EBM Experimental Biology and Medicine

TYPE Original Research
PUBLISHED 02 November 2026
DOI 10.3389/ebm.2026.11213

Check for updates

OPEN ACCESS

*CORRESPONDENCE
Minhaj Nur Alam,
malam8@charlotte.edu
Sally S. Ong,
Sally.Ong@wfusm.edu

RECEIVED 29 May 2026
REVISED 08 October 2026
ACCEPTED 13 October 2026
PUBLISHED 02 November 2026

CITATION
Siraz S, Kamanda H, Gholami S, Nabil AS, Rao NT, Ong SS and Alam MN (2026) Multi-class classification of central and non-central geographic atrophy using optical coherence tomography. *Exp. Biol. Med.* 251:11213. doi: 10.3389/ebm.2026.11213

COPYRIGHT
© 2026 Siraz, Kamanda, Gholami, Nabil, Rao, Ong and Alam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Multi-class classification of central and non-central geographic atrophy using optical coherence tomography

Sadia Siraz[1], Hindolo Kamanda[2], Sina Gholami[1], Ahammed Sakir Nabil[1], Nitya Tangada Rao[2], Sally S. Ong[2]* and Minhaj Nur Alam[1]*

[1]Department of Electrical and Computer Engineering, University of North Carolina at Charlotte, Charlotte, NC, United States, [2]Department of Ophthalmology, Wake Forest University School of Medicine, Winston-Salem, NC, United States

## Abstract

This retrospective study aimed to develop and validate deep learning (DL)-based models for classifying geographic atrophy (GA) subtypes using Optical Coherence Tomography (OCT) scans across four clinical classification tasks. A total of 455 OCT volumes from 91 patients at Wake Forest University School of Medicine were evaluated with 258 Central GA (CGA), 74 Non-Central GA (NCGA), and 123 no GA (NGA) cases. For GA classification, the dataset was supplemented with early and intermediate non-neovascular AMD and neovascular AMD cases, which were sourced from four public repositories. Two experimental approaches were tested: utilizing all B-scans within each OCT volume and selectively using B-scans containing foveal regions. Models were trained using transfer learning with patient-level data splitting at a 70:15:15 ratio, and performance was evaluated using area under the receiver operating characteristic curve (AUC-ROC), F1 score, and accuracy as primary outcome measures. Vision Transformers (ViTs) demonstrated the most consistent performance across the evaluated architectures. For CGA versus NCGA classification, ViT-B/16 achieved a B-scan-level AUC-ROC of 0.728 $\pm$ 0.058 in Case 2, while volume/visit-level AUC-ROC was 0.837 $\pm$ 0.011 in Case 1 and 0.656 $\pm$ 0.070 in Case 2. NCGA recognition remained challenging, particularly in Case 2. In contrast, GA-versus-NGA and GA-versus-non-GA AMD classification showed substantially stronger discrimination. These findings support the feasibility of OCT-based automated GA phenotyping while highlighting NCGA recognition as an important area for further optimization and external validation.

KEYWORDS
**age-related macular degeneration, deep learning, geographic atrophy, optical coherence tomography, vision transformers**

## Impact statement

Geographic atrophy (GA) is a leading cause of irreversible vision loss in older adults, and recently approved treatments provide greater benefit when atrophy has not yet reached the center of vision. Distinguishing central from non-central GA is therefore clinically important, yet assessment remains subjective and can vary between clinicians, limiting consistency in patient care and research. This study presents the first deep learning

framework for automated classification of central and non-central GA using optical coherence tomography (OCT), a noninvasive and widely available imaging modality. Using real-world clinical data, our models accurately identify GA and distinguish clinically meaningful disease subtypes. These findings provide a foundation for objective, reproducible, and scalable GA assessment. By enabling earlier and more consistent identification of patients who may benefit most from treatment, this work has the potential to improve clinical decision-making, patient management, and future research in age-related macular degeneration.

## Introduction

Age-related macular degeneration (AMD) is the leading cause of irreversible vision loss in developed nations, affecting approximately one in five individuals over 65, with global prevalence projected to reach 288 million by 2040 [1, 2]. Geographic atrophy (GA), the advanced stage of non-neovascular AMD, is characterized by progressive degeneration of the retinal pigment epithelium (RPE), photoreceptors, and choriocapillaris, accounting for 20% of AMD-related legal blindness [3, 4]. GA typically begins in the parafoveal region (non-central GA; NCGA) before extending toward the foveal center, at which point it is classified as central GA (CGA). Early subtype distinction (identifying NCGA for early intervention) is critical for clinical management particularly in light of recent therapeutic breakthroughs [5]. In 2023, the FDA approved two complement inhibition therapies--pegcetacoplan and avacincaptad pegol--the first pharmacological interventions capable of slowing GA progression [6, 7].Crucially, these therapies exhibit location-dependent efficacy. In the OAKS and DERBY trials, pegcetacoplan showed greater benefit in NCGA (22–26% reduction at 24 months) than CGA (16–19%), a pattern reinforced by long-term GALE data (28–42% vs. 25–31%) [8]. This 15–40% greater relative effect in NCGA underscores the need for reliable, early classification.

In addition to clinical examination, ophthalmologists rely on imaging modalities such as fundus autofluorescence (FAF) and optical coherence tomography (OCT) to diagnose and monitor GA [9, 10]. While FAF remains the gold standard for GA area measurement, OCT offers superior precision for determining foveal involvement through its cross-sectional, depth-resolved imaging enabling layer-by-layer structural assessment and precise measurement of atrophy borders relative to the foveal center [10]. OCT's widespread availability and volumetric 3D imaging make it ideally suited for CGA versus NCGA distinction.

However, even with OCT, CGA/NCGA classification remains challenging in practice. The foveal center is difficult to define when anatomical landmarks are distorted by advanced atrophy [11]. Thus inter-grader agreement for GA identification and boundary delineation varies widely across imaging modalities and grading protocols ($\kappa = 0.28–0.92$) [12–14]. These limitations point to a clear need for automated, objective classification. AI-driven tools offer a practical solution, enabling scalable, consistent phenotyping by reducing interobserver variability and providing objective metrics.

Recent deep learning (DL) advances have achieved expert-level GA segmentation, clinical validation under complement inhibitor treatment (Dice 0.86–0.91), progression prediction (AUC 0.77–0.81), and nascent GA detection with 95–97% sensitivity [15–18]. Despite this progress, no study has addressed multi-class phenotyping distinguishing CGA from NCGA - a striking gap given therapies with location-dependent efficacy. This gap persists because: (1) foveal localization requires integrating anatomical landmarks with lesion characteristics beyond segmentation alone; (2) subjective boundaries have discouraged reliably labeled CGA/NCGA datasets; (3) most prior work used FAF, lacking depth resolution; and (4) clinical needs emerged only with recent FDA approval of complement inhibitors.

In this study, we directly address this gap by developing and validating DL-based models using OCT for four classification tasks: i) CGA vs. NCGA; ii) CGA vs. NCGA vs. no GA (NGA); iii) GA vs. NGA; and iv) GA vs. other AMD forms (namely non-neovascular AMD that has not progressed to GA or wet AMD). Our key contributions are [19]:

1. We use a real-world privately curated dataset from Wake Forest University School of Medicine to address the need for automated CGA classification due to the recent FDA approval of complement inhibitors.
2. This is the first DL study specifically designed to distinguish CGA from NCGA based on foveal involvement.
3. We implement and compare two distinct methodological approaches for leveraging OCT data: Case 1 (Full Volume): Utilizing all B-scans within each OCT volume to capture the complete spectrum of retinal pathology across different patient visits; Case 2 (Selective B-scans): an expert-guided, semi-automated approach using B-scans in which the foveal region was most clearly visualized and/or prominent GA-associated structural abnormalities were present, enabling evaluation of diagnostically focused OCT input.

## Materials and methods

This section outlines the methods adopted for GA classification tasks. Two experimental approaches were evaluated. Case 1 utilized all B-scans within each OCT volume. Case 2 represented an expert-guided, semi-automated analysis in which selected B-scans were retrospectively identified by the primary clinical grader, with emphasis on scans in which the foveal region was most clearly visualized and/or prominent GA-associated structural abnormalities were present. Similar model-development procedures were applied across both approaches (Figure 1).

