Automated Diagnosis of Diabetic Retinopathy Using Clinical Biomarkers, Optical Coherence Tomography, and Optical Coherence Tomography Angiography

HARPAL SINGH SANDHU, MOHAMMED ELMOGY, AHMED TAHER SHARAFELDEEN, MOHAMED ELSHARKAWY, NABILA EL-ADAWY, AHMED ELTANBOLY, AHMED SHALABY, ROBERT KEYNTON, AND AYMAN EL-BAZ

• PURPOSE: To determine if combining clinical, demographic, and imaging data improves automated diagnosis of nonproliferative diabetic retinopathy (NPDR).

• DESIGN: Cross-sectional imaging and machine learning study.

• METHODS: This was a retrospective study performed at a single academic medical center in the United States. Inclusion criteria were age >18 years and a diagnosis of diabetes mellitus (DM). Exclusion criteria were non-DR retinal disease and inability to image the macula. Optical coherence tomography (OCT) and OCT angiography (OCTA) were performed, and data on age, sex, hypertension, hyperlipidemia, and hemoglobin A1c were collected. Machine learning techniques were then applied. Multiple pathophysiologically important features were automatically extracted from each layer on OCT and each OCTA plexus and combined with clinical data in a random forest classifier to develop the system, whose results were compared to the clinical grading of NPDR, the gold standard.

• RESULTS: A total of 111 patients with DM II were included in the study, 36 with DM without DR, 53 with mild NPDR, and 22 with moderate NPDR. When OCT images alone were analyzed by the system, accuracy of diagnosis was 76%, sensitivity 85%, specificity 87%, and area under the curve (AUC) was 0.78. When OCT and OCTA data together were analyzed, accuracy was 92%, sensitivity 95%, specificity 98%, and AUC 0.92. When all data modalities were combined, the system achieved an accuracy of 96%, sensitivity 100%, specificity 94%, and AUC 0.96.

• CONCLUSIONS: Combining common clinical data points with OCT and OCTA data enhances the power of computer-aided diagnosis of NPDR. (Am J Ophthalmol 2020;216:201–206. © 2020 Published by Elsevier Inc.)

AJO.com Supplemental Material available at AJO.com.
Accepted for publication Jan 10, 2020.

From the Department of Ophthalmology and Visual Sciences, University of Louisville School of Medicine (H.S.S.), and the Department of Bioengineering, University of Louisville Speed School of Engineering (M.E., N.E.-A., A.E., A.S., R.K., A.E.-B.), Louisville, Kentucky, USA.

Inquiries to Harpal Singh Sandhu, Department of Ophthalmology and Visual Sciences, University of Louisville School of Medicine, 301 E. Muhammad Ali Blvd, Louisville, KY 40202, USA; e-mail: harpal.sandhu@louisville.edu

With an estimate prevalence of over 450 million, diabetes mellitus (DM) has reached epidemic proportions in the United States and worldwide.[1] Annual screening of diabetic patients presents a major public health problem both in the United States and around the world. The demand for annual screening and closer monitoring of those who develop diabetic retinopathy (DR) is taxing, particularly in resource-poor settings with less access to care. A second, unrelated problem in ophthalmology is the inherent subjectivity of optical coherence tomography (OCT) and OCT angiography (OCTA) interpretation. Objective metrics from these imaging modalities are limited, and interpretation thus suffers from some degree of subjectivity.

Machine learning is a branch of artificial intelligence in which programs learn to classify data, such as medical images, based on repeated exposure to labeled datasets. The applications to medical imaging are legion, and the field has exploded of late in ophthalmology, with particular foci on retinal imaging. However, image analysis and diagnosis are not the alpha and omega of machine learning in medicine. These techniques can be applied to a range of data types, including demographic and clinical data. Indeed, fundus images have already been used to predict clinical metrics like blood pressure or demographic ones like age with fairly high fidelity.[2] Moreover, these 3 types of data—imaging, clinical, and demographic—can also be combined to enhance automated diagnosis via machine learning techniques. Such an approach was undertaken herein.

The purpose of this study was to apply machine learning techniques to multiple modes of data, namely OCT, OCTA, and standard clinical and demographic data, in order to improve upon existing computer-aided diagnostic (CAD) systems that automate the diagnosis and grading of nonproliferative diabetic retinopathy (NPDR).

