



A Deep Learning Model for Segmentation of Geographic Atrophy to Study Its Long-Term Natural History

Bart Liefers, MSc,^{1,2} Johanna M. Colijn, MD, MSc,^{3,4} Cristina González-Gonzalo, MSc,^{1,2} Timo Verzijden, MSc,^{3,4} Jie Jin Wang, PhD,^{5,6} Nichole Joachim, BSc,⁵ Paul Mitchell, MD, PhD,⁵ Carel B. Hoyng, MD, PhD,^{2,7} Bram van Ginneken, PhD,¹ Caroline C.W. Klaver, MD, PhD,^{2,3,4,7,8} Clara I. Sánchez, PhD^{1,2,7}

Purpose: To develop and validate a deep learning model for the automatic segmentation of geographic atrophy (GA) using color fundus images (CFIs) and its application to study the growth rate of GA.

Design: Prospective, multicenter, natural history study with up to 15 years of follow-up.

Participants: Four hundred nine CFIs of 238 eyes with GA from the Rotterdam Study (RS) and Blue Mountain Eye Study (BMES) for model development, and 3589 CFIs of 376 eyes from the Age-Related Eye Disease Study (AREDS) for analysis of GA growth rate.

Methods: A deep learning model based on an ensemble of encoder–decoder architectures was implemented and optimized for the segmentation of GA in CFIs. Four experienced graders delineated, in consensus, GA in CFIs from the RS and BMES. These manual delineations were used to evaluate the segmentation model using 5-fold cross-validation. The model was applied further to CFIs from the AREDS to study the growth rate of GA. Linear regression analysis was used to study associations between structural biomarkers at baseline and the GA growth rate. A general estimate of the progression of GA area over time was made by combining growth rates of all eyes with GA from the AREDS set.

Main Outcome Measures: Automatically segmented GA and GA growth rate.

Results: The model obtained an average Dice coefficient of 0.72 ± 0.26 on the BMES and RS set while comparing the automatically segmented GA area with the graders' manual delineations. An intraclass correlation coefficient of 0.83 was reached between the automatically estimated GA area and the graders' consensus measures. Nine automatically calculated structural biomarkers (area, filled area, convex area, convex solidity, eccentricity, roundness, foveal involvement, perimeter, and circularity) were significantly associated with growth rate. Combining all growth rates indicated that GA area grows quadratically up to an area of approximately 12 mm², after which growth rate stabilizes or decreases.

Conclusions: The deep learning model allowed for fully automatic and robust segmentation of GA on CFIs. These segmentations can be used to extract structural characteristics of GA that predict its growth rate. *Ophthalmology* 2020;■:1–11 © 2020 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Supplemental material available at www.aaojournal.org.

Geographic atrophy (GA) occurs in the advanced stage of age-related macular degeneration (AMD). It is characterized by progressive atrophy of the retinal pigment epithelium, overlying photoreceptors, and underlying choriocapillaris.¹ Areas of GA often initially appear extrafoveal, which may result in their causing difficulties in reading or dim-light vision.² Over time, the atrophic area may grow, and when it reaches the fovea, visual acuity is severely diminished. Prevalence of GA increases exponentially with age³ and is highest in people of European ancestry.⁴ The number of people affected by GA is expected to increase further in the near future because of the aging population.⁵

Currently, no approved treatment exists to prevent progression of GA.^{6,7} However, clinical trials of several potential therapies are underway.⁸ For evaluation of these trials, reliable anatomic end points are required, because visual acuity alone provides insufficient insight in the severity of the disease.⁹ Growth rate of the atrophic area has been suggested as an important indicator of disease progression.^{9–11} However, the speed at which GA progresses varies greatly between patients.^{12–14} Therefore, understanding the patterns associated with progression and the variability between patients is important for the design and interpretation of clinical trials.

To assess growth rate, accurate delineation of the GA area is required. However, because manual delineation can be challenging and time consuming,^{15,16} automatic segmentation could provide a scalable and reproducible alternative. Deep learning has emerged as a powerful technique for the automatic analysis of medical images.¹⁷ Deep learning models require labeled examples (training data) to tune their internal parameters. The model then learns to extract features that are important for the segmentation task without further need for explicit domain knowledge from experts. It has been successfully applied to color fundus images (CFIs) for classification of severity stages in AMD^{18,19} and diabetic retinopathy^{20–22} and recently also for the detection of GA.²³ Although manually labeled examples are still required for training and validation, the model thereafter can be applied to large data sets, opening up new possibilities for studies and possibly reducing the overall effort that is required from experienced graders.

These automatic methods also have the potential to extract structural characteristics of GA efficiently and accurately, as seen in imaging that has been demonstrated to correlate with growth rate. For example, multifocal lesions grow faster than unifocal lesions,²⁴ and extrafoveal lesions grow faster than foveal lesions.¹³ Circular lesions have been demonstrated to grow at a slower rate than more irregularly shaped lesions.²⁵ Baseline lesion area has been consistently associated with future growth, with larger lesions growing faster than smaller lesions.^{11,13,26,27} However, applying a square root transformation to the lesion size may remove this dependency.^{16,28} Therefore, it has been hypothesized that lesions with approximate circular shape grow at a constant radial speed, thus leading to a quadratic growth of the area.^{16,29}

Various imaging methods have been used to assess GA. Color fundus imaging has been used most widely historically, particularly in large epidemiologic studies.¹² More recently, fundus autofluorescence (FAF) and OCT have also become popular for the study of GA and GA progression.^{13,16,27} Several lesion characteristics visible with those methods can be linked to progression of GA. For example, banded or diffuse perilesional patterns on FAF and structural abnormalities at the junctional zone on OCT have been associated with faster GA progression.^{13,27,30} Although GA may be detected earlier on FAF images than CFIs,³¹ good agreement on quantification of GA area on CFIs between 2 independent reading centers has been demonstrated,¹¹ and progression rates assessed from both FAF images and CFIs are highly correlated.^{13,31} Color fundus imaging has the advantage that it is widely available, often over longer periods, making it suitable for the study of long-term progression of GA.

Previous work on automatic methods for segmentation of GA focuses mainly on OCT^{32–34} or FAF.³⁵ Feeny et al³⁶ proposed a method based on a random forest classifier in color fundus imaging. In contrast, herein we present a model that is based on deep learning.

The purpose of this study is twofold: (1) to develop and validate a fully automatic model for segmentation of GA on

CFIs and (2) to demonstrate its usefulness in a longitudinal setting for the study of GA progression. The performance of the developed model was compared against the work of 4 graders on a challenging dataset to evaluate its robustness. Next, the automatically segmented GA areas provided measures of structural characteristics related to lesion size, location, and morphologic features. We investigated the associations between those structural characteristics at baseline and the subsequent growth rate of GA. Finally, we combined GA growth rates across patients to obtain an estimate of average progression of GA area over time.

Methods

Data

Data for development and evaluation of the deep learning model for GA segmentation were collected from the Blue Mountains Eye Study (BMES)³⁷ and the Rotterdam Study (RS) cohorts I, II, and III.³⁸ The developed model was applied to CFIs from the Age-Related Eye Disease Study (AREDS)¹¹ for the assessment of GA growth rate.

