



Development and Validation of an Automated Diabetic Retinopathy Screening Tool for Primary Care Setting

Diabetes Care 2020;43:e147–e148 | <https://doi.org/10.2337/dc19-2133>

Alauddin Bhuiyan,^{1,2}
 Arun Govindaiah,¹ Avnish Deobhakta,²
 Meenakshi Gupta,² Richard Rosen,²
 Sophia Saleem,² and
 R. Theodore Smith²

We report the construction of a fully automated artificial intelligence deep learning (DL) software on a secure Health Insurance Portability and Accountability Act–compliant cloud-based platform for effective and efficient screening of referable and nonreferable diabetic retinopathy (DR) in the primary care setting. DR is one of the leading causes of blindness in the U.S. and other developed countries. Early detection is the key for prevention. Currently, screening for DR is done by ophthalmologists or optometrists with limited catchment areas and also requires time-consuming referral. Many patients (some studies suggest almost one out of three) do not get these exams (1).

The critical goal is the screening of “referable” DR for referral to an ophthalmologist. A 5-point scale (2) (no DR and mild, moderate, severe, and proliferative DR) may be used for grading DR based on the presence and extent of microaneurysms, exudates, hemorrhages, and other abnormalities. By definition, no DR and mild DR are considered nonreferable, and the other categories are considered referable. Previous automated DR screening techniques are of varying accuracy and performance, but recently DL approaches have had greater success.

DL refers to a class of machine learning techniques that learns features directly from images without feature labels and usually

requires very large training data sets such as the publicly available kaggle data set (KG-set) (88,702 subjects) (3). DL has also been applied for detecting diseases such as macular degeneration and melanoma.

Our DL screening system was initially built on 70% of the KG-set and validated prospectively. Complete details will be presented in a forthcoming publication. Conceptually, a model for retinal fundus image classification must be robust to learning features on a wide scale, from small microaneurysms to large fronds of neovascularization. Therefore, three DL neural networks were used (Xception, Inception-V3, and Inception-ResNet-V2), each operating at one or two image resolutions to give five networks total and increase robustness to image features of different sizes. We have also used this ensemble approach in age-related macular degeneration (4). Each of the five networks then produces a set of five probabilities of a given image belonging to each of the five DR classes: none and early, intermediate, severe, and proliferative DR. The 25 total probabilities are then input to a logistic model tree (LMT), combining logistic regression and decision tree learning. The LMT is trained to decide the DR class based on the totality of the DL inputs and is the final classifier. (See Fig. 1.)

We then validated our system on the remaining 30% of the KG-set and the

complete publicly available MESSIDOR data set (3) of 1,748 high-resolution fundus images. We achieved high sensitivities and specificities (98–99% in each category)—equivalent to those achieved by human experts. Significantly, we compared our model with a well-known DL algorithm developed by IDx (5) on MESSIDOR, which achieved a sensitivity of 97% and specificity of 87.0% for referral-level DR. Our algorithm’s specificity was 99% compared with IDx’s 87%. Indeed, to the best of our knowledge, our system performs at least comparably with other existing screening systems on these data sets.

To test real-world performance, we then examined prospectively acquired nonmydriatic retinal images from 974 patients with diabetes with approval by the Institutional Review Board of Mount Sinai in a primary care setting at the New York Eye and Ear Infirmary of Mount Sinai on a Topcon TRC-NW400 camera between 1 January 2017 and 31 December 2017. A total of 814 patients were judged by two retina specialists (S.S. and M.G.) to have no DR; 83 had mild DR, 12 had moderate DR, 2 had severe DR, and 5 had proliferative DR. These gold standard gradings were compared with analysis by our DL system, which achieved a sensitivity of 82.6%, adequate for U.S. Food and Drug Administration

¹iHealthScreen Inc, Richmond Hill, NY

²New York Eye and Ear Infirmary of Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY

Corresponding author: R. Theodore Smith, rts1md@gmail.com

Received 24 October 2019 and accepted 15 July 2020

© 2020 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/content/license>.

