DSC	R^2	Pearson r^2	DSC	R^2	Pearson r^2	DSC	R^2	Pearson r^2
Grader agreement	0.71 ± 0.12	0.88	0.88	0.72 ± 0.14	0.8	0.85	0.73 ± 0.10	0.9	0.90

DSC = Dice score coefficient; G1 = Grader 1; G2 = Grader 2; r^2 = correlation coefficient; R^2 = coefficient of determination; ROG = region of growth; T = time point; T1 = baseline; T2 = 6 months; T3 = 1 year; T4 = 1.5 years.

lesion) can be found in Supplementary Fig S10 (available at https://www.ophthalmologyscience.org/).

Discussion

This study demonstrates the feasibility of using standardized FAF images to predict the 1-year GA ROG using a deep learning approach. Prediction of GA growth could be used in clinical trials for enrichment, stratification, or covariate adjustment and in clinical practice for patient counseling. In addition to GA growth, the location of GA lesions has an impact on vision.[12] Consequently, predicting the future ROG of GA lesions may be useful for identifying patients with a higher risk of vision loss. All the CNN models developed in this study were able to predict the 1-year ROG (with and without postprocessing) of GA lesions, although with somewhat variable performance, and the

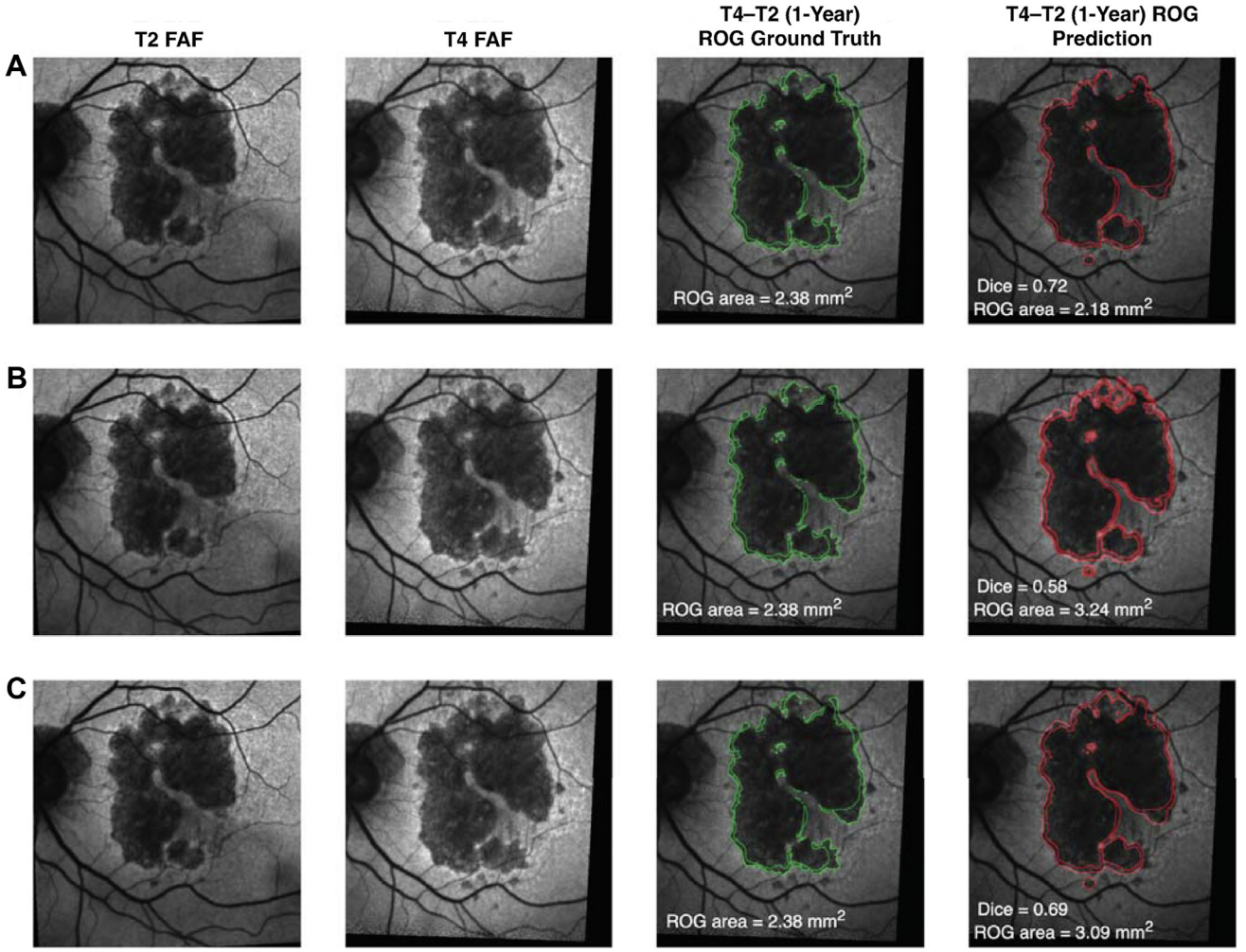

Figure 9. Overlays of ground truth (Grader 1 [G1] annotations) and predicted 1-year ROG on an example from test set of (**A**) Model 2; (**B**) Model 5 multiclass; and (**C**) Model 5 multiclass whole lesion. DSC = Dice score coefficient; FAF = fundus autofluorescence; ROG = region of growth; T = time point; T1 = baseline; T2 = 6 months; T3 = 1 year; T4 = 1.5 years; T4–T2 = 1-year ROG.

incorporation of temporal information in a channel-wise fashion was more successful at predicting 1-year ROG than other models. On comparing the 1-year ROG across the manual annotations of 2 expert graders (G1 and G2) in a subset of study eyes (Table 6), the DSC was comparable to the CNN models. However, because graders did not predict the ROG, r^2 was better across graders than the CNN models and G1 were.

