AUTO IDENTIFICATION OF DIABETIC RETINOPATHY IN DEEP LEARNING

Kunda Prashanth Kumar[1]., Nelluri Navya Sree[2]., Gade Vaishnavi[3].,Yallu Chandana Priya[4].,Ziya Samreen[5]

1 Assistant Professor, Department Of CSE., Malla Reddy College Of Engineering For Women., Maisammaguda.,

Medchal.,Ts, India (✉kundaas.web@gmail.com)

2, 3, 4, 5 B.Tech CSE, (17RG1A05G2, 19RG1A05E3, 19RG1A05H9, 19RG1A05J0),

Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India

Abstract

As the prevalence of diabetes increases, millions of people need to be screened for diabetic retinopathy (DR). Remarkable advances in technology have made it possible to use artificial intelligence to screen DR from retinal images with high accuracy and reliability, resulting in reducing human labor by processing large amounts of data in a shorter time. We developed a fully automated classification algorithm to diagnose DR and identify referable status using optical coherence tomography angiography (OCTA) images with convolutional neural network (CNN) model and verified its feasibility by comparing its performance with that of conventional machine learning model. Ground truths for classifications were made based on ultra-widefield fluorescein angiography to increase the accuracy of data annotation. The proposed CNN classifier achieved an accuracy of 91–98%, a sensitivity of 86–97%, a specificity of 94–99%, and an area under the curve of 0.919–0.976. In the external validation, overall similar performances were also achieved. The results were similar regardless of the size and depth of the OCTA images, indicating that DR could be satisfactorily classified even with images comprising narrow area of the macular region and a single image slab of retina. The CNN-based classification using OCTA is expected to create a novel diagnostic workflow for DR detection and referral.

Introduction

About 80% of patients with diabetes develop DR within 20 years of diagnosis1, 2, and DR is a leading cause of blindness worldwide. The onset of potentially irreversible vision loss due to DR complications treatment with laser therapy or intravitreal injection of anti-vascular endothelial growth factor or steroids4-7 may postpone or halt progression of the disease. Timely diagnosis is particularly important when the condition has progressed to the point that treatment is necessary. Despite the gravity of this issue and the sharp rise in the incidence of diabetes, treating diabetic retinopathy is a laborious and costly process8, as it necessitates the participation of specialists and highly trained ophthalmologists classification of DR severity and early illness identification is typically subjective as well, depending on the expert clinical interpretation of the data9,10. Several machine learning and deep learning-based automated solutions were proposed to deal with this issue. However, models trained using fundus images can be limited because vascular features deep in the retina are not represented in two-dimensional images11-15, and most studies have applied these learning-based methods to fundus photographs to achieve high performance for various DR classification tasks. Subsequently, automated methods based on optical coherence tomography (OCT) and OCT angiography (OCTA) for DR classification were proposed and also showed its feasibility16–24. However, DR classification method that applies deep learning using only OCTA data, not the combined data of OCT and OCTA, has not been published yet.Moreover, classification models presented to date have been trained and tested based on incomplete ground truths created on the basis of traditional fundus images that rely only on the limited angle of view without taking account retinal lesions that may be present outside the imaged area, making it difficult to evaluate its performance. As the vascular alterations caused bydiabetes are widely distributed, more than 50% of DR lesions are known to be located outside seven-standard Early Treatment Diabetic Retinopathy Study (ETDRS) fields25,26.

	Accuracy, % (SD)		Sensitivity, % (SD)		Specificity, % (SD)		AUC (SD)	
	3 × 3 mm²	6 × 6 mm²	3 × 3 mm²	6 × 6 mm²	3 × 3 mm²	6 × 6 mm²	3 × 3 mm²	6 × 6 mm²
CNN-based classifier								
Superficial capillary plexus layer	92.9 (4.8)	94.2 (3.6)	95.0 (2.8)	95.6 (1.8)	88.8 (11.2)	91.3 (9.0)	0.950 (0.038)	0.949 (0.031)
Deep capillary plexus layer	94.2 (1.9)	91.3 (5.3)	97.5 (4.8)	95.0 (4.5)	87.5 (13.4)	83.8 (10.3)	0.966 (0.047)	0.937 (0.047)
Full-thickness retina layer	90.4 (2.5)	90.4 (2.5)	93.1 (7.4)	90.6 (3.7)	85.0 (20.3)	90.0 (8.3)	0.946 (0.021)	0.928 (0.017)
All	93.8 (3.2)	95.4 (3.4)	96.3 (4.1)	96.9 (2.1)	88.8 (7.4)	92.5 (9.0)	0.960 (0.009)	0.967 (0.029)
Machine learning-based classifier								
All	77.5 (2.5)	77.5 (2.5)	89.4 (6.0)	86.9 (5.7)	53.8 (6.5)	58.8 (8.9)	0.713 (0.023)	0.742 (0.022)

Table 1. Performance of classification task for detecting diabetic retinopathy using the proposed deep learning- and machine learning-based classifiers. AUC area under the curve, CNN convolutional neural network, SD standard deviation.