0002-9394/$36.00
https://doi.org/10.1016/j.ajo.2020.01.016

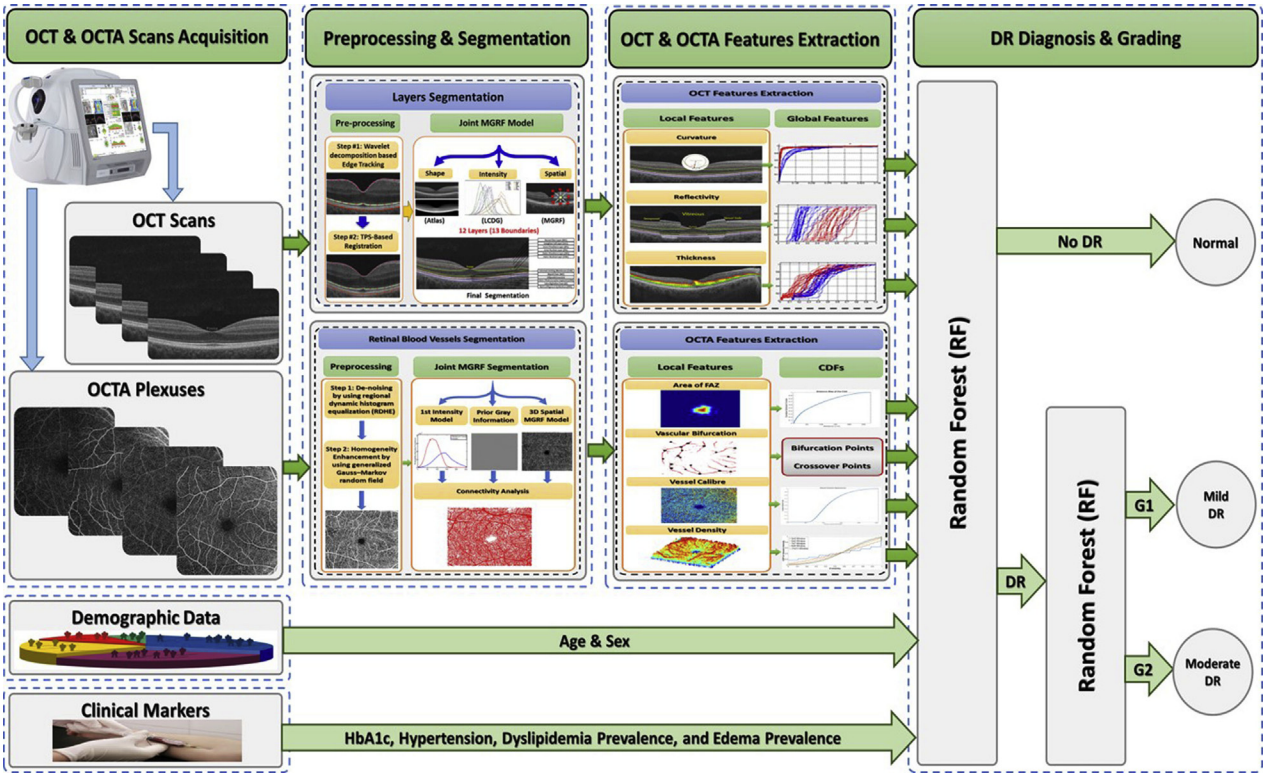

FIGURE 1. The pipeline of the computer-aided diagnostic system that incorporates clinical biomarkers, optical coherence tomography (OCT), and optical coherence tomography angiography (OCTA) data to diagnose and grade diabetic retinopathy (DR). OCT and OCTA images as well as demographic and clinical data are collected from patients in the clinic. The imaging data are then processed and segmented. Key OCT features (curvature, reflectivity, and thickness of the individual layers) and OCTA features (area of the foveal avascular zone, vessel density, vessel caliber, and number of bifurcation points) are extracted. These data are combined with clinical and demographic data by a random forest classifier to arrive at a diagnosis of no DR, mild nonproliferative DR, or moderate nonproliferative DR.

METHODS

THIS WAS A SINGLE-CENTER, RETROSPECTIVE, CROSS-sectional imaging and machine learning study. This study was conducted in accordance with the Declaration of Helsinki and was approved by the institutional review board of the University of Louisville. Data were collected between November 2016 and January 2019 and analyzed from February 1, 2019 to April 15, 2019. The need for signed informed consent documents was waived by the institutional review board because all data were collected retrospectively as part of routine clinical practice.