### Data

#### Data acquisition

We utilized OCT scans from two sources. For primary classification tasks (CGA vs. NCGA, CGA vs. NCGA vs. NGA, and GA vs. NGA), we employed a private clinical dataset from Wake Forest University School of Medicine. For differential diagnosis between GA and other AMD forms, we supplemented this with four publicly available datasets (DS1-DS4). This study was approved by

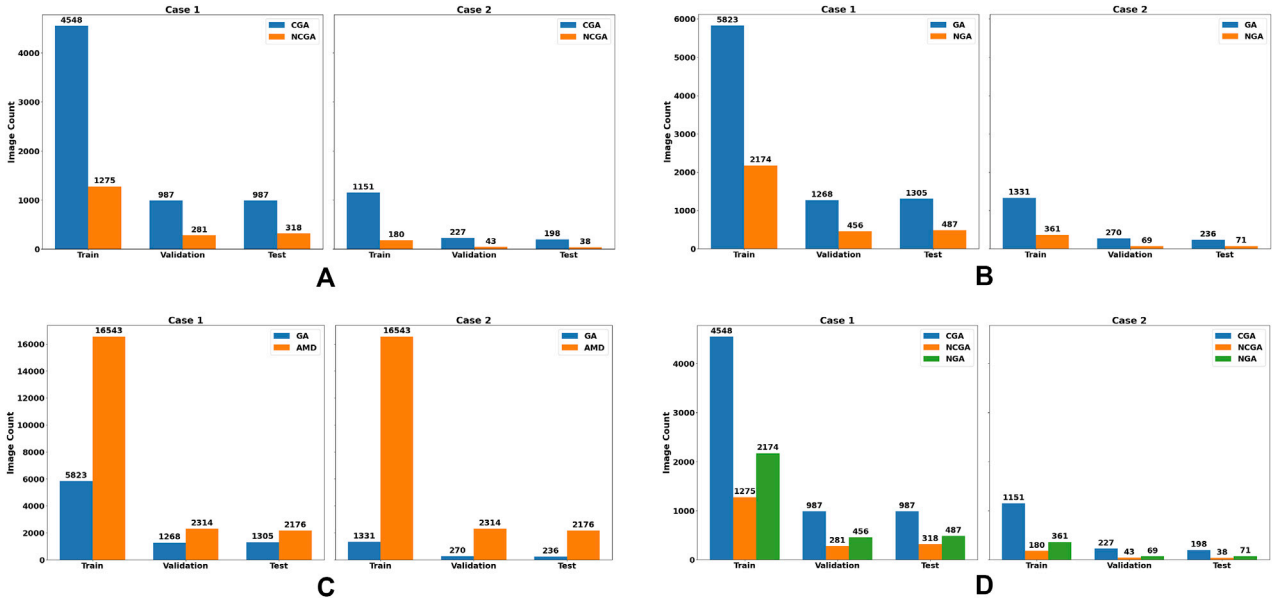



FIGURE 1
Dataset distribution across training, validation, and test sets for the four classification tasks. **(A)** Binary classification of CGA vs. NCGA using all B-scans (Case 1) and selected B-scans of interest (Case 2). **(B)** Binary classification of GA vs. no NGA. **(C)** Binary classification of GA vs. other forms of AMD. **(D)** Multi-class classification of CGA vs. NCGA vs. NGA.

the Institutional Review Board (IRB) at Wake Forest University School of Medicine and adhered to the Declaration of Helsinki.

## Wake forest dataset

A retrospective chart review (Jan 1, 2013 - Jan 28, 2023) was conducted using ICD9 codes 362.50/362.51 and ICD10 codes H35.30/H35.31. Patients were included if they had longitudinal OCT imaging with at least two visits with OCTs available for assessment of GA progression. Exclusion criteria included: non-neovascular AMD without evidence of GA development; foveal GA at baseline; absence or non-retrievability of longitudinal OCT imaging; fewer than two available visits with OCTs; coexisting ocular pathologies likely to confound OCT interpretation including neovascular AMD, diabetic retinopathy, glaucoma, retinal vascular occlusions, retinal detachment, and macular dystrophies; and prior retinal treatments. OCT volumes were acquired using Heidelberg HRA + OCT Spectralis Version 2.5.9 with Heyex software during routine clinical encounters. The final cohort consisted of 91 patients. GA was operationally defined in this retrospective cohort by loss of the outer retina and retinal pigment epithelium with associated choroidal hypertransmission. A minimum lesion diameter of 250 μm was not required; consequently, the cohort may include smaller or earlier atrophic changes that do not meet a conventional minimum size threshold. CGA was defined as involvement of the central 1-mm ETDRS circle by the identified atrophic process, whereas NCGA was defined as atrophy outside this circle. Two experienced graders independently assigned ordinal severity grades to each patient visit using a three-point scale (1 = NGA, 2 = NCGA, 3 = CGA). Inter-rater reliability was assessed at the visit level using percent agreement, Cohen's kappa (κ), and quadratic-weighted kappa to account for the ordered grading structure. Discordant cases were resolved through a joint adjudication session, where consensus review established the final reference standard; concordant grades required no modification. Primary model development was performed using labels assigned by Grader 1. Independent grading by Grader 2 and subsequent consensus adjudication were conducted later in the study, after completion of the original model-development experiments, to assess inter-rater reliability and establish a consensus reference standard for discrepant visits. Accordingly, consensus-adjudicated labels were not available during the original model-development stage.

We incorporated 258 OCT volumes for CGA, 74 for NCGA, and 123 for NGA, all captured across multiple clinical visits as shown in Figure 2. The vast majority of OCT volumes consisted of 25 B-scans (512 × 496 × 25 voxels), reflecting the standard acquisition protocol. In the CGA cohort, 242 volumes contained 25 B-scans, while 14 and 2 volumes contained 31 and 19 B-scans respectively. Similarly, 70 (NCGA) and 116 (NGA) volumes contained 25 B-Scans while 4 (NCGA) and 7 (NGA) volumes contained 31 B-scans. Thus, we had 6522 B-scans for CGA, 1874 B-scans for NCGA, and 3117 B-scans for NGA.

For Case 2, B-scan selection was performed retrospectively by the primary clinical grader after diagnostic grading. Across the 455 OCT volumes, 2,338 of 11,513 B-scans (20.3%) were retained. The mean number of selected B-scans per original volume was 5.14 ± 4.22, with a median of 4 (IQR, 3–5.5; range, 0–23). Retention differed by class, with 24.2% of CGA, 13.9% of NCGA, and 16.1% of NGA B-scans retained. Because selection was performed retrospectively after diagnostic grading, potential selection bias associated with expert-guided B-scan choice is acknowledged. This clinically informed benchmark can be

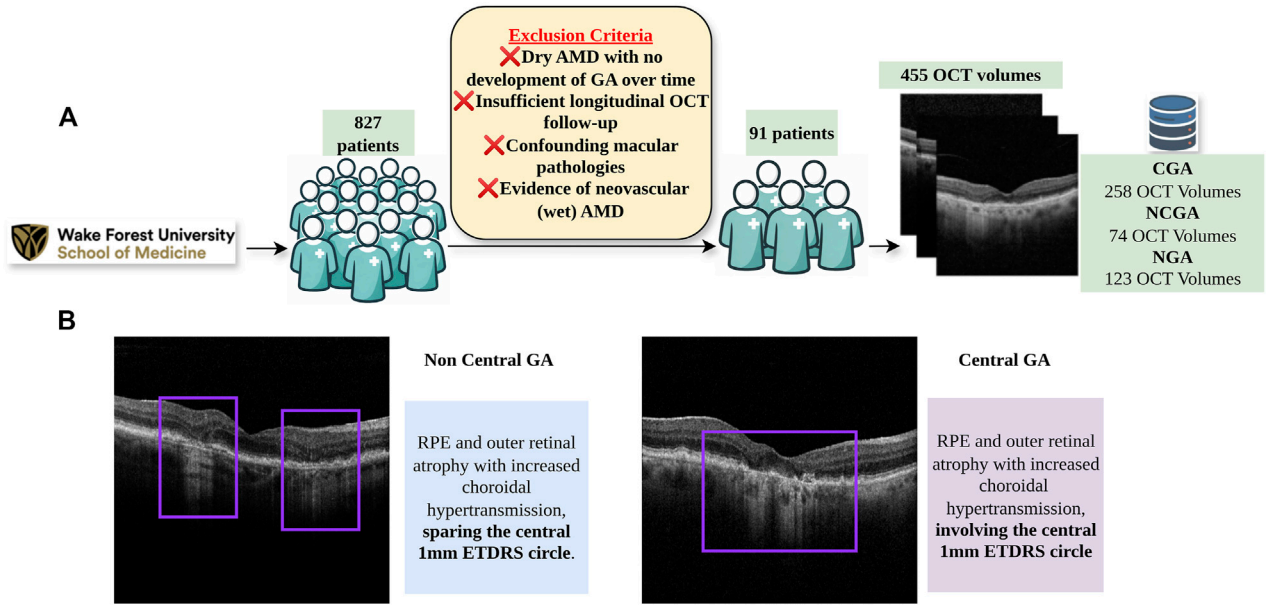



FIGURE 2
Study dataset overview and OCT characteristics of GA subtypes. **(A)** Workflow of dataset selection from the Wake Forest University School of Medicine cohort (total 11513 B-Sans altogether) **(B)** Representative OCT B-scan examples showing morphological differences between NCGA and CGA. Red boxes indicate regions of retinal pigment epithelium (RPE) and outer retinal atrophy with choroidal hypertransmission.

extended in future work using automated foveal localization or informative B-scan selection to enable a fully automated workflow.