The BMES is a population study from the Blue Mountains region in Australia that started between 1992 and 1994 and included 3654 participants 49 years of age or older. For the first 3 visits, 30° macula-centered (field 2) CFIs were obtained with a Zeiss fundus camera (Carl Zeiss AG, Oberkochen, Germany). For the fourth visit, 40° macula-centered digital CFIs were obtained with a Canon CF-60 DSi with DS Mark II body (Canon, Tokyo, Japan). The BMES was approved by the University of Sydney and the Sydney West Area Health Service Human Research Ethics Committees.

The RS is a population study from a suburb in Rotterdam, The Netherlands. The RS cohort I started in 1990 and included 7983 participants 55 years of age or older. Cohort II started in 2000 and included 3011 participants 55 years of age or older. Cohort III started in 2006 and included 3932 participants 45 years of age or older. The CFIs for the first examinations were obtained with a Topcon TRV-50VT (Topcon Optical Company, Tokyo, Japan), and those from the last 2 examinations were obtained with a Topcon TRC 50EX and a Sony DXC-950P (Sony Electronics Inc., New York, NY) digital camera. All CFIs were 35° and macula centered (field 2). The RS was approved by the Medical Ethics Committee of the Erasmus Medical Center and by the Netherlands Ministry of Health, Welfare and Sport.

The AREDS is a long-term, multicenter, prospective study of the clinical course of AMD and cataract. Starting between 1992 and 1998, 11 clinics in the United States enrolled 4757 participants between 55 and 80 years of age. Stereoscopic CFIs (30° macula centered) were acquired with a Zeiss FF-series camera (Carl Zeiss AG). The AREDS was approved by an independent institutional review board at each clinical center.

The follow-up interval for RS and BMES was 5 years. The AREDS had follow-ups at 6-month intervals, although the typical interval between available CFIs was 1 year. The BMES, RS, and AREDS all adhered to the tenets of the Declaration of Helsinki, and all participants provided informed consent.

A total of 504 CFIs of patients diagnosed with AMD and signs of GA were included from the BMES and RS sets. Twenty-six images with mixed signs of AMD (neovascularization, bleedings, scars) were excluded to disambiguate overlapping areas. Furthermore, no GA was delineated in 43 images because it was either not present or ungradable, and 26 images were excluded because of poor image quality. The remaining 409 images were included for

development of the model and evaluation of its performance. This set contained 87 images from the BMES (26 participants, 43 eyes) and 322 images from the RS (149 participants, 195 eyes). The 409 images represent 315 unique visits (some visits had 2 CFIs available).

Images for the study of GA progression were selected from the AREDS set, following the grading available from the database of genotype and phenotype 2014 table. Inclusion criteria were presence of GA or central GA and at least 2 years of follow-up. Images with neovascular disease co-occurring with GA were excluded. A total of 3589 images of 376 eyes were included. Most of these images were stereoscopic, so this accounted for 1826 unique acquisitions (eye-visit). Pixel-to-millimeter conversion was fixed for all images, based on the average distance between the fovea and center of the optic disc measured in a subset of the images. This distance was assumed to be 4.5 mm.³⁹

Delineations of GA area were made by 4 graders (3 of them with more than 20 years of experience), using an in-house created software platform for manual annotations (https://www.a-eye-research.nl/software/ophthalmology_workstation/).⁴⁰ For the RS, additional multimodal imaging (infrared, FAF, OCT, or a combination thereof) was available for some of the visits, and the platform allowed images of the same eye (both multimodal and longitudinal) to be aligned manually by identifying corresponding landmarks. The graders could view images of the same eye simultaneously using a synchronized cursor on multiple screens. Geographic atrophy was identified as the absence of the retinal pigment epithelium and increased visibility of the choriocapillaris on CFIs. Additional evidence from other methods was used whenever available. Areas of macular and peripapillary atrophy were delineated as separate classes, but for this study, only macular GA was used.

Each grader annotated the entire BMES set, whereas the RS set was divided in such a way that each grader annotated approximately half of the entire set and every image was graded by at least 2 graders. Finally, a consensus grading was made for all images in both sets. During the consensus grading, all graders decided together which of the individual gradings was most accurate and updated this grading, if necessary, until consensus was reached. If 2 CFIs of the same visit were present, both were included for model development, and the delineated GA area was propagated from one image to the other by using the affine transformation calculated from the manual landmarks. For evaluation, only the CFI that was used to make the consensus grading was used. Additionally, for external validation of the model, we randomly selected 100 CFIs from 100 participants in the subset of AREDS images that contain GA (including 32 CFIs with co-occurring neovascular lesions). A single grader delineated the GA area in all of these images, whereas an additional 2 graders delineated GA in 50 of the 100 selected CFIs.

Deep Learning Model

The proposed deep learning model for GA segmentation consisted of an ensemble of several models, each trained with partly overlapping training sets. The network architecture (the topology of connections between internal parameters of the deep learning model) for each model consisted of a deep encoder–decoder structure with residual blocks and shortcut connections, similar to that of De Fauw et al,⁴¹ but adapted to work with CFIs. This architecture, and its variations, can be characterized by a contracting path in which the high-resolution input image is converted to a low-resolution abstract representation, followed by an expanding path in which the original resolution is reconstructed. The contracting and expanding path are connected by shortcut connections. This approach has been shown to be very effective for

semantic segmentation in medical imaging for which large contextual information is required.^{17,42}

Input to each model was both the original color (RGB) image and a contrast-enhanced version of the same image, both resampled to 512×512 pixels. The contrast-enhanced image was obtained by subtracting a blurred image from the original image.⁴³ The input was transformed through the many layers of artificial neurons in the contracting and expanding path and ultimately yielded a new image in which the value of every pixel represented a likelihood of being part of an area of GA. A schematic overview of the model can be found in Figure 1. More details about the model and the training procedure used for this study can be found in the Appendix (available at www.aaojournal.org).

Geographic Atrophy Segmentation

For the development and validation of the model, we applied a 5-fold cross-validation scheme. Data from the BMES and RS were merged into 1 dataset and split randomly at the patient level into 5 approximately equal folds. In a rotating scheme, 4 folds were used for model training and validation (development set), whereas the remaining fold was used for performance evaluation (test set). Furthermore, 4 separate models were created within each development set. Each model used 3 folds for tuning of the internal parameters (training) and 1 for validation. An ensemble of these 4 models was then evaluated on the respective test set. The output of the ensemble model was obtained by taking the average output of the individual models for every pixel, after correcting for differences in sensitivity between models. This procedure is explained in more detail in the Appendix (available at www.aaojournal.org). Ultimately, an ensemble of the 20 obtained models (4 models developed for each of the 5 rounds) constituted the final model. Performance of this model was validated on the selection of 100 CFIs from the AREDS set.

The performance of the model and the agreement between graders were assessed using the Dice coefficient, which is defined as 2 times the intersection of 2 areas divided by the sum of the individual areas. Hence, a value of 0 represents disjoint areas (no overlap), whereas a value of 1 represents perfect agreement. Dice coefficients were calculated between graders to assess the inter-observer agreement, whereas the areas delineated in the consensus grading were used as reference for the model. Note that the consensus grading was not independent of the individual gradings and therefore could not be used as a reference to estimate graders' performances. Furthermore, the intraclass correlation coefficient of the GA area and the square root of the GA area were used to measure agreement between graders and the model.