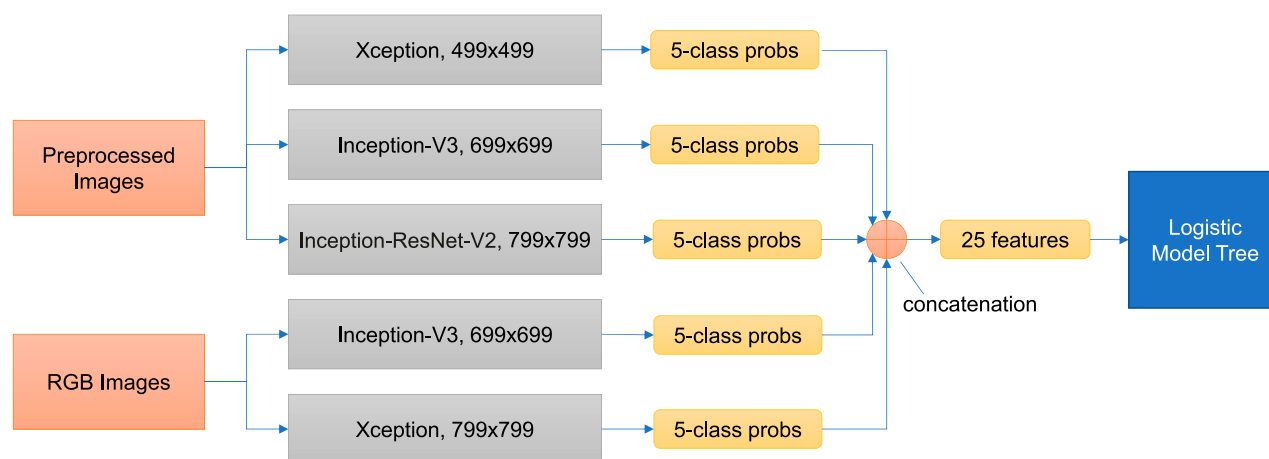


Figure 1—Ensemble framework of DL-based DR screening system. The preprocessed and the original RGB images are input to ensembles of three and two DL models, respectively, differing in type of architecture and input image size. Each model then produces a set of five probabilities (probs) belonging to each of the five DR classes: none and early, intermediate, severe, and proliferative DR. The 25 total probabilities are then concatenated (grouped) to form a vector of 25 features, which is input to an LMT. The LMT has been trained to decide the DR class based on the totality of the DL inputs and is the final classifier.

approval of a screening system, and a specificity of 93.7% for referral-level DR.

A telemedicine platform was also created that integrates the server-side programs (DL modules) and local remote devices for collecting patient data and images. A report on the automated analysis of the patient's DR status (referable or nonreferable) can be sent within a minute to the primary care provider.

The validations of our DR screening tool on the KG-set and MESSIDOR data set, and the high accuracy on the Mount Sinai data set taken prospectively, suggest that it is suitable for the primary care setting. The system should be tested prospectively on a large scale in clinical settings on the integrated telemedicine platform. If the excellent clinical performance of the tool thus far is confirmed, deployment in primary

care settings for early diagnosis of DR will be warranted.

Funding. This study was funded by the National Eye Institute, National Institutes of Health (R01 EY015520 [to R.T.S.] and SBIR R43EY02684 [to A.B.]).

The sponsor or funding organization had no role in the design or conduct of this research. **Duality of Interest.** A.B. and A.G. receive salary support from iHealthScreen Inc. iHealthScreen Inc. has ownership of a patent on this technology. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. S.S. and M.G. researched data. A.D. and R.R. contributed to discussion and reviewed and edited the manuscript. A.B. and A.G. researched data and contributed to discussion. R.T.S. wrote the manuscript. R.T.S. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Kant K. 2017 update on diabetic retinopathy Preferred Practice Pattern by AAO, 2017.

Accessed 18 September 2019. Available from <https://specialty.medicaldialogues.in/2017-update-on-diabetic-retinopathy-preferred-practice-pattern-by-aa/>.

2. Wilkinson CP, Ferris FL 3rd, Klein RE, et al.; Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;110:1677–1682

3. MESSIDOR-2 DR grades: adjudicated DR severity, DME, and gradability for the MESSIDOR-2 fundus dataset. Accessed 1 February 2019. Available from <https://www.kaggle.com/google-brain/messidor2-dr-grades>.

4. Govindaiah A, Smith RT, Bhuiyan A. A new and improved method for automated screening of age-related macular degeneration using ensemble deep neural networks. *Conf Proc IEEE Eng Med Biol Soc* 2018;2018:702–705

5. 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