AMERICAN ACADEMY OF OPHTHALMOLOGY®

A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography

Felix Grassmann, PhD,[1,*] *Judith Mengelkamp, PhD,*[1,2,*] *Caroline Brandl, PhD,*[1,3,4] *Sebastian Harsch,*[1] *Martina E. Zimmermann, PhD,*[3] *Birgit Linkohr, PhD,*[5] *Annette Peters, PhD,*[5] *Iris M. Heid, PhD,*[3] *Christoph Palm, PhD,*[2,6] *Bernhard H.F. Weber, PhD*[1]

Purpose: Age-related macular degeneration (AMD) is a common threat to vision. While classification of disease stages is critical to understanding disease risk and progression, several systems based on color fundus photographs are known. Most of these require in-depth and time-consuming analysis of fundus images. Herein, we present an automated computer-based classification algorithm.

Design: Algorithm development for AMD classification based on a large collection of color fundus images. Validation is performed on a cross-sectional, population-based study.

Participants: We included 120 656 manually graded color fundus images from 3654 Age-Related Eye Disease Study (AREDS) participants. AREDS participants were >55 years of age, and non-AMD sight-threatening diseases were excluded at recruitment. In addition, performance of our algorithm was evaluated in 5555 fundus images from the population-based Kooperative Gesundheitsforschung in der Region Augsburg (KORA; Cooperative Health Research in the Region of Augsburg) study.

Methods: We defined 13 classes (9 AREDS steps, 3 late AMD stages, and 1 for ungradable images) and trained several convolution deep learning architectures. An ensemble of network architectures improved prediction accuracy. An independent dataset was used to evaluate the performance of our algorithm in a population-based study.

Main Outcome Measures: κ Statistics and accuracy to evaluate the concordance between predicted and expert human grader classification.

Results: A network ensemble of 6 different neural net architectures predicted the 13 classes in the AREDS test set with a quadratic weighted κ of 92% (95% confidence interval, 89%−92%) and an overall accuracy of 63.3%. In the independent KORA dataset, images wrongly classified as AMD were mainly the result of a macular reflex observed in young individuals. By restricting the KORA analysis to individuals >55 years of age and prior exclusion of other retinopathies, the weighted and unweighted κ increased to 50% and 63%, respectively. Importantly, the algorithm detected 84.2% of all fundus images with definite signs of early or late AMD. Overall, 94.3% of healthy fundus images were classified correctly.

Conclusions: Our deep learning algoritm revealed a weighted κ outperforming human graders in the AREDS study and is suitable to classify AMD fundus images in other datasets using individuals >55 years of age. *Ophthalmology 2018;125:1410-1420 © 2018 by the American Academy of Ophthalmology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).*

Supplemental material available at www.aaojournal.org.

Age-related macular degeneration (AMD) is the leading cause of severe vision impairment among people 50 years of age and older in Western countries.[1] It is a multifactorial trait influenced by both genetic and environmental effects. The underlying mechanisms of AMD pathologic features remain elusive.[2] Age, smoking, and—to a lesser extent—diet and sunlight exposure are among the most commonly reported individual risk factors for disease onset. A genetic contribution to AMD is well established by familial aggregation analyses, twin studies, as well as genome-wide association studies.[3–5]

Age-related macular degeneration typically progresses in a sequence of different stages from an early to a late form, where atrophic and neovascular subtypes are distinguished.[6] The early stages are characterized by the appearance of yellowish deposits called *drusen*. Although few, small,

https://doi.org/10.1016/j.ophtha.2018.02.037
ISSN 0161-6420/18

distinct drusen are also typical age-related changes in the outer retina, soft confluent drusen as well as a large number of drusen are risk factors for the progression to late stages of AMD.[7] In addition, pigmentary changes in the retinal pigment epithelium layer can occur and also are regarded as independent risk factors for late-stage AMD.[7] The neovascular or wet form of AMD is described by the growth of new, leaky blood vessels into the retina causing widespread photoreceptor loss and ultimately rapid decline in visual acuity. Geographic atrophy (GA) is characterized by a gradual degeneration and disappearance of retinal pigment epithelium, photoreceptor cells, and the choriocapillaris layer in the central retina.[8] Both late-stage forms can occur in the same eye or in different eyes at the same time or in succession.

To classify patients according to their disease status, several classification systems have been developed. Most of those systems were derived from the Wisconsin Age-Related Maculopathy Grading System, which is based on the presence and extent of AMD features like drusen, pigmentary changes, GA, and neovascularization.[9] Among the most recently established used grading systems is the 9-step Age-Related Eye Disease Study (AREDS) severity scale from AREDS report number 17,[10] the 5-step AREDS simplified severity scale from AREDS report number 18,[7] the Three-Continent AMD Consortium severity scale,[11] the Rotterdam system,[12] as well as the clinical classification proposed in 2013 by Ferris et al.[13] Any classification system requires trained graders to measure and quantify the funduscopic changes to create a grading for the eye or the individual. This is time consuming and also error prone. For many AMD classification systems, the intergrader performance expressed as a quadratic weighted κ is between 22% and 86%.[13–18]

So far, automated classification systems have relied on the use of hand-designed feature-based approaches by extracting features from a preprocessed image and then using those features to classify the images using various methods, for example, by automated drusen area and number quantification.[19,20] Recent advances in the field of image recognition and classification have seen a shift toward deep learning approaches, leveraging new algorithms as well as increased computational capacities. The most successful deep learning approaches are based on convolution filters that allow automated feature extraction and learning.[21] Convolution deep learning uses convolution filters to scan images with small perceptive fields. This approach reduces the computational load because only the weights of the small filter are trained as opposed to a fully connected layer. This enables the networks to contain more layers and thus to be deeper and more comprehensive in the classification task. In addition, the perceptive fields are able to evaluate and perceive higher-level structures (such as textures, structure, color, and lightning gradients), and therefore are able to generalize many observed features. This has led to improved accuracies for various image classification and detection tasks such as classification of real-life images (e.g., cars, houses, and animals), reading and processing of license plates, as well as classifying clinical images according to disease status.[22]

In this study, we developed an automated classification strategy based on training deep learning models to predict the AMD stage in color fundus images from the AREDS study, a prospective study of the clinical course of AMD. For classification we applied a scheme consisting of 13 classes including 9 classes based on the ARED 9-step severity scale, 3 late-stage classes, and 1 class for ungradable images. We also applied our algorithm to an independent study to assess the algorithm's performance in a population-based study for future epidemiologic studies and, potentially, for harmonizing different existing studies.

Methods

Overview

The proposed deep learning classification strategy consists of 4 steps (Fig 1). In the first step, the color fundus images are preprocessed. They are used in the second step to train multiple convolution neural nets (CNNs) independently. In general, the aim of training a CNN is to optimize an evaluation metric by comparing the CNN output with the true class iteratively and then adjusting the weights to minimize the loss between CNN output and actual label. In the third step, a random forest algorithm is trained to build a model ensemble based on the results of the single CNNs. In the last step, the final model is applied to predict AREDS testing data and the Kooperative Gesundheitsforschung in der Region Augsburg (KORA; Cooperative Health Research in the Region of Augsburg) study dataset.[23] The individual steps are explained in more detail below.

Thirteen Classes of Age-Related Macular Degeneration Based on the Age-Related Eye Disease Study 9-Step Severity Scale

We adopted a system with 13 classes based on the AREDS 9-step severity scale. The AREDS 9-step grading aims at quantifying AMD-related features on fundus images.[10] Age-Related Eye Disease Study grade 1 indicates fundus images with little or no AMD-related changes, whereas fundus images with AREDS grades 2 through 9 present changes associated with early or intermediate AMD.[10] In addition, AREDS grades 10 through 12 represent late-stage AMD, namely GA,[10] neovascular AMD,[11] and images with both late-stage forms.[12] Furthermore, we added a new category to indicate fundus images that are not suitable to grade AMD severity, ungradable.