• **PATIENT SELECTION:** Inclusion criteria were age >18 years and a diagnosis of DM. Exclusion criteria were a diagnosis of a retinal disease other than DR, media opacity precluding imaging, high myopia, and proliferative diabetic retinopathy (PDR). PDR was excluded because of the low prevalence of non-high-risk PDR in our population with media sufficiently clear of vitreous hemorrhage to allow for high-quality OCTA imaging. Clinical and demographic data collected were age, sex, history of hypertension, history of hyperlipidemia, and hemoglobin A1c. Patients underwent a full dilated eye examination and OCT and OCTA imaging with the Zeiss Cirrus AngioPlex (Carl Zeiss AG, Oberkochen, Germany).

• **THE COMPUTER-AIDED DIAGNOSTIC SYSTEM:** Machine learning techniques were then applied. Three pathophysiologically important features were extracted from each layer on OCT: reflectivity, curvature, and thickness. For OCTA, 4 features were extracted: blood vessel caliber, vessel density, the size of the foveal avascular zone (FAZ), and the number of bifurcation and crossover points. Details of the previously validated segmentation process are available elsewhere.[3] The size of all OCTA images was 1024 × 1024 pixels, spanning a 6 mm × 6 mm square centered about the fovea. The OCT data were similarly 1024 × 1024 pixels, captured as raw grayscale files with 5 slices. Only the superficial vascular plexus and deep vascular plexus were analyzed (Figure). Established image processing methods, including regional dynamic histogram equalization and an adaptive gray level threshold estimation with a generalized Gauss-Markov random field model,

TABLE 1. Performance of the Computer-Aided Diagnostic System for Stage 1, Distinguishing Diabetic Retinopathy From No Diabetic Retinopathy, by Type of Imaging and/or Clinical Data Used

Features	ACC (%)	Sens (%)	Spec (%)	DSC (%)	AUC
OCT	86.5	92.3	84.8	76.2	0.885
OCTA	94.6	87.8	98.6	92.3	0.932
OCT + OCTA	95.5	100	93.7	92.5	0.968
All features	98.2	100	97.4	97.2	0.987

ACC = accuracy; AUC = area under the curve; DSC = dice similarity coefficient; OCT = optical coherence tomography; OCTA = optical coherence tomography angiography; Sens = sensitivity; Spec = specificity.

The first task of the system was to distinguish between diabetic retinopathy and no diabetic retinopathy. The system was tested 4 times, first with data from OCT images alone (OCT), second from OCTA images alone (OCTA), third from images of both modalities (OCT + OCTA), and finally with all imaging, clinical, and demographic data (all features). All features performed the best in all metrics.

TABLE 2. Performance of the Computer-Aided Diagnostic System for Stage 2, Grading of Diabetic Retinopathy for Images Previously Identified as Having Diabetic Retinopathy, by Type of Imaging and/or Clinical Data Used

Classifiers	ACC (%)	Sens (%)	Spec (%)	DSC (%)	AUC
OCT	88.3	92.5	86.9	79.3	0.897
OCTA	94.7	91.3	96.2	91.3	0.937
OCT + OCTA	97.3	95.4	98.1	95.4	0.967
All features	98.7	100	97.8	99.0	0.981

ACC = accuracy; AUC = area under the curve; DSC = dice similarity coefficient; OCT = optical coherence tomography; OCTA = optical coherence tomography angiography; Sens = sensitivity; Spec = specificity.

The second task of the system was to grade the level of nonproliferative diabetic retinopathy in those images identified as having diabetic retinopathy in stage 1. The system was again tested 4 times, first with data from OCT images alone (OCT), second from OCTA images alone (OCTA), third from images of both modalities (OCT + OCTA), and finally with all imaging, clinical, and demographic data (all features). All features performed the best in four out of five all metrics.

were applied to improve the signal-to-noise ratio of the OCTA images, detailed elsewhere.[4] All imaging features were automatically extracted.

Data from these 7 extracted features as well as the 5 clinical and demographic data points for each patient were then fed into a 2-stage random forest classifier to train the system. In the first step, the system classified the image as having DR. In the second step, images already classified as DR in step 1 were further classified into mild or moderate NPDR.

• **STATISTICAL ANALYSIS:** The results of the system were tested by 4-fold cross-validation and leave-1-subject-out (LOSO) validation. Although only 1 testing method is necessary to show that there has been no overfitting of the data, both methods were performed in order to address the effect of the size of the training data on system performance. These results were then compared to the clinical grading of DR, which was considered the gold standard. The accuracy, sensitivity, specificity, dice similarity coefficient (DSC), and area under the curve (AUC) of the system were calculated with use of OCT data alone, OCTA data alone, combined OCT and OCTA data, and finally combined OCT, OCTA, clinical, and demographic data.