Previous studies reported that the presence of peripheral retinal lesions may suggest increased DR severity in9 to 15% of eyes26–29.We present an end-to-end deep convolutional neural network (CNN)-based method to detect DR and identify referable status (severe non-proliferative DR or worse) automatically from the OCTA images with more accurate,ground truths using ultra-widefield (UWF) fluorescein angiography (FA). Further, we investigated the ability of the algorithm across different image sizes and retinal slabs to identify the most appropriate mode of OCTA imageacquisition for DR classification. Moreover, we confirmed the feasibility of the proposed model by quantitative comparison of model performance against a machine learning-based classifier that uses handcrafted features extracted from OCTA images.

Results

Eyes from 301 participants were photographed; 51 were considered healthy controls, 51 had been diagnosed with diabetic mellitus (DM) but showed no signs of DR, 53 had mild non-proliferative DR, 49 had moderate non-proliferative DR, 48 had severe non-proliferative DR, and 49 were considered to have proliferative DR. After removing low-quality scans, we were left with a total of a total of 240 datasets were collected, consisting of 40 datasets for each DR stage, for both 3 mm2 and 6 mm2 scan pictures. The external validation included the recruitment and imaging of an additional 195 eyes (33 normal, 36 with DM but no DR, 34 with mild DR, 31 with moderate DR, 31 with severe DR, and 30 with proliferative DR). After excluding low-quality scans, we were left with a total of 120 datasets, 20 for each DR stage and covering both 3x3 and 6x6 mm scans. Our CNN-based classifier trained on 3 3 mm2 OCTA images had an area under the curves (AUC) of 0.950 (with 95% sensitivity and 89% specificity), 0.966 (with 98% sensitivity and 88% specificity), 0.946 (with 93% sensitivity and 85% specificity), and 0.960 (with 96% sensitivity and 89% specificity) for only superficial capillary plexus (SCP), only deep capillary plexus (D Using 6 mm x 6 mm OCTA images, the CNN-based classifier achieved AUCs of 0.949 (96% sensitivity and 91% specificity), 0.937 (95% sensitivity and 84% specificity), 0.928 (91% sensitivity and 90% specificity), and 0.967 (97% sensitivity and 93% specificity) for just SCP, just DCP, just full-thickness retina, and the combined data, respectively. In contrast, the machine learning-based classifier that relied on local characteristics derived from OCTA pictures had AUCs of 0.713 (with 89% sensitivity and 54% specificity) for the 3 3 mm2 images and 0.742 (with 87% sensitivity and 59% specificity) for the 6 6 mm2 OCTA images (Table 1, Fig. 1). Supplemental Fig. S1 depicts the confusion matrix between the ground-truth labels and the predictions of the proposed technique. The CNN-based classifier maintained a similar level of performance on the external dataset, with an AUC of 0.02 to 0.04, a sensitivity of 3% to 8%, and a specificity of 2.5% to 8% for 3 x 3 mm2 OCTA images, and an AUC of 0 to + 0.02, a sensitivity of 0% to 2%, and a specificity of 1% to 9% (Table 2). The AUC of the CNN-based classifier trained on 3 3 mm2 OCTA images for the classification task of diagnosing referable DR was 0.919 (86% sensitivity, 93% specificity), 0.967 (91% sensitivity, 98% specificity), 0.942 (88% sensitivity, 97% specificity), and 0.940 (89% sensitivity, 97% specificity) for only SCP, only DCP, only full-thickness retina, and the combined data, respectively. The 6 mm x 6 mm OCTA images used to train the CNN-based classifier had AUCs of 0.975 (95% sensitivity, 98% specificity), 0.975 (94% sensitivity, 99% specificity), 0.970 (95% sensitivity, 99% specificity), and 0.976 (96% sensitivity, 98%) for SCP-only, DCP-only, full-thickness retina-only, and combined data, respectively. In contrast, Table 3 and Figure 1 show that AUCs for machine learning-based classification utilizing local features derived from OCTA pictures ranged from 0.795 (with 66% sensitivity and 86% specificity) to 0.837 (76% sensitivity and 88%

specificity). Supplemental Fig. S2 depicts the confusion matrix between the ground-truth labels and the predictions of the proposed technique. Comparable results were achieved using the CNN-based classifier on the external dataset. − 0 to − 0.05 AUC, − 3%

