



Automated Identification of Diabetic Retinopathy Using Deep Learning

Rishab Gargeya,¹ Theodore Leng, MD, MS²

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

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,

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

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

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

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,

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

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

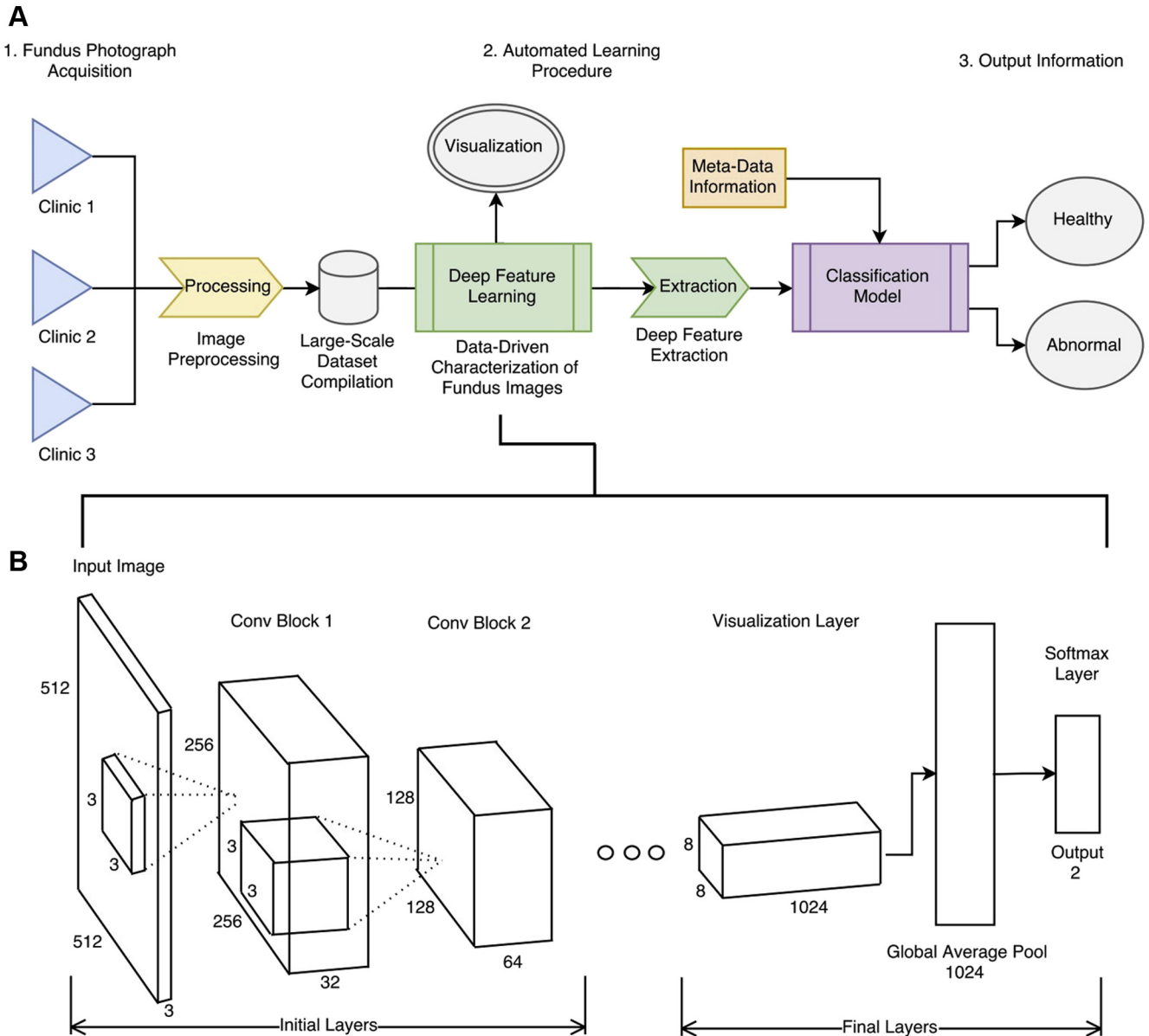


Figure 1. Abstraction of the proposed algorithmic pipeline. **A**, Integration of our algorithm in a real diagnostic workflow. **B**, Abstraction of the deep neural network. We extracted features from the global average pool layer for a total of 1024 deep features. Conv = convolutional.