RESULTS

THE PROPOSED CAD SYSTEM WAS TRAINED AND TESTED ON data collected from 111 subjects (48% female, age range 20-82 years). Thirty-six had DM without DR, 53 mild NPDR, and 22 moderate NPDR.

The first stage of the system was simply to classify images as demonstrating DR or no DR. In this first stage, the system was tested with 4 different sets of data inputs: OCT data alone, OCTA data alone, OCT and OCTA data combined, and OCT, OCTA, demographic, and clinical data combined. Combining all data produced the best results, with diagnostic accuracy of 97%-98% and an AUC of 0.981 by LOSO and 0.987 by 4-fold cross-validation (Table 1). AUCs for OCT data alone were approximately 0.89, for combined OCT and OCTA 0.968, and for OCT, OCT, and clinical and demographic data 0.987.

The second stage of the system was to grade the level of NPDR. No cases of severe NPDR were included in the dataset, so the 2 outputs were either mild or moderate NPDR. By both LOSO and 4-fold cross-validation, the system's accuracy for the grading stage was 98.7%, sensitivity 100%, specificity 97.8%, DSC 99%, and AUC 0.981 (Table 2). Again, a steady improvement in almost all metrics was observed when the system was given only OCT data, to OCT and OCTA data, to OCT, OCTA, clinical, and demographic data, with AUCs increasing from 0.897, to 0.967, to 0.981 (Table 2).

Of course, the final diagnosis involves performing both stage 1 and stage 2 of the system in sequence. When running the whole system on all data inputs, final accuracy was 96%, sensitivity 100%, specificity 94%, DSC 98%, and AUC 0.960 (Table 3). When the system was given OCT data only, AUC was 0.783, increasing to 0.921 with combined OCT and OCTA data, and finally to 0.960 when given all data modalities (OCT, OCTA, clinical, and demographic data).

The random forest classifier has the capability to scrutinize the effect of each feature on diagnostic accuracy and

TABLE 3. Overall Performance of the Computer-Aided Diagnostic System in Correctly Identifying and Grading Diabetic Retinopathy by Type of Imaging and/or Clinical Data Used

Classifiers	ACC (%)	Sens (%)	Spec (%)	DSC (%)	AUC
OCT	75.6	84.6	86.9	72.2	0.783
OCTA	88.3	79.2	94.1	83.7	0.865
OCT + OCTA	92.2	95.4	98.1	91.1	0.921
All features	96.0	100	94.1	98.0	0.960

ACC = accuracy; AUC = area under the curve; DSC = dice similarity coefficient; OCT = optical coherence tomography; OCTA = optical coherence tomography angiography; Sens = sensitivity; Spec = specificity.

The overall performance of the automated diagnostic system involves the combined results of stage 1 (diabetic retinopathy [DR] vs no DR) and stage 2 (grading of nonproliferative DR). As before, the system was tested with 4 different data sets: OCT images alone (OCT), OCTA images alone, both OCT and OCTA images (OCT + OCTA), and all imaging, clinical, and demographic data (all features). All features performed the best in 4 out of 5 metrics.

thereby assign relative weights or degree of importance to each feature. Performing such experiments allows one to understand the contribution that each of these features makes to diagnosis. The obtained feature importance, expressed as a contribution percentage, from the highest to lowest, was as follows: blood vessel caliber, 16.1%; vessel density, 15.9%; thickness, 15.8%; reflectivity, 14.6%; size of the FAZ, 13.2%; number of bifurcations, 12.6%; curvature, 11.7%.

Finally, the random forest classifier was also compared to other state-of-the-art classifiers (classification tree, support vector machines with linear kernel, and K-nearest neighbor), and it outperformed them in all outcome measures.

DISCUSSION

IN THIS STUDY, WE DESIGNED A NOVEL CAD SYSTEM FOR THE diagnosis and grading of NPDR that integrates imaging data from both OCT and OCTA with basic clinical and demographic data for 111 patients. The AUC of the final diagnosis was a mere 0.76 when analyzing structural OCT data alone, improved dramatically to 0.92 with the addition of OCT angiographic data, and improved incrementally still with the addition of clinical and demographic data to 0.96.