Ethics Statement

The AREDS of the National Eye Institute, National Institutes of Health, is a long-term multicenter, prospective study. The study protocol was approved by an independent institutional review board at each clinical center involved in the AREDS. Written informed consent was obtained from all participants before enrollment. The corresponding author (B.H.F.W) was granted access to the AREDS data by the AREDS data access committee through the database of genotypes and phenotypes, and our analyses are in accordance with the approved research use statement (data access request no. 48440). The study was adherent to the tenets of the Declaration of Helsinki and was HIPAA compliant.[24]

The KORA study is a research platform to survey the development and course of chronic diseases. The ethics committee of the Bavarian Medical Association (Bayerische Landesärztekammer) and the Bavarian commissioner for data protection and privacy

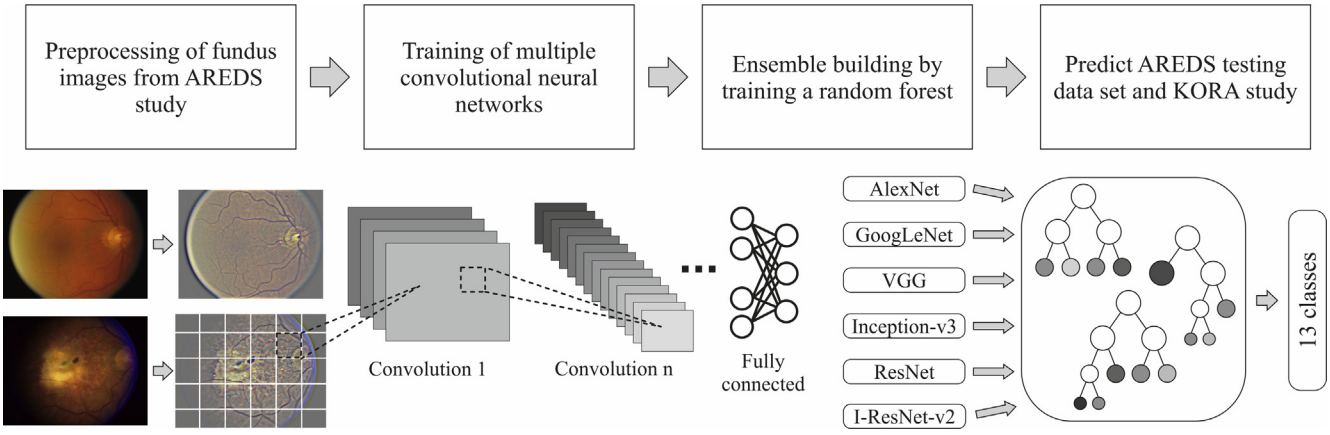

Figure 1. Processing outline of the proposed classification scheme. First, fundus images are normalized to have equal illumination and color balance. Next, 6 different convolution neural net models are trained on 86770 normalized images, namely, AlexNet,[33] GoogLeNet,[34] VGG with 11 convolution layers,[35] Inception-V3,[36] ResNet with 101 layers,[37] and Inception-ResNet-V2 (I-ResNet-v2).[37] The class prediction from each individual network then was used to train a random forest classifier to improve classification accuracy. Finally, the random forest ensemble model was used to predict the Age-Related Eye Disease Study (AREDS) 9-step plus 3 scale and to identify ungradable images from 12 019 fundus images from the unrelated AREDS testing dataset, as well as of 5555 fundus images from the Cooperative Health Research in the Region of Augsburg (KORA) study.

(Bayerischer Datenschutzbeauftragter) approved the study, which complied with the tenets of the 1964 Declaration of Helsinki and its later amendments. Informed written consent was obtained from all individual participants included in the study.

Data Acquisition

In the AREDS study, color stereoscopic fundus images from mydriatic eyes were obtained with a Zeiss (Carl Zeiss AG, Oberkochen, Germany) FF series 30° camera in field 2 (centered above the macula) as previously described.[25] In the KORA study, the full macular region and optic disc of nonmydriatic eyes were acquired with a 45° degree Topcon (Topcon Corporation, Tokyo, Japan) TRC-NW5S fundus camera.[23] We extracted 120 656 fundus images and their respective previously performed manual gradings (AREDS 9 steps plus 3 late-stage steps) from the AREDS from the database of Genotypes and Phenotypes (dbGaP) (accession: phs000001.v3.p1; Table 1). The AREDS data has been described previously in more detail.[26] Briefly, AREDS is a long-term, multicenter, prospective study of AMD and age-related cataract to study risk factors and to understand the clinical progression of both diseases. Eligible participants were between 55 and 80 years of age at recruitment and free of sight-threatening conditions other than cataract or AMD. Each patient was assigned either to the training set (70% of the patients), validation set (20% of the patients), or testing set (10% of the patients), and all fundus images from the patient were included in the respective dataset. Thus, the training, validation, and testing datasets consisted of 86 770, 21 867, and 12 019 fundus images, respectively. Approximately 5% of the images were estimated to be ungradable

Table 1. Number of Fundus Images in the Training, Validation, and Test Datasets from the Age-Related Eye Disease Study and from the Cooperative Health Research in the Region of Augsburg Study

Age-Related Eye Disease Study Scale	No. of Fundus Images (% Total)					
	Age-Related Eye Disease Study				*Cooperative Health Research in the Region of Augsburg Study*	
	All	*Training*	*Validation*	*Testing*	*All Ages*	*Age >55 Years*
UG	4158 (3.45)	3117 (3.59)	712 (3.26)	329 (2.74)	322 (5.8)	155 (5.43)
1	41770 (34.62)	30278 (34.89)	7416 (33.91)	4076 (33.91)	4829 (86.93)	2409 (84.41)
2	12133 (10.06)	8585 (9.89)	2234 (10.22)	1314 (10.93)	226 (4.07)	150 (5.26)
3	5070 (4.2)	3570 (4.11)	1002 (4.58)	498 (4.14)	75 (1.35)	61 (2.14)
4	8985 (7.45)	6437 (7.42)	1471 (6.73)	1077 (8.96)	60 (1.08)	45 (1.58)
5	6012 (4.98)	4392 (5.06)	1008 (4.61)	612 (5.09)	24 (0.43)	18 (0.63)
6	7953 (6.59)	5755 (6.63)	1426 (6.52)	772 (6.42)	10 (0.18)	7 (0.25)
7	6916 (5.73)	4991 (5.75)	1374 (6.28)	551 (4.58)	1 (0.02)	1 (0.04)
8	6634 (5.5)	4734 (5.46)	1295 (5.92)	605 (5.03)	0 (0)	0 (0)
9	2539 (2.1)	1855 (2.14)	453 (2.07)	231 (1.92)	1 (0.02)	1 (0.04)
10	4128 (3.42)	2952 (3.4)	831 (3.8)	345 (2.87)	2 (0.04)	2 (0.07)
11	13260 (10.99)	9357 (10.78)	2445 (11.18)	1458 (12.13)	5 (0.09)	5 (0.18)
12	1098 (0.91)	747 (0.86)	200 (0.91)	151 (1.26)	0 (0)	0 (0)
Total	120656 (100.0)	86770 (100.0)	21867 (100.0)	12019 (100.0)	5555 (100.0)	1967 (100.0)

UG = ungradable.

because of technical issues such as overexposure, blurring resulting from an out-of-focus image, or dirt on the lenses. Hence, a trained ophthalmologist identified images that could not be graded, and we flagged those images in the datasets. Thus, we trained a classification system consisting of 13 classes in total (12 AREDS steps and 1 additional class comprising ungradable pictures). In addition, 5555 fundus images were provided by the cross-sectional KORA study[23] and were used as an independent testing dataset to assess classification accuracy.