## Public dataset (DS1-DS4)

For the GA vs. AMD classification task, AMD cases were drawn from four public OCT datasets. DS1 [20] contained retinal images from 45 subjects (Normal, AMD, DME; 15 per category), acquired using Heidelberg Spectralis OCT. DS2 [21] (OCTA-500) comprised data from 500 subjects, including volumetric OCT and OCTA acquisitions with 3-mm and 6-mm fields of view, acquired using a spectral-domain RTVue-XR system (Optovue, 840 nm). DS3 [22] included over 500 high-resolution images from a Cirrus HD-OCT machine (Carl Zeiss Meditec, 840 nm). DS4 [23], drawn from the OCTDL dataset, contained over 2,000 images from patients aged 20–93 (male-to-female ratio 3:2, mean age 63), acquired using Optovue Avanti RTVue XR (840 nm). Together, these datasets provided diverse imaging protocols and equipment for a comprehensive AMD comparison cohort. In this classification task, the term “other AMD forms” refers to non-GA AMD cases; specifically, early or intermediate non-neovascular AMD and neovascular AMD sourced from these four public datasets (DS1-DS4). Given that subtype-specific annotations (e.g., neovascular versus nonneovascular AMD) were not consistently available across all datasets, these cases represent a heterogeneous mix of AMD phenotypes and were grouped into a single comparison category for classification purposes.

In the assembled non-GA AMD cohort, the four public datasets contributed 21,033 images across the training, validation, and test partitions: 723 from DS1/Srinivasan, 19,024 from DS2/OCTA-500, 55 from DS3/OCTID, and 1,231 from DS4/OCTDL. Within the held-out test set, 1,904 images originated from DS2, 55 from DS3, and 217 from DS4; DS1 contributed no images to the held-out test partition. Based on the available dataset naming conventions, we were able to infer 15, 49, and 421 unique subject-level identifiers for DS1/Srinivasan, DS2/OCTA-500, and DS4/OCTDL, respectively. For DS3/OCTID, 55 unique identifiers were available, but the accompanying metadata did not allow us to verify that these represented 55 distinct subjects; therefore, we do not report a definitive subject count for DS3. We also did not sum subject counts across datasets because cross-dataset identity linkage was unavailable.

## Dataset partitioning

To prevent patient-level data leakage, we implemented a strict patient-exclusive splitting strategy. All OCT B-scans were grouped by a uniquely assigned patient identifier, ensuring all scans from the same patient remained in a single partition. Patient groups were randomly assigned to training, validation, and test sets at a 70%-15%-15% distribution. We confirmed that all 91 unique patients appeared in exactly one split. The training set optimized model parameters, the validation set guided hyperparameter tuning and early stopping, and the test set remained fully isolated until final evaluation.

Evaluation was performed at both the B-scan and OCT volume/visit levels. For volume/visit-level analysis, class probabilities were averaged across the evaluated B-scans belonging to each OCT volume, and the class with the highest mean probability was assigned as the volume-level prediction. As multiple clinical visits could originate from the same patient, uncertainty for the principal volume-level metrics was estimated using 2,000 patient-cluster bootstrap replicates. Patients were resampled with replacement, and all held-out visits from each sampled patient were retained

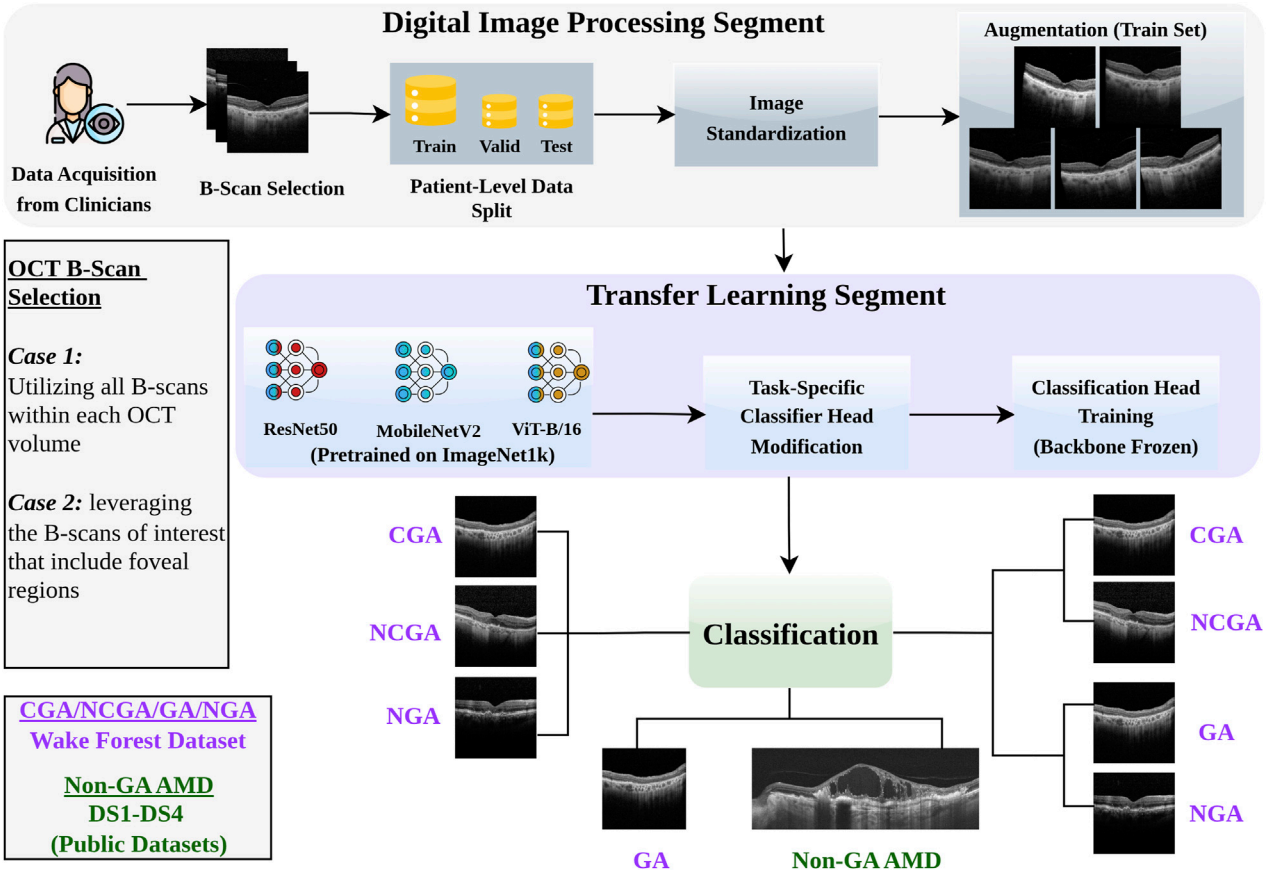



FIGURE 3
Step by step methodology of DL based GA classification task- Digital Image processing segment; Transfer learning segment; Classification task.

together in each bootstrap replicate, thereby preserving within-patient correlation. The OCT volume/visit was used as the primary clinically interpretable prediction unit, while the B-scan-level analysis was retained as a complementary evaluation.

All OCT B-scan images underwent a standardized preprocessing pipeline. Each image was converted to 3-channel RGB format and resized to 224 × 224 pixels to match ImageNet-pretrained model input requirements. Training images were augmented using horizontal flipping (50% probability), random rotation (±15°), color jitter (brightness, contrast, saturation ±20%; hue ±0.1), random translation (up to 10%), and random resized cropping (scale 0.8–1.0). All images were normalized using ImageNet statistics (mean = [0.485, 0.456, 0.406]; SD = [0.229, 0.224, 0.225]). Validation and test sets received only resizing and normalization, with no stochastic augmentation, to ensure deterministic and reproducible evaluation. Data were loaded in batches of 16 with shuffling enabled for training and four parallel workers.

## Model architecture

### Baseline model selection

We implemented and compared three neural network architectures: ResNet50, MobileNetV2, and Vision Transformer (ViT-B/16) [24–26]. ResNet50 is a 50-layer residual network with skip connections that mitigate vanishing gradients, enabling deep feature extraction well-suited to medical image classification. MobileNetV2 is a lightweight architecture using depth-wise separable convolutions, included to evaluate performance-efficiency tradeoffs relevant to clinical deployment. ViT-B/16 processes images as sequences of fixed-size patches and captures long-range spatial dependencies, which may be particularly beneficial for detecting subtle GA patterns distributed across OCT B-scans. The overall model training framework is described in Figure 3.

### Model implementation

All models were initialized with ImageNet-pretrained weights using feature-extraction transfer learning. Backbone parameters were frozen, with gradient updates restricted to a custom classification head consisting of a 512-neuron fully connected layer with ReLU activation, 30% dropout, and a final linear output layer mapping to the appropriate number of output classes. Partial or full backbone fine-tuning was not evaluated as a separate experimental condition in the present study.

### Training configuration

Training parameters were standardized across all architectures and tasks. We used the Adam optimizer [27] with an initial learning

rate of $1 \times 10^{-3}$ and a cosine annealing scheduler that reduced the learning rate to $1 \times 10^{-6}$ over the training duration. Standard, unweighted cross-entropy loss was used as the objective function for all classification tasks; no class-weighted loss, oversampling, undersampling, or class-balanced batch sampling was applied. A maximum of 50 training epochs was permitted, with early stopping triggered if validation accuracy did not improve for 20 consecutive epochs; most models converged before the 50-epoch limit. All models were trained on NVIDIA GeForce GTX 1080 Ti GPUs (11 GB VRAM) under CUDA 12.2 using a consistent batch size of 16.