Few groups so far have attempted to predict the future ROG of GA lesions. One study used a random forest classifier to determine future ROG of GA lesions from OCT scans of 38 eyes/29 patients using leave-1-out cross-validation to evaluate 3 different scenarios.[18] Considering only the regions without evidence of GA at baseline, predicted ROG showed relatively high Dice scores, ranging from 0.72 to 0.74. Another study developed an integrated time adaptive prediction model for identifying the location of future GA growth using OCTs from 25 patients with a minimum of 3 visits per patient.[31] The average Dice indexes of the predicted whole GA lesions in 10 scenarios ranged from 0.86 to 0.92. By integrating time factors into the bidirectional long short-term memory (LSTM) models, the prediction accuracy was improved by almost 10%. More recently, longitudinal OCT data from 129 eyes of 119 patients was used to develop and validate an automatic RPE and outer retinal atrophy (RORA) progression prediction model in GA.[17] The average Dice score for segmentations at baseline was 0.85. When predicting progression from baseline OCTs, Dice scores ranged from 0.73 to 0.80 for total RORA area and from 0.46 to 0.72 for RORA growth region. It is important to note that all these studies were performed using OCT scans, whereas our models were developed using FAF images. FAF is a 2-dimensional imaging modality, whereas OCT can capture 3-dimensional images of the retina, allowing better structural identification of the lesion. This structural information can help characterize GA precursors, onset, and progression,[32] and the Classification of Atrophy Meetings group recommends OCT as the reference method for defining different atrophy subphenotypes.[33] However, multiple publications show that lesion shape-descriptive features and surrounding abnormal autofluorescence patterns in FAF are prognostic of GA progression.[16,34] In addition, FAF is the imaging modality of choice to measure the primary endpoint in GA clinical trials. The performance of the models described in our study using FAF is similar to that observed in previous studies using OCT.