Preprocessing

The aim of the preprocessing step was to reduce the influence of lighting variations such as brightness and incident angle of the fundus camera between the images. We normalized the color balance as well as local illumination of each fundus image by using a Gaussian filtering to subtract the local average color (Fig 1).[27] The short edge of each image was reduced to 512 pixels while keeping the aspect ratio constant. The training, validation, and test datasets were encoded as binary .rec files for fast access.[28]

Loss Functions and Other Metrics

As the main loss metric to maximize during training, we applied a custom weighted κ metric (κ_c; Fig S1, available at www.aaojournal.org). The κ metric is especially suited for classification task with unbalanced class distributions and, in addition, reflects the ordinal scaled nature of the AREDS score (higher scores relate to a more severe phenotype). We derived κ_c from Cohen's quadratic weighted κ metric, κ_w,[29] by adapting the weights to impose a larger penalty for a misclassification between gradable and ungradable images as well as between late AMD stages[10–12] and among no AMD, early AMD, and intermediate AMD classes.[1–9] The weights for disagreements between the AREDS classes 1 through 9 are identical to the quadratic weighted κ, κ_c (Fig S1, available at www.aaojournal.org), because small deviations between these phenotypically similar classes are tolerable with a quadratic decay. The same applies for misclassifications among late-stage AMD classes 10 through 12. However, disagreements between ungradable fundus images and any other class receive the maximum possible penalty. Thus, the weights are set to 0 in the first row and the column of the weighting matrix, except when ungradable images are identified correctly as such. Similarly, misclassifications among any of the classes, including the 9-step AREDS scale (classes 1–9) and the classes covering late-stage AMD (classes 10–12), receive maximum penalties. Training with other loss functions, such as cross-entropy, log-loss, as well as accuracy, top 2 accuracy, and an unweighted κ, revealed comparable model accuracies over all classes. However, those models performed more poorly in separating early- from late-stage AMD as well as in identifying ungradable images.

We also reported linear weighted and unweighted κ measures, overall accuracy, as well as top 2 accuracy, which indicates that the true class of a fundus image is among the 2 classes that are predicted by the CNN with the highest confidence. In addition, we reported the balanced accuracy for all models and model ensembles, representing the accuracy that would be expected in case all classes are balanced, that is, that have the same number of observations. Furthermore, we calculated the F1 metric,[30] which is the harmonic mean of precision and recall, to evaluate the CNN performance in different AREDS classes. Finally, for each AREDS class, we reported different measures such as sensitivity, specificity, positive and negative predictive values using the confusion Matrix function provided by the caret package[31] in R software.[32]

Training

We trained 6 state-of-the-art convolution neural networks implemented in the MXNet Deep Learning Framework[28] separately on 86 770 images from the training data. The applied CNNs were AlexNet (SuperVision group, consisting of Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever at the University of Toronto).[33] GoogLeNet (Google),[34] Visual Geometry Group (VGG), Department of Engineering Science, University of Oxford with 11 convolution layers,[35] Inception-V3,[36] Deep Residual Learning for Image Recognition (ResNet, Microsoft Research) with 101 layers,[37] and Inception-ResNet-V2.[37] We trained each model for 10 epochs or iterations using a Cohen's quadratic weighted κ metric as the main loss function, because directly training with the custom weighted κ metric resulted in delayed convergence in the training accuracy and unstable validation accuracies. Using higher epochs for initialization resulted in delayed or poorer convergence when training with the custom weighted κ metric. The loss function was minimized using a stochastic gradient descent optimizer with momentum (Table S1, available at www.aaojournal.org). Using the initialized models, we trained each model for at least another 30 epochs with the custom weighted κ metric (Fig S1, available at www.aaojournal.org). Training was stopped as soon as the validation κ values no longer increased or started to decrease. We then selected the most suitable epoch to use for prediction based on a high validation κ and a lower training κ, because overfit models performed poorly in model ensemble computation. The training parameters for each neural net architecture are listed in Table S1 (available at www.aaojournal.org).

We performed the computations an a single system equipped with two Nvidia GTX 1080 Ti graphics processing units (GPUs) having each 11.2 GB of random access memory (RAM) memory as well as a single Nvidia Titan X Pascal with 12.1GB of RAM memory. This allowed multi-GPU training in which the images in each training batch are partitioned evenly among the GPUs and are processed simultaneously. The results are averaged across the GPUs. The available RAM, the size of the images, as well as the model architecture, that is, the number of convolutional layers suitably filtered, pose a limit on the maximum possible batch size that can be used for training. In addition, we used backward propagation mirroring as implemented in the MXNet framework, which reduces the required RAM by approximately 50% for most architectures, at the cost of increased computation time (by approximately 15%).

Data Augmentation

To increase the diversity of the dataset, and thus to reduce the risk of overfitting the CNNs, we applied data augmentation methods. Classification should be independent of image transformations as long as the transformations do not alter the diagnostically decisive characteristics, that is, the macula and surrounding region are not affected. During training, each image was scaled to 512×512 pixels, and we applied several augmentations to each image: (1) images were cropped randomly on both height and width for approximately 10% of their size, (2) images also were randomly mirrored or flipped, (3) images were rotated randomly between 1° and 180°, and (4) the aspect ratio was adjusted randomly between 0% and 15%.

Random Forest for Ensemble

Initially, we evaluated the performance of various methods to compute the network ensemble in the validation data set. In particular, we applied Support Vector Machines (implemented in R

software[32]), multilayer perceptrons (implemented in MXNet), Gradient Boosting Classifiers (implemented in Scikit-learn in Python[38]), majority voting, as well as random forest classification (implemented in R software) and found that a random forest classifier showed the best classification metrics and was less prone to overfitting. We therefore constructed the model ensemble by training a random forest classifier. Generally in a random forest classification task, a number of decision trees are computed and the final classification corresponds to the majority vote among the individual trees.[39] Using the predicted class probabilities of the 6 CNN models, we trained the random forest using 1000 trees on the training dataset. To optimize the control parameters of the random forest algorithm (such as the maximum number of trees and convergence parameters), we used the validation dataset consisting of 21 867 color fundus images and reported the performance of the best model ensemble on an independent test dataset consisting of 12 019 images. In the KORA study data of 5555 images, we applied the resulting ensemble to evaluate the performance in a dataset that is completely independent without any overlap of images, different image acquisition conditions, and a different study design (population-based, cross-sectional study including many controls, some early AMD and few late AMD cases, as well as other sight-threatening conditions) compared with the CNN-generating dataset (clinic-based, prospective cohort without confounding blinding conditions). Based on the confusion matrix comparing the physician-based classification with the CNN-based classification, we derived different κ statistics and proportions of agreement.

Assessing Important Features in Fundus Images Using Convolution Neural Nets

To evaluate how the CNNs assessed the fundus images, we randomly masked 10 000 100×100-pixel fields in 1 fundus image from AREDS classes 3 through 12 with the mean pixel value of the respective image.[40] This approach effectively masks important AMD-related features in the fundus image and allows assessment of the importance of those features. In addition, images from AREDS classes 1 (healthy) and 2 (few drusen) were masked by 10 000 40×40-pixel fields with a pixel value of 0 (black). By adding small black areas to otherwise healthy fundus images, we could assess whether the addition of unexpected features in certain areas of the fundus image influences the classification confidence. Next, we allowed GoogLeNet predict the AREDS class of the masked images and calculated the confidence of the CNN for the true AREDS class. A significant drop in confidence indicates that this area was indeed an important feature for the respective true AREDS class. Next, for each pixel in the fundus image, we calculated the average confidence for the true class by averaging the confidence observed for all those images that masked the respective pixel. Finally, we overlaid the original fundus image with areas that show a significant drop in confidence for the true AREDS class.