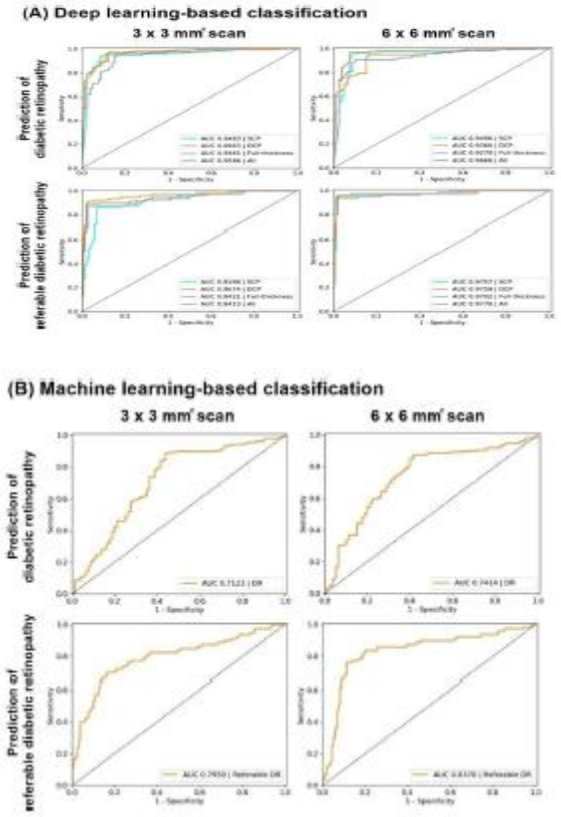


Figure 1. Receiver operating characteristic (ROC) curves illustrating classification performances for the prediction of onset and referable status of diabetic retinopathy (DR): (A) ROC curve using a deep learningbased classifier, and (B) ROC curve using a machine learning-based classifier. The dotted line represents the tradeoff resulting from random chance, and the solid curved line represents the tradeoff of the model.

	Accuracy, %		Sensitivity, %		Specificity, %		AUC	
	3×3 mm²	6×6 mm²	3×3 mm²	6×6 mm²	3×3 mm²	6×6 mm²	3×3 mm²	6×6 mm²
CNN-based classifier								
Superficial capillary plexus layer	90.8	93.3	92.5	93.8	87.5	92.5	0.931	0.949
Deep capillary plexus layer	89.2	94.2	91.3	95.0	85.0	92.5	0.918	0.956
Full-thickness retina layer	87.5	90.8	87.5	88.8	87.5	95.0	0.903	0.928
All	90.8	95.0	88.8	95.0	95.0	95.0	0.928	0.962
Machine learning-based classifier								
All	70.8	70.0	73.8	88.8	65.0	32.5	0.771	0.798

Table 2. Performance of classification task on external dataset for detecting diabetic retinopathy using the proposed deep learning- and machine learning-based classifiers. AUC area under the curve, CNN convolutional neural network, SD standard deviation.