TABLE 1 Demographic and ocular characteristics at baseline.

| Demographic features | Patient characteristics (91 patients) |
|---|---|
| Age, years, mean (standard deviation) | 81.58 (10.21) |
| Gender, female | 74.04% |
| Race, white/other | 92%/8% |
| Smoking Status, yes/no | 44.23%/55.77% |
| Study eye, right | 48.08% |
| Best corrected visual acuity, logMAR, mean (standard deviation) | 0.83 (0.57) |

# Results

## Cohort characteristic and inter-grader reliability

The Wake Forest University School of Medicine cohort, used for the primary GA classification tasks, consisted of 91 patients contributing 455 OCT volumes acquired across multiple clinical visits. Baseline demographic and ocular characteristics are summarized in Table 1.

Across 686 visits from 91 patients with valid paired grades, graders achieved 82.8% exact agreement ($\kappa = 0.70$; weighted $\kappa = 0.79$), reflecting substantial inter-rater reliability. Discordant cases were resolved through a joint adjudication session, where consensus review established the final reference standard; concordant grades required no modification. Grader 1 showed 93.6% agreement with the adjudicated standard ($\kappa = 0.89$; weighted $\kappa = 0.90$), with 6.4% of visits revised post-adjudication. Grader 2 agreed with the reference standard in 89.0% of visits ($\kappa = 0.81$; weighted $\kappa = 0.90$), with 11.0% requiring modification.

## Binary classification of CGA vs. NCGA

For distinguishing CGA from NCGA, ViT-B/16 demonstrated the strongest discrimination among the evaluated architectures (Table 2). At the B-scan level, ViT-B/16 achieved a mean AUC-ROC of 0.728 ± 0.058 in Case 2. Class-specific performance differed substantially between the two GA subtypes. For the representative ViT-B/16 checkpoint shown in the confusion matrices, Case 1 yielded CGA sensitivity and specificity of 96.5% and 24.2%, respectively, whereas NCGA sensitivity and specificity were 24.2% and 96.5%. In Case 2, CGA sensitivity was 99.5%, whereas NCGA sensitivity was 5.3%; corresponding NCGA specificity was 99.5%. Across the five ViT-B/16 runs in Case 2, mean NCGA sensitivity was 4.2% ± 3.0% and mean NCGA F1-score was 7.1% ± 4.9%, confirming that NCGA recognition remained the more challenging component of the binary task. The marked asymmetry between CGA and NCGA performance may partly reflect the substantial class imbalance in the cohort, which contained 258 CGA versus 74 NCGA OCT volumes. This imbalance was further accentuated in the selective B-scan analysis, in which 24.2% of CGA B-scans but only 13.9% of NCGA B-scans were retained. Because no class-weighting or resampling strategy was used during training, lower representation of NCGA may have contributed to the model's tendency to predict the majority CGA class. Complete class-specific sensitivity, specificity, PPV/precision, NPV, and F1-score values for the representative ViT-B/16 checkpoints are provided in Table 3.

At the OCT volume/visit level, ViT-B/16 achieved an AUC-ROC of 0.8368 ± 0.0114 in Case 1 and 0.6562 ± 0.0704 in Case 2, with corresponding accuracies of 0.7882 ± 0.0088 and 0.7830 ± 0.0095 (Table 4). For the representative ViT-B/16 checkpoint, the volume/visit-level AUC-ROC was 0.831 in Case 1, with a patient-clustered bootstrap 95% CI of 0.633–0.958, and 0.646 in Case 2, with a corresponding 95% CI of 0.381–0.842. These confidence intervals were estimated by resampling patients rather than individual B-scans or visits, thereby accounting for within-patient correlation among repeated OCT volumes. Case 2 volume-level NCGA sensitivity remained low across repeated runs, indicating that volume aggregation did not eliminate the class-specific difficulty in NCGA recognition. Accordingly, the CGA-versus-NCGA findings are interpreted as feasibility-level evidence for automated subtype classification and identify NCGA recognition as an important target for further optimization.

## Binary classification of GA vs. NGA

For distinguishing GA from NGA, ViT-B/16 achieved the highest AUC-ROC among the evaluated architectures in both cases. At the B-scan level, mean ViT-B/16 AUC-ROC was 0.8521 ± 0.0068 in Case 1 and 0.9500 ± 0.0050 in Case 2, with corresponding macro-F1 scores of 0.7262 ± 0.0100 and 0.8343 ± 0.0156 and accuracies of 0.7912 ± 0.0072 and 0.8730 ± 0.0098 (Table 2). At the OCT volume/visit level, AUC-ROC was 0.9052 ± 0.0088 in Case 1 and 0.9631 ± 0.0073 in Case 2 (Table 3). Thus, selective B-scan analysis was associated with higher ViT-B/16 discrimination for this task, whereas the magnitude and direction of the Case-2 effect differed across architectures.

## Binary classification of GA vs. Non-GA AMD

All three architectures demonstrated high AUC-ROC for distinguishing GA from the assembled non-GA AMD comparison

TABLE 2 B-scan-level comparative performance of deep-learning architectures across OCT classification tasks.

| Classification task | Case | Classifier name | Evaluation metrics | | |
|---|---|---|---|---|---|
| | | | AUC-ROC | F1 score | Accuracy |
| CGA vs. NCGA | 1 | ResNet50 | 0.545 ± 0.004 | 0.431 ± 0.0 | 0.756 ± 0.0 |
| | | MobileNetV2 | 0.521 ± 0.016 | 0.432 ± 0.002 | 0.756 ± 0.0 |
| | | ViT-B/16 | **0.718 ± 0.002** | **0.602 ± 0.004** | **0.780 ± 0.005** |
| | 2 | ResNet50 | 0.574 ± 0.015 | 0.456 ± 0.000 | 0.839 ± 0.000 |
| | | MobileNetV2 | 0.619 ± 0.050 | 0.456 ± 0.000 | **0.837 ± 0.002** |
| | | ViT-B/16 | **0.728 ± 0.058** | **0.492 ± 0.024** | 0.831 ± 0.017 |
| GA vs. NGA | 1 | ResNet50 | 0.804 ± 0.003 | 0.612 ± 0.011 | 0.760 ± 0.003 |
| | | MobileNetV2 | 0.777 ± 0.007 | 0.643 ± 0.021 | 0.754 ± 0.007 |
| | | ViT-B/16 | **0.852 ± 0.007** | **0.726 ± 0.010** | **0.791 ± 0.007** |
| | 2 | ResNet50 | 0.873 ± 0.006 | 0.638 ± 0.030 | 0.786 ± 0.003 |
| | | MobileNetV2 | 0.823 ± 0.010 | 0.653 ± 0.015 | 0.786 ± 0.009 |
| | | ViT-B/16 | **0.950 ± 0.005** | **0.834 ± 0.016** | **0.873 ± 0.010** |
| GA vs. Non-GA AMD | 1 | ResNet50 | **0.999 ± 0.0** | **0.998 ± 0.000** | **0.998 ± 0.000** |
| | | MobileNetV2 | 0.999 ± 0.0 | 0.945 ± 0.036 | 0.947 ± 0.035 |
| | | ViT-B/16 | 0.996 ± 0.003 | 0.811 ± 0.109 | 0.813 ± 0.110 |
| | 2 | ResNet50 | 0.999 ± 0.0 | 0.979 ± 0.006 | 0.980 ± 0.006 |
| | | MobileNetV2 | **0.999 ± 0.0** | **0.994 ± 0.003** | **0.994 ± 0.003** |
| | | ViT-B/16 | 0.996 ± 0.004 | 0.9748 ± 0.144 | 0.838 ± 0.139 |
| CGA vs. NCGA vs. NGA | 1 | ResNet50 | 0.704 ± 0.005 | 0.407 ± 0.013 | 0.618 ± 0.007 |
| | | MobileNetV2 | 0.671 ± 0.003 | 0.399 ± 0.006 | 0.598 ± 0.004 |
| | | ViT-B/16 | **0.803 ± 0.002** | **0.593 ± 0.007** | **0.694 ± 0.006** |
| | 2 | ResNet50 | 0.737 ± 0.014 | 0.462 ± 0.024 | 0.702 ± 0.018 |
| | | MobileNetV2 | 0.759 ± 0.019 | 0.437 ± 0.018 | 0.685 ± 0.013 |
| | | ViT-B/16 | **0.885 ± 0.017** | **0.619 ± 0.062** | **0.790 ± 0.025** |

Bold values indicate the best performance per group.

cohort (Table 2). In Case 1, mean AUC-ROC was 0.9999 ± 0.0001 for ResNet50, 0.9990 ± 0.0006 for MobileNetV2, and 0.9959 ± 0.0028 for ViT-B/16. In Case 2, mean AUC-ROC was 0.999 ± 0.0000 for both ResNet50 and MobileNetV2 and 0.9964 ± 0.0035 for ViT-B/16. Threshold-dependent metrics varied across architectures, with the CNN models achieving higher accuracy and macro-F1 than ViT-B/16 in this task. Because the comparison cohort was assembled from multiple public sources acquired using different OCT systems and protocols, additional source-stratified testing was performed to assess performance across the available held-out datasets. Confusion matrices demonstrated strong discrimination between GA and non-GA AMD. Because the comparison cohort was assembled from multiple public sources acquired using different OCT systems and protocols, additional source-stratified testing was performed to assess performance across the available held-out datasets.

