AMERICAN ACADEMY™
OF OPHTHALMOLOGY

Automated Identification of Diabetic Retinopathy Using Deep Learning

Rishab Gargeya,[1] *Theodore Leng, MD, MS*[2]

Purpose: Diabetic retinopathy (DR) is one of the leading causes of preventable blindness globally. Performing retinal screening examinations on all diabetic patients is an unmet need, and there are many undiagnosed and untreated cases of DR. The objective of this study was to develop robust diagnostic technology to automate DR screening. Referral of eyes with DR to an ophthalmologist for further evaluation and treatment would aid in reducing the rate of vision loss, enabling timely and accurate diagnoses.

Design: We developed and evaluated a data-driven deep learning algorithm as a novel diagnostic tool for automated DR detection. The algorithm processed color fundus images and classified them as healthy (no retinopathy) or having DR, identifying relevant cases for medical referral.

Methods: A total of 75 137 publicly available fundus images from diabetic patients were used to train and test an artificial intelligence model to differentiate healthy fundi from those with DR. A panel of retinal specialists determined the ground truth for our data set before experimentation. We also tested our model using the public MESSIDOR 2 and E-Ophtha databases for external validation. Information learned in our automated method was visualized readily through an automatically generated abnormality heatmap, highlighting subregions within each input fundus image for further clinical review.

Main Outcome Measures: We used area under the receiver operating characteristic curve (AUC) as a metric to measure the precision–recall trade-off of our algorithm, reporting associated sensitivity and specificity metrics on the receiver operating characteristic curve.

Results: Our model achieved a 0.97 AUC with a 94% and 98% sensitivity and specificity, respectively, on 5-fold cross-validation using our local data set. Testing against the independent MESSIDOR 2 and E-Ophtha databases achieved a 0.94 and 0.95 AUC score, respectively.

Conclusions: A fully data-driven artificial intelligence–based grading algorithm can be used to screen fundus photographs obtained from diabetic patients and to identify, with high reliability, which cases should be referred to an ophthalmologist for further evaluation and treatment. The implementation of such an algorithm on a global basis could reduce drastically the rate of vision loss attributed to DR. *Ophthalmology 2017;■:1–8 © 2017 by the American Academy of Ophthalmology*

Supplemental material is available at www.aaojournal.org.

Diabetes affects more than 415 million people worldwide, or 1 in every 11 adults.[1] Diabetic retinopathy (DR) is a vasculopathy that affects the fine vessels in the eye and is a leading cause of preventable blindness globally.[2] Forty to 45% of diabetic patients are likely to have DR at some point in their life; however, fewer than half of DR patients are aware of their condition.[3] Thus, early detection and treatment of DR is integral to combating this worldwide epidemic of preventable vision loss.

Although DR is prevalent today, its prevention remains challenging. Ophthalmologists typically diagnose the presence and severity of DR through visual assessment of the fundus by direct examination and by evaluation of color photographs. Given the large number of diabetes patients globally, this process is expensive and time consuming.[4] Diabetic retinopathy severity diagnosis and early disease detection also remain somewhat subjective, with agreement statistics between trained specialists varying substantially, as recorded in previous studies.[5,6] Furthermore, 75% of DR patients live in underdeveloped areas, where sufficient specialists and the infrastructure for detection are unavailable.[7] Global screening programs have been created to counter the proliferation of preventable eye diseases, but DR exists at too large a scale for such programs to detect and treat retinopathy efficiently on an individual basis. Consequently, millions worldwide continue to experience vision impairment without proper predictive diagnosis and eye care.

To address the shortfalls of current diagnostic workflows, automated solutions for retinal disease diagnoses from screened color fundus images have been proposed in the past.[8,9] Such a tool could alleviate the workloads of trained specialists, allowing untrained technicians to screen and process many patients objectively, without dependence on clinicians. However, previous approaches to automated DR detection have significant drawbacks that hinder usability in large-scale screenings. Because most of these algorithms have been derived from small data sets of approximately 500

http://dx.doi.org/10.1016/j.ophtha.2017.02.008
ISSN 0161-6420/17

fundus images obtained in isolated, singular clinical environments, they struggle to detect DR accurately in large-scale, heterogeneous real-world fundus data sets.[8–10] Indeed, methods derived from a singular data set may not generalize to fundus images obtained from other clinical studies that use different types of fundus cameras, eye dilation methods, or both, hindering clinical impact in real-world workflows.[8,9] Moreover, many of these algorithms depend on manual feature extraction for DR characterization, aiming to characterize prognostic anatomic structures in the fundus, such as the optic disc or blood vessels, through detailed hand-tuned features. Although these hand-tuned features may perform well on singular fundus data sets, they again struggle to characterize DR accurately in fundus images from varying target demographics by overfitting to the original sample. General-purpose features, such as Speeded Up Robust Features (SURF) and Histogram of Oriented Gradients (HOG) descriptors, have been investigated as a nonspecific method for DR characterization, but these methods tend to underfit and learn weaker features unable to characterize subtle differences in retinopathy severity.[11–13]