We used multiple approaches to predict the 1-year ROG of GA lesions on FAF. The Ts used for modeling were: T1 (baseline), T2 (6 months), T3 (1 year), and T4 (1.5 years). Thus, T4−T2 is the 1-year ROG and T3−T2 is the 6-month ROG. The first 3 models (also called whole-lesion models as they predicted the whole GA lesion) predicted T4 (1.5 years) whole lesion using T2 (6 months) FAF only (Model 1); T4 (1.5 years) whole lesion using T1 (baseline) and T2 (6 months) FAF images (Model 2); and T3 (1 year) and T4 (1.5 years) whole lesions using T2 (6 months) FAF only (Model 3). The aim of Model 2 was to provide temporal information to the model in a channel-wise fashion to help the model learn the rate and pattern of GA lesion growth across 2 visits. The aim of Model 3 was to provide longitudinal labels to the model to help with the training process, with the thought that it will be easier for the model to learn to predict T3 (a nearer timepoint), which in turn will help with the T4 prediction. The 1-year ROG for these models was calculated by subtracting the T2 annotation by the grader from the predicted T4 whole lesions. The reason for using T2 as input for Models 1 and 3 was to provide a fair comparison with Model 2 on predicting 1-year ROG in the future (T4).

We were also interested in directly predicting the 1-year ROG instead of deriving it by subtracting the grader annotation. Consequently, we developed 3 multiclass models that predicted T2 whole lesion and T4−T2 ROG from T2 FAF only (Model 4); T2 whole lesion and T4−T2 ROG from T1 and T2 FAF images (Model 5); and T2 whole lesion, T3−T2 ROG, and T4−T3 ROG using T2 FAF only (Model 6). On comparing the predicted 1-year ROG (T4−T2) with the ground truth, we found that the performance was worse for this model than for the whole-lesion models that used T2 grader annotation to derive the ROG. We speculate that the reduced performance was due to variability in the predicted T2 lesion. To remove this variability, we derived the T4 lesion for the 3 multiclass models from the predictions and then generated the 1-year ROG using the T2 grader annotation. However, we did not see any advantage for multiclass models over whole-lesion models and found that the performance was similar in terms of DSC and other metrics. In all cases (whole lesion and multiclass/multiclass whole lesion), the model using T1 and T2 FAF images as input had the best performance, indicating that the temporal information has added value. We chose to predict T4 and not T3 because our initial explorations demonstrated that it was much easier to predict a nearer timepoint compared with a farther timepoint; e.g., predicting T2 (6 months) from T1 was easier than predicting T3 (1 year) from T1. Our goal was to have all models predict 1-year ROG from the given input. The multichannel U-Net required 2 timepoints (T1 + T2), whereas the simple and sequential U-Net models required just 1 input; therefore, to keep the comparison fair across all the models, we used T4 as the label/output. It follows that if we used T3 as output and T1 as input for simple and sequential U-Nets and T1 and T2 as inputs for multichannel U-Net, then it would have been easier for multichannel U-Net to predict T3.

There were some limitations with this study. Firstly, we observed that it was possible for the grader to annotate a few lesion-associated pixels present in a previous annotation that were not present in the follow-up annotation, assuming accurate registration. The percentage of such pixels was very small (∼2%). In other words, there was a small discrepancy between annotations from the same grader over time. However, because we were only interested in lesion growth, not potential areas of lesion shrinkage, we defined our ROG algorithm to ignore these areas when generating our labels or predictions. The timepoint combination W24-W48-W72-W96 was most available, which could have biased the models toward baseline GA area observed at the W24 timepoint as compared to SCR or other timepoints. We observed that the GA growth was very linear over the period

of 2 years, which gave us the confidence to pool data from different timepoints and treat them as equivalent; however, this could be a potential limitation for future evaluations in nonlinear datasets. In the current study, we could not quantify the location of future ROG because we did not have the fovea annotations available. Also, we did not perform any experiments to evaluate the impact of misregistration of images across different timepoints on the model performance. Next, the images included in this study were captured using the same vendor's devices (Heidelberg Engineering, Inc) had similar automatic real-time function values, and had clear enough media and sufficient image quality, as required for clinical trial eligibility screening determined by a central reading center. Further, patient distribution outside of the clinical trial data used for model development (e.g., race, FAF pattern, focality, etc) limits the generalizability of these models. Therefore, any use case for these models outside of the image vendor, image quality, and patient distribution of the lampalizumab clinical trial data needs additional assessment.