For ViT-B/16 in Case 1, source-specific AUC-ROC remained high for DS2/OCTA-500 (0.9956 ± 0.0029), DS3/OCTID (0.9890 ± 0.0132), and DS4/OCTDL (0.9999 ± 0.0001). In Case 2, corresponding AUC-ROC values were 0.9961 ± 0.0039 for DS2, 0.9931 ± 0.0055 for DS3, and approximately 1.000 for DS4. Thus, strong discrimination was observed across each public source represented in the held-out test set, although some variation in threshold-dependent metrics was observed across datasets.

## Multi-class classification of CGA vs. NCGA vs. NGA

Simultaneous classification of CGA, NCGA, and NGA presented the greatest challenge for all models (Table 2, 4). In Case 1, the five-run ViT-B/16 B-scan-level analysis yielded an

TABLE 3 Class-specific B-scan-level performance of representative ViT-B/16 checkpoints for CGA-versus-NCGA and multiclass CGA-versus-NCGA-versus-NGA classification.

| Classification task | Case | Class | Sensitivity/ Recall | Specificity | PPV/ Precision | NPV | F1-score | Balanced accuracy | OvR AUC |
|---|---|---|---|---|---|---|---|---|---|
| CGA vs. NCGA | 1 | CGA | 0.965 | 0.242 | 0.798 | 0.688 | 0.873 | 0.603 | 0.720 |
| | | NCGA | 0.242 | 0.965 | 0.688 | 0.798 | 0.358 | 0.603 | 0.720 |
| | 2 | CGA | 0.995 | 0.053 | 0.845 | 0.667 | 0.914 | 0.524 | 0.701 |
| | | NCGA | 0.053 | 0.995 | 0.667 | 0.845 | 0.098 | 0.524 | 0.701 |
| CGA vs. NCGA vs. NGA | 1 | CGA | 0.882 | 0.616 | 0.738 | 0.810 | 0.804 | 0.749 | 0.852 |
| | | NCGA | 0.258 | 0.969 | 0.646 | 0.858 | 0.369 | 0.614 | 0.707 |
| | | NGA | 0.616 | 0.858 | 0.619 | 0.857 | 0.617 | 0.737 | 0.857 |
| | 2 | CGA | 0.899 | 0.771 | 0.877 | 0.808 | 0.888 | 0.835 | 0.924 |
| | | NCGA | 0.211 | 0.981 | 0.615 | 0.898 | 0.314 | 0.596 | 0.798 |
| | | NGA | 0.887 | 0.881 | 0.692 | 0.963 | 0.778 | 0.884 | 0.958 |

TABLE 4 Volume/visit-level performance of evaluated deep-learning architectures.

| Classification task | Case | Classifier name | Evaluation metrics | | |
|---|---|---|---|---|---|
| | | | AUC-ROC | F1 score | Accuracy |
| CGA vs. NCGA | 1 | ResNet50 | 0.572 ± 0.026 | 0.433 ± 0.000 | 0.765 ± 0.000 |
| | | MobileNetV2 | 0.545 ± 0.015 | 0.433 ± 0.000 | 0.765 ± 0.000 |
| | | ViT-B/16 | **0.837 ± 0.011** | **0.582 ± 0.026** | **0.788 ± 0.009** |
| | 2 | ResNet50 | 0.490 ± 0.021 | 0.441 ± 0.000 | 0.787 ± 0.000 |
| | | MobileNetV2 | 0.563 ± 0.068 | 0.441 ± 0.000 | **0.787 ± 0.000** |
| | | ViT-B/16 | **0.656 ± 0.070** | **0.439 ± 0.003** | 0.783 ± 0.009 |
| GA vs. NGA | 1 | ResNet50 | 0.899 ± 0.003 | 0.480 ± 0.081 | 0.746 ± 0.024 |
| | | MobileNetV2 | 0.861 ± 0.010 | 0.701 ± 0.022 | 0.811 ± 0.012 |
| | | ViT-B/16 | **0.905 ± 0.009** | **0.753 ± 0.013** | **0.817 ± 0.012** |
| | 2 | ResNet50 | 0.882 ± 0.005 | 0.651 ± 0.070 | 0.803 ± 0.026 |
| | | MobileNetV2 | 0.814 ± 0.015 | 0.591 ± 0.069 | 0.764 ± 0.033 |
| | | ViT-B/16 | **0.963 ± 0.007** | **0.870 ± 0.024** | **0.897 ± 0.022** |
| CGA vs. NCGA vs. NGA | 1 | ResNet50 | 0.778 ± 0.008 | 0.426 ± 0.030 | 0.649 ± 0.019 |
| | | MobileNetV2 | 0.734 ± 0.002 | 0.443 ± 0.025 | 0.660 ± 0.019 |
| | | ViT-B/16 | **0.867 ± 0.003** | **0.614 ± 0.027** | **0.734 ± 0.016** |
| | 2 | ResNet50 | 0.704 ± 0.024 | 0.473 ± 0.010 | 0.681 ± 0.013 |
| | | MobileNetV2 | 0.736 ± 0.022 | 0.389 ± 0.040 | 0.610 ± 0.031 |
| | | ViT-B/16 | **0.863 ± 0.016** | **0.620 ± 0.090** | **0.761 ± 0.026** |

Bold values indicate the best performance per group.

AUC-ROC of 0.8029 ± 0.0020, accuracy of 0.6916 ± 0.0063, and macro-F1 score of 0.5902 ± 0.0071. In Case 2, the corresponding AUC-ROC increased to 0.8845 ± 0.0165, with accuracy of 0.7896 ± 0.0245 and macro-F1 score of 0.6192 ± 0.0622 (Table 2). At the OCT volume/visit level, ViT-B/16 achieved a mean AUC-ROC of 0.867 ± 0.003 in Case 1 and 0.863 ± 0.016 in Case 2 (Table 4). Class-specific analysis continued to identify NCGA as the most challenging class. Detailed class-specific sensitivity, specificity, precision, NPV, and