Two points stand out from the results of this study. First, OCT angiography imaging clearly adds significant value to an automated diagnostic system for DR. This result is intuitive; DR is primarily a disease of the retinal vasculature, and OCTA provides instructive information about the status of the vasculature that structural OCT does not. In our particular system, the size of the FAZ and density of capillaries within the macula are both known to have diagnostic value in diagnosing DR, consistent with the pathophysiology of the disease, driving capillary nonperfusion and eventually macular ischemia.[5,6] OCTA's ability to image both the deep and superficial vascular plexuses (both affected in DR), its noninvasive nature, ease of acquisition, and presence on the same imaging platforms as OCT are distinct advantages over fluorescein angiogram; thus, it is a natural complement to conventional OCT. The relative importance of each of the 7 features was also striking in 2 ways. First, vessel caliber and vessel density made the greatest contributions. This is consistent with the pathologic microvascular changes in diabetic retinopathy, which causes capillary nonperfusion and frequently a degree of arteriolar attenuation, culminating in ghost vessels. Second, the range in contribution percentage was very narrow, spanning only 4.4%. Thus, all of these features made a significant contribution and added to the accuracy of the system.

Second, the addition of simple clinical and demographic data also had added value for the system, improving the first-stage, second-stage, and overall performance by 2%-4% of AUC. Systemic hypertension and hemoglobin A1c are perhaps the oldest and most reliable predictors of DR onset and progression.[7,8] While these risk factors are well known, what was unclear was to what extent, if any, this additional information might add diagnostic value. If the local effects of poor blood glucose and blood pressure control are already indirectly captured by OCTA imaging of the retinal vasculature, for instance, one would not expect these data to add any significant value. Indeed, providing inputs of the patient's age, sex, last hemoglobin A1c, and history of systemic hypertension and/or hyperlipidemia provided a small but appreciable improvement in diagnostic performance. While a 4% improvement in overall accuracy and AUC might not appear important *prima facie*, a diagnostic system of clinical significance would ideally have an AUC in the high 90s. With an improvement in AUC from 0.92 to 0.96, the addition of clinical and demographic data helps us move closer to that goal.

While many studies have applied machine learning, including deep learning, to fundus photographs to diagnose DR, many fewer have focused on OCT and OCTA. Multiple groups have applied machine learning to OCT to identify macular edema of various etiologies.[9–12] Automated systems for diagnosis of exudative age-related macular degeneration and geographic atrophy on OCT have also shown good results.[13,14] One particularly impressive deep learning system trained on over 14,000 OCT images in the United Kingdom can provide probabilities of 10 different common OCT diagnoses and produced a correct referral recommendation, the primary outcome, 95% of the time.[15] However, it also simply identified macular

edema rather than diagnosing a particular etiology of edema, and DR itself was not 1 of the 10 output diagnoses.

Despite tremendous advances in machine learning for ophthalmic imaging, automated systems that diagnose and grade DR via OCT are at present very rare. Kuwayama and associates[16] reported a deep learning system to diagnose multiple different macular diseases by OCT, including DR. However, they explicitly tested their system on only 7 DR cases, and grading was not performed. To the best of our knowledge, this is the first study to diagnose NPDR using simultaneous OCT and OCTA data.

This study's greatest limitation is that it deals with only lower grades of nonproliferative diabetic retinopathy, namely no DR, mild NPDR, and moderate NPDR. Any clinically meaningful system with immediate real-world application must be able to reliably diagnose severe NPDR and PDR as well. Clearly, this system is not yet ready for clinical use. Insufficient numbers of severe NPDR patients and PDR patients with sufficiently clear media for OCT and OCTA imaging were available in the existing database to validate the system on those higher grades of retinopathy. Collection of more data and validation of the system on severe NPDR and PDR are critical next steps in this work. Presently, the point of this study is proof of concept that combining multiple modes of data is complementary and additive in improving machine diagnosis.

In summation, automated diagnosis of NPDR using a combination of OCT, OCTA, and clinical and demographic data is feasible and highly accurate. Combining easily collectable clinical data points with imaging data enhances the power of computer-aided diagnosis and could be considered as additional inputs into existing machine learning and deep learning systems. As OCT technology becomes more accessible and ubiquitous, these methods could potentially be applied to screening.