Several steps can potentially be taken to improve these models. For example, using multimodal data (e.g., near-infrared data) in addition to FAF images could help the models distinguish between the fovea and surrounding lesions. In addition, OCT images could be added to the model to see if the structural information provided by OCT can improve model performance. We currently have human grader annotations for FAF images only because FAF is the standard endpoint for clinical trials. However, we are developing in-house OCT segmentation algorithms that would enable the use of OCT images for these approaches. For the multiclass and multiclass whole-lesion models (Table 4), the (uncalibrated) R^2 value is much lower than the r^2, which indicates that the models are miscalibrated. The reason for this could be the use of the Dice coefficient to train and tune the models. The miscalibration was fixed by estimating a calibration function using the validation set, which can then be assessed together with the model in the test set. Table 4 shows that R^2 improves after recalibration because the miscalibration has decreased (discrimination remains unchanged by the calibration). Because Model 2 was already calibrated in the validation and test sets, R^2 is a little smaller in the test set after recalibration. Future segmentation has often been applied for marking objects within complex urban scenes. For example, authors described an autoregressive network,[35] which predicts semantic segmentation maps of unobserved future frames from past sequences of videos. Researchers reported a temporal encoder-decoder network architecture,[36] which derives features from past frames and later constructs the future semantic segmentation. A separate encoder-decoder architecture[37] was leveraged to identify future trajectory points of pedestrians in these urban scenes. These models often contain an LSTM component. Despite adopting a simpler modeling approach, in the future, we could build our own autoregressive segmentation model underpinned on LSTMs to predict future, and potentially concurrent, lesion growth across patients' visits. Further, we are also interested in implementing a "maximal" model with all possible input and output combinations using LSTM. As a next step to the prediction of region of growth, we are looking into identifying patterns or features in FAF images that could be predictive of GA future ROG or progression. We used the same data splitting strategy[14] as prior work because our next step is to compare the results of the models developed in this study with our previous progression prediction models.

In summary, we demonstrated the feasibility of using FAF images to predict the 1-year ROG of GA lesions. The performance of the CNN models was similar across the training, validation, and test sets and was comparable to the intergrader reproducibility in terms of DSC. This is a proof-of-concept study, and validation on independent and external datasets is required to determine whether the performance is clinically meaningful or if there is a need for improvement. This work can potentially enable model use in clinical research, for example, to increase the efficacy of clinical trials through covariate adjustment, stratification, and, potentially, enrichment.[11] In the future, this work can also be used to inform the development of a useful tool to support patient counseling in clinical practice.

Data Sharing Statement

For up-to-date details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here: https://go.roche.com/data_sharing. For eligible studies, qualified researchers may request access to the clinical data through a data request platform. At the time of writing, this request platform is Vivli. https://vivli.org/ourmember/roche/. For the imaging data underlying this publication, requests can be made by qualified researchers, subject to a detailed, hypothesis-driven proposal and necessary agreements.

Footnotes and Disclosures

Originally received: March 27, 2024.
Final revision: October 8, 2024.
Accepted: October 17, 2024.
Available online: October 23, 2024. Manuscript no. XOPS-D-24-00018.

[1] Genentech, Inc., South San Francisco, California.

[2] Department of Ophthalmology, University of Washington, Seattle, Washington.

[3] National Institute for Health Research, Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust, UCL Institute of Ophthalmology, London, UK.

[4] Doheny Image Reading Center, Doheny Eye Institute, Los Angeles, California.

[5] Department of Ophthalmology, University of Bonn, Bonn, Germany.

Portions of these data were previously presented at the Association for Research in Vision and Ophthalmology Annual Meeting; New Orleans,

LA; April 23−27, 2023, and the Ophthalmic Artificial Intelligence Summit; Virtual Meeting, May 6, 2023.

Disclosures

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s):

A.S.: Employee − Genentech, Inc.

J.C., S.S.G., C.R., C.S., Q.Y., D.F., N.A.: Employee − Genentech, Inc.; Stocks/stock options − F. Hoffmann-La Roche Ltd.