Data Availability Statement

Phenotypes and fundus images from the AREDS are available at dbGAP (http://dbgap.ncbi.nlm.nih.gov/; accession, phs000001.v3.p1). The KORA data are available on an individual project agreement with KORA at https://epi.helmholtz-muenchen.de/. The individual models (architecture and the trained weights) as well as the model ensemble are publicly available at https://github.com/RegensburgMedicalImageComputing/ARIANNA. The authors possess no intellectual property on the methods or the models themselves.

Results

Training the Individual Convolution Neural Net Models and Model Ensemble

In total, we used 86 770 fundus images from the AREDS to train 6 different CNNs and used 21 867 fundus images to estimate the model's validation accuracy after each iteration (Fig S2, available at www.aaojournal.org). Initially, we trained each network for 10 iterations using a quadratic weighted κ loss function, because training directly with the custom metric resulted in a less stable training process and in a higher variance in validation accuracy. After those iterations, we trained each model using the custom κ metric, which heavily penalizes misclassification errors between ungradable and gradable images, as well as between fundus images from healthy patients and early or intermediate AMD patients and fundus images showing signs of late-stage AMD (Figs S1 and S2, available at www.aaojournal.org). The different network architectures had overall accuracies ranging between 57.7% and 61.7% and with quadratic weighted κ values ranging from 89.7% to 91.1% in an independent test set of 12 019 fundus images (Table 2). By combining the different network architectures in a model ensemble using random forests, we were able to improve the overall accuracy to 63.3% and weighted κ to 92.1% in the AREDS test dataset (Table 2; Table S2, available at www.aaojournal.org). In addition, the model ensemble showed an improved Cohen's κ value of 55.5% and a Cohen's linear weighted κ of 83.3% as well as a top 2 accuracy of 85.7%. The gain in classification accuracies can be attributed to increased precision as well as recall over all AREDS classes, because the F1 metric of the model ensemble outperforms the individual

Table 2. Comparison of Different Metrics for Different Model Architectures in the Age-Related Eye Disease Study Test Dataset

Name of Convolution Neural Net	κ_0	κ_l	κ_q	κ_c	ac	ac_2	ac_b
AlexNet	48.73	79.27	89.7	80.43	58.3	80.8	70.3
GoogLeNet	48.06	78.93	89.45	80.93	57.7	81.1	71.1
VGG	50.53	80.66	90.5	81.52	59.8	83.3	71.3
Inception-v3	52.19	81.72	91.11	82.84	60.7	84.2	72.9
ResNet	53.06	81.66	90.79	82.39	61.7	84.5	74.2
Inception-ResNet-v2	52.44	81.04	90.98	83.06	61.1	83.7	71.6
Ensemble: random forest	55.47	83.32	92.14	84.03	63.3	85.7	74.7

ac = accuracy; ac_b = average balanced accuracy over all classes; ac_2 = top 2 accuracy; κ_c = custom weighted Cohen's κ; κ_l = linear weighted Cohen's κ;κ_q = quadratic weighted Cohen's κ; κ_0 = unweighted Cohen's κ.
Data are percent.

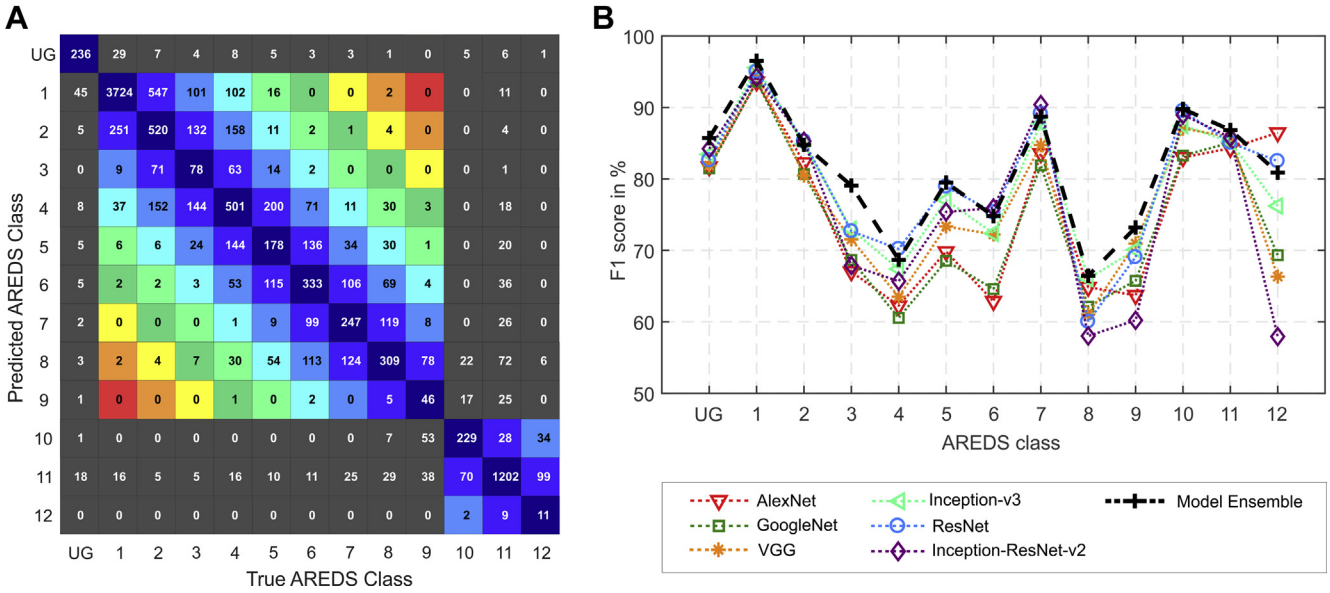

Figure 2. Confusion matrix and F1 scores over all classes in the Age-Related Eye Disease Study (AREDS) test set. **A**, Confusion matrix of the test set depicting the true versus the predicted class of 12 019 fundus images. The colors represent the custom κ weighting scheme used in the training process (Fig S1, available at www.aaojournal.org). **B**, The F1 score of the 6 neural net architectures as well as the F1 score of the model ensemble were calculated as the harmonic mean of precision and recall of each AREDS class.

models in almost all classes (Fig 2). Importantly, the different model architectures showed a high rate of agreement with each other with an average quadratic weighted κ of 91% between the models. Taken together, the model ensemble showed the best classification performance across all investigated metrics and thus was used for further steps.