We created a fully automated algorithm for DR detection in red, green, and blue fundus photographs using deep learning methods and addressed the above limitations in previously published DR detection algorithms. Deep learning recently has gained traction in a variety of technological applications, including image recognition and semantic understanding, and has been used to characterize DR in the past.[14–17] In this study, we adapted scalable deep learning methods to the domain of medical imaging, accurately classifying the presence of any DR in fundus images from a data set of 75 137 DR images. Our algorithm used these images as inputs and predicted a DR classification of 0 or 1. These classes corresponded to no retinopathy and DR of any severity (mild, moderate, severe, or proliferative DR). This solution was fully automated and could process thousands of heterogeneous fundus images quickly for accurate, objective DR detection, potentially alleviating the need for the resource-intensive manual analysis of fundus images across various clinical settings and guiding high-risk patients for referral to further care. In addition, all information learned in our algorithmic pipeline was visualized readily through an abnormality heatmap, intuitively highlighting subregions within the classified image for further clinical review.

Methods

Figure 1A represents an abstraction of our algorithmic pipeline. We compiled and preprocessed fundus images across various sources into a large-scale data set. Our deep learning network learned data-driven features from this data set, characterizing DR based on an expert-labelled ground truth. These deep features were propagated (along with relevant metadata) into a tree-based classification model that output a final, actionable diagnosis.

Fundus Image Data set and Preprocessing

We derived our predictive algorithm from a data set of 75 137 color fundus images obtained from the EyePACS public data set (EyePACS LLC, Berkeley, CA).[17] The images represented a heterogeneous cohort of patients with all stages of DR. This data set contained a comprehensive set of fundus images obtained with varying camera models from patients of different ethnicities, amalgamated from many clinical settings.[8,9] Each image was associated with a diagnostic label of 0 or 1 referring to *no retinopathy* or *DR of any severity*, respectively, determined by a panel of medical specialists.

Because of the large-scale nature of our data set and the wide number of image sources, images often demonstrated environmental artifacts that were not diagnostically relevant. To account for image variation within our data set, we performed multiple preprocessing steps for image standardization before deep feature learning. First, we scaled image pixel values to values in the range of 0 through 1. Images then were downsized to a standard resolution of 512×512 pixels by cropping the inner retinal circle and padding it to a square.

To preprocess images further before learning, we used data set augmentation methods to encode multiple invariances in our deep feature learning procedure. Data set augmentation is a method of applying image transformations across a sample data set to increase image heterogeneity while preserving prognostic characteristics in the image itself. One important principle of fundus diagnosis is that disease detection is rotationally invariant; identification and characterization of pathologic structures are determined locally relative to major anatomic structures, regardless of orientation. We encoded rotational invariance into our predictions by randomly rotating each image before propagating these images into our model. By enforcing similar predictions for randomly rotated images, we improved our model's ability to generalize and correctly classify fundus images of various orientations across different types of fundus imaging devices without a loss of accuracy. Other important characteristics were the color and brightness of the image. To encode invariance to varying color contrast between images, we introduced brightness adjustment with a random scale factor α per image, sampled from a uniform distribution over [−0.3, 0.3], through equation 1,

$$y = (x - \text{mean}) \times (1 + \alpha) \tag{1}$$

and contrast adjustment with a random scale factor β per image, sampled from a uniform distribution over [−0.2, 0.2], using equation 2.

$$y = (x - \text{mean}) \times (\beta) \tag{2}$$

These image transformations aimed to improve our model's ability to classify varieties of retinal images obtained in unique lighting settings with different camera models.

Deep Feature Learning

Our novel approach to feature learning for DR characterization leveraged deep learning methods for automated image characterization. Specifically, we used customized deep convolutional neural networks for automated characterization of fundus photography because of their wide applicability in many image recognition tasks and robust performance on tasks with large ground truth data sets.[18,19] These networks used convolutional parameter layers to learn iteratively filters that transform input images into hierarchical feature maps, learning discriminative features at varying spatial levels without the need for manually tuned parameters. These convolutional layers were positioned successively, whereby each layer transformed the input image, propagating output information into the next layer. We used the principle of deep residual learning to develop a custom convolutional network, learning discriminative features for DR detection, as defined by equation 3,

$$x_l = \text{conv}_l(x_{l-1}) + x_{l-1}\#, \tag{3}$$

where conv_l represents a convolutional layer l, which returns the sum of both its output volume and the previous convolutional