4049921

Automated Segmentation of Geographic Atrophy Using Machine Learning on Real-World Fundus Autofluorescence Ophthalmic Images in the IRIS® Registry

Kim Le[1], Durga Borkar[1,2], Theodore Leng[1,3], Carlos Candano[1], Aishwarya Ramakrishnan[1], Ketki Khapare[1], Michael Mbagwu[1,3], Aracelis Z Torres[1], Zhongdi Chu[1]

[1]Verana Health, San Francisco, California, US
[2]Duke University Department of Ophthalmology, Durham, NC, USA
[3]Byers Eye Institute at Stanford, Stanford University School of Medicine, Palo Alto, CA, USA

Purpose

To develop a machine learning (ML) pipeline to automatically segment and assess the size of atrophic lesions in patients with geographic atrophy (GA) using real-world fundus autofluorescence (FAF) images linked to the American Academy of Ophthalmology IRIS® Registry (Intelligent Research in Sight).

Methods

- The IRIS Registry is the nation's first and largest comprehensive eye disease clinical database
- An ML pipeline was developed using de-identified FAF images from GA patients
  - Patients with images were linked to their clinical records within the IRIS Registry
  - GA was identified through: 1) prior developed ML imaging models; and 2) ICD-10 codes in clinical records from 2006-2022
  - FAF images were included if fovea centered and had gradable image quality (assessed by prior developed ML models)
- Training, validation, and testing sets were randomly selected and labeled by an expert grader for model training
  - Sets were selected using stratified sampling on imaging based characteristics using a 60:20:20 split
  - Labels from the expert grader were validated against those from two fellowship-trained retinal specialists
- A deep learning model was trained to segment atrophic lesions using the training and validation sets with model performance measured via Dice coefficient scores (DSCs) using the test set
- The trained model was applied to new patient eyes, with multiple imaging visits, to extract clinically relevant metrics over time

Results

Dataset

- A total of **214 FAF images with GA** were labeled from 209 unique patient eyes and 208 unique patients

Table 1. Training, validation and test set sizes

Set	Patients	Patient Eyes	Images
Training	122	122	127
Validation	44	45	45
Testing	42	42	42

- The expert grader (G1) had similar labeling accuracy against two fellowship-trained retinal specialists (G2, G3)
  - Compared DSCs for 50 randomly sampled FAF images labeled by all three graders (*Figure 1*)
  - **Labeled outputs are reliable representations of GA lesions from real-world images**

Table 2. DSCs for labeled FAF images across graders

Grader Comparison	Average DSC (std)
G1 v. G2	0.896 (0.172)
G1 v. G3	0.892 (0.145)
G2 v. G3	0.888 (0.132)

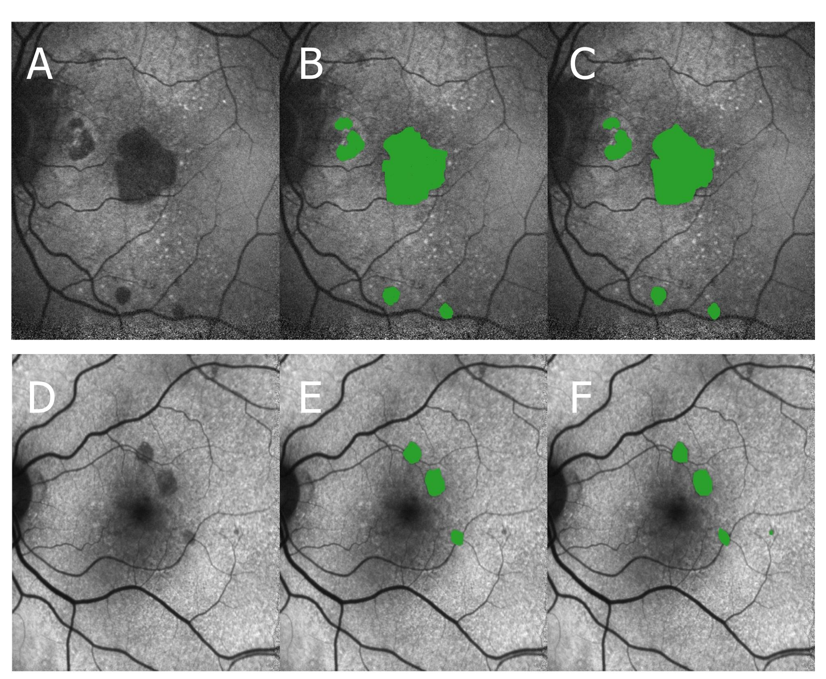

Figure 1. Representative labels across graders. A) original FAF image B) label from G1 C) label from G2 (DSC for G1 v. G2 = 0.96) D) original FAF image E) label from G1 F) label from G3 (DSC for G1 v. G3 = 0.92)