A.Y.L.: Financial support − Genentech, Inc.; Grants − Amazon, iCare World, Meta, Regeneron, Santen, Topcon, ZEISS, Research to Prevent Blindness, NIH/NEI K23EY029246, NIH OT2OD032644; Personal fees − Boehringer Ingelheim, Genentech, Inc., Gyroscope, Johnson & Johnson, Janssen, Alcon, Regeneron, US FDA; Speaker fees and travel expenses − Alcon; Nonfinancial support − Microsoft, Heidelberg, iCareWorld, Optomed. This article does not reflect the views of the US FDA.

P.A.K.: Financial support − Genentech, Inc. Consultant − Apellis, Bitfount, DeepMind, Novartis, Roche; Equity owner − Big Picture Medical; Speaker fees − Allergan, Bayer, Heidelberg Engineering, Topcon; Stock − Bitfount, Big Picture Medical Supported by a Moorfields Eye Charity Career Development Award (R190028A) and a UK Research & Innovation Future Leaders Fellowship (MR/T019050/1); Participation on a Data Safety Monitoring Board or Advisory Board − Topcon, Bayer, Boehringer-Ingleheim, RetinAI, and Novartis; Patents − Active: Generalizable medical image analysis using segmentation and classification neural networks https://patents.google.com/patent/US10198832B2/en; Pending: Predicting disease progression from tissue images and tissue segmentation maps https://patents.google.com/patent/US20220301152A1/en; Honararia − Zeiss, Topcon, Novartis, Boehringer-Ingleheim, Apellis, Roche, Abbvie.

S.R.S.: Consultant − 4D Molecular Therapeutics, AbbVie/Allergan, Alexion, Amgen, Apellis, Astellas, Bayer, Biogen, Boehringer Ingelheim, Catalyst, CenterVue, Gyroscope, Heidelberg Engineering, Iveric Bio, Janssen, Nanoscope, Notal, Novartis, Optos, Oxurion, Pfizer, Regeneron, Roche/Genentech, Inc.; Speaker fees − Bayer, Heidelberg Engineering, Nidek, Novartis, Roche; Received research instruments from CenterVue, Heidelberg Engineering, Meditec, Nidek, Optos, Topcon, ZEISS; Grants − Carl Zeiss Meditec; Travel expenses − Optos, Roche, Samsung Bioepis; Participation on a Data Safety Monitoring Board or Advisory Board − Regeneron, RegenxBio; Leadership − Macula Society, ARVO, International Retinal Imaging Society, RIMR AMD Consortium.

F.G.H.: Consultant − 4D Molecular Therapeutics, Acucela, Alcon, Alexion, Alzheon, Annexon, Astellas, Bayer, Boehringer Ingelheim, Genentech, Inc./Roche, Heidelberg Engineering, Lin BioScience, Novartis, OcuTerra, Pixium Vision, ZEISS; Financial support − Acucela, Allergan (an AbbVie company), Astellas, Bayer, CenterVue, Genentech, Inc./Roche, Heidelberg Engineering, Novartis, ZEISS; Leadership − Executive Board German Ophthalmological Society (DOG).

Genentech, Inc., a member of the Roche Group, provided financial support for the study and participated in the study design, conducting the study, data collection, management, analysis, interpretation, preparation, review, and approval of the manuscript. Third-party writing assistance was provided by Jackie Read, PhD, of Envision Pharma Group and was funded by Genentech, Inc., a member of the Roche Group.

Author Contributions:

Conception and design: Salvi, Cluceru, Gao, Rabe, Anegondi.

Data collection: Salvi, Gao, Rabe, Ferrara, Anegondi.

Analysis and interpretation: Salvi, Cluceru, Gao, Rabe, Schiffman, Yang, Lee, Keane, Sadda, Holz, Ferrara, Anegondi.

Obtained funding: N/A.

Overall responsibility: All authors.

SriniVas R Sadda, MD, an editor of this journal, and Aaron Y Lee, MA, MSCI, an associate editor of this journal, were recused from the peer-review process of this article and had no access to information regarding its peer-review.

HUMAN SUBJECTS: Human subjects were included in this study. The trials adhered to the Declaration of Helsinki and were Health Insurance Portability and Accountability Act compliant. Protocols were approved by the institutional review board at each study site before the trials started. All patients provided written informed consent for future medical research and analyses.