layer's output volume.²⁰ These summations between convolutional layers facilitated incremental learning of an underlying polynomial function, allowing us to train deep networks with many parameter layers for enhanced characterization of fundus images. A full diagram of all layers in this residual network can be viewed in Figure S2 (available at www.aaojournal.org). We separated blocks of convolutional layers based on size, yielding 5 residual blocks of 4, 6, 8, 10, and 6 layers, respectively. We increased the number of filters in each convolutional block, yielding 32, 64, 128, 256, and 512 filters in successive blocks. Leading convolutional layers in each block used a stride of 2 to transition between spatial levels, where successive blocks analyze the input image at reduced spatial dimensions in a coarse-to-fine fashion. An abstraction of this feature learning architecture is represented

in Figure 1B. As is standard in deep convolutional networks, each convolutional layer used batch normalization and the ReLU nonlinearity function to ensure smooth training and prevent overfitting, while using 2-class categorical cross-entropy loss for class discrimination.^{21,22}

Visualization Heatmap Embedding

To visualize the learning procedure of our network, we implanted a convolutional visualization layer at the end of our network. This layer was followed by an average pooling layer and a traditional softmax layer, generating class probabilities for error optimization.²³ The visualization layer was functionally a convolutional layer with a large width of 1024 filters, encapsulating all

previous information into a large layer at the end of the network. Using this final layer, we generated a visualization heatmap by applying equation 4, as detailed in Zhou et al²⁴:

$$\text{Map}_c = \sum_k W_k^c \times x_k \quad (4)$$

This visualization highlighted highly prognostic regions in the input image for future review and analysis, potentially aiding real-time clinical validation of automated diagnoses at the point of care.

Feature Extraction

We chose to extract learned features from the global average pooling layer in our residual network. This layer represented the average activations of each unit in the final visualization parameter layer, yielding 1024 features. This layer generated the most comprehensive, discriminative features because it averaged all activations of the final, largest convolutional layer. We used these features both to construct a visualization heatmap and to generate a final image diagnosis through second-level classification models.

Metadata Appendage and Feature Vector Construction

To enhance the diagnostic accuracy of our final prediction, we appended multiple metadata features related to the original fundus image to our feature vector. We defined 3 metadata variables useful in characterizing the original image: original pixel height of the image, original pixel width of the image, and field of view of the original image. These variables augmented our input feature vector into a final vector of 1027 features. Inserting image metadata was important in accounting for environmental variables related to the original fundus photograph before preprocessing, which may influence the model's predictions.

Decision Tree Classification Model

To generate a final diagnosis, we trained a second-level gradient boosting classifier on our representative feature vector of 1027 values. Gradient boosting classifiers are tree-based classifiers known for capturing fine-grained correlations in input features based on intrinsic tree ensembles and bagging.²⁵ We choose this classifier because of its speed of implementation and robustness against overfitting. This classifier was trained using the categorical cross-entropy loss function, yielding the probability that the input image was indeed pathologic.

Results

We tested the model using 5-fold stratified cross-validation on our local data set of 75 137 images, preserving the percentage of samples of each class per fold. This testing procedure trained 5 separate models, each holding out a distinct validation bucket of approximately 15 000 images. Average metrics were derived from 5 test runs on respective held-out data by comparing the model's predictions with the gold standard determined by the panel of specialists. A final, complete model was trained on all 75 137 images before external validation on public data set.

Local Cross-Validation Results

Our algorithm scored an average area under the receiver operating characteristic curve (AUC) of 0.97 during cross-validation. This metric indicated excellent performance on a large-scale data set. The algorithm also achieved an average 94% sensitivity and a 98% specificity. This statistic represented the highest point on the receiver operating characteristic curve with minimal trade-off

