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;■:1−11 © 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