Model Training

- A modified U-Net architecture was developed and trained
  - Trained with PyTorch 1.8 and Python 3.6 GPU
  - Model was trained from scratch using Dice loss, with 150 maximum epochs, batch size of 8, and Adam's optimizer with learning rate of 5e-4
  - Image augmentation was done on the training set
- Model achieved a **DSC of 0.887** for the hold-out test set (*Figure 2*)

Table 3. Model accuracy (DSCs) for training, validation, and test sets

Cohort	DSC
Training set (n = 127 images)	0.861
Validation set (n = 45 images)	0.905
Test set (n = 42 images)	0.887

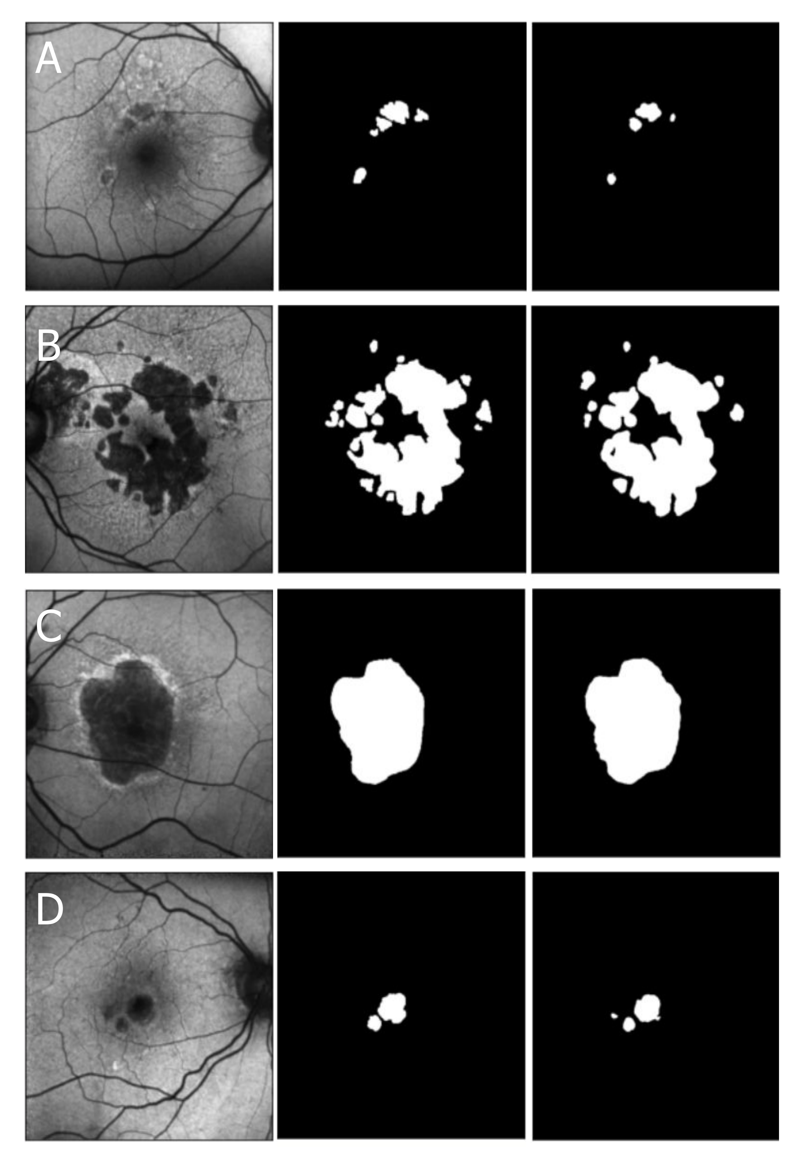

Figure 2. Examples of segmented masks from the deep learning model. (*left to right*) original FAF image, ground truth mask, the predicted mask