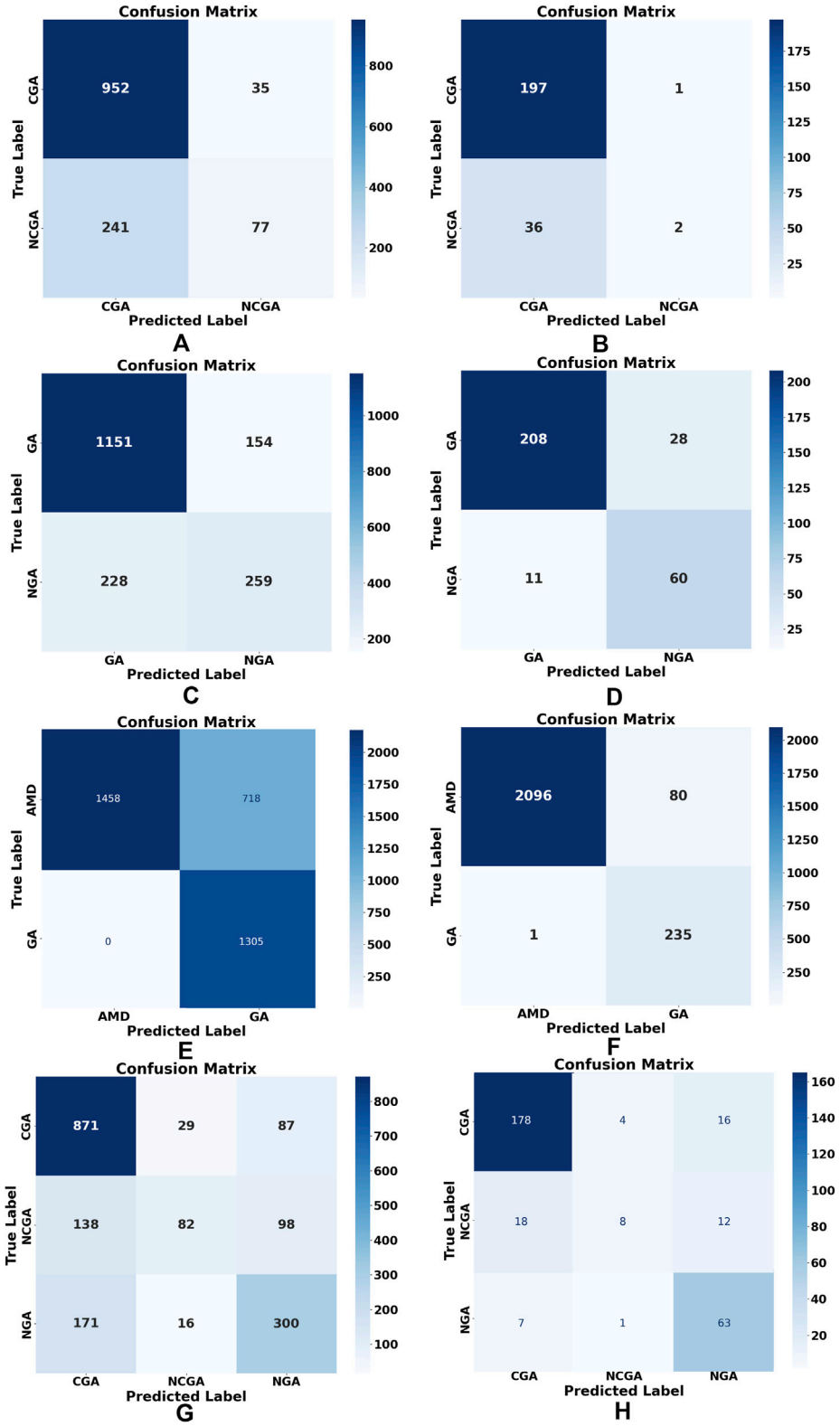



FIGURE 4
Confusion matrices for representative ViT-B/16 checkpoints across the evaluated classification tasks. Case 1: **(A)** CGA versus NCGA, **(C)** GA versus NGA, **(E)** GA versus non-GA AMD, and **(G)** CGA versus NCGA versus NGA. Case 2: **(B)** CGA versus NCGA, **(D)** GA versus NGA, **(F)** GA versus non-GA AMD, and **(H)** CGA versus NCGA versus NGA. Rows indicate true class labels and columns indicate predicted class labels.

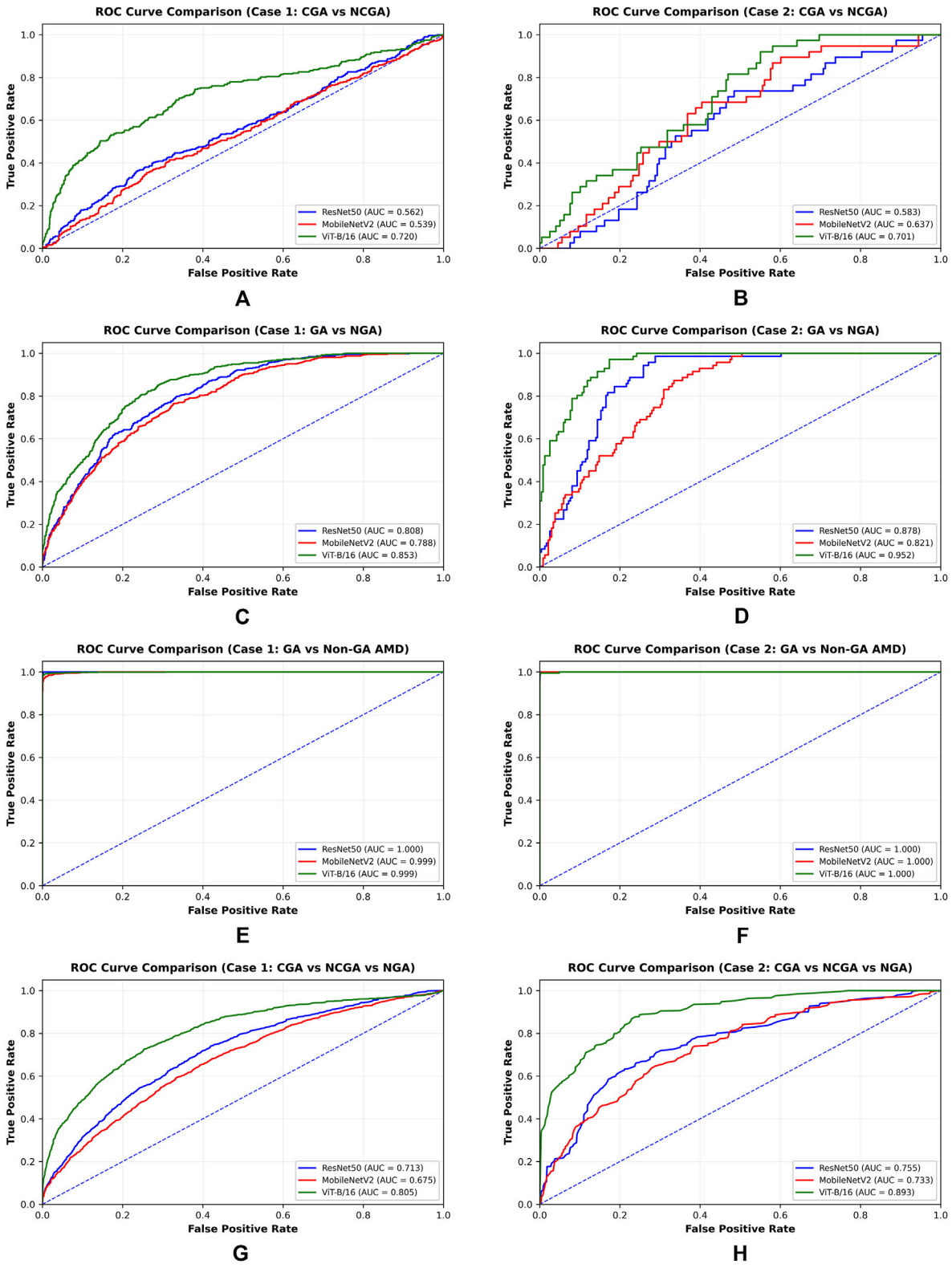



FIGURE 5
ROC Curves for different DL Models for GA classification tasks- Case 1: **(A)** CGA versus NCGA, **(C)** GA versus NGA, **(E)** GA versus non-GA AMD, and **(G)** CGA versus NCGA versus NGA. Case 2: **(B)** CGA versus NCGA, **(D)** GA versus NGA, **(F)** GA versus non-GA AMD, and **(H)** CGA versus NCGA versus NGA.

F1-score values for the representative ViT-B/16 checkpoints are provided in Table 3, with the corresponding confusion matrices shown in Figure 4. The ROC-AUC plots are demonstrated in Figure 5.

## Discussion

Recent advances in AI have enabled a growing number of DL approaches for analyzing GA in AMD. Nevertheless, most prior work has concentrated on GA detection, segmentation, or progression prediction rather than the classification of clinically meaningful phenotypic subtypes. Previous works include automated GA detection and quantification from OCT (Zhang et al.) [18] and Deep-GA-Net, an explainable Convolutional Neural Network (CNN) framework for OCT-based GA detection (Elsawy et al.) [28]. Others examined FAF-based segmentation, growth-rate prediction, and morphological quantification under complement-inhibitor therapy [29–31]. Although these studies collectively demonstrate the potential of AI for detecting and monitoring GA, they address lesion detection or progression assessment rather than the distinction between visually significant phenotypic subtypes defined by involvement of the fovea. To our knowledge, automated classification of CGA versus NCGA using OCT imaging has not previously been systematically investigated, and no prior studies have attempted CGA classification in clinically derived OCT datasets.

The distinction between CGA and NCGA has become increasingly important in clinical practice, as therapeutic response and visual prognosis differ substantially depending on foveal involvement [5]. Emerging complement-inhibitor therapies have demonstrated greater benefit in extrafoveal than foveal disease, underscoring the need for reliable and reproducible lesion localization [8]. While FAF remains the gold standard for GA lesion area measurement, OCT provides superior structural resolution for assessing foveal involvement [32, 33]. Nevertheless, OCT-based CGA versus NCGA classification is challenging due to poorly defined lesion boundaries, anatomical distortion, and significant inter-grader variability [34].

We developed and evaluated DL models across four clinically relevant OCT classification tasks: CGA versus NCGA; CGA versus NCGA versus NGA; GA versus NGA; and GA versus other non-GA AMD subtypes. Among the architectures tested, ResNet50 served as a residual-network baseline, MobileNetV2 evaluated lightweight efficiency, and ViT-B/16 was selected for its self-attention capacity to model long-range spatial dependencies relevant to foveal localization. ViT-B/16 demonstrated the most consistent discrimination across the evaluated tasks. Its architecture is well suited to modeling spatial relationships between retinal structural features; however, the present study was not designed to isolate the specific architectural mechanism responsible for the observed performance differences. MobileNetV2 remained competitive on selected tasks, particularly under selective B-scan analysis.

CGA-versus-NCGA classification remained the most challenging binary task, particularly for recognition of the NCGA class. This difficulty should be interpreted in the context of both the class distribution within the cohort and the clinical complexity of defining foveal involvement in advanced atrophy. Identification of the central 1-mm ETDRS region can be challenging when retinal anatomy is distorted, and prior OCT studies have reported variability in inter-grader agreement for atrophy-related features. The asymmetric class performance observed in our study may therefore reflect both dataset composition and the underlying complexity of the phenotyping task. NCGA was substantially less represented than CGA at the volume level (74 vs. 258 volumes) and among the selectively retained Case-2 B-scans (13.9% vs. 24.2% retention). Because training used unweighted cross-entropy without class-specific rebalancing, this unequal representation may have contributed to the observed bias toward CGA predictions. Evaluation in larger, more balanced cohorts, together with investigation of imbalance-aware training strategies, will be important for determining whether NCGA recognition can be improved.