Geographic Atrophy Growth Rate

The final deep learning model (the ensemble of 20 models) was applied to CFIs from AREDS for the analysis of GA progression. It is well documented that GA area increases faster for larger lesions. To remove the dependency of baseline lesion size on growth rate, many researchers apply a square root transformation to the GA area.²⁸ Similarly, we calculated the square root annual growth in millimeters per year for each eye to assess progression in the AREDS set.³⁹ This value was obtained from the slope of a linear regression through the square root of the GA area for a selected set of time points. The selected set consisted of all available CFIs within a window of 2 years for which the number of available CFIs was highest for the respective eye. The window was limited at 2 years because growth rate and lesion characteristics may change over time.²⁵ We calculated the correlation of square root annual growth rate between fellow eyes and compared the growth rate between groups using an

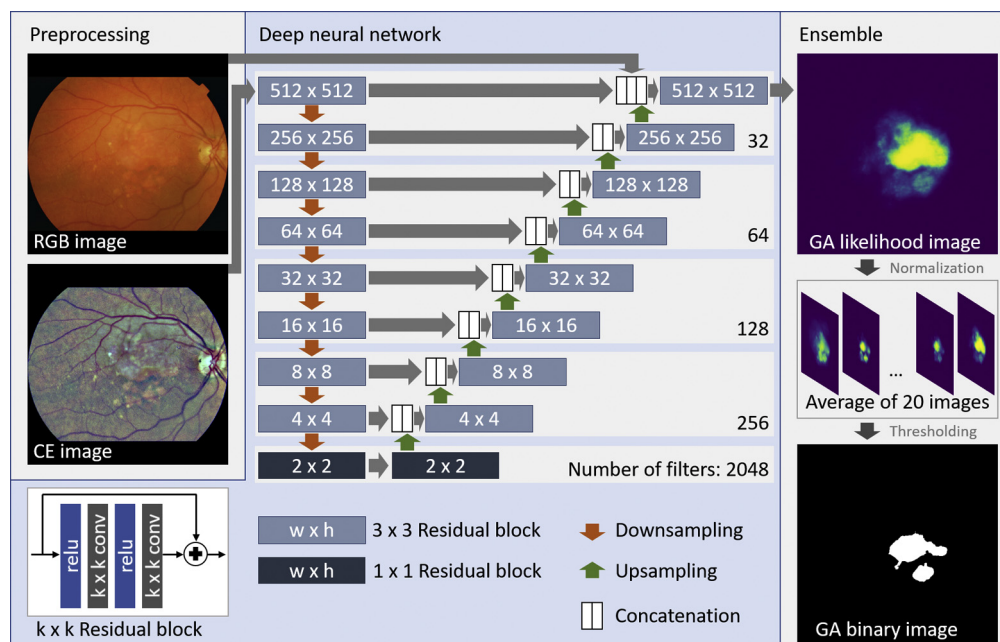


Figure 1. Schematic overview of the model. On the left, the preprocessed input: a 512×512 color image (RGB) and a contrast-enhanced (CE) version. In the middle, the model, with the downsampling path (orange arrows), upsampling path (green arrows), and shortcut connections (gray arrows). The ensemble model combines multiple outputs into a single binary image. The geographic atrophy (GA) area in the chosen example is intentionally ambiguous to highlight how the ensemble handles differences in predicted GA between the individual models.

unpaired t test for unilateral versus bilateral cases, unifocal versus multifocal cases, and foveal versus extrafoveal cases.

To identify structural characteristics or features that may be predictive for growth rate, we built a linear model based on features that were extracted from the segmented GA area at baseline (the first image within the selected window). Candidate features were area, perimeter, convex area, filled area, convex solidity (area divided by convex area), filled area (area divided by filled area), number of lesions, eccentricity, circularity, roundness, and foveal involvement. Details on how these features were calculated can be found in the [Appendix](http://www.aojournal.org) (available at www.aojournal.org). Associations between individual features and square root annual growth rate were calculated using univariate linear regression. Because the features were not independent, a multivariate linear model was created to investigate further which features best explain variation in square root annual growth rate. The multivariate model was built using forward selection by iteratively adding the feature that yielded the highest increase in adjusted R^2 value, until it increased no further. When stereoscopic images were available, lesion characteristics were represented by the mean of the 2 calculated values. To obtain a more homogeneous set for the prediction model, we discarded images in which the relative difference in GA area between the left and right stereoscopic image was more than 50% and included only eyes with at least 2 years of follow-up images.

Finally, we combined all estimates of GA growth in a single figure. Geographic atrophy growth in square millimeters per year (not square root transformed) was estimated as a function of GA area, again using a linear regression for each eye through the GA area in a window of 2 years. This resulted in an estimate of GA growth (the slope of the regression), bounded by a minimum and maximum GA area. The estimated general GA growth for a given GA area was then represented by the mean of all growth estimates for which this GA area fell within the respective area bounds. Confidence intervals were estimated using bootstrapping.

Results

Geographic Atrophy Segmentation

The deep learning model reached a Dice coefficient of 0.72 ± 0.26 ($n = 315$), measured in cross-validation in the BMES and RS data sets, where each test fold was evaluated by the ensemble of 4 models. Dice coefficients between 2 independent graders ranged from 0.72 ± 0.26 to 0.82 ± 0.21 (0.78 ± 0.24 on average). See [Table 1](#) for more details. The intraclass correlation coefficient between the model and the consensus was 0.83 for GA area and 0.84 for the square root of the GA area. Consistency in those values is visualized further in [Figure 2](#) using Bland-Altman plots. The mean value of the differences between consensus and model did not differ significantly from 0 on the basis of a 1-sample t test for either GA area ($P = 0.82$) or square root GA area ($P = 0.22$). Examples of manually and automatically segmented GA areas can be found in [Figure 3](#).

Table 1. Dice Coefficients between Model and Consensus Grading and between Individual Graders

	No.	Dice coefficient
Model – consensus	315	0.72 ± 0.26
Grader 1 – grader 2	146	0.80 ± 0.27
Grader 1 – grader 3	138	0.78 ± 0.27
Grader 1 – grader 4	90	0.72 ± 0.26
Grader 2 – grader 3	91	0.82 ± 0.21
Grader 2 – grader 4	134	0.78 ± 0.22
Grader 3 – grader 4	130	0.78 ± 0.19

The average Dice coefficient on the AREDS set was 0.66 ± 0.27 ($n = 50$) for the model, compared with 0.73 ± 0.24 (grader 1) and 0.73 ± 0.27 (grader 2). The intra-class correlation coefficient between the model and reference grader was 0.77 for GA area and 0.80 for the square root of the GA area ($n = 100$). The mean value of the differences between reference and model did not differ significantly from 0 on the basis of a 1-sample t test for either GA area ($P = 0.59$) or square root GA area ($P = 0.54$). Examples of automatic segmentation results on the AREDS set can be found in Figure S1 (available at www.aaojournal.org).

Geographic Atrophy Growth Rate

After excluding visits at which the difference between left and right stereoscopic images in the automatically segmented area was more than 50%, 335 of the 376 eyes in the AREDS with at least 2 years of follow-up remained. Square root annual growth of GA for those eyes was 0.25 ± 0.40 mm/year. This value was significantly higher for eyes with small (< 5 mm²) baseline GA area (0.31 ± 0.34 mm/year; $n = 194$) compared with eyes with large (≥ 5 mm²) baseline GA area (0.16 ± 0.46 ; $n = 141$ mm/year; $P < 0.001$). Table 2 shows differences in growth rate between groups. We observed that multifocal and extrafoveal lesions grow faster than unifocal or foveal lesions. Patients with bilateral GA showed faster progression than patients with unilateral GA, although this was not significant in our analysis ($P = 0.58$). Growth rates between fellow eyes were correlated ($r = 0.58$; $P < 0.001$). Figure 4 highlights progression of GA for selected individual eyes.