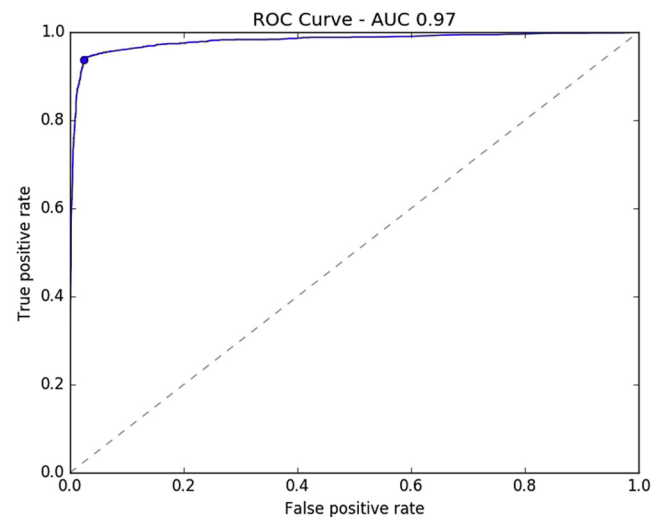


Figure 3. The mean receiver operating characteristic (ROC) curve derived from 5-fold cross-validation. The dotted line represents the trade-off resulting from random chance. The blue curve represents the model's trade-off, with the blue dot marking the threshold point yielding a sensitivity and specificity of 94% and 98%, respectively. AUC = area under the receiver operating characteristic curve.

between precision and recall. This receiver operating characteristic curve is plotted in Figure 3, with the AUC representing the AUC metric.

Public Data set Test Results and Prior Study Performance Comparison

Because our data set was compiled independently, we also evaluated the performance of our best-performing algorithm on 2 separate and independent data sets (MESSIDOR 2 and E-Ophtha) for model evaluation and comparison to prior algorithms in the field.^{26–28} These data sets contained fundus images with various pathologic signs, indicating a wide variety of DR cases.

The public MESSIDOR 2 database contains 1748 fundus images from 4 French eye institutions monitoring retinal complications in diabetics. These images were graded already for the existence of pathologic signs, separating images into healthy and diseased. All images were used to make real clinical diagnoses in their respective institutions and were released online in a de-identified format for algorithm evaluation. This data set is the largest public data set used in multiple prior studies for disease detection, allowing for robust performance comparison between our model and these published reports. Our algorithm achieved a 93% sensitivity and 87% specificity on this external data set, with an AUC of 0.94, reporting comparable or better results in comparison with previously published studies on disease detection in similar data set^{29–32} (Table 1). It is worth noting that our model did not train on any MESSIDOR 2 fundus images before validation, unlike previous studies that used this database alone to create predictive models. In addition to evaluating our model's ability to predict the presence of DR, we also tested the ability of our model to discern healthy images from those with mild DR specifically, using a subset of 1368 healthy and mild DR images from the MESSIDOR 2 database (MESSIDOR 2 is provided by the LaTIM laboratory, Available at: <http://latim.univ-brest.fr>; and the Messidor program partners, Available at: <http://messidor.crihan.fr>). This data set was important for evaluation as it provided deeper insight into

Table 1. Average Sensitivity, Specificity, and Area under the Receiver Operating Characteristic Curve Measures of No Diabetic Retinopathy versus Any Stage of Diabetic Retinopathy Using the MESSIDOR 2 Public Data set

	Sensitivity	Specificity	Area under the Receiver Operating Characteristic Curve
Our model	0.93	0.87	0.94
Antal et al (2014) ²⁹	0.90	0.91	0.99
Sánchez et al (2011) ³⁰	0.92	0.50	0.88
Seoud et al (2016) ³¹	0.94	0.50	0.90
Roychowdhury et al (2014) ³²	1.00	0.53	0.90

our model’s potential as a preliminary screening method for mild DR detection and its ability to detect fine microaneurysms in retinal images, which make up less than 1% of the entire image. Our algorithm achieved 74% sensitivity and 80% specificity, with an overall AUC of 0.83.

The public E-Ophtha database (provided by the ANR-TEC-SAN-TELEOPHTA project funded by the French Research Agency [ANR]) contained 463 images separated into 268 healthy images and 195 abnormal images. We evaluated our algorithm on a subset of the E-Ophtha database consisting of 405 images of healthy eyes and images with early DR. Like the MESSIDOR 2 mild DR subset, all pathologic images in this subset exhibited only mild-stage DR with mild signs, specifically the presence of microaneurysms or small hemorrhages. We achieved a 90% sensitivity and a 94% specificity with this data set, with an AUC of 0.95 (Table 2). These results show high potential for early detection of mild DR symptoms, demonstrating the ability for our network to classify early cases of DR. Figure 4 represents a t-distributed stochastic neighbor embedding visualization of this data set by our automated method, clearly showing 2 clusters of fundus images and indicating the ability of our model to discern images of mild retinopathy with fine microaneurysms and small hemorrhages.³³ Unfortunately, no previous studies have published performance statistics using this data set for normal versus abnormal evaluation because the data set is relatively new.