Predicting Age-Related Eye Disease Study Scale in the Cooperative Health Research in the Region of Augsburg Dataset

Many CNN models fail to perform accurately in unrelated datasets because of technical differences (e.g., camera setup, illumination, further processing) and clinical differences (e.g., inclusion or exclusion criteria, grading of non-AMD–specific features).[22] To evaluate our algorithm, we used 5555 fundus images from the KORA study[23] and predicted their AREDS scale using the model ensemble that had excellent classification accuracies in the AREDS test dataset. The observed accuracies were higher than the accuracies observed in the AREDS data, because most of the fundus images were of individuals without any changes related to AMD (Table 3; Table S2, available at www.aaojournal.org). However, we observed reduced weighted and unweighted κ values in the KORA dataset, particularly because of 313 misclassified fundus images that were classified to show features of neovascular AMD (AREDS class 11), but that were actually images from healthy individuals (Fig 3). Most of those fundus images were from young individuals (40 years of age and younger) who demonstrated dominant macular reflexes, which were absent in fundus images in the AREDS because of the increased inclusion age of 55 years or older. Therefore, by restricting the analysis to fundus images from individuals 55 years of age and older—as corresponding to the AREDS age range—we observed significantly increased accuracies as well as increased unweighted and weighted κ values (Table 3; Table S2, available at www.aaojournal.org). Importantly, 76 of 1677 fundus images from individuals 55 years of age and older showed definite features of intermediate AMD (AREDS classes 4–9), and our classification scheme would correctly identify 63 of those and place them in a category of more than 4 (82.2% sensitivity and 97.1% specificity). Furthermore, our algorithm successfully identified all fundus images with late-stage AMD (AREDS classes 10–12; sensitivity, 100%; specificity, 96.5%). Importantly, 1504 images showing a healthy fundus (AREDS class 1) were identified correctly as such, resulting a specificity of 84.2% and sensitivity of 94.3% for this class.

Table 3. Comparison of Evaluation Metrics for the Cooperative Health Research in the Region of Augsburg Study Dataset Using the Model Ensemble

Metric	κ_0	κ_l	κ_q	κ_c	ac	ac_2	ac_b
All ages	34.90	19.55	11.01	38.75	83.1	90.5	63.8
Age >55 years	45.80	41.87	34.96	55.11	79.4	91.3	64.9
Age >55 years*	50.07	58.38	63.30	61.38	82.5	94.8	65.3

ac = accuracy; ac_b = average balanced accuracy over all classes; ac_2 = top 2 accuracy; κ_c = custom weighted Cohen's κ; κ_l = linear weighted Cohen's κ;κ_q = quadratic weighted Cohen's κ; κ_0 = unweighted Cohen's κ.
Data are percent.
*Excluding non–age-related macular degeneration pathologic features (e.g., synchysis, central serous chorioretinopathy, myopia related changes), artifacts and inadequately illuminated images.

A Closer Look at Misclassified Images in the Cooperative Health Research in the Region of Augsburg Dataset

Among the 44 images that were classified erroneously to be in AREDS class 11, although they were actually class 1 or 2, 15 had photographic and digital artifacts, 6 fundus images were either too bright or too dark for accurate assessment by CNNs, 4 images showed signs of epiretinal membranes, and 5 showed severe fundus changes associated with myopia (Fig S3, available at www.aaojournal.org). Similarly, 7 of 29 fundus images from classes 1 and 2 mistakenly classified in AREDS

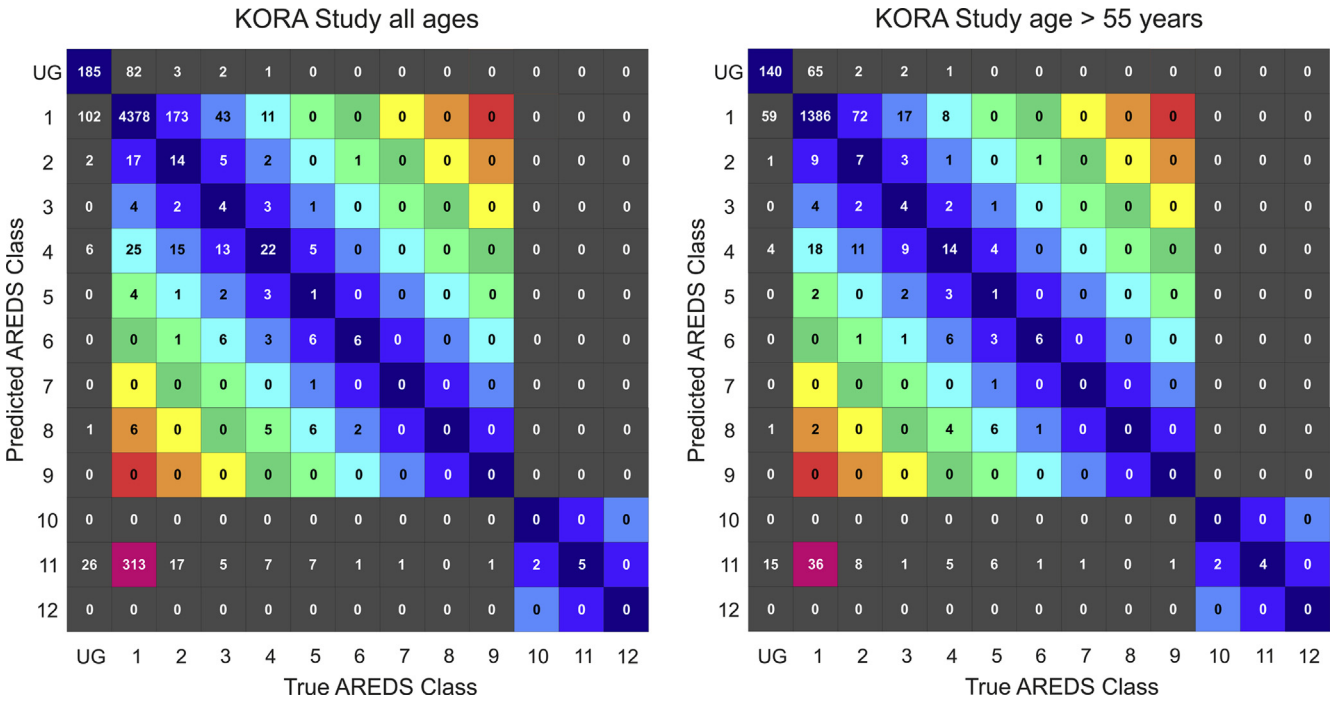

Figure 3. Confusion matrix of the true and predicted classes in the Cooperative Health Research in the Region of Augsburg (KORA) study. The true Age-Related Eye Disease Study (AREDS) class is plotted against the AREDS class predicted by the algorithm. The colors denote the weighting scheme that was used to train the neural networks. Fundus images that were classified mistakenly as AREDS class 11 but that did not have AMD-associated changes are highlighted in pink. Most of those images are from young individuals with visible macular reflexes. By restricting the analysis to images from individuals 55 years of age and older, the occurrence of misclassification can be reduced drastically.

class 4 showed changes associated with central serous chorioretinopathy, and an additional 10 images showed technical artifacts or other pathologic features that were recognized by the algorithm as disease-associated changes (Fig S3, available at www.aaojournal.org). Excluding those images increased the quadratic weighted κ measure to 63.2% (from 11.0%) and the unweighted κ to 50.1% (from 34.9%; Table 3).

A Peek into the Black Box of Deep Neural Networks

Neural networks often are regarded as black boxes, which makes interpretation of results difficult.[22] To visualize the important areas in the fundus images that are perceived and integrated by the perceptive fields, we randomly masked fundus images and let the CNN predict the AREDS class. In case the mask covers important phenotypic features for the respective class, a significant drop in confidence of prediction can be observed (Fig 4). Generally, the CNN recognizes important features in the macular region. Masking parts of the fovea further reduced the confidence of the CNNs. However, as soon as the pathologic changes are spread across the entire fundus image (e.g., in case the neovascular lesions encompass the entire macula as observed in AREDS class 11), masking a small portion of the image does not result in a reduced confidence.