Correlations between baseline lesion characteristic and square root annual growth are summarized in Table 3. Nine of 11 features were significantly correlated with GA growth rate (after Bonferroni correction). Correlations for the subset with baseline lesions size smaller than 12 mm² are analyzed in Table 4. Features included in the multivariate model were area, circularity, convex area, eccentricity, foveal

involvement, and number of lesions. The coefficient of determination of this model was 0.18. A visualization that summarizes growth over time for all eyes with GA in the AREDS set can be found in Figure 5. The red dashed line in these graphs represent a quadratic model that best fitted the data for GA area of less than 12 mm².

Discussion

A deep learning model for segmentation of GA on CFIs was developed and evaluated. We demonstrated how the automatically obtained segmentations of the model can be used to study the growth rate of GA on an independent set and reproduced several previously reported associations with growth rate. The model can also be applied to datasets for which GA measurements are not yet available, providing a fast alternative to manual delineation.

The performance of the deep learning model in terms of Dice coefficient on the BMES and RS set approached that of human experts. The model was able to identify GA even when image quality or contrast were relatively poor, as demonstrated in Figure 3. Nevertheless, some failure cases remained, which was the main reason for the lower average Dice coefficient. We suspect that more training data may solve this issue, because each of the models used only 60% of the data (approximately 245 images) for training, which may not be enough given the inherent difficulty of the problem and the variability in the data. For application to the AREDS set, this problem was circumvented partly by using an ensemble model, which indirectly made use of all training data.

Generalization ability of the model to the AREDS set was assessed on a subset of 50 CFIs. We separately analyzed the performance of the model for cases of pure GA and mixed late AMD, with co-occurring neovascular lesions (Table 5). Performance on the pure GA cases, in terms of Dice coefficient, was comparable with that on the BMES and RS set (0.71 ± 0.26 versus 0.72 ± 0.26). However, performance on mixed cases was significantly worse, as

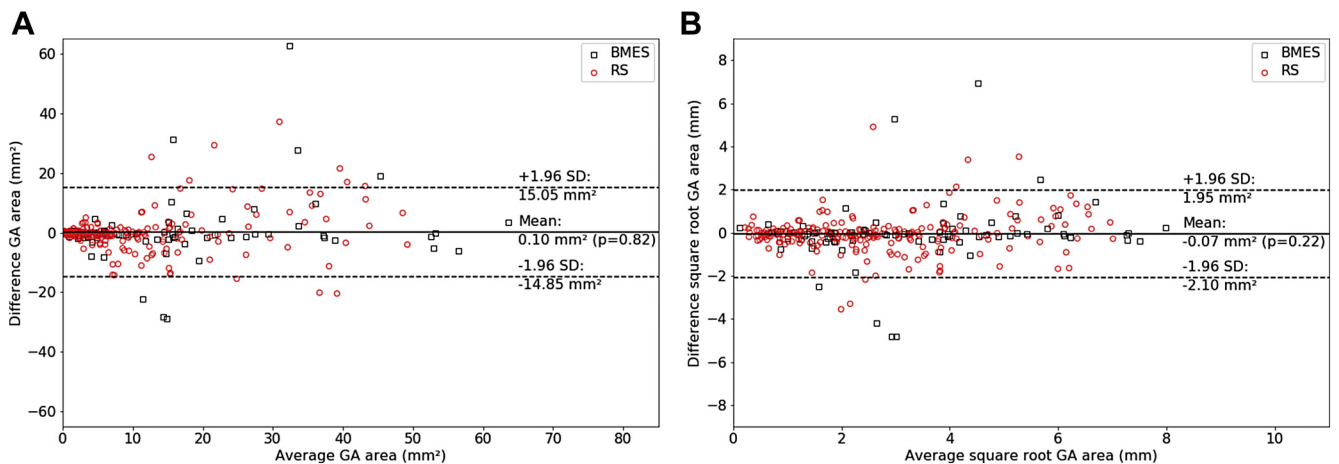


Figure 2. Bland-Altman plot showing (A) geographic atrophy (GA) area and (B) square root GA area. Differences are calculated as the area or the square root area of the consensus grading minus the automatic segmentation. SD = standard deviation.

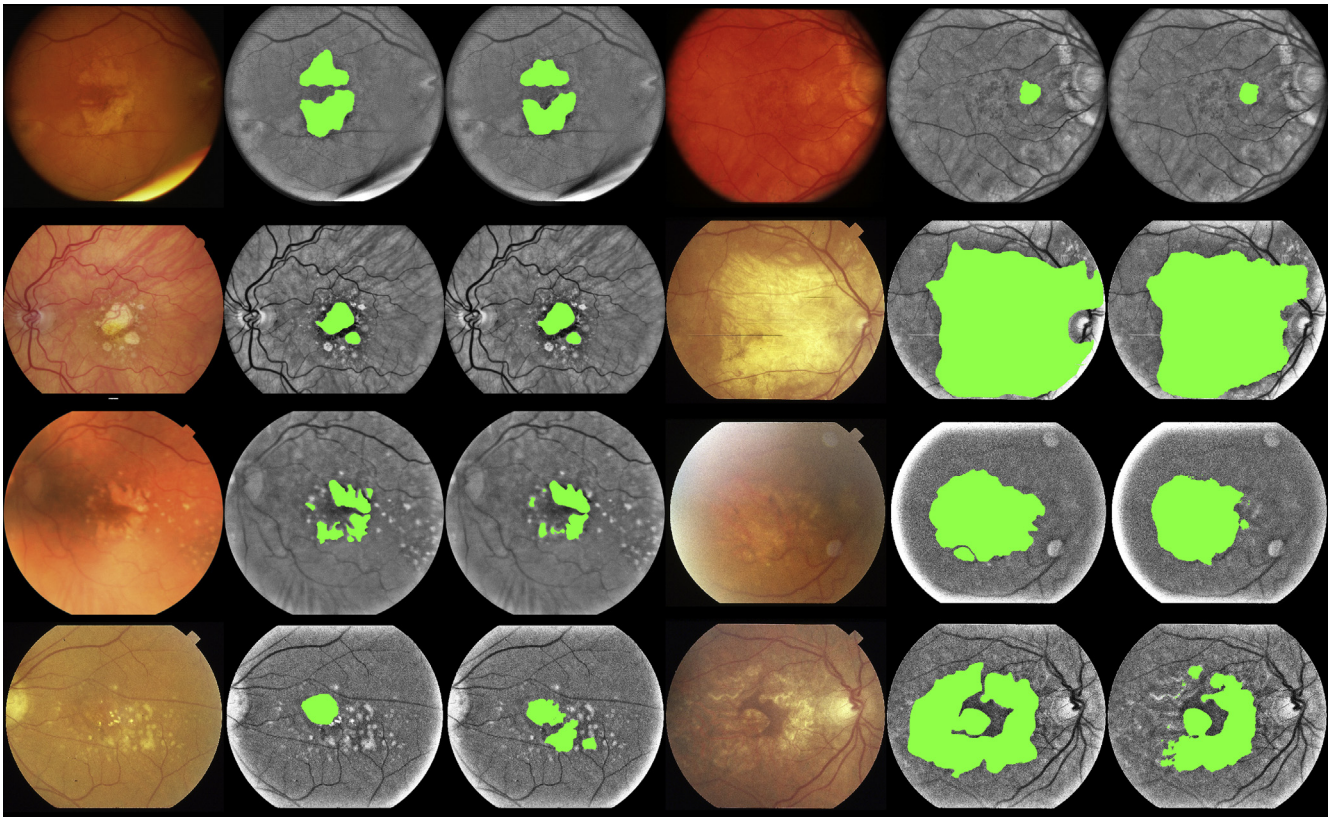


Figure 3. Images showing examples of automatic geographic atrophy (GA) segmentation. The green area corresponds to either the (left) consensus or (right) model output. The top 3 rows show accurate segmentation results for various configurations of GA differing in area, shape, and number of lesions and variable image quality and contrast. The bottom row shows examples of inaccurate model output.