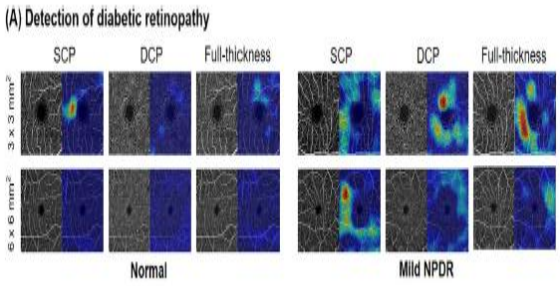
	Accuracy, % (SD)		Sensitivity, % (SD)		Specificity, % (SD)		AUC (SD)	
	3×3 mm²	6×6 mm²	3×3 mm²	6×6 mm²	3×3 mm²	6×6 mm²	3×3 mm²	6×6 mm²
CNN-based classifier								
Superficial capillary plexus layer	90.8 (4.9)	97.1 (2.6)	86.3 (6.5)	95.0 (10.6)	93.1 (5.4)	98.1 (4.7)	0.919 (0.054)	0.975 (0.024)
Deep capillary plexus layer	95.4 (1.9)	97.1 (4.1)	91.3 (5.4)	93.8 (9.4)	97.5 (4.8)	98.8 (6.5)	0.967 (0.005)	0.975 (0.020)
Full-thickness retina layer	94.2 (2.2)	97.5 (3.0)	87.5 (2.5)	95.0 (7.4)	97.5 (3.7)	98.8 (1.8)	0.942 (0.029)	0.970 (0.018)
All	94.2 (2.7)	97.5 (2.6)	88.8 (7.5)	96.3 (10.6)	96.9 (5.7)	98.1 (2.8)	0.940 (0.022)	0.976 (0.021)
Machine learning-based classifier								
All	79.6 (2.7)	84.2 (4.2)	66.3 (8.9)	76.3 (15.6)	86.3 (7.2)	88.2 (3.7)	0.795 (0.031)	0.837 (0.092)

Table 3. Performance of classification task for detecting referable status of diabetic retinopathy using the proposed deep learning- and machine learning-based classifiers. AUC area under the curve, CNN convolutional neural network, SD standard deviation.

to + 9% sensitivity, and − 2% to + 5% specificity for 3 × 3 mm2 OCTA images, and − 0.04 to − 0.06 AUC, − 1% to − 7% sensitivity, and − 0% to − 2% specificity for 6 × 6 mm2 OCTA images (Table 4). Figure 2 illustrates examples of OCTA images with the class activation maps (CAM) obtained using the proposed model for detecting DR and referable DR cases. In the case of no DR, the whole image was weakly activated, with the exception of the foveal avascular zone (FAZ) and the region around the large blood vessel

	Accuracy, %		Sensitivity, %		Specificity, %		AUC	
	3×3 mm²	6×6 mm²	3×3 mm²	6×6 mm²	3×3 mm²	6×6 mm²	3×3 mm²	6×6 mm²
CNN-based classifier								
Superficial capillary plexus layer	94.2	94.2	85.0	87.5	98.8	97.5	0.907	0.909
Deep capillary plexus layer	95.8	95.0	95.0	87.5	96.3	98.8	0.938	0.921
Full-thickness retina layer	93.3	94.2	85.0	87.5	97.5	97.5	0.895	0.907
All	95.8	95.8	97.5	95.0	95.0	96.3	0.940	0.938
Machine learning-based classifier								
All	80.0	77.5	82.5	72.5	78.8	80.2	0.733	0.682

Table 4. Performance of classification task on external dataset for detecting referable status of diabetic retinopathy using the proposed deep learning- and machine learning-based classifiers. AUC area under the curve, CNN convolutional neural network, SD standard deviation.



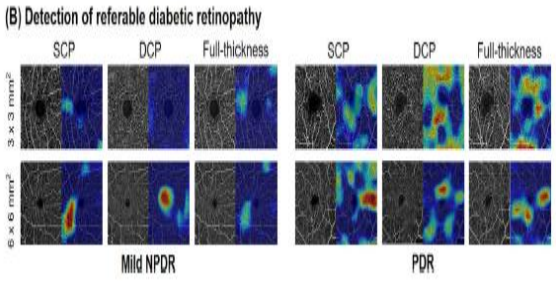


Figure 2. Representative Class activation map (CAM) images highlighting pathologic regions, shown in pairs of original optical coherence tomography angiography (OCTA) images (left) and corresponding CAM overlaid on the original images (right) of representative cases. CAM images were generated using the superficial capillary plexus, deep capillary plexus, and full-thickness retinal slabs from 3 × 3 mm2 to 6 × 6 mm2 macular OCTA scans. (A) CAM images of normal (left) and mild non-proliferative DR (right) cases for DR detection, accompanied by the corresponding original image. (B) CAM images of mild non-proliferative DR (left) and proliferative DR (right) for referable DR detection, accompanied by the corresponding original image

.being occasionally activated. On the other hand, in the case of referable DR, the overall area around the FAZ and vicinity of the large blood vessels were strongly activated. The activation increased in regions where the density of blood vessels significantly changed compared to regions of even spread, which is related to the non-uniformity in the blood vessel region. As the activation map visualizes areas where the network is used for decision making, rather than the specific abnormal area, the activated area was not the same for the 3 × 3 mm2 images and 6 × 6 mm2 images.