Discussion

High Classification Accuracy in the Age-Related Eye Disease Study Dataset

Herein, we present an automated classification scheme based on the AREDS 9-step plus 3 severity scale and ungradable fundus images with high classification accuracy, outperforming a human grader in the AREDS dataset by reporting classification metrics (unweighted κ of 55.2% for AREDS 9-step plus 3 classes), similar to those observed for regraded images of the same grader (intragrader unweighted κ value of 58%).[10] Therefore, the model ensemble is well suited to classifying fundus images according to their AMD-associated changes. Interestingly, most model architectures show a stronger agreement on the assessment of the individual fundus images between each other (average quadratic weighted κ, 91%) than compared with intergrader agreement between physicians,[19] in line with the observation from other studies aimed at grading fundus images for diabetic retinopathy.[27] As other investigators have noted, the neural network was provided with only the image and associated grade, without explicit definitions of features (e.g., drusen, pigmentary changes, or atrophic areas). Because the network learned the features that were most predictive for the respective class, it is possible that the algorithm is using features previously unknown to or ignored by humans that may be highly predictive of certain AREDS classes. Alternatively, the algorithm was able to grade the respective features more consistently within the areas defined by the grading grid.

Interestingly, the F1 score was similar for all 6 neural net architectures across all AREDS classes, but differed significantly for class 12. We believe that this is the result of 3 factors: (1) class 12 contained the fewest samples, so the networks were not able to learn from many different training examples; (2) AREDS class 12 includes individuals who

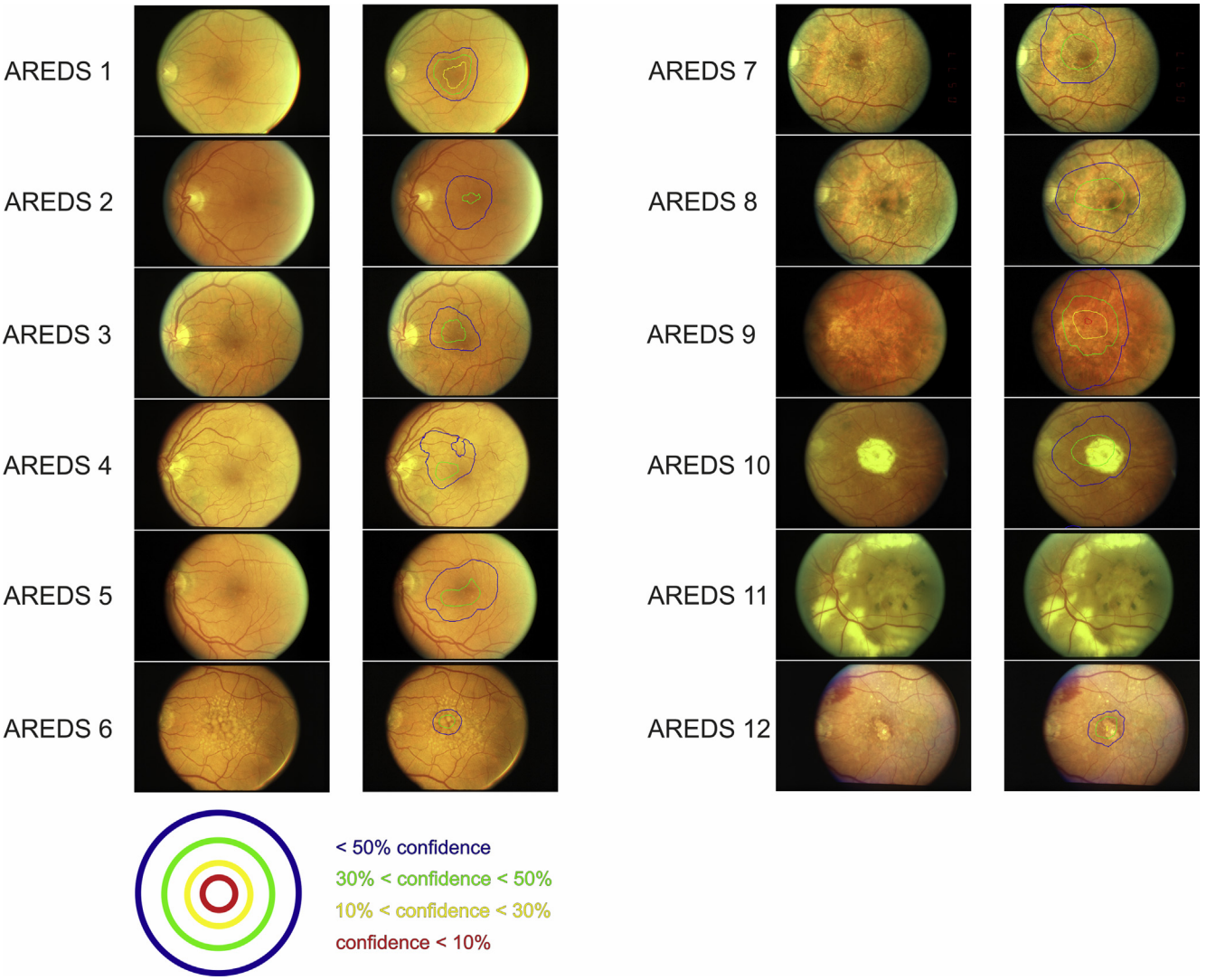

Figure 4. A peek into the black box of convolution neural nets. A representative image from each Age-Related Eye Disease Study (AREDS) class was masked randomly 10 000 times by a small rectangular field. We then used GoogLeNet to predict the true AREDS class of the masked images and recorded the confidence of the algorithm for the respective class. We then mapped the confidence onto the respective fundus image to highlight areas that are important for the convolution neural net (CNN) to predict the true AREDS class correctly. Masking areas within the blue borders reduces the CNN's confidence to less than 50%. Even stronger reduction in confidence is highlighted in green (confidence between 30% and 50%), yellow (confidence between 10% and 30%), and red (confidence less than 10%).

have both features of GA (class 10) as well as features of neovascular AMD (class 11), which further increases complexity for prediction because the quantifiable features overlap with other classes; and (3) the different model architectures may have to make tradeoffs between predicting certain classes. For instance, AlexNet showed the best F1 score for detecting class 12 (GA plus neovascular AMD); however, the price was that it performed worse at detecting classes 10 (GA) and 11 (neovascular AMD).

Previous efforts primarily have aimed at automated grading of fundus images of fewer classes and have shown reliable classification metrics outperforming human graders. For instance, recently, a deep learning approach was used to classify fundus images from the AREDS according to a 2-, 3-, or 4-step AMD severity scale, and the performance of this approach was compared with the accuracy of a physician.[41] The automated grading deep learning approach showed similar κ statistics, because the physician and was able to outperform her at certain tasks. Notably, the unweighted and linear weighted κ values for the 4-class classification problem were 70% and 79%, respectively, for the algorithm and 66% and 79%, respectively, for the physician. Our approach revealed even higher classification accuracies in a more complex setting. When broken down to a 4-class problem as described previously,[23] comprising no AMD (AREDS class 1), early AMD (AREDS classes 2 and 3), intermediate AMD (AREDS classes 4–9), and late AMD (AREDS classes 10–12), our algorithm revealed an

unweighted and linear weighted κ of 74% and 83%, respectively, in the AREDS and 42% and 51%, respectively, in the full KORA study.

Using Convolution Neural Nets to Classify Fundus Images from Other Population-Based Cohorts

In the AREDS dataset, the different metrics were correlated highly, and better-performing models generally outperformed worse-performing models in all evaluated metrics. In the KORA study, however, the model ensemble accuracy decreased when restricting the analysis to individuals 55 years of age and older, probably because many healthy fundus images (which are easier to identify with our scheme) were removed from the analysis. However, the balanced and top 2 accuracy, as well as all investigated κ metrics, were increased, indicating that the model performed better on the data. The proposed automated classification scheme therefore is limited largely to classification of images from individuals 55 years of age and older, because the algorithm was trained on fundus images from individuals who were at least 55 years of age at enrollment. When evaluating our algorithm on a population-based study including general adults 25 to 75 years of age, the quadratic weighted and unweighted κ values were markedly lower. We were able to show that this was the result of the presence of macular reflexes, particularly in younger individuals, that the algorithm mistakenly identified as disease-associated changes. The algorithm was not able to learn that macular reflexes are not disease-associated changes because the youngest individual in the AREDS was 55 years of age at recruitment, and macular reflexes are rarely observed in older individuals.