was agreement between graders. Hence, these cases were not included in the analysis of GA growth. We did not observe any bias in the automatic assessment of GA area or square root GA area.

The obtained mean square root annual growth rate on the AREDS set (0.25 ± 0.40 mm/year) was slightly lower than previously reported values. For example, Domalpally et al³¹ observed 0.30 mm/year, and Keenan et al⁴⁴ observed 0.28 mm/year. A reason for this may be the dependence of

growth rate on baseline area. When we split the dataset on baseline lesion size, we observed that small lesions have larger square root growth rates (Table 2). This phenomenon was analyzed in more detail in Figure 5. A quadratic curve seemed to fit the observed GA progression very well up to an area of approximately 12 mm². For larger areas, the growth rate seemed to stabilize or even decrease, whereas the variability between patients also increased. Similar observations were made by

Table 2. Square Root Annual Growth of the Geographic Atrophy Area

	Square Root Annual Growth (mm/year)		
	All	Small (<5 mm ²)	Large (≥ 5 mm ²)
Overall	0.25 ± 0.40 (n = 335)	0.31 ± 0.34 (n = 194)	0.16 ± 0.46 (n = 141)
Unifocal	0.22 ± 0.39 (n = 251)	0.28 ± 0.33 (n = 142)	0.14 ± 0.44 (n = 109)
Multifocal	0.33 ± 0.43 (n = 84)	0.39 ± 0.35 (n = 52)	0.23 ± 0.52 (n = 32)
P value	0.028	0.039	0.339
Foveal	0.21 ± 0.41 (n = 258)	0.27 ± 0.31 (n = 120)	0.16 ± 0.47 (n = 138)
Extrafoveal	0.36 ± 0.36 (n = 77)	0.37 ± 0.37 (n = 74)	0.14 ± 0.17 (n = 3)
P value	0.006	0.066	0.934
Unilateral	0.21 ± 0.32 (n = 41)	0.23 ± 0.30 (n = 29)	0.15 ± 0.36 (n = 12)
Bilateral	0.25 ± 0.41 (n = 128)	0.29 ± 0.34 (n = 72)	0.19 ± 0.48 (n = 56)
P value	0.58	0.446	0.76

Values represent mean \pm standard deviation. P values are calculated using an unpaired *t* test.

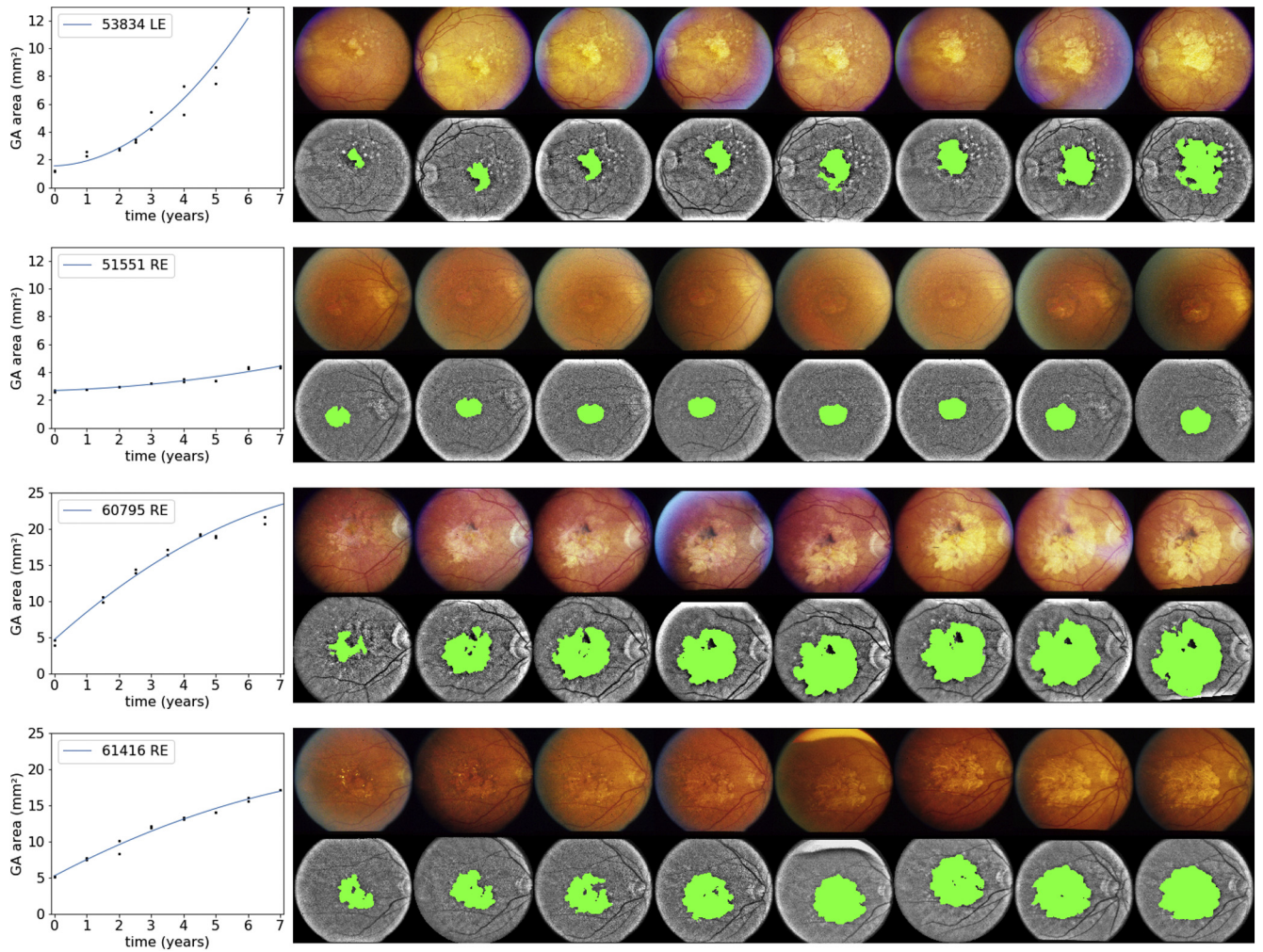


Figure 4. Graphs and images showing progression of geographic atrophy (GA) over time for 4 selected eyes. The graphs represent area measurements over time (2 points per time point for the left and right stereoscopic images). The blue line is a quadratic fit through the points. For the top 2 cases, an increment in growth rate can be observed: 53834 left eye (LE) shows a more irregular shape than 51551 right eye (RE) and progressed faster. In the bottom 2 cases, we observe that the growth decreased as the GA area increased.