Discussion

In this investigation, we developed a comprehensive OCTA-based deep learning-based classification system for DR and referable DR diagnoses utilizing UWF FA-determined ground realities. Accuracy between 90% and 95%, sensitivity between 91% and 98%, and specificity between 85% and 100% were all achieved using the suggested CNN classifier. Accuracy ranged from 91 to 98%, sensitivity from 86 to 96%, specificity from 93 to 99%, and area under the curve (AUC) from 0.94 to 0.98 were reported for identifying the beginning of DR; for referable DR, these figures were 86 to 96%, 93%, 93%, and 99%, respectively. Similar results were obtained in the external validation. As a result, the suggested classifier consistently demonstrated dependable performance across all OCTA slabs (both separately and collectively) and throughout all examined scan sizes. Recently, scientists have been trying to find ways to automatically identify and categorize retinal disorders. For the diagnosis of DR stages in the training set for automated models, however, past research relied on either medical records or fundus pictures with very restricted viewing angles. Due to its subjectivity, this method cannot accurately reflect the widespread retinal changes brought on by diabetes. While it is possible to train models using large amounts of data using traditional fundus photographs, there is a risk of misclassification of DR. So, it's hard to say if these models will actually work well with other gadgets in a medical setting. One of the major benefits of this research is that we were able to accurately diagnose DR in all instances by reflecting diabetic alterations over the whole retina using UWF FA. Using less than 250 samples, the deep CNN algorithm in this study with OCTA achieved accuracy on par with previous CNN-based DR grading algorithms. Through the use of OCTA, pathological vascular changes associated with diabetes can be visualized and analyzed quantitatively across multiple retinal layers. Since DR is primarily a disease of the retinal vasculature, optical coherence tomography angiography (OCTA) may be a more appropriate imaging modality for the automated classification of DR than fundus photographs. Previous deep learning classification studies with OCT and OCTA have shown that the detailed information of the macular region extracted through OCT/OCTA images is sufficient to diagnose DR at a level similar to fundus photographs23,24, so these findings are consistent with those of those studies. UWF FA provides reliable ground truth, hence further research is needed to directly compare the outcomes of deep learning algorithms trained on OCTA with those trained on standard fundus photography. Since DR can cause extensive and full-depth damages to the retinal microvasculature, we looked into the ability of the CNN algorithm to determine which images are best suited for DR classification across a range of image sizes and retinal slabs. The diagnostic performance of DR evaluation based on OCTA scan size or slab depth utilizing quantitative microvascular metrics of OCTA has been studied in a number of cross-sectional investigations, however the findings are still up for debate30–39. Results may have been influenced by image collection settings such as oversampling density, filters, and algorithms employed to quantify vascular metrics40,41, since those investigations relied on handcrafted feature extraction for DR characterization.

Overfitting on the original samples might make it hard to generalize to new datasets42,43, even while techniques utilizing handmade features may perform