In addition, several other pathologic features as well as technical artifacts in the KORA images, which the algorithm was not trained to detect, further attenuated our classification accuracies in a population-based setting. This is not surprising because patients with other sight-threatening diseases were excluded in the prospective setting of the AREDS, and thus, the algorithm was not able to learn that those diseases are not associated with AMD. Nevertheless, the algorithm had an acceptable specificity and sensitivity to detect healthy and diseased fundus images in the KORA dataset restricted to fundus images from individuals older than 55 years (unweighted and quadratic weighted κ of 50.1% and 63.2%, respectively). Both manual and automated AREDS grading have inherited inaccuracies. However, misclassification of images by 1 or 2 steps on the scale is not that severe for most subsequent applications, especially because many steps are quite similar in appearance. Thus, our proposed algorithm is suitable to be applied to other population-based studies to expand the number of currently available datasets to dissect risk factors for early AMD and late AMD.

To increase the specificity further (i.e., to reduce the number of false-positive findings), a trained ophthalmologist can identify erroneously classified images as well as images demonstrating other pathologic features. Alternatively, other retinal diseases can be excluded by the imaging center, for instance, using a questionnaire or medical records. However, because the large majority of images in a population-based study are expected to be of healthy eyes, the proposed scheme already will reduce the burden of labor by correctly classifying most healthy fundus images as such. Importantly, the proposed classification scheme may be useful to harmonize the classification of different epidemiologic studies that use color fundus images for AMD classification, a challenge that is usually time consuming,[11] but necessary, to be able to compare studies or to conduct meta-analyses. In such a setting, confounding pathologic features and low-quality fundus images usually are excluded already, which should result in acceptable AREDS grading in those studies, similar to the accuracies observed in the KORA study after exclusion of other pathologic features and restricted to individuals 55 years of age and older.

Convolution Neural Nets Primarily Detect Changes in the Macular Region

By randomly masking regions in the fundus image, we were able to assess the importance of certain regions for classification confidence of the CNNs. The CNNs try to assess features in the macular region of the fundus images, which is to be expected because the AREDS 9-step plus 3 scale was developed to quantify changes in this area.[10] Importantly, masking the foveal region of the fundus images further decreased the confidence of the algorithm, potentially reflecting the decision to count only pigmentary changes in the central grid and not in the outer grid to contribute to the final AREDS class.[10] In case the AMD-associated changes are too numerous and cover most of the fundus image, masking a small portion of the fundus image did not decrease the confidence in the prediction. This further highlights that the CNNs are able to detect local as well as global features to predict the AREDS scale. Although it is possible that CNNs can find a way to "game the system" and be accurate in identifying diseases by using novel features or a combination of features, we did not observe this in our study. Future studies may be able to include additional phenotypic, genetic, or other image information (such as information about the second eye) in the training process to improve the classification accuracy further.

In conclusion, taken together, our new algorithm showed high classification accuracies outperforming human graders in the AREDS data and can be used to grade fundus images from population-based studies in individuals 55 years of age and older. Other pathologic features have to be excluded either before classification or afterward by a trained ophthalmologist. Nevertheless, our algorithm should reduce the financial burden and workload significantly by correctly identifying images with any pathologic features present. Importantly, our algorithm learned to extract important features from the macular region of the fundus images, further reinforcing the notion that automated grading systems indeed can be trained to detect specific disease-related changes in fundus images. At the current stage, we do not suggest that this classification system should be used by in an eye clinic or ophthalmologist practice. Although the model ensemble is acceptable at identifying healthy fundus

images, fundus images with features of AMD or other disease will causes the model ensemble to make a prediction for a certain AREDS class. In the future, we hope that we can improve the model by incorporating images from population-based surveys that contain other phenotypes to enable a CNN that can be used to prescreen patients, to aid diagnosis in day-to-day patient care, or both.

Acknowledgments

The authors thank the Age-Related Eye Disease Study participants and the Age-Related Eye Disease Study Research Group for their valuable contribution to this research, and all study participants for contributing to the Cooperative Health Research in the Region of Augsburg study.

References

1. Stark K, Olden M, Brandl C, et al. The German AugUR study: study protocol of a prospective study to investigate chronic diseases in the elderly. *BMC Geriatr*. 2015;15:130.
2. Grassmann F, Fauser S, Weber BHF. The genetics of age-related macular degeneration (AMD)—novel targets for designing treatment options? *Eur J Pharm Biopharm*. 2015;95:194–202.
3. Fritsche LG, Igl W, Bailey JNC, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet*. 2016;48:134–143.
4. Grassmann F, Ach T, Brandl C, et al. What does genetics tell us about age-related macular degeneration? *Annu Rev Vis Sci*. 2015;1:73–96.
5. Grassmann F, Kiel C, Zimmermann ME, et al. Genetic pleiotropy between age-related macular degeneration (AMD) and 16 complex diseases and traits. *Genome Med*. 2017;9:29.
6. Swaroop A, Branham KE, Chen W, Abecasis G. Genetic susceptibility to age-related macular degeneration: a paradigm for dissecting complex disease traits. *Hum Mol Genet*. 2007;16(spec no):R174–R182.
7. Ferris FL, Davis MD, Clemons TE, et al. A simplified severity scale for age-related macular degeneration: AREDS report no. 18. *Arch Ophthalmol*. 2005;123:1570–1574.
8. 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.
9. Klein R, Davis MD, Magli YL, et al. The Wisconsin Age-Related Maculopathy Grading System. *Ophthalmology*. 1991;98:1128–1134.
10. Davis MD, Gangnon RE, Lee L-Y, et al. The Age-Related Eye Disease Study severity scale for age-related macular degeneration: AREDS report no. 17. *Arch Ophthalmol*. 2005;123:1484–1498.
11. Klein R, Meuer SM, Myers CE, et al. Harmonizing the classification of age-related macular degeneration in the Three-Continent AMD Consortium. *Ophthalmic Epidemiol*. 2014;21:14–23.
12. Klaver CC, Assink JJ, van Leeuwen R, et al. Incidence and progression rates of age-related maculopathy: the Rotterdam Study. *Invest Ophthalmol Vis Sci*. 2001;42:2237–2241.
13. Ferris FL, Wilkinson CP, Bird A, et al. Clinical classification of age-related macular degeneration. *Ophthalmology*. 2013;120:844–851.
14. Tikellis G, Robman LD, Dimitrov P, et al. Characteristics of progression of early age-related macular degeneration: the Cardiovascular Health and Age-Related Maculopathy Study. *Eye*. 2007;21:169–176.
15. Sallo FB, Peto T, Leung I, et al. The International Classification system and the progression of age-related macular degeneration. *Curr Eye Res*. 2009;34:238–240.
16. Cachulo M, da L, Lobo C, Figueira J, et al. Prevalence of age-related macular degeneration in Portugal: the Coimbra Eye Study—report 1. *Ophthalmologica*. 2015;233:119–127.
17. Adams MKM, Simpson JA, Aung KZ, et al. Abdominal obesity and age-related macular degeneration. *Am J Epidemiol*. 2011;173:1246–1255.
18. Danis RP, Domalpally A, Chew EY, et al. Methods and reproducibility of grading optimized digital color fundus photographs in the Age-Related Eye Disease Study 2 (AREDS2 report number 2). *Invest Opthalmol Vis Sci*. 2013;54:4548.
19. Verma L, Prakash G, Tewari HK, et al. Screening for diabetic retinopathy by non-ophthalmologists: an effective public health tool. *Acta Ophthalmol Scand*. 2003;81:373–377.
20. Wintergerst MWM, Schultz T, Birtel J, et al. Algorithms for the automated analysis of age-related macular degeneration biomarkers on optical coherence tomography: a systematic review. *Transl Vis Sci Technol*. 2017;6:10.
21. Ciresan D, Meier U, Schmidhuber J. *Multi-column deep neural networks for image classification*. 2012 IEEE Conf. Comput. Vis. Pattern Recognit. IEEE; 2012:3642–3649.
22. Castelvecchi D. Can we open the black box of AI? *Nature*. 2016;538:20–23.
23. Brandl C, Breinlich V, Stark KJ, et al. Features of age-related macular degeneration in the general adults and their dependency on age, sex, and smoking: results from the German KORA study. *PLoS One*. 2016;11:e0167181.
24. Chew EY, Klein ML, Clemons TE, et al. No clinically significant association between *CFH* and *ARMS2* genotypes and response to nutritional supplements. AREDS Report Number 38. *Ophthalmology*. 2014;121:2173–2180.
25. The Age-Related Eye Disease Study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the Age-Related Eye Disease Study report number 6. *Am J Ophthalmol*. 2001;132:668–681.
26. Group TA-REDSR. The Age-Related Eye Disease Study (AREDS): design implications AREDS report no. 1. *Control Clin Trials*. 1999;20:573–600.
27. Graham B. *Kaggle Diabetic Retinopathy Detection competition report*. Department of Statistics and Centre for Complexity Science at the University of Warwick, Coventry, UK; 2015.
28. Chen T, Mu L, Li Y, et al. Mxnet: a flexible and efficient machine learning library for heterogeneous distributed systems. *arXiv Prepr*. 2015;arXiv:1512.
29. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70:213–220.
30. Powers DMW. Evaluation: from precision, recall and f-measure to roc., informedness, markedness & correlation. *J Mach Learn Technol*. 2011;2:37–63.
31. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28.
32. Aiello S, Eckstrand E, Fu A, et al. Fast scalable R with H20. In: Grün B, Hothorn T, Pebesma E, et al., eds. *Foundation for Open Access Statistics*. ISSN; 2015:1548–7660.
33. Krizhevsky A, Sutskever I, Hinton GE. *ImageNet Classification with Deep Convolutional Neural Networks*. Curran Associates