Table 3. Correlations between Baseline Lesion Characteristics (Features) and Square Root Annual Growth Rate (in Millimeters per Year)

Feature	R ² Value	Slope	Intercept	R Value	P Value	Standard Error
Area	0.101	-0.015	0.351	-0.318	< 0.001	0.002
Filled area	0.100	-0.015	0.351	-0.316	< 0.001	0.002
Convex area	0.081	-0.012	0.347	-0.285	< 0.001	0.002
Convex solidity	0.078	-0.743	0.849	-0.279	< 0.001	0.140
Eccentricity	0.073	0.647	-0.167	0.271	< 0.001	0.126
Roundness	0.073	-0.697	0.723	-0.270	< 0.001	0.136
Foveal involvement	0.050	-2.950	0.370	-0.225	< 0.001	0.701
Perimeter	0.029	-0.007	0.336	-0.170	0.002	0.002
Circularity	0.025	-0.282	0.384	-0.159	0.004	0.096
No. of lesions	0.024	0.078	0.130	0.154	0.005*	0.027
Filled solidity	0.001	0.646	-0.396	0.025	0.649*	1.419

Features are sorted in decreasing order of strength of association. A *P* value of less than 0.0045 (0.05, Bonferroni corrected) was considered significant. *Not significant.

Table 4. Correlations between Baseline Lesion Characteristics (Features) and Square Root Annual Growth Rate (in Millimeters per Year) for Baseline Lesion Size of Less than 12 mm²

Feature	R ² Value	Slope	Intercept	R Value	P Value	Standard Error
Convex solidity	0.076	-0.553	0.741	-0.276	< 0.001	0.117
Circularity	0.066	-0.365	0.486	-0.258	< 0.001	0.083
Roundness	0.060	-0.491	0.628	-0.245	< 0.001	0.118
Eccentricity	0.055	0.446	0.005	0.234	< 0.001	0.113
Foveal involvement	0.032	-1.858	0.368	-0.179	0.003	0.621
No. of lesions	0.031	0.076	0.188	0.175	0.004	0.026
Perimeter	0.007	0.005	0.252	0.086	0.156*	0.003
Area	0.001	-0.004	0.312	-0.034	0.574*	0.007
Filled area	0.001	-0.004	0.312	-0.034	0.580*	0.006
Convex area	0.001	0.002	0.289	0.026	0.671*	0.005
Filled solidity	0.000	0.091	0.209	0.004	0.948*	1.380

Features are sorted in decreasing order of strength of association. A *P* value of less than 0.0045 (0.05, Bonferroni corrected) was considered significant.

*Not significant.

Keenan et al,⁴⁴ whose reported values are included in Figure 5 for comparison.

The importance of baseline area for assessing growth rate also became apparent in the regression analysis, where area, filled area, and convex area were correlated most strongly with square root annual growth rate. However, when we included only lesions with baseline area of less than 12 mm² in the regression analysis, no features related to lesion size were significantly associated with square root annual growth rate (Table 4). On an individual level, we also observed a quadratic growth of the area of GA in many cases in the AREDS set, some of them highlighted in Figure 4, in which we fitted a quadratic curve through the GA area over time. Again, the decrease in growth rate for larger lesions was visible (bottom 2 cases in Fig 4).

Of the features that are invariant to lesion size, convex solidity was associated most significantly with square root annual growth rate. Convex solidity is low for irregularly shaped lesions but also for multifocal lesions. Hence, this feature captures multiple previously reported associations. Of note, the association of square root annual growth rate with circularity was much stronger in the subset of images with baseline area of less than 12 mm². We observed that the average value for circularity of large lesions (≥ 12 mm²) was significantly lower: 0.40 ± 0.19 versus 0.51 ± 0.23 for small lesions ($P < 0.001$). For large lesions in particular, the model may have produced a segmentation with a very jagged border for lesions with indistinct borders of the atrophic area. This could have led to a relatively large perimeter and hence a lower value for circularity. In those cases, roundness will be a better representation of how well the lesion

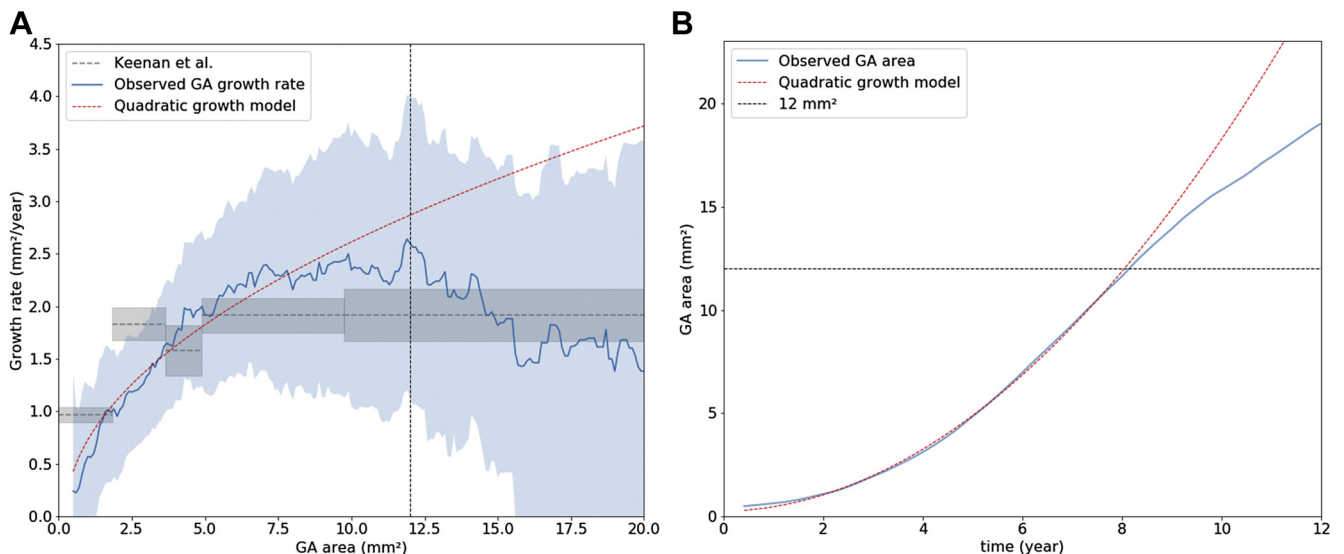


Figure 5. Graphs showing geographic atrophy (GA) growth over time. **A**, Geographic atrophy growth rate (in square millimeters per year) as a function of GA area. The blue line represents growth rates estimated from the segmentations of the deep learning model. The shaded area represents the 95% confidence interval (estimated using bootstrapping). The dashed red line represents the growth rate of a quadratic model, as visualized in **(B)**. **B**, Blue line represents the evolution of GA area over time, obtained by numerically integrating the estimated growth rates from **(A)** using a GA area of 0.5 mm² at $t = 0$. The red dashed line represents the best quadratic fit to the plot for GA area of less than 12 mm². Above this area, the observed GA area diverges from the quadratic fit.