FUNDING/SUPPORT: THIS RESEARCH PROJECT WAS SUPPORTED IN PART BY AL JALILA FOUNDATION, UNITED ARAB EMIRATES (grant number AJF2018053). Financial Disclosures: Harpal Singh Sandhu, Mohammed Elmogy, Nabila El-Adawy, Ahmed Eltanboly, Ahmed Shalaby, Robert Keynton, and Ayman El-Baz hold intellectual property in 1 of the segmentation algorithms employed by this system. Ayman El-Baz also holds intellectual property in computer-aided diagnostic systems for malignant lung nodules on computed tomography (CT) scan, for autism based on magnetic resonance imaging (MRI), for mild cognitive impairment based on MRI, for early detection of prostate cancer based on CT scan or MRI, for early diagnosis of kidney allograft rejection based on CT or MRI, for detection of systemic hypertension based on MRI, and in computerized methods for controlling the speed of cardiac ventricular assist devices. Robert Keynton holds intellectual property in capillary electrophoresis systems, microfluidic devices, neurostimulators, self-assembly of nanoparticles, and, along with Ayman El-Baz, in computer-aided diagnosis of autism on brain MRI. Robert Keynton also holds equity in 2 biomedical companies: PromiSight, which has developed a hydrogel to prevent vitrectomy-induced cataracts, and CoulSense, which has developed analytical instrumentation to provide remote, autonomous, calibration-free heavy metal sensing capabilities. All authors attest that they meet the current ICMJE criteria for authorship.

Other Acknowledgments: The authors would like to acknowledge post-doctoral researchers Ahmed Taher Sharafeldeen, PhD, and Mohamed Elsharkawy, PhD, for performing additional experiments on the random forest classifier in order to determine relative feature importance.

REFERENCES

1. Cho NH, Shaw JE, Karuranga S, et al. IDF diabetes atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract* 2018;138:271–281.
2. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018;2(3):158–164.
3. Sandhu HS, Eladawi N, Elmogy N, et al. Automated diabetic retinopathy detection using optical coherence tomography angiography: a pilot study. *Br J Ophthalmol* 2018;102(11):1564–1569.
4. Eltanboly A, Ismail M, Shalaby A, et al. A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images. *Med Phys* 2017;44(3):914–923.
5. Lu Y, Simonett JM, Wang J, et al. Evaluation of automatically quantified avascular zone metrics for diagnosis of diabetic retinopathy using optical coherence tomography angiography. Invest Ophthalmol Vis Sci 2018;59(6):2212–2221.
6. Samara WA, Shahlaee A, Adam MK, et al. Quantification of diabetic macular ischemia using optical coherence tomography angiography and its relationship with visual acuity. *Ophthalmology* 2017;124(2):235–244.
7. Diabetes Control and Complications Trial Research Group. The relationship of glycemic exposure to the risk of development and progression of retinopathy in the Diabetes Control and Complications Trial. *Diabetes* 1995;44(8):968–983.
8. Stratton IM, Kohner EM, Aldington SJ, et al. UKPDS 50: risk factors for incidence and progression of retinopathy in type II diabetes over 6 years from diagnosis. *Diabetologia* 2001;44(2):156–163.
9. Samagio G, Estevez A, Moura J, et al. Automatic macular edema identification and characterization using OCT images. *Comput Methods Programs Biomed* 2018;163:47–63.
10. Alsaih K, Lemaitre G, Rastgoo M, et al. Machine learning techniques for diabetic macular edema classification on SD-OCT images. *Biomed Eng Online* 2017;16(1):68.
11. Syed AM, Hassan T, Akram MU, et al. Automated diagnosis of macular edema and central serous retinopathy through robust reconstruction of 3D retinal surfaces. *Comput Methods Programs Biomed* 2016;137:1–10.
12. Li F, Chen H, Liu Z, et al. Fully automated detection of retinal disorders by image-based deep learning. *Graefes Arch Clin Exp Ophthalmol* 2019;257(3):495–505.
13. Treder M, Lauermann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral

domain optical coherence tomography using deep learning. *Graefes Arch Clin Exp Ophthalmol* 2018;256(2):259–265.
14. Treder M, Lauermann JL, Eter N. Deep learning-based detection and classification of geographic atrophy using a deep convolutional neural network classifier. *Graefes Arch Clin Exp Ophthalmol* 2018;256(11):2053–2060.
15. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24(9):1342–1350.
16. Kuwayama S, Ayatsuka Y, Yanagisono D, et al. Automated detection of macular diseases by optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images. *J Ophthalmol* 2019;6319581.