A key methodological finding was the benefit of selective B-scan analysis. Restricting input to diagnostically informative B-scans containing the foveal region or prominent GA-associated structural features improved performance in several tasks, with the largest gain observed for ViT-B/16 in GA-versus-NGA classification. Case 2 was designed as an expert-guided, semi-automated analysis, and the retrospective selection of informative B-scans was performed after clinical grading. Accordingly, potential selection bias associated with expert-guided slice choice should be considered when interpreting this analysis. Future work incorporating automated foveal localization or informative-slice selection could extend this clinically informed strategy into a fully automated workflow. Grad-CAM analysis revealed distinct attention patterns across architectures (Figure 6). ResNet50 and MobileNetV2 generally demonstrated more diffuse activation across retinal regions, whereas ViT-B/16 showed more spatially localized activation in representative correctly classified CGA cases, frequently corresponding to regions of RPE and outer-retinal disruption near the foveal center. These qualitative findings complement the quantitative performance results by illustrating distinct attention patterns across the evaluated architectures. Because quantitative foveal or lesion-localization annotations were not available in the present dataset, the Grad-CAM findings were assessed qualitatively. Future studies incorporating spatial annotations will allow quantitative evaluation of the relationship between model attention, lesion location, and the foveal center. From a clinical standpoint, automated GA phenotyping could support risk stratification, patient counseling, individualized monitoring, treatment selection, and consistent clinical trial eligibility criteria.

The strong GA-versus-non-GA AMD results were obtained using a comparison cohort assembled from multiple public datasets with heterogeneous acquisition systems and imaging protocols. To examine whether performance was dependent on an individual public source, we performed source-stratified

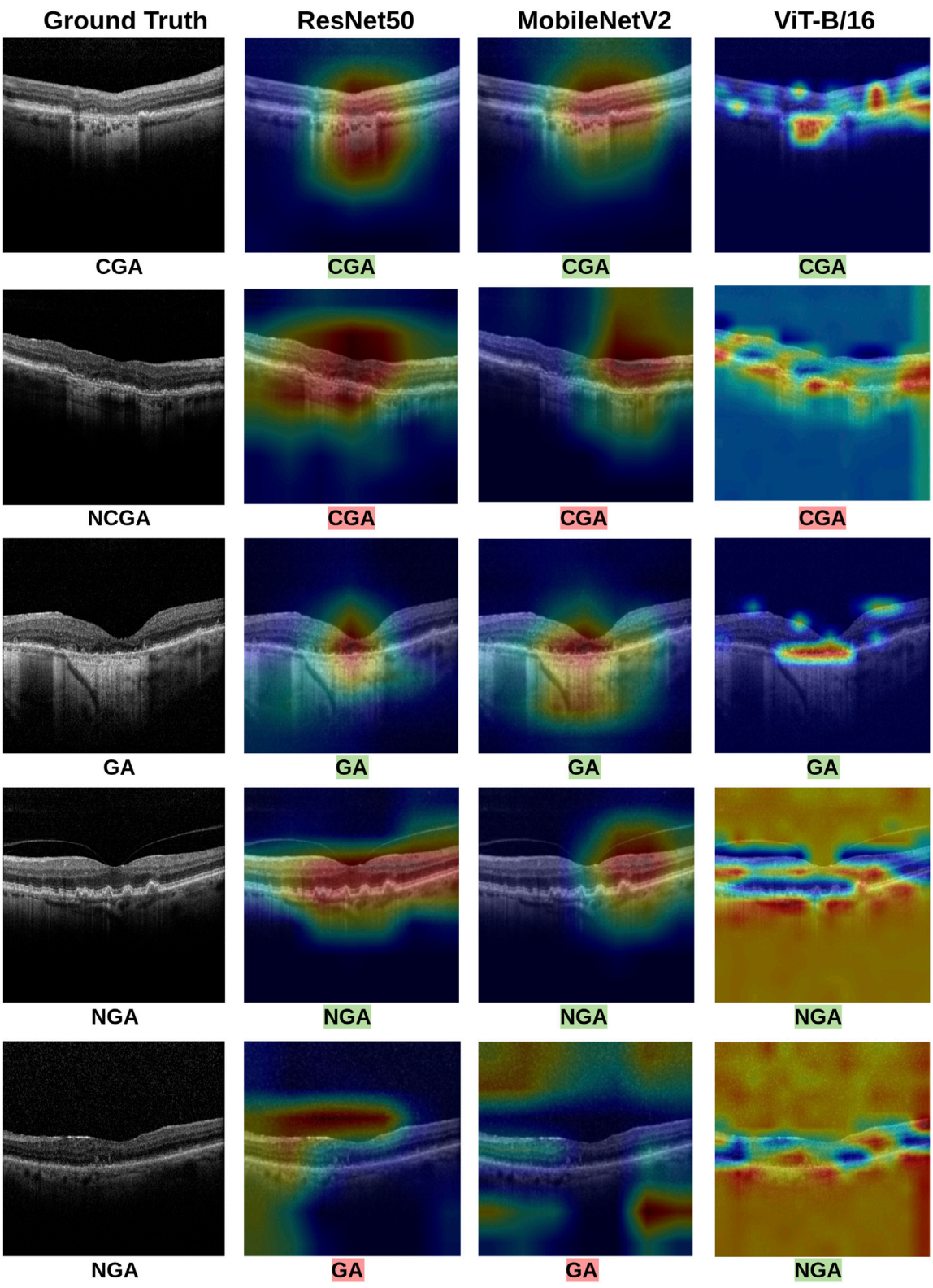



FIGURE 6
Grad-CAM visualizations revealing feature attention patterns across DL architectures for GA classification (Predictions with green highlighted text indicate correct classifications, while red text indicates misclassifications).

testing using the same held-out GA images paired separately with each public dataset represented in the test set. AUC-ROC remained consistently high across DS2, DS3, and DS4, indicating that the strong overall discrimination was not driven exclusively by a single comparison dataset. Some variation in threshold-dependent metrics was observed across sources. These findings demonstrate consistent discrimination across heterogeneous comparison datasets, while further validation under harmonized or matched acquisition conditions will help characterize cross-device and cross-dataset generalizability.

In conclusion, this study demonstrates the feasibility of OCT-based deep learning for automated GA phenotyping across several clinically relevant classification tasks. ViT-B/16 demonstrated the most consistent discrimination among the evaluated architectures, while selective B-scan analysis improved performance in several settings. CGA-versus-NCGA classification remained the most challenging task, particularly for NCGA recognition, identifying an important direction for further optimization. The performance of ViT-B/16 and its observed Grad-CAM patterns are consistent with the potential benefit of self-attention for OCT classification tasks requiring integration of pathological features with broader anatomical context. With evaluation in larger, more balanced and externally validated cohorts, automated GA phenotyping has the potential to support reproducible disease characterization and clinical research.

These promising findings should, however, be interpreted in the context of several limitations. This single-center, retrospective dataset was acquired exclusively on the Heidelberg Spectralis OCT system, necessitating external validation across diverse platforms and populations. The modest cohort size and class imbalance between CGA and NCGA may have contributed to the lower recognition of NCGA, and larger multicenter datasets with more balanced subtype representation may further strengthen class-specific performance. The inclusion of smaller or earlier atrophic lesions broadens the spectrum of atrophic changes represented in the cohort, including cases closer to the CGA/NCGA boundary, which should be considered when interpreting subtype-classification performance. Finally, published inter-grader agreement coefficients ranging from 0.28 to 0.92 even among experienced retinal specialists suggest that the moderate AUC-ROC for CGA versus NCGA classification may partly reflect an irreducible upper bound imposed by diagnostic subjectivity, underscoring both the clinical need for and the inherent challenge of automated foveal localization in advanced atrophic disease. With validation on larger, multicenter datasets, automated GA classification systems have the potential to standardize phenotyping, support clinical decision-making, and improve patient management.

## Author contributions

SS contributed to methodology, data analysis, model development, visualization, and writing of the original draft. HK contributed to data collection. SG and AN contributed to validation and writing – review and editing. NR contributed to second-round data label grading and writing – review and editing. SO contributed to first-round data label grading, supervision, and writing – review and editing. MA contributed to conceptualization, supervision, funding acquisition, and writing – review and editing. All authors contributed to the article and approved the submitted version.

## Data availability

The datasets presented in this article are not readily available because the dataset analyzed in this study consists of retrospective clinical OCT images obtained from patients at Wake Forest University School of Medicine. Due to patient privacy protections, IRB restrictions, and institutional policies governing clinical data, the dataset is not publicly available and cannot be shared outside the study team. Requests to access the datasets should be directed to MA, malam8@charlotte.edu; SO, sally.ong@wfusm.edu.