No animal subjects were used in this study.

Abbreviations and Acronyms:
AMD = age-related macular degeneration; **CNN** = convolution neural network; **DSC** = Dice score coefficient; **FAF** = fundus autofluorescence; **G1** = Grader 1; **G2** = Grader 2; **GA** = geographic atrophy; **LSTM** = long short-term memory; $\mathbf{r^2}$ = Pearson correlation coefficient; $\mathbf{R^2}$ = coefficient of determination; **ROG** = region of growth; **RORA** = retinal pigment epithelial and outer retinal atrophy; **RPE** = retinal pigment epithelium; **SCR** = screening; **T** = time point; **W** = week.

Keywords:
Age-related macular degeneration, Artificial intelligence, Deep learning, Fundus autofluorescence imaging, Geographic atrophy.

Correspondence:
Neha Anegondi, MTech, Genentech, Inc., 1 DNA Way, MS-44 #3B, South San Francisco, CA 94080. E-mail: anegondi.neha@gene.com.

References

1. Fleckenstein M, Mitchell P, Freund KB, et al. The progression of geographic atrophy secondary to age-related macular degeneration. *Ophthalmology*. 2018;125:369−390.
2. Klein R, Wang Q, Klein BEK, et al. The relationship of age-related maculopathy, cataract, and glaucoma to visual acuity. *Invest Ophthalmol Vis Sci*. 1995;36:182−191.
3. American Foundation for the Blind. Key definitions of statistical terms. www.afb.org/research-and-initiatives/statistics/key-definitions-statistical-terms. Accessed April 3, 2023.
4. Rees A, Zekite A, Bunce C, Patel PJ. How many people in England and Wales are registered partially sighted or blind because of age-related macular degeneration? *Eye (Lond)*. 2014;28:832−837.
5. Sunness JS, Gonzalez-Baron J, Applegate CA, et al. Enlargement of atrophy and visual acuity loss in the geographic atrophy form of age-related macular degeneration. *Ophthalmology*. 1999;106:1768−1779.
6. Göbel AP, Fleckenstein M, Schmitz-Valckenberg S, et al. Imaging geographic atrophy in age-related macular degeneration. *Ophthalmologica*. 2011;226:182−190.
7. Nattagh K, Zhou H, Rinella N, et al. OCT angiography to predict geographic atrophy progression using choriocapillaris flow void as a biomarker. *Transl Vis Sci Technol*. 2020;9:6.
8. Schmitz-Valckenberg S, Fleckenstein M, Göbel AP, et al. Optical coherence tomography and autofluorescence findings