well on specific OCTA datasets. To counter this, a fully automated CNN algorithm can process heterogeneous images quickly regardless of the size and slab for accurate and objective DR classification, potentially reducing the need for resource-intensive manual analysis and helping to direct high-risk patients toward further treatment. OCTA contains more unlabeled information. Regardless of the resolution or depth of the OCTA pictures, we found that the findings were consistent across all sizes. Our findings demonstrated that DR, including macular pathology, could be identified successfully using just a 3 mm2 macular region and a single slab of SCP or DCP pictures. Foveal avascular zone (FAZ) area, blood vessel density, bone vessel density, and/or fractal dimension are only few of the local variables that are reflected in the emphasized areas of the CAM pictures. Without DR, the whole picture was faintly active, but in the presence of DR, activation was limited to the FAZ and the area surrounding the major blood artery. The area surrounding the FAZ and close to the major blood arteries, on the other hand, were substantially active in cases with referable DR. In areas where there were large variations in the density of blood vessels, CAM activation was higher than in areas where the number of blood vessels was uniformly distributed. This led us to postulate that the density of blood vessels and the foveal avascular zone (FAZ) were critical in CNN's categorization of DR. In this research, we employed CAM to verify the prediction process; nonetheless, its potential utility in clinical practice remains an open question. This fully-convolutional neural network (CNN) classifier outperformed a machine learning classifier trained on the extracted local features following vessel and FAZ segmentation. This further demonstrates that the commonly-used human criteria in this work are inadequate for DR characterisation, and that the necessary key characteristics may be recovered efficiently by means of end-to-end deep learning. The machine-learning model in this research performed worse than those in earlier studies16-18, which is not surprising given the wide variety of experimental conditions used. There may be limits to how much the aforementioned enhancements may help the machine-learning model's performance. Since deep learning makes use of unlabeled information to achieve the highest accuracy, it can do well even with a limited training dataset. Despite our similar performance findings, this study is limited by the still-small sample size of employed patients. Since OCTA is still not widely used in ophthalmology, the sample size of this study is similar to others that have used it (16-18 patients). Furthermore, normal subjects' lack of FA may undermine confidence in the ground truth label as a whole. Due to ethical concerns, FA could not be conducted on the healthy patients. However, a thorough medical history was taken to rule out any underlying systemic or ocular conditions. Finally, although we conducted external validation using additional data, the system still needs to be proven on a fully independent, larger, de novo set of images that also includes images with macular edema, artifacts, or low-quality before it can be used in actual clinical practice. Nonetheless, the findings of this study provide crucial evidence for the development of comprehensive deep learning models for DR classification from OCTA images. UWF imaging includes diabetic alterations from the whole retina, covering more area than traditional FA and fundus photography, hence this research has the added benefit of providing the ground truth for categorization of DR stages based on UWF FA. Next, we must clarify how well the deep learning algorithm can use OCTA to distinguish between eyes with referrable DR and eyes with peripheral dominant lesions or PDR. Using only OCTA images, we developed a fully automated deep CNN DR classification method. OCTA is quickly becoming a standard part of clinical practice, yet there are still gaps in our ability to interpret the data we collect. Even in cases when making a diagnosis or referral is challenging, invasive FA may be avoided if OCTA provided diagnostic value on par with UWF FA. Automatic categorization framework based on optical coherence tomography allows for more precise DR screening than traditional fundus photography. When used in a clinical setting, this technology has the potential to significantly cut down on vision loss caused by DR, enhance clinical management, and provide a brand-new diagnostic procedure for the early diagnosis and referral of illness. Further testing and improvement of the sensitivity metrics, such as genetic variables, hemoglobin A1C, length of diabetes, and other clinical data, may be necessary to achieve a low falsenegative rate for optimal practical use of our technique. It is possible to further enhance accuracy by combining data from several imaging modalities, such as FA and fundus photography. Detailed severity levels of DR should be classified with a larger patient population, which will require expanding the algorithm in future studies.

Methods

Dataset. This cross-sectional study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (IRB) of Yeungnam University Medical Center (approval number: 2020-02-003). The requirement for written

consent was waived by the IRB because of the retrospective naturl of the study. Data were collected between January 2018 and January 2019. Data between February 2019 and January 2020 were additionally collected to validate our method on external dataset. Subjects who had visited the hospital for visual floater and ocular discomfort, and who had undergone detailed examination including OCTA (Optovue RTVue XR AVANTI, Optovue Inc., Fremont, CA, USA), but had no systemic disease or ocular disease were retrospectively included. Subjects who had previously been diagnosed with diabetes mellitus (DM) and undergone comprehensive ophthalmic examinations including UWF FA (Optos California, Optos plc, Dunfermline, UK) and OCTA were also included. UWF FA was performed limitedly after explaining possible side effects if the patient desires a full-examination in spite of absence of DR.A small number of diabetic patients without DR wanted FA examination, accounting for less than 10% of those diabetic patients without DR who visited our clinic during the year. Only diabetic patients underwent fluorescein angiography, but not the healthy control participants. As the normal control group, patients without systemic disease who had undergone several ophthalmic examinations including mydriatic examination and OCTA for health-screening purposes, but had no definite ocular diseases were included. In the case of diabetes without retinopathy, only cases where FA was performed were included. This is retrospective study, hence, no patients received FA to participate this study. Indication of FA is not related with the protocol of this study. Exclusion criteria included the presence of glaucoma or retinal disorders affecting retinal capillary changes other than DR. Eyes with macular edema were excluded because it can obscure retinal microvasculature on OCTA. Images with low signal strength (≤ 6), excessive motion artifacts, and projection artifacts caused by media opacities were also excluded. OCTA images were obtained as volume scans of 3 × 3 mm2 and 6 × 6 mm2 sizes centered on the macula, and images of the SCP, DCP, and full-thickness retina slab were used for analysis. The ground truth for determining the accuracy of each diagnosis and grading DR was determined by two masked expert retinal specialists (G.H. and D.P.) reviewing all phases of central/axial UWF FA images, which were recorded up to 15 min after dye injection. Grading was performed based on the International Clinical DR Severity Scale44, which was adapted by means of extending the grading quadrants to the periphery of the entire image while maintaining the original grading nomenclature for simplicity29,45. When there was a disagreement between the graders, the supervising grader (M.S.) confirmed the final decision.

