AMERICAN ACADEMY OF OPHTHALMOLOGY®

Predicting Visual Acuity by Using Machine Learning in Patients Treated for Neovascular Age-Related Macular Degeneration

Markus Rohm, MSc,[1,2] Volker Tresp, PhD,[2] Michael Müller, PhD,[1] Christoph Kern, MD,[1] Ilja Manakov, MSc,[1,2] Maximilian Weiss,[1] Dawn A. Sim, MD, PhD,[3] Siegfried Priglinger, MD,[1] Pearse A. Keane, MD, PhD,[3] Karsten Kortuem, MD, MHBA[1,3]

Purpose: To predict, by using machine learning, visual acuity (VA) at 3 and 12 months in patients with neovascular age-related macular degeneration (AMD) after initial upload of 3 anti–vascular endothelial growth factor (VEGF) injections.

Design: Database study.

Participants: For the 3-month VA forecast, 653 patients (379 female) with 738 eyes and an average age of 74.1 years were included. The baseline VA before the first injection was 0.54 logarithm of the minimum angle of resolution (logMAR) (±0.39). A total of 456 of these patients (270 female, 508 eyes, average age: 74.2 years) had sufficient follow-up data to be included for a 12-month VA prediction. The baseline VA before the first injection was 0.56 logMAR (±0.42).

Methods: Five different machine-learning algorithms (AdaBoost.R2, Gradient Boosting, Random Forests, Extremely Randomized Trees, and Lasso) were used to predict VA in patients with neovascular AMD after treatment with 3 anti-VEGF injections. Clinical data features came from a data warehouse (DW) containing electronic medical records (41 features, e.g., VA) and measurement features from OCT (124 features, e.g., central retinal thickness). The VA of patient eyes excluded from machine learning was predicted and compared with the ground truth, namely, the actual VA of these patients as recorded in the DW.

Main Outcome Measures: Difference in logMAR VA after 3 and 12 months upload phase between prediction and ground truth as defined.

Results: For the 3-month VA forecast, the difference between the prediction and ground truth was between 0.11 logMAR (5.5 letters) mean absolute error (MAE)/0.14 logMAR (7 letters) root mean square error (RMSE) and 0.18 logMAR (9 letters) MAE/0.2 logMAR (10 letters) RMSE. For the 12-month VA forecast, the difference between the prediction and ground truth was between 0.16 logMAR (8 letters) MAE/0.2 logMAR (10 letters) RMSE and 0.22 logMAR (11 letters) MAE/0.26 logMAR (13 letters) RMSE. The best performing algorithm was the Lasso protocol.

Conclusions: Machine learning allowed VA to be predicted for 3 months with a comparable result to VA measurement reliability. For a forecast after 12 months of therapy, VA prediction may help to encourage patients adhering to intravitreal therapy. *Ophthalmology 2018;■:1–9 © 2017 by the American Academy of Ophthalmology*

Supplemental material available at www.aaojournal.org.

In computer science, machine learning is a term that covers diverse approaches to artificial intelligence in computers and is currently a common phenomenon. The original aim, which was stated as early as 1959, was for computers to have the ability to learn without being explicitly programmed.[1] Many methods were tried over the years, but it took until the 1990s when machines started to beat humans. Such stories were then broadcast in the mass media and gained a great deal of publicity. In 1996 and 1997, Garry Kasparov lost a chess match against the IBM super-computer Deep Blue; this was the first time that a long-standing world champion (1985–2000) was defeated by a computer. Another well-publicized milestone was reached when IBM Watson beat 2 expert human players in the television show *Jeopardy* in 2011.[2]

In the medical sciences involving vision, several machine-learning techniques have been applied in various subspecialties.[3] Mostly, glaucoma and retinal imaging–related problems have been addressed. As the number of patients with diabetic retinopathy is continuing to increase,[4] efforts have recently been made to detect early forms of diabetic retinopathy by means of machine learning.[5–7] These programs have been shown to provide a highly sensitive tool for detecting such disease complications.[8] First approaches have been published for the automatic imaging analysis of OCT by using machine learning.[9,10] Additional studies are under way to automate the detection and classification of pathologic features in eye imaging including fundus photography and OCT.[11]

https://doi.org/10.1016/j.ophtha.2017.12.034
ISSN 0161-6420/18

Age-related macular degeneration (AMD) is still one of the leading causes of legally defined blindness in industrial countries.[12] However, the use of anti−vascular endothelial growth factor (VEGF) agents such as bevacizumab, ranibizumab, or aflibercept in AMD and other retinal diseases such as diabetic macular edema or venous occlusions have allowed, for the first time, the sustainable improvement, stabilization, or slow-down of disease progression.[13−20] Nevertheless, even after several years of experience, real-life results in patients vary greatly from results obtained in clinical trials.[21,22]

To improve insights into our own clinical results and to facilitate clinical research, a data warehouse (DW) has been set up in our institution.[23] It comprises data, such as diagnoses, medication, and surgery undergone, from electronic medical records (EMRs), as developed in our institution on the basis of the hospital-wide EMR system SAP i.s.h.med, from more than 320 000 patients. It also contains measurement data from OCT (e.g., central retinal thickness [CRT], macular volume) and other measurement devices such as IOLMaster (Zeiss, Oberkochen, Germany) and Pentacam (Oculus, Wetzlar, Germany).