Visualization Heatmap Analysis

In clinical use, interpreting the output of diagnosis-guiding software is important for triaging referrals and focusing one’s clinical examination. Toward that end, we created a heatmap visualization

method to represent intuitively the learning procedure of our deep learning network.

Figure 5 ties the mathematical learning of the network to the domain of clinical ophthalmology by highlighting the regions important to our model’s prediction. The retinal image in Figure 5A has highlighted regions of retinal hemorrhage and neovascularization, as well as hard exudate, in the nasal and temporal quadrants, indicating proliferative DR. The retinal image in Figure 5B highlights retinal findings in the upper and lower left quadrants. These features are what ophthalmologists use to make a diagnosis, and the highlighted importance of these regions corroborates the domain-guided learning procedure of our model.

Model Runtime Performance Test

To evaluate the realistic performance of our software as a preliminary screening tool, we found it important to evaluate the runtime performance on common computer hardware found in an average clinic. We tested the performance on 2 devices: a computer desktop with an Intel Dual-Core Processor (Intel, Santa Clara, CA) running at 2.4 GHz and an iPhone 5 (Apple Inc., Cupertino, CA). The real-time runtime performance yielded an average of 6 and 8 seconds, respectively, per evaluated image, indicating broad usability in a variety of medical screening situations.

Discussion

This study proposed a novel automated-feature learning approach to DR detection using deep learning methods. It provides a robust solution for DR detection within a large-scale data set, and the results attained indicate the high efficacy of our computer-aided model in providing efficient, low-cost, and objective DR diagnostics without depending on clinicians to examine and grade images manually. Our method also does not require any specialized, inaccessible, or costly computer equipment to grade images; it can be run on a common personal computer or smartphone with average processors. In addition to image classification, our pipeline accurately visualized abnormal regions in the input fundus images, enabling clinical review and verification of the automated diagnoses.

We validated our algorithm against multiple public databases, yielding competitive results without having trained on fundus images from the same clinic. Our method delivered competitive results compared with multiple published

Table 2. Summary of Experimental Results

	Sensitivity	Specificity	Area under the Receiver Operating Characteristic Curve
EyePacs public data set 5-fold stratified cross-validation of our model	0.94	0.98	0.97
Performance of our model on the MESSIDOR 2 data set (no DR vs. any stage of DR)	0.93	0.87	0.94
Performance of our model on the MESSIDOR 2 subset with mild DR (no DR vs. mild DR)	0.74	0.80	0.83
Performance of our model on the E-Ophtha subset with mild DR (no DR vs. mild DR)	0.90	0.94	0.95

DR = diabetic retinopathy.

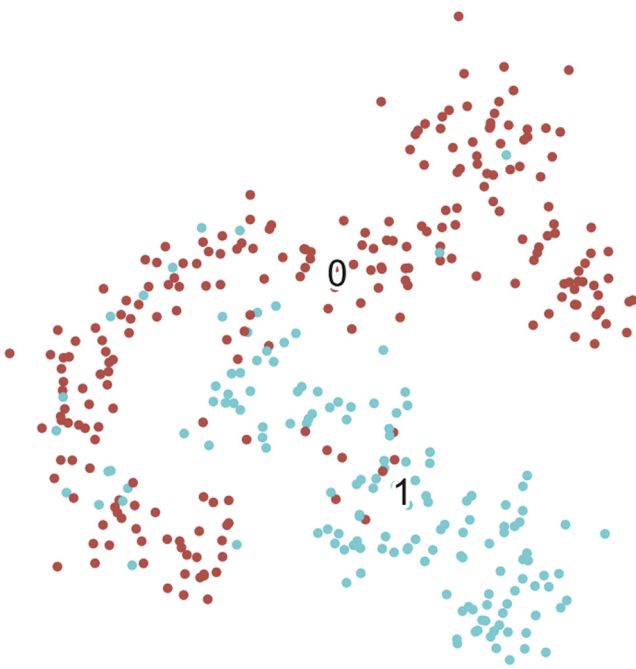


Figure 4. t -Distributed stochastic neighbor embedding visualization of the E-Ophtha data set, clustered based on deep features. *Red dots* represent healthy fundus images, whereas *blue dots* represent images with mild retinopathy. This visualization represents the ability of our method objectively to separate normal patients from those with early cases of diabetic retinopathy for referral.