Table 5. Dice Coefficients on the 50 Images Selected from the Age-Related Eye Disease Study Dataset for the Model and 2 Graders Compared with the Reference Grader

	All (n = 50)	Pure Geographic Atrophy (n = 30)	Mix (n = 20)
Model	0.66±0.27	0.71±0.26	0.59±0.27
Grader 2	0.73±0.24	0.78±0.22	0.64±0.25
Grader 3	0.73±0.27	0.81±0.21	0.61±0.31

approaches a circular shape, because it represents the ratio of the area of an enclosing circle and the area of the lesion and hence is less sensitive to irregular borders.⁴⁵ Indeed, contrary to circularity, roundness was significantly higher for large lesions ($\geq 12 \text{ mm}^2$): 0.76 ± 0.10 versus 0.67 ± 0.16 for small lesions ($P < 0.001$).

The multivariate model, which included 6 features related to the baseline shape and size of the GA area, was able to explain 18% of the variation in square root annual growth rate. It is unclear how much of the actual variation is explainable, because there may be factors influencing the growth rate that are not expressed in the image, such as genetics,⁴⁶ and possibly lifestyle or environmental factors. Moreover, inclusion of follow-up information or lesion patterns around the border of the GA area may improve the model further, because those have been previously demonstrated to contain additional information and are not captured by the presented multivariate model.¹² The model also does not take into account nonlinear interactions between features. Finally, errors in the automatically segmented GA may have resulted in inaccuracies, both in the estimation of the features and in the estimation of square root annual growth rate. Therefore, it is to be expected that in future work, building a model that captures more of the observed variation in growth rate will be achievable.

A limitation of our study is that the conversion from pixels to millimeters may have been inaccurate. This conversion was based on the average distance between fovea and optic disc in a subset of images. Although it is unlikely that this inaccuracy was a source for bias in reported associations with growth rate, reported values for area and growth rate may be slightly larger or smaller in reality.

Furthermore, direct application of the model in a setting where small errors are detrimental, such as clinical trials with GA area or progression rates as the end point, is currently beyond reach. The model may still fail in some cases, and additionally, color fundus imaging may not be the main method for assessment of GA in such a setting, where OCT and FAF are preferred. Nevertheless, the output of the model could still be valuable as a secondary measurement for identification of cases that may need further adjudication.^{13,26}

In the future, we will extend the model to other methods, specifically FAF and OCT. This may give more accurate measurements of the atrophic area and hence more reliable assessment of growth rate. In this study, only morphologic features of the atrophic area were considered. A next step would be to include associations between growth rate and other lesions patterns, especially those visible on FAF or

OCT. Finally, we are investigating the capabilities of deep learning models to predict directly areas where GA may develop. This will provide predictions of both the extent and location of future GA area.

In conclusion, we have presented and validated a robust segmentation model based on deep learning for GA on CFIs. The model was capable of reproducing known associations between current GA status and future growth. Moreover, we indicated novel structural biomarkers that are predictive for future growth rate, such as solidity, eccentricity, or roundness of the lesion. We demonstrated how deep learning can help in the automation of grading, allowing for analysis of larger datasets and helping to understand progression of GA.

Acknowledgments

The authors thank Johanna Colijn, Caroline Klaver, Corina Brussee, and Ada Hooghart, EyeNED Reading Center, for performing manual delineations of geographic atrophy.

References

1. Lim LS, Mitchell P, Seddon JM, et al. Age-related macular degeneration. *Lancet*. 2012;379:1728–1738.
2. Sunness JS, Rubin GS, Applegate CA, et al. Visual function abnormalities and prognosis in eyes with age-related geographic atrophy of the macula and good visual acuity. *Ophthalmology*. 1997;104:1677–1691.
3. Owen CG, Jarrar Z, Wormald R, et al. The estimated prevalence and incidence of late stage age related macular degeneration in the UK. *Br J Ophthalmol*. 2012;96:752–756.
4. Wong WL, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health*. 2014;2:e106–e116.
5. Colijn JM, Buitendijk GH, Prokofyeva E, et al. Prevalence of age-related macular degeneration in Europe: the past and the future. *Ophthalmology*. 2017;124:1753–1763.
6. Gehrs KM, Anderson DH, Johnson LV, et al. Age-related macular degeneration—emerging pathogenetic and therapeutic concepts. *Ann Med*. 2006;38:450–471.
7. Boyer DS, Schmidt-Erfurth U, van Lookeren Campagne M, et al. The pathophysiology of geographic atrophy secondary to age-related macular degeneration and the complement pathway as a therapeutic target. *Retina*. 2017;37:819–835.
8. Hanus J, Zhao F, Wang S. Current therapeutic developments in atrophic age-related macular degeneration. *Br J Ophthalmol*. 2016;100:122–127.
9. Holz FG, Strauss EC, Schmitz-Valckenberg S, et al. Geographic atrophy: clinical features and potential therapeutic approaches. *Ophthalmology*. 2014;121:1079–1091.
10. Sunness JS, Applegate CA, Bressler NM, et al. Designing clinical trials for age-related geographic atrophy of the macula: enrollment data from the geographic atrophy natural history study. *Retina*. 2007;27:204–210.
11. Lindblad AS, Lloyd PC, Clemons TE, et al. Change in area of geographic atrophy in the Age-Related Eye Disease Study: AREDS report number 26. *Arch Ophthalmol*. 2009;127:1168–1174.
12. Fleckenstein M, Mitchell P, Freund KB, et al. The progression of geographic atrophy secondary to age-related macular degeneration. *Ophthalmology*. 2018;125:369–390.