Inc.; 2012. Available at: https://www.google.de/search?dcr=0&source=hp&ei=UDSxWr3ONs3XkwWxhISACg&q=Aiello+S%2C+Eckstrand+E%2C+Fu+A%2C+et+al.+Fast+scalable+R+with+H20.+2015&oq=Aiello+S%2C+Eckstrand+E%2C+Fu+A%2C+et+al.+Fast+scalable+R+with+H20.+2015&gs_l=psy-ab.3...751.2269.0.2686.3.2.0.0.0.0.162.310.0j2.2.0....0...1c.1.64.psy-ab..1.0.0.0...0.MiyQB2ZmWew.

34. Szegedy C, Wei Liu, Yangqing J, et al. *Going deeper with convolutions*. 2015 IEEE Conf. Comput. Vis. Pattern Recognit. IEEE, Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Press, ISSN: 1063-6919; 2015:1–9.
35. Simonyan K, Zisserman A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. CoRR; 2014. abs/1409.1.
36. Szegedy C, Vanhoucke V, Ioffe S, et al. *Rethinking the Inception Architecture for Computer Vision*. Computing Research Repository (CoRR); 2015. abs/1512.0; Available at: https://arxiv.org/corr/home.
37. He K, Zhang X, Ren S, Sun J. *Identity Mappings in Deep Residual Networks*. Computing Research Repository (CoRR); 2016. abs/1603.0.
38. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.
39. Ho TK. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell*. 1998;20:832–844.
40. Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. *Ophthalmol Retina*. 2017;1:322–327.
41. Burlina P, Pacheco KD, Joshi N, et al. Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis. *Comput Biol Med*. 2017;82:80–86.

Footnotes and Financial Disclosures

Originally received: October 5, 2017.
Final revision: February 20, 2018.
Accepted: February 27, 2018.
Available online: April 10, 2018. Manuscript no. 2017-2295.

[1] Institute of Human Genetics, University of Regensburg, Regensburg, Germany.

[2] Regensburg Medical Image Computing, Ostbayerische Technische Hochschule Regensburg (OTH Regensburg), Regensburg, Germany.

[3] Department of Ophthalmology, University Hospital Regensburg, Regensburg, Germany.

[4] Department of Genetic Epidemiology, Institute of Epidemiology, University of Regensburg, Regensburg, Germany.

[5] Institute of Epidemiology II, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany.

[6] Regensburg Center of Biomedical Engineering (RCBE), OTH Regensburg and Regensburg University, Regensburg, Germany.

*Both authors contributed equally as first authors.

Financial Disclosure(s):
The author(s) have no proprietary or commercial interest in any materials discussed in this article.

Funding support for Age-Related Eye Disease Study was provided by the National Eye Institute, National Institutes of Health, Bethesda, Maryland (grant no.: N01-EY0-2127). The work was funded in part by grants from the German Federal Ministry of Education and Research (grant nos.: BMBF 01ER1206 [I.M.H] and 01ER1507 [I.M.H. and B.H.F.W]); by the Institutional Budget for Research and Teaching from the Freestate of Bavaria and the German Research Foundation (grant no.: WE 1259/19-2 [B.H.F.W.]).

HUMAN SUBJECTS: Human subjects were included in this study. The Age-Related Eye Disease Study was approved by an independent institutional review board at each clinical center, and informed consent to participate in the study was obtained from all patients. The Cooperative Health Research in the Region of Augsburg project (Kooperative Gesundheitsforschung in der Region Augsburg [KORA]) was approved by the ethics committee of the Bavarian Medical Association (Bayerische Landesärztekammer) and the Bavarian commissioner for data protection and privacy (Bayerischer Datenschutzbeauftragter), and informed consent to participate in the study was obtained from all patients. The study was performed in accordance with the tenets of the Declaration of Helsinki.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Grassmann, Mengelkamp, Heid, Palm, Weber

Analysis and interpretation: Grassmann, Mengelkamp, Brandl, Harsch, Zimmermann, Linkohr, Peters, Heid, Palm, Weber

Data collection: Grassmann, Mengelkamp, Brandl, Harsch, Zimmermann, Linkohr, Peters

Obtained funding: None

Overall responsibility: Grassmann, Mengelkamp, Brandl, Heid, Palm, Weber

Abbreviations and Acronyms:
AMD = age-related macular degeneration; **AREDS** = Age-Related Eye Disease Study; **CNN** = convolution neural net; **GA** = geographic atrophy; **KORA** = Cooperative Health Research in the Region of Augsburg.

Correspondence:
Bernhard H. F. Weber, PhD, Institute of Human Genetics, University of Regensburg, Franz-Josef-Strauss-Allee 11, D-93053 Regensburg, Germany. E-mail: bweb@klinik.uni-regensburg.de; and Christoph Palm, PhD, Regensburg Medical Image Computing, Ostbayerische Technische Hochschule Regensburg, Universitätsstr. 31, D-93053 Regensburg, Germany. E-mail: christoph.palm@oth-regensburg.de.