methods for DR detection in the external MESSIDOR 2 database, showing the potential of an automated data-driven system over manual counterparts. Our results with the MESSIDOR 2 mild DR subset and E-Ophtha database were particularly insightful, evaluating the capability of our

method in distinguishing between normal and early cases of DR that exhibit only fine microaneurysms and small hemorrhages. Although we were able to achieve an AUC of 0.95 on the E-Ophtha database, our algorithm struggled to differentiate between healthy and very early cases of DR in the MESSIDOR 2 database, missing cases that demonstrated only a few, fine microaneurysms. Microaneurysm detection is difficult even for human graders because of its small appearance and poses an important limitation in future DR detection systems for accurate, robust early detection. We expect that a combination of manual features, targeting specific characteristics of microaneurysms for mild DR detection, with the robust potential of deep learning systems to characterize accurately all other stages of DR without confusion from brightness and capture artifacts, will yield more robust results in future early DR detection studies.

Our experiments on held-out data indicate the potential of deep learning systems to model and predict disease accurately in fundus images. We found our performance to corroborate the findings of a recent deep learning investigation in automated DR assessment by Gulshan et al.³⁴ Although we also evaluated our model using the same external MESSIDOR 2 data set as Gulshan et al, that group detected only referable DR, whereas we also evaluated the ability of our algorithm to detect mild DR. Given our results, there is a high potential for automated machine learning systems to predict the presence of early-stage DR, as well as referable DR.

It is also important to note the background ethnicity and geographic location of the target demographics in the 3 analyzed data sets. Although our method trained only using the EyePACS data set, consisting of images from the geographic region of California, it could generalize sufficiently to both the MESSIDOR 2 and E-Ophtha databases, containing images of diabetic patients from France. Further