Convolutional neural network‑based classifier using raw images from OCTA .

The overall structure of the proposed method for detecting early signs of DR and referable DR is shown in Fig. 3. SCP, DCP, and full-retina OCTA images were concatenated and used as input of the CNN. Using the ResNet101 model46, images were passed through residual blocks with 101 layers, which repeatedly performed a summation of the input and output feature maps from the convolution layers each with batch normalization, rectified linear unit (ReLU) activation functions, and max pooling. After the residual blocks, each feature map was averaged in the global average pooling (GAP) layer with the probability of each stage obtained through a fully connected layer with a softmax function. CAM were derived from the GAP layer by summating the feature maps with the weights from the last layer in order to visualize regions that show high correlation with the task of interest. It is worth nothing that parameters of the network were initially transferred from the pre-trained parameters of the ImageNet dataset, excluding the first and last layer parameters. Subsequently, all parameters were retrained using our OCTA dataset, which was optimized based on the cross-entropy loss with an Adam optimizer and a learning rate of 0.000147.

Machine learning‑based classifier using local features extracted from OCTA images.

The machine learning-based classifier consisted of three stages: segmentation, feature-extraction, and classification stages (Fig. 4). In the segmentation step, U-Net48 was used to segment the blood vessels and FAZ from the OCTA images. The combined data from each layer of the OCTA image was used as input, given that prior machine learning based studies demonstrated that the best DR classification results were obtained when local features from the combined data was used i.e. both SCP and DCP16–18. In the U-Net model, a contracting path extracts high-level features from the input images by repeatedly using convolution layers, batch normalization, ReLU activation function, and max pooling; while the expanding path generates a segmentation map of the same size as the input image by repeatedly using upsampling, convolution layers, batch normalization, and ReLU activation functions on the

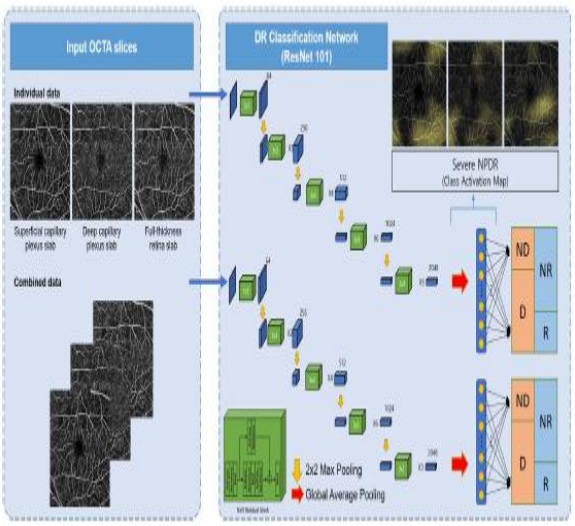


Figure 3. Convolutional neural network-based classifier for DR. This end-to-end classification task was performed to detect DR (D; mild non-proliferative DR or worse)

and identify referable status (R; severe mild non-proliferative DR or worse) from OCTA images using a ResNet model. The OCTA images are passed through 16 residual blocks, while increasing the number of filters successively. After the residual blocks, each feature map is averaged in the global average pooling layer, and the probability for each stage is then obtained through a fully connected layer with softmax function. CAMs are derived from the global average pooling layer in the last part of the network to visualize the regions that significantly affect the tasks.

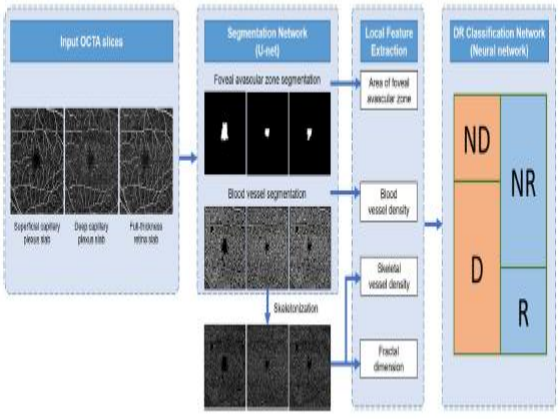


Figure 4. Machine learning-based classification networks for DR. Blood vessels and the FAZ were segmented from OCTA images using a U-net model. The segmented FAZ and vessels were processed to extract four significant retinal features, i.e., FAZ area, blood vessel density, skeletal vessel density, and fractal dimension. These features are supplied to a neural network classifier to detect the onset (D) and referable status (R) of DR