in areas with geographic atrophy due to age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2011;52:1−6.
9. Schmitz-Valckenberg S, Brinkmann CK, Alten F, et al. Semiautomated image processing method for identification and quantification of geographic atrophy in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2011;52:7640−7646.
10. Holz FG, Sadda SR, Staurenghi G, et al. Imaging protocols in clinical studies in advanced age-related macular degeneration: recommendations from Classification of Atrophy Consensus Meetings. *Ophthalmology*. 2017;124:464−478.
11. Schiffman C, Friesehahn M, Rabe C. How to get the most out of prognostic baseline variables in clinical trials. https://www.stats4datascience.com/posts/covariate_adjustment/. Accessed April 3, 2023.
12. Sadda SR, Chakravarthy U, Birch DG, et al. Clinical endpoints for the study of geographic atrophy secondary to age-related macular degeneration. *Retina*. 2016;36:1806−1822.
13. Shen LL, Sun M, Ahluwalia A, et al. Relationship of topographic distribution of geographic atrophy to visual acuity in nonexudative age-related macular degeneration. *Ophthalmol Retina*. 2021;5:761−774.
14. Anegondi N, Gao SS, Steffen V, et al. Deep learning to predict geographic atrophy area and growth rate from multimodal imaging. *Ophthalmol Retina*. 2023;7:243−252.
15. Shen LL, Sun M, Ahluwalia A, et al. Geographic atrophy growth is strongly related to lesion perimeter: unifying effects of lesion area, number, and circularity on growth. *Ophthalmol Retina*. 2021;5:868−878.
16. Holmen IC, Aul B, Pak JW, et al. Precursors and development of geographic atrophy with autofluorescence imaging: age-related eye disease study 2 report number 18. *Ophthalmol Retina*. 2019;3:724−733.
17. Gigon A, Mosinska A, Montesel A, et al. Personalized atrophy risk mapping in age-related macular degeneration. *Transl Vis Sci Technol*. 2021;10:18.
18. Niu S, de Sisternes L, Chen Q, et al. Fully automated prediction of geographic atrophy growth using quantitative spectral-domain optical coherence tomography biomarkers. *Ophthalmology*. 2016;123:1737−1750.
19. Holz FG, Sadda SR, Busbee B, et al. Efficacy and safety of lampalizumab for geographic atrophy due to age-related macular degeneration: chroma and Spectri phase 3 randomized clinical trials. *JAMA Ophthalmol*. 2018;136:666−677.
20. Holekamp N, Wykoff CC, Schmitz-Valckenberg S, et al. Natural history of geographic atrophy secondary to age-related macular degeneration: results from the prospective Proxima A and B clinical trials. *Ophthalmology*. 2020;127:769−783.
21. Patro SGK, Sahu KK. Normalization: a preprocessing stage. *arXiv*. 2015:1−4. https://doi.org/10.48550/arXiv.1503.06462.
22. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *aeXiv*. 2015:1−7. https://doi.org/10.48550/arXiv.1505.04597.
23. Iakubovskii P. Segmentation models PyTorch. https://github.com/qubvel/segmentation_models.pytorch. Accessed June 25, 2022.
24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *arXiv*; 2015:1−12. https://arxiv.org/abs/1512.03385.
25. Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2009:248−255.
26. Loshchilov I, Hutter F. Decoupled weight decay regularization. *arXiv*. 2019:1−8. https://doi.org/10.48550/arXiv.1711.0510.
27. Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. *arXiv*. 2019. https://doi.org/10.48550/arXiv.1912.01703.
28. Cardoso M, Li W, Brown R, et al. MONAI: an open-source framework for deep learning in healthcare. *arXiv*. 2022:1−25. https://doi.org/10.48550/arXiv.2211.02701.
29. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern B Cybern*. 1979;9:62−66.
30. Friesenhahn M, Rabe C, Schiffman C. Everything you wanted to know about R^2 but were afraid to ask. Part 1: Using a fundamental decomposition to gain insights into predictive model performance. https://stats4datascience.com/posts/three_metrics/. Accessed March 4, 2024.
31. Zhang Y, Zhang X, Ji Z, et al. An integrated time adaptive geographic atrophy prediction model for SD-OCT images. *Med Image Anal*. 2021;68:101893.
32. Qu J, Velaga SB, Hariri AH, et al. Classification and quantitative analysis of geographic atrophy junctional zone using spectral domain optical coherence tomography. *Retina*. 2018;38:1456−1463.
33. Sadda SR, Guymer R, Holz FG, et al. Consensus definition for atrophy associated with age-related macular degeneration on OCT: classification of atrophy report 3. *Ophthalmology*. 2018;125:537−548.
34. Pfau M, Lindner M, Goerdt L, et al. Prognostic value of shape-descriptive factors for the progression of geographic atrophy secondary to age-related macular degeneration. *Retina*. 2019;39:1527−1540.
35. Luc P, Neverona N, Couprie C, et al. Predicting deeper into the future of semantic segmentation. *arXiv*. 2017:1−10. https://doi.org/10.48550/arXiv.1703.07684.
36. Chiu Hk, Adeli E, Niebles J. Segmenting the future. *arXiv*. 2019:1−10. https://doi.org/10.48550/arXiv.1904.10666.
37. Syed A, Morris BT. SSeg-LSTM: semantic scene segmentation for trajectory prediction. presented at: 2019 IEEE Intelligent Vehicles Symposium (IV). Paris, France; 2019. https://ieeexplore.ieee.org/abstract/document/8813801; 2019. Accessed June 25, 2022.