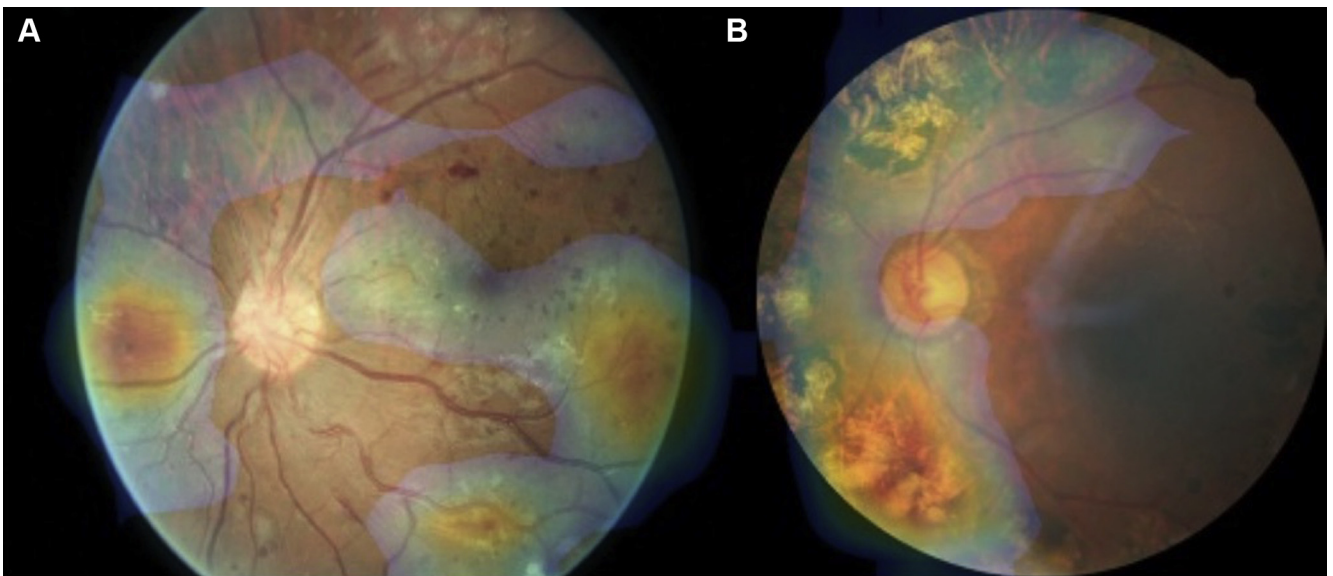


Figure 5. Visualization maps generated from deep features. **A**, Fundus heatmap overlaid on a fundus image, highlighting pathologic regions in the nasal and temporal quadrants. **B**, Pathologic findings in the upper and lower left quadrants. These visualizations are generated automatically, locating regions for closer examination after a patient is seen by a consultant ophthalmologist.

work may be needed to analyze the impact of geographic variation within training and testing data sets on model performance with regard to pigmentation of the retina and prevalence of different DR severity stages, as well as the impact of pupil dilation during the screening of target demographics.

For proper clinical application of our method, further testing and optimization of the sensitivity metric may be necessary to ensure a minimum false-negative rate. Because a false-negative result represents potentially denying a patient necessary eye care, computer-aided models for disease detection must prioritize this metric. To increase our sensitivity metric further, it may be important to control specific variances in our data set, such as ethnicity or age, to optimize our algorithm for certain demographics during clinical use. In the future, it also may be important to investigate different types of common patient metadata, such as genetic factors, patient history, duration of diabetes, hemoglobin A1C value, and other clinical data that may influence a patient's risk for retinopathy. Adding this information into the classification model may yield insightful correlations into underlying DR risk factors outside of strictly imaging information, potentially enhancing diagnostic accuracy.

Overall, our algorithm's results show the potential of automated feature-learning systems in streamlining current retinopathy screening programs in a cost-effective and time-efficient manner. The implementation of such an algorithm on a global basis could drastically reduce the rate of vision loss attributed to DR, improving clinical management and creating a novel diagnostic workflow for disease detection and referral.

References

- International Diabetes Federation (IDF). *IDF Diabetes Atlas 7th edition*; 2015. <http://www.diabetesatlas.org/>. Accessed October 20, 2016.
- Wilkinson CP, Ferris FL, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110:1677-1682.
- Beagley J, Guariguata L, Weil C, Motala AA. Global estimates of undiagnosed diabetes in adults. *Diabetes Res Clin Pract*. 2015;103:150-160.
- Ozieh MN, Bishu KG, Dismuke CE, Egede LE. Trends in health care expenditure in U.S. adults with diabetes: 2002–2011. *Diabetes Care*. 2015;38:1844-1851.
- Sellahewa L, Simpson C, Maharajan P, et al. Grader agreement, and sensitivity and specificity of digital photography in a community optometry-based diabetic eye screening program. *Clin Ophthalmol*. 2014;8:1345-1349.
- Ruamviboonsuk P, Wongcumchang N, Surawongsin P, et al. Screening for diabetic retinopathy in rural area using single-field, digital fundus images. *J Med Assoc Thai*. 2005;88:176-180.
- Guariguata L, Whiting DR, Hambleton I, et al. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Res Clin Pract*. 2014;103:137-149.
- Mookiah MRK, Acharya UR, Chua CK, et al. Computer-aided diagnosis of diabetic retinopathy: a review. *Comput Biol Med*. 2013;43:2136-2155.
- Winder RJ, Morrow PJ, McRitchie IN, et al. Algorithms for digital image processing in diabetic retinopathy. *Comput Med Imaging Graph*. 2009;33:608-622.
- 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.
- Sidibé D, Sadek I, Mériaudeau F. Discrimination of retinal images containing bright lesions using sparse coded features and SVM. *Comput Biol Med*. 2015;62:175-184.
- Sadek I, Sidibé D, Meriaudeau F. Automatic discrimination of color retinal images using the bag of words approach. In: Hadjiiski LM, Tourassi GD, eds. Proc. SPIE 9414, Medical Imaging 2015: Computer-Aided Diagnosis, 94141J (March 20, 2015). <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2075824>; Accessed October 20, 2016.
- Veras R, Silva R, Araujo F, Medeiros F. SURF descriptor and pattern recognition techniques in automatic identification of pathological retinas. 2015 Brazilian Conference on Intelligent Systems (BRACIS). *IEEE*. 2015:316-321.
- LeCun Y, Kavukcuoglu K, Farabet C. *Convolutional networks and applications in vision*. ISCAS 2010—2010 IEEE International Symposium on Circuits and Systems. Nano-Bio Circuit Fabrics and Systems; 2010:253-256.
- Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.
- Kaggle, Inc. *Diabetic retinopathy detection*; 2015. <https://www.kaggle.com/c/diabetic-retinopathy-detection>. Accessed 1.12.2016.
- Cuadros J, Bresnick G, Affiliations A. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *J Diabetes Sci Technol*. 2009;3:509-516.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012:1-9.
- Karpathy A, Toderici G, Shetty S, et al. *Large-scale video classification with convolutional neural networks*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2014:1725-1732.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *ArxivOrg*. 2015;7:171-180. <http://arxiv.org/pdf/1512.03385v1.pdf>; Accessed October 20, 2016.
- Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *ICML*. 2015.
- Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. Proc 27th International Conference on Machine Learning. 2010:807-814.
- Dunne R, Campbell N. On the pairing of the Softmax activation and cross-entropy penalty functions and the derivation of the Softmax activation function. Proc 8th Aust Conf Neural Networks 1997:1-5. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.49.6403&rep=rep1&type=pdf>; 1997 Accessed October 20, 2016.
- Zhou B, Khosla A, Lapedriza A, et al. *Learning deep features for discriminative localization*. arXiv.org. 151204150 [cs]. 2015. 2015:2921-2929. <http://arxiv.org/abs/1512.04150/n>. <http://www.arxiv.org/pdf/1512.04150.pdf>. Accessed October 20, 2016.
- Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29:1189-1232.