extracted high-level features. In the expanding path, intermediate feature maps of the contracting path were concatenated with the feature maps of the previous expanding path and used as input in the next expanding path. The parameters in the network were optimized using the Adam optimizer47 with a dice similarity coefficient loss and learning rate of 0.0001. In the feature-extraction stage, four local features (blood vessel density, skeletal vessel density, fractal dimension, and size of the FAZ) were extracted from the segmented OCTA images. Finally, in the classification step, data from these four extracted features were fed into a neural network classifier to classify the OCTA images into DR and normal cases as well as referable DR and non-referable states.

We divided the data into four distinct subsets with an even class distribution and performed four-fold cross validation. Specifically, a classifier was trained using three subsets and then tested on the remaining subset. The above operation was repeated four times with different combinations of subsets so that the DR stages of the entire dataset was obtained. The results were then compared to the ground truth determined by retinal specialists using the UWF FA images. Accuracy, sensitivity, specificity, and AUC of the system was calculated to evaluate overall performance. These metrics were calculated based on the average value obtained by four test runs. To validate our CNN method on external dataset, we trained our model using all training data used for fourfold cross validation and tested on the external dataset. The machine learning model was also trained with the same setup to compare with the CNN model.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Cheung, N., Mitchell, P. & Wong, T. Diabetic retinopathy. *Lancet* **376**, 124–136 (2010).
2. David, A., Klein, R. & Gardner, T. Mechanisms of disease diabetic retinopathy. *N. Engl. Med. J.* **366**, 1227–1239 (2012).
3. Kobrin Klein, B. E. Overview of epidemiologic studies of diabetic retinopathy. *Ophthalmic Epidemiol.* **14**, 179–183 (2007).

4. Massin, P. *et al.* Safety and efficacy of ranibizumab in diabetic macular edema (RESOLVE Study): A 12-month, randomized, controlled, double-masked, multicenter phase II study. *Diabetes Care* **33**, 2399–2405 (2010).
5. Elman, M. J. *et al.* Randomized trial evaluating ranibizumab plus prompt or deferred laser or triamcinolone plus prompt laser for diabetic macular edema. *Ophthalmology* **117**, 1064–1077 (2010).
6. Michaelides, M. *et al.* A prospective randomized trial of intravitreal bevacizumab or laser therapy in the management of diabeticmacular edema (BOLT study): 12-month data: Report 2. *Ophthalmology* **117**, 1078–1086 (2010).
7. Mitchell, P. *et al.* The RESTORE study: Ranibizumab monotherapy or combined with laser versus laser monotherapy for diabetic macular edema. *Ophthalmology* **118**, 615–625 (2011).
8. Ozieh, M. N., Bishu, K. G., Dismuke, C. E. & Egede, L. E. Trends in health care expenditure in US adults with diabetes: 2002–2011. *Diabetes Care* **38**, 1844–1851 (2015).
9. Sellahewa, L., Simpson, C., Maharajan, P., Duffy, J. & Idris, I. Grader agreement, and sensitivity and specificity of digital photography in a community optometry-based diabetic eye screening program. *Clin. Ophthalmol.* **8**, 1345 (2014).
10. Ruamviboonsuk, P., Wongcumchang, N., Surawongsin, P., Panyawatananukul, E. & Tiensuwan, M. Screening for diabetic retinopathy in rural area using single-field, digital fundus images. *J. Med. Assoc. Thai* **88**, 176–180 (2005).
11. Gulshan, V. *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402–2410 (2016).
12. Abramoff, M. D. *et al.* Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest. Ophthalmol. Vis. Sci.* **57**, 5200–5206 (2016).
13. Gargeya, R. & Leng, T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* **124**, 962–969 (2017).
14. Akram, M. U., Khalid, S., Tariq, A., Khan, S. A. & Azam, F. Detection and classification of retinal lesions for grading of diabetic retinopathy. *Comput. Biol. Med.* **45**, 161–171 (2014).
15. Ruamviboonsuk, P. *et al.* Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *NPJ Digit. Med.* **2**, 1–9 (2019).