13. Schmitz-Valckenberg S, Sahel J, Danis R, et al. Natural history of geographic atrophy progression secondary to age-related macular degeneration (Geographic Atrophy Progression Study). *Ophthalmology*. 2016;123:361–368.
14. Danis RP, Lavine JA, Domalpally A. Geographic atrophy in patients with advanced dry age-related macular degeneration: current challenges and future prospects. *Clin Ophthalmol*. 2015;9:2159.
15. Sunness JS, Bressler NM, Tian Y, et al. Measuring geographic atrophy in advanced age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 1999;40:1761–1769.
16. Yehoshua Z, Rosenfeld PJ, Gregori G, et al. Progression of geographic atrophy in age-related macular degeneration imaged with spectral domain optical coherence tomography. *Ophthalmology*. 2011;118:679–686.
17. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
18. Burlina PM, Joshi N, Pekala M, et al. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol*. 2017;135:1170–1176.
19. Peng Y, Dharssi S, Chen Q, et al. DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. *Ophthalmology*. 2019;126:565–575.
20. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–2410.
21. Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57:5200–5206.
22. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318:2211–2223.
23. Keenan T, Dharssi S, Peng Y, et al. A deep learning approach for automated detection of geographic atrophy from color fundus photographs. *Ophthalmology*. 2019;126(11):1533–1540.
24. Klein R, Meuer SM, Knudtson MD, et al. The epidemiology of progression of pure geographic atrophy: the Beaver Dam Eye Study. *Am J Ophthalmol*. 2008;146:692–699.
25. Domalpally A, Danis RP, White J, et al. Circularity index as a risk factor for progression of geographic atrophy. *Ophthalmology*. 2013;120:2666–2671.
26. Sunness JS, Margalit E, Srikumaran D, et al. The long-term natural history of geographic atrophy from age-related macular degeneration: enlargement of atrophy and implications for interventional clinical trials. *Ophthalmology*. 2007;114:271–277.
27. Holz FG, Bindewald-Wittich A, Fleckenstein M, et al. Progression of geographic atrophy and impact of fundus autofluorescence patterns in age-related macular degeneration. *Am J Ophthalmol*. 2007;143:463–472.
28. Feuer WJ, Yehoshua Z, Gregori G, et al. Square root transformation of geographic atrophy area measurements to eliminate dependence of growth rates on baseline lesion measurements: a reanalysis of Age-Related Eye Disease Study report no. 26. *JAMA Ophthalmol*. 2013;131:110–111.
29. Shen L, Liu F, Nardini HG, et al. Natural history of geographic atrophy in untreated eyes with nonexudative age-related macular degeneration: a systematic review and meta-analysis. *Ophthalmol Retina*. 2018;2:914–921.
30. Fleckenstein M, Schmitz-Valckenberg S, Martens C, et al. Fundus autofluorescence and spectral-domain optical coherence tomography characteristics in a rapidly progressing form of geographic atrophy. *Invest Ophthalmol Vis Sci*. 2011;52:3761–3766.
31. Domalpally A, Danis R, Agrón E, et al. Evaluation of geographic atrophy from color photographs and fundus autofluorescence images: Age-Related Eye Disease Study 2 report number 11. *Ophthalmology*. 2016;123:2401–2407.
32. Chiu SJ, Izatt JA, O'Connell RV, et al. Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images. *Invest Ophthalmol Vis Sci*. 2012;53:53–61.
33. Hu Z, Medioni GG, Hernandez M, et al. Segmentation of the geographic atrophy in spectral-domain optical coherence tomography and fundus autofluorescence images. *Invest Ophthalmol Vis Sci*. 2013;54:8375–8383.
34. Niu S, de Sisternes L, Chen Q, et al. Automated geographic atrophy segmentation for SD-OCT images using region-based CV model via local similarity factor. *Biomed Opt Express*. 2016;7:581–600.
35. Hu Z, Medioni GG, Hernandez M, et al. Automated segmentation of geographic atrophy in fundus autofluorescence images using supervised pixel classification. *J Med Imaging*. 2015;2:014501.
36. Feeny AK, Tadarati M, Freund DE, et al. Automated segmentation of geographic atrophy of the retinal epithelium via random forests in AREDS color fundus images. *Comput Biol Med*. 2015;65:124–136.
37. Mitchell P, Smith W, Attebo K, et al. Prevalence of age-related maculopathy in Australia. *Ophthalmology*. 1995;102:1450–1460.
38. Ikram MA, Brusselle GG, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol*. 2017;32:807–850.
39. Grunwald JE, Pistilli M, Ying G, et al. Growth of geographic atrophy in the comparison of age-related macular degeneration treatments trials. *Ophthalmology*. 2015;122:809–816.
40. van Zeeland H, Meakin J, Liefers B, et al. EyeNED workstation: development of a multi-modal vendor-independent application for annotation, spatial alignment and analysis of retinal images. *Invest Ophthalmol Vis Sci*. 2019;60:6118.
41. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24:1342–1350.
42. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Med Image Comput Assist Interv*. 2015:234–241.
43. Graham B. Kaggle diabetic retinopathy detection competition report. 2015. <https://kaggle-forum-message-attachments.storage.googleapis.com/88655/2795/competitionreport.pdf>. Accessed March 3, 2020.
44. Keenan TD, Agron E, Domalpally A, et al. Progression of geographic atrophy in age-related macular degeneration: AREDS2 report number 16. *Ophthalmology*. 2018;125:1913–1928.
45. Zdilla MJ, Hatfield SA, McLean KA, et al. Circularity, solidity, axes of a best fit ellipse, aspect ratio, and roundness of the foramen ovale: a morphometric analysis with neurosurgical considerations. *J Craniofac Surg*. 2016;27:222–228.
46. Grassmann F, Harsch S, Brandl C, et al. Assessment of novel genome-wide significant gene loci and lesion growth in geographic atrophy secondary to age-related macular degeneration. *JAMA Ophthalmol*. 2019;137:867–876.

Footnotes and Financial Disclosures

Originally received: August 15, 2019.

Final revision: January 17, 2020.

Accepted: February 7, 2020.

Available online: ■■■■.

Manuscript no. D-19-00180.

¹ Diagnostic Image Analysis Group, Department of Radiology, Radboud University Medical Center, Nijmegen, The Netherlands.

² Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands.

³ Department of Ophthalmology, Erasmus University Medical Center, Rotterdam, The Netherlands.

⁴ Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands.

⁵ Centre for Vision Research, Department of Ophthalmology, The Westmead Institute for Medical Research, The University of Sydney, Sydney, Australia.

⁶ Health Services and Systems Research, Duke-NUS Medical School, National University of Singapore, Singapore, Republic of Singapore.

⁷ Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands.

⁸ Institute for Molecular and Clinical Ophthalmology, Basel, Switzerland.

Financial Disclosure(s):

The author(s) have made the following disclosure(s): B.vG.: Royalties and Equity owner – Thirona

C.C.W.K.: Consultant – Bayer, Thea Pharma

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, The Netherlands; the Netherlands Organization for the Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly; the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (grant no.: DG XII); and the Municipality of Rotterdam, Rotterdam, The Netherlands. The ophthalmic research within the Rotterdam Study was supported by Oogfonds; Landelijke Stichting voor Blinden en Slechtzienden; Novartis Foundation; and MaculaFonds that contributed through

UitZicht (grant nos.: 2015-36 and 2018-34) and the Royal Dutch Academy of Sciences (Koninklijke Nederlandse Akademie van Wetenschappen) through the Ammodo Award (C.C.W.K.). Other funding was obtained from the automation in medical imaging (AMI) project, a collaborative project of the Fraunhofer-Gesellschaft and the Radboud University and University Medical Center; the National Health and Medical Research Council, Australia (grant nos.: 211069, 457349, and 512423 [J.J.W., N.J.]); and the European Union (C.C.W.K.). The sponsor or funding organization had no role in the design or conduct of this research.

HUMAN SUBJECTS: Human subjects were included in this study. The human ethics committees at the University of Sydney, the Sydney West Area Health Service, the Erasmus Medical Center, and the Netherlands Ministry of Health, Welfare and Sport approved the study. All research adhered to the tenets of the Declaration of Helsinki.

No animal subjects were included in this study.

Author Contributions:

Conception and design: Liefers, Klaver, Sánchez

Analysis and interpretation: Liefers, Colijn, González-Gonzalo, Mitchell, Hoyng, van Ginneken, Klaver, Sánchez

Data collection: Liefers, Colijn, Verzijden, Wang, Joachim, Klaver, Sánchez

Obtained funding: van Ginneken, Klaver

Overall responsibility: Liefers, Colijn, Verzijden, Klaver, Sánchez

Abbreviations and Acronyms:

AMD = age-related macular degeneration; **AREDS** = Age-Related Eye Disease Study; **BMES** = Blue Mountains Eye Study; **CFI** = color fundus image; **FAF** = fundus autofluorescence; **GA** = geographic atrophy; **RS** = Rotterdam Study.

Correspondence:

Bart Liefers, MSc, Diagnostic Image Analysis Group, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, The Netherlands. E-mail: Bart.Liefers@radboudumc.nl.