## Ethics statement

The studies involving humans were approved by Institutional Review Board (IRB) at Wake Forest University School of Medicine. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the National Eye Institute (NEI) through grants 1R01EY037828-01, R21EY035271, and R15EY035804, which support the development and evaluation of AI/ML approaches for ophthalmic imaging, disease classification, progression prediction, and generalizable learning from OCT/OCTA data.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. During the preparation of this work the authors used generative artificial intelligence tools for English language editing and grammar correction. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

# References


1. Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng C-Y, 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:e106–116. doi: 10.1016/S2214-109X(13)70145-1

2. Fleckenstein M, Mitchell P, Freund KB, Sadda S, Holz FG, Brittain C, et al. The progression of geographic atrophy secondary to age-related macular degeneration. *Ophthalmology* (2018) 125:369–90. doi: 10.1016/j.ophtha.2017.08.038

3. Evolution of geographic atrophy of the retinal pigment epithelium. *Eye*. doi: 10.1038/eye.1988.106

4. Sunness JS, Rubin GS, Applegate CA, Bressler NM, Marsh MJ, Hawkins BS, et al. Visual function abnormalities and prognosis in eyes with age-related geographic atrophy of the macula and good visual acuity. *Ophthalmology* (1997) 104:1677–91. doi: 10.1016/s0161-6420(97)30079-7

5. Schmitz-Valckenberg S, Fleckenstein M, Scholl HPN, Holz FG. Fundus autofluorescence and progression of age-related macular degeneration. *Surv Ophthalmol* (2009) 54:96–117. doi: 10.1097/IAE.0b013e318164a907

6. Liao DS, Grossi FV, El Mehdi D, Gerber MR, Brown DM, Heier JS, et al. Complement C3 inhibitor pegcetacoplan for geographic atrophy secondary to age-related macular degeneration: a randomized phase 2 trial. *Ophthalmology* (2020) 127:186–95. doi: 10.1016/j.ophtha.2019.07.011

7. Jaffe GJ, Westby K, Csaky KG, Monés J, Pearlman JA, Patel SS, et al. C5 inhibitor avacincaptad pegol for geographic atrophy due to age-related macular degeneration: a randomized pivotal phase 2/3 trial. *Ophthalmology* (2021) 128:576–86. doi: 10.1016/j.ophtha.2020.08.027

8. Heier JS, Lad EM, Holz FG, Rosenfeld PJ, Guymer RH, Boyer D, et al. OAKS and DERBY study investigators. Pegcetacoplan for the treatment of geographic atrophy secondary to age-related macular degeneration (OAKS and DERBY): two multicentre, randomised, double-masked, sham-controlled, phase 3 trials. *Lancet* (2023) 402:1434–48. doi: 10.1016/S0140-6736(23)01520-9

9. Schmitz-Valckenberg S, Bindewald-Wittich A, Dolar-Szczasny J, Dreyhaupt J, Wolf S, Scholl HPN, et al. Correlation between the area of increased autofluorescence surrounding geographic atrophy and disease progression in patients with AMD. *Invest Ophthalmol Vis Sci* (2006) 47:2648–54. doi: 10.1167/iovs.05-0892

10. Keane PA, Patel PJ, Liakopoulos S, Heussen FM, Sadda SR, Tufail A. Evaluation of age-related macular degeneration with optical coherence tomography. *Surv Ophthalmol* (2012) 57:389–414. doi: 10.1016/j.survophthal.2012.01.006

11. Fleckenstein M, Charbel Issa P, Helb H-M, Schmitz-Valckenberg S, Finger RP, Scholl HPN, et al. High-resolution spectral domain-OCT imaging in geographic atrophy associated with age-related macular degeneration. *Invest Ophthalmol Vis Sci* (2008) 49:4137–44. doi: 10.1167/iovs.08-1967

12. Schmitz-Valckenberg S, Sahel J-A, Danis R, Fleckenstein M, Jaffe GJ, Wolf S, et al. Natural history of geographic atrophy progression secondary to age-related macular degeneration (geographic atrophy progression study). *Ophthalmology* (2016) 123:361–8. doi: 10.1016/j.ophtha.2015.09.036

13. Wu Z, Schmitz-Valckenberg S, Blodi BA, Holz FG, Jaffe GJ, Liakopoulos S, et al. Reticular pseudodrusen: interreader agreement of evaluation on OCT imaging in age-related macular degeneration. *Ophthalmol Sci* (2023) 3:100325. doi: 10.1016/j.xops.2023.100325

14. Patel PJ, Browning AC, Chen FK, Da Cruz L, Tufail A. Interobserver agreement for the detection of optical coherence tomography features of neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci* (2009) 50:5405–10. doi: 10.1167/iovs.09-3505

15. Spaide T, Rajesh AE, Gim N, Blazes M, Lee CS, Macivannan N, et al. Estimating uncertainty of geographic atrophy segmentations with bayesian deep learning. *Ophthalmol Sci* (2025) 5:100587. doi: 10.1016/j.xops.2024.100587

16. Chatzara A, Maliagkani E, Mitsopoulou D, Katsimpris A, Apostolopoulos ID, Papageorgiou E, et al. Artificial intelligence approaches for geographic atrophy segmentation: a systematic review and meta-analysis. *Bioengineering* (2025) 12:475. doi: 10.3390/bioengineering12050475

17. Mai J, Lachinov D, Reiter GS, Riedl S, Grechenig C, Bogunovic H, et al. Deep learning-based prediction of individual geographic atrophy progression from a single baseline OCT. *Ophthalmol Sci* (2024) 4:100466. doi: 10.1016/j.xops.2024.100466

18. Zhang G, Fu DJ, Liefers B, Faes L, Glinton S, Wagner S, et al. Clinically relevant deep learning for detection and quantification of geographic atrophy from optical coherence tomography: a model development and external validation study. *Lancet Digit Health* (2021) 3:e665–e675. doi: 10.1016/S2589-7500(21)00134-5

19. Siraz S, Kamanda H, Gholami S, Nabil AS, Yee Ong SS, Alam MN. Multi-class classification of central and non-central geographic atrophy using optical coherence tomography. *Medrxiv Prepr Serv Health Sci* (2025) 2025:2025.05.27.25328446. doi: 10.1101/2025.05.27.25328446

20. Srinivasan PP, Kim LA, Mettu PS, Cousins SW, Comer GM, Izatt JA, et al. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed Opt Express* (2014) 5:3568–77. doi: 10.1364/BOE.5.003568

21. Li M, Huang K, Xu Q, Yang J, Zhang Y, Ji Z, et al. OCTA-500: a retinal dataset for optical coherence tomography angiography study. *Med Image Anal* (2024) 93:103092. doi: 10.1016/j.media.2024.103092

22. Gholami P, Roy P, Parthasarathy MK, Lakshminarayanan V. OCTID: optical coherence tomography image database. (2019). doi: 10.48550/arXiv.1812.07056

23. Kulyabin M, Zhdanov A, Nikiforova A, Stepichev A, Kuznetsova A, Ronkin M, et al. OCTDL: optical coherence tomography dataset for image-based deep learning methods. *Sci Data* (2024) 11:365. doi: 10.1038/s41597-024-03182-7

24. *resnet50 — Torchvision Main Documentation* (2026). doi: 10.1109/CVPR.2016.90

25. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. MobileNetV2: inverted residuals and linear bottlenecks. (2025). doi: 10.48550/arXiv.1801.04381

26. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. (2026). doi: 10.48550/arXiv.1706.03762

27. Kingma DP, Ba J. Adam: a method for stochastic optimization. (2026). doi: 10.48550/arXiv.1412.6980

28. Elsawy A, Keenan TDL, Chen Q, Shi X, Thavikulwat AT, Bhandari S, et al. Deep-GA-Net for accurate and explainable detection of geographic atrophy on OCT scans. *Ophthalmol Sci* (2023) 3:100311. doi: 10.1016/j.xops.2023.100311

29. Arslan J, Samarasinghe G, Sowmya A, Benke KK, Hodgson LAB, Guymer RH, et al. Deep learning applied to automated segmentation of geographic atrophy in fundus autofluorescence images. *Transl Vis Sci Technol* (2021) 10:2. doi: 10.1167/tvst.10.8.2

30. Cluceru J, Anegondi N, Gao SS, Lee AY, Lad EM, Chakravarthy U, et al. Topographic clinical insights from deep learning–based geographic atrophy progression prediction. *Transl Vis Sci Technol* (2024) 13:6. doi: 10.1167/tvst.13.8.6

31. Prediction of individual disease conversion in early AMD using artificial intelligence | IOVS. *ARVO Journals* (2026) doi: 10.1167/iovs.18-24106

32. Schmitz-Valckenberg S, Holz FG, Bird AC, Spaide RF. Fundus autofluorescence imaging: review and perspectives. *Retina* (2008) 28:385–409.

33. Keane PA, Sadda SR. Retinal imaging in the twenty-first century: state of the art and future directions. *Ophthalmology* (2014) 121:2489–500. doi: 10.1016/j.ophtha.2014.07.054

34. Holz FG, Bindewald-Wittich A, Fleckenstein M, Dreyhaupt J, Scholl HPN, Schmitz-Valckenberg S, et al. Progression of geographic atrophy and impact of fundus autofluorescence patterns in age-related macular degeneration. *Am J Ophthalmol* (2007) 143:463–72. doi: 10.1016/j.ajo.2006.11.041