For this study, we have used 5 modern machine-learning algorithms to predict the outcome of visual acuity (VA) after 1 year of anti-VEGF treatment in patients with newly diagnosed neovascular AMD. The base for the prediction was our DW with its clinical and measurement data. The forecasting of VA might help to reduce the psychologic pressure on patients, because many fear sight loss as a consequence of the diagnosis of wet AMD.[24] In particular, before the first injections, patients are extremely anxious and nervous, because they do not know what the outcome of the therapy will be and whether it is worth pursuing.[25]

Methods

Data Warehouse as Repository

To estimate VA after an initial upload of 3 anti-VEGF injections in patients with neovascular AMD after 3 and 12 months of intravitreal therapy, we applied machine-learning algorithms on real-life data from our DW. This is updated every night with new and altered information from the EMR (i.s.h.med, Cerner AG, Erlangen, Germany). The nightly transfer from EMR includes diagnoses, clinical data such as VA and intraocular pressure, intravitreal injections, medications, appointments, and surgical operations. Up until December 2016, it incorporated data from 330 801 patients, including 44 134 recorded intravitreal injections and 402 001 VA measurements. Our Ethics Committee ruled that approval was not required for this study. This study adheres to the tenets of the Declaration of Helsinki.

In addition to clinical data, 75 750 Spectralis OCT (Heidelberg Engineering, Heidelberg, Germany) measurements, but no images, were extracted from the software Heidelberg Eye Explorer. It provides an Extensible Markup Language (XML) export, which is a structured file format of measurement values, as in the company's software. This export contains data such as CRT, retinal volume, and measurements in a 9-zone grid over the macular region (Fig 1). The XML files have been incorporated into the DW with a custom-made extract, transform, load script, which is also programmed in Java. The DW itself was built on a Microsoft SQL database server running in a database cluster in the hospital's

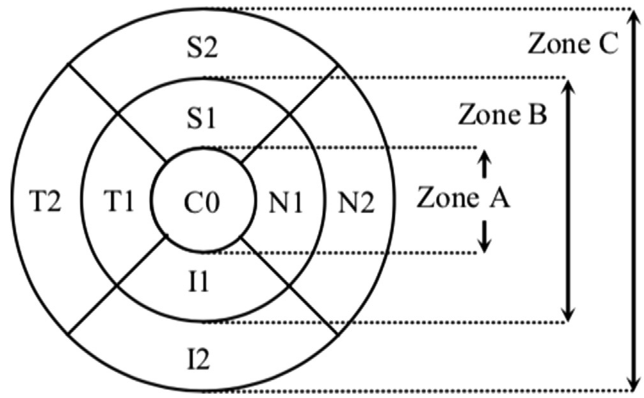

Figure 1. Nine sectors are shown, which are placed over the macula within the Heidelberg Eye Explorer software (Heidelberg Engineering, Heidelberg, Germany), C0 being the fovea. Data measurements in the Extensible Markup Language (XML) files were given for the whole grid and each individual sector. Therefore, features of an individual sector in Table S1 (available at www.aaojournal.org) were counted 9 times. The letters T (temporal), I (inferior), N (nasal), and S (superior) indicate the anatomic position of the sector.

datacenter. The PC used for the machine-learning algorithms was a quad core Intel i5 (3.30 GHz) PC with 32 gigabytes of RAM, running on Ubuntu 16.04 LTS and Python (3.6.0). In Python, the following libraries were used: scipy (Version: 0.18.1), numpy (Version: 1.12.0), pandas (Version: 0.19.2), jupyter (Version: 1.0.0), scikit-learn (Version: 0.18.1), and Theano (Version 0.8.2).

Data Preprocessing

For our test, we identified all patient eyes in the DW that, at any stage, received the International Classification of Diseases 10th Revision code H35.3, which stands for AMD, and which received at least 3 injections of anti-VEGF medication (aflibercept, bevacizumab, or ranibizumab) as noted in the surgical report. The standard operating procedure for treating neovascular AMD in our institution requires monthly injections for the first 3 months. After the initial upload of 3 injections, patients were treated according to the most recent treatment recommendations of the German Ophthalmological Society (Deutsche Ophthalmologische Gesellschaft [DOG]).[26,27] Accordingly, a pro re nata scheme was recommended until 2015. From 2015 onward, a treat-and-extend scheme was recommended alongside the pro re nata regimen. This led to 456 patients with 508 eyes for the long-term prediction goal (365 days) and 653 patients with 738 eyes for the short-term prediction goal (90 days); these were the patients who remained after the removal of all patients with invalid entries. The patients used for the long-term prediction represent a subgroup of the short-term prediction group.

Before algorithms can be trained on these patients, data need to be preprocessed and joined into a single dataset. The data came from the EMR and the OCT measurements. Most of the machine-learning algorithms expect a dataset with no missing values; this is difficult to achieve in real-life datasets such as ours. Therefore, we centered all the other values around VA measurements. Consequently, we associated VA values with the most recently available measurements from previous (but not future) records to rule out potential data mix-ups shortly before and after an intravitreal medication application.

In the data-collection process, we treated every eye as a separate patient and associated it with a generic eye identifier (GEID). The newly created dataset included not only the most recent