A) Image with 0.741 DSC
B) Image with 0.942 DSC
C) Image with 0.978 DSC
D) Image with 0.886 DSC

Examining GA Lesion Size Across Time

- Examined 20 new patients with multiple imaging visits over time *(Figure 3)*
  - The cohort had an average of 3.0 images per eye across an average of 2.9 years
  - Used the developed ML pipeline to automatically segment the GA lesions
- The area of the segmented lesion (mm^2) and growth rate (mm/year) was derived from the mask
  - Area extracted using the scikit-image library
- **Average annual growth rate (and SD) was 0.310 (0.283) mm/yr**

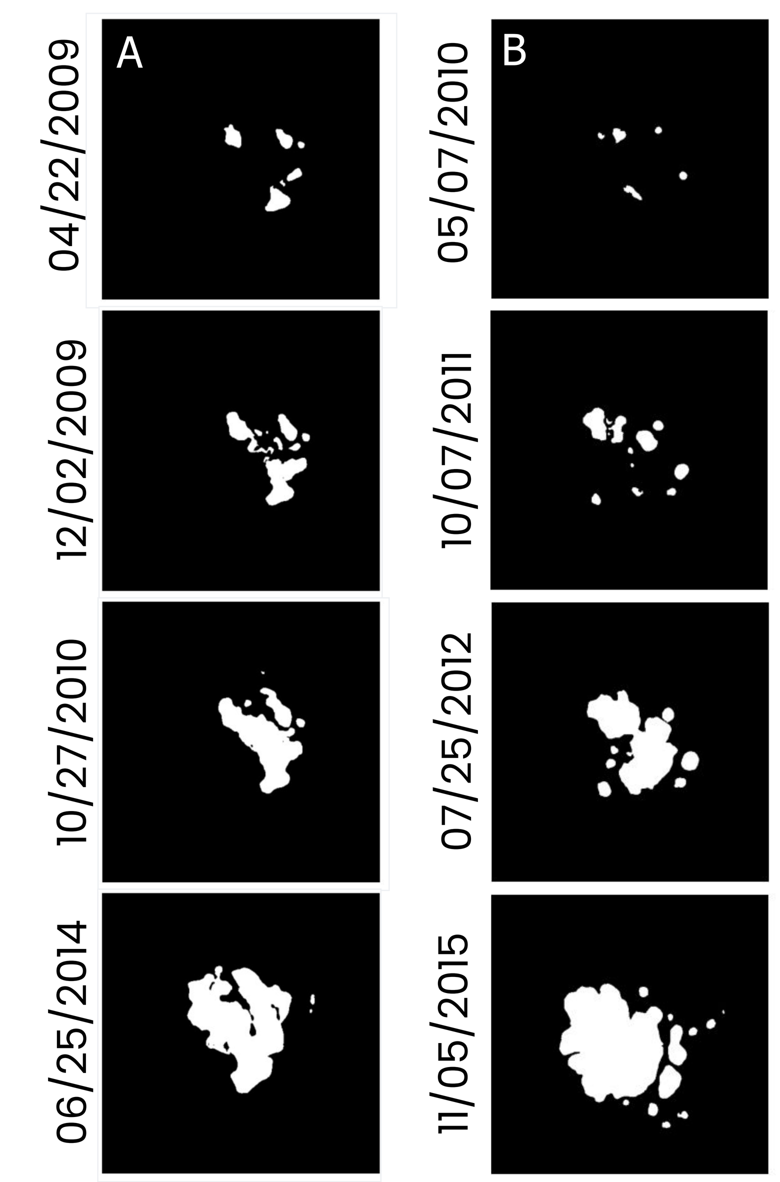

Figure 3. Examples of predicted segmented masks across two unique patient journeys

Conclusions

- The proposed pipeline demonstrates satisfactory accuracy for segmenting atrophic lesions in real-world FAF images
- This developed ML pipeline could be leveraged to assess real-world GA disease progression at scale to help inform monitoring and treatment decisions

Funding

This study was funded by Verana Health and all authors are either former or current employees or consultants of Verana Health.

Contact Information: kim.le@veranahealth.com