26. Decencière E, Zhang X, Cazuguel G, et al. Feedback on a publicly distributed image database: the Messidor database. *Image Anal Stereol.* 2014;33:231-234.
27. Decencière E, Cazuguel G, Zhang X, et al. TeleOphta: machine learning and image processing methods for tele-ophthalmology. *IRBM.* 2013;34:196-203.
28. Quellec G, Lamard M, Josselin PM, et al. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Trans Med Imaging.* 2008;27:1230-1241.
29. Antal B, Hajdu A. An ensemble-based system for automatic screening of diabetic retinopathy. *Knowledge-Based Syst.* 2014;60:20-27.
30. Sánchez CI, Niemeijer M, Dumitrescu AV, et al. Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data. *Invest Ophthalmol Vis Sci.* 2011;52:4866-4871.
31. Seoud L, Hurtut T, Chelbi J, et al. Red lesion detection using dynamic shape features for diabetic retinopathy screening. *IEEE Trans Med Imaging.* 2016;35:1116-1126.
32. Roychowdhury S, Koozekanani DD, Parhi KK. DREAM: diabetic retinopathy analysis using machine learning. *IEEE J Biomed Heal Informatics.* 2014;18:1717-1728.
33. van der Maaten L. Accelerating t-SNE using tree-based algorithms. *J Mach Learn Res.* 2014;15:3221-3245.
34. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;304:649-656. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27898976>.

Footnotes and Financial Disclosures

Originally received: October 27, 2016.

Final revision: February 8, 2017.

Accepted: February 8, 2017.

Available online: ■■■.

Manuscript no. 2016-645.

¹ The Harker School, San Jose, California.

² Byers Eye Institute at Stanford, Stanford University School of Medicine, Palo Alto, California.

Financial Disclosure(s):

The author(s) have made the following disclosure(s): R.G.: Patent - (Patent Application Number: 62383333); Patent Filing date: September 2, 2016.

Author Contributions:

Conception and design: Gargeya, Leng

Analysis and interpretation: Gargeya, Leng

Data collection: Gargeya, Leng

Obtained funding: none

Overall responsibility: Gargeya, Leng

Abbreviations and Acronyms:

AUC = area under the receiver operating characteristic curve;

DR = diabetic retinopathy.

Correspondence:

Theodore Leng, MD, MS, Byers Eye Institute at Stanford, Stanford University School of Medicine, 2452 Watson Court, Palo Alto, CA 94303.

E-mail: tedleng@stanford.edu.