tvst

Article

Differentiation of Diabetic Status Using Statistical and Machine Learning Techniques on Optical Coherence Tomography Angiography Images

Tariq Mehmood Aslam[1,2], David Charles Hoyle[1], Vikram Puri[1], and Goncalo Bento[1,2]

[1] School of Pharmacy and Optometry, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
[2] Manchester Royal Eye Hospital, Manchester University NHS Foundation Trust, Manchester, UK

Correspondence: Tariq Mehmood Aslam, Manchester Royal Eye Hospital, Oxford Road, Manchester, UK. e-mail: tariq.aslam@mft.nhs.uk

Received: July 16, 2019
Accepted: December 16, 2019
Published: March 9, 2020

Keywords: machine learning; optical coherence tomography; image analysis; artificial intelligence; diabetes

Citation: Aslam TM, Hoyle DC, Puri V, Bento G. Differentiation of diabetic status using statistical and machine learning techniques on optical coherence tomography angiography images. Trans Vis Sci Tech. 2020;9(4):2, https://doi.org/10.1167/tvst.9.4.2

Purpose: To investigate the potential of statistical and machine learning approaches to determine the diabetic status of patients from optical coherence tomography angiography (OCT-A) images.

Methods: This was a retrospective cross-sectional observational study based at Manchester Royal Eye Hospital, United Kingdom. OCT-A scans were sequentially selected from one eye of each of 182 patients who were either not diabetic, diabetic without retinopathy, or diabetic with retinopathy requiring hospital follow-up. Eligible images were analyzed by expert purpose-built automated algorithms to calculate clinically relevant outcome measures. These were used in turn as inputs to machine learning and statistical procedures to derive algorithms to perform clinically relevant classifications of patient images into the clinical groups. Receiver operating characteristic curves for the classifiers were evaluated and predictive accuracy assessed using area under curve (AUC).

Results: For distinguishing diabetic patients from those without diabetes, the Random Forest classifier provided the highest AUC (0.8). For distinguishing diabetic patients with significant retinopathy from those with no retinopathy, the highest AUC was represented by logistic regression (0.91).

Conclusions: The study demonstrates the potential of novel techniques using automated analysis of OCT-A scans to diagnose patients with diabetes, or when diabetic status is known, to automatically determine those that require hospital input.

Translational Relevance: This work advances the concept of a rapid and noninvasive clinical screening tool using OCT-A to determine a patient's diabetic status.

Introduction

The impact of diabetes on retinal vasculature has for many years been investigated with traditional techniques involving photography and fluorescent intravenous dyes.[1] Such traditional invasive techniques may show leakage and be sensitive to subtle features, such as microaneurysms, but require intravenous injection of contrast agents, a process that is time-consuming and can have potentially serious side effects.[2] Optical coherence tomography angiography (OCT-A) is a modern imaging innovation that allows for detailed but noninvasive visualization of retinal vasculature. The advent of OCT-A technology has spawned numerous studies including research on the automated segmentation and analysis of OCT-A images for identifying characteristics of diabetic disease.

Eladawi et al.[3] used spatial models to segment blood vessels, and Ting et al.[4] investigated the impact of diabetes on such segmented retinal microvasculature. Other articles corroborate the finding that changes in OCT-A vasculature occur even in patients with diabetes who have no discernible diabetic retinopathy.[5] Nesper et al.[6] showed that higher intensity in the superficial capillary plexus may be an early marker of diabetic microvascular changes before clinical signs of diabetic retinopathy. Durbin et al.[7] have reported variable diagnostic potential when individual features

of OCT-A are used to distinguish eyes with diabetic retinopathy from normal eyes. They found that superficial plexus vessel density had the highest area under the receiver operating characteristic (ROC) curve (area under curve [AUC]) of 0.893, compared with features such as foveal avascular zone (FAZ) area (0.472) and vessel density in the deep plexus (0.703). More recent studies by Eladawi et al.[8] have progressed to demonstrate the ability to distinguish normal patients from those who were diabetic with retinopathy (DR) by combining OCT-A features using a machine learning algorithm.

We propose to expand on work in this field using a wider variety of OCT-A features that have been devised by a retinal expert for their diagnostic potential including novel characteristics that have not previously been studied. A greater number of modern statistical and machine learning techniques will thereafter be applied to optimize the diagnostic potential of these OCT-A features. Finally, to equate findings to real-world use, the algorithms will be designed to differentiate two sets of less diverse patient groups that better represent clinical need. Two principal scenarios are envisaged, which would be most appropriate to demonstrate the potential clinical utility of this technique. First, the capacity to reliably diagnose diabetes using a single noninvasive test (distinguishing diabetic patients from those without diabetes). Second, the capacity to determine if patients with known diabetes have significant retinal disease that requires hospital care (distinguishing diabetics with significant retinopathy from those without significant retinopathy).The aim of this study was therefore to investigate the potential of statistical and machine learning approaches to distinguish diabetic status between these specific closer related groups of patients through clinically driven OCT-A analysis.

Methods

This was a retrospective clinical case–control analysis of anonymized OCT-A scans of one eye of patients seen at Manchester Royal Eye Hospital, United Kingdom, between April 2017 and June 2018. The research was prospectively approved to study anonymized patient images by the Central Manchester University Foundation Trust Research and Development Office and adhered to the Declaration of Helsinki.

Patients were included in the study if they had no significant ocular condition that might affect quality of an OCT-A image or introduce other inappropriate clinical cues, including macular edema, macular degeneration, retinal vein or artery occlusion, glaucoma, and significant cataract. Patients recruited were required to have had an OCT-A scan with good segmentation free of artefacts in the examined eye as determined by the researcher. If both eyes were eligible, a random number generator was used to determine which eye was selected for analysis. The clinical characteristics and best corrected visual acuity (BCVA) on the day of the scan were determined from electronic records and paper notes for patients and visual acuity converted to ETDRS (Early Treatment Diabetic Retinopathy Study) letter score.

We included three cohorts of patients in the study. The first cohort consisted of patients with no diagnosis current nor pending of diabetes, nor signs or symptoms of the disease (normal controls, [N]). They were patients who had been attending Manchester Royal Eye Hospital and had screening scans for a variety of other conditions that had ultimately returned diagnoses that did not involve any posterior segment disease.

The second cohort consisted of patients who were diabetic (insulin and noninsulin dependent) but who had been referred into the hospital eye service for other reasons and did not have any signs of diabetic retinopathy on clinical examination by an ophthalmologist nor on color imaging (diabetic no retinopathy [DnR]).

Finally, we included a cohort of patients who had been diagnosed with diabetic retinopathy and were currently in the hospital eye service for active or stable disease requiring hospital observation or management (DR).

OCT-A Imaging

Patients were imaged using a Topcon swept-source OCT-A machine (Triton DRI-OCT, Topcon, Inc., Topcon, Tokyo, Japan) with a field of view of 12 × 9 mm with 7-layer automated segmentation and a refresh rate of 100,000 A- scans per second. Depth of field of view was set to 3 × 3 mm with axial sampling density of 320 × 320 pixels and number of repetitions set at 4. Scans were obtained aligned to the central point of the fovea. En face OCT angiograms were segmented automatically using the built-in software to define the superficial capillary plexus. Superficial capillary plexus images were exported in the PNG format from ImageNet6—the central OCT-A viewer for the site.

Image Processing and Analysis

The software algorithm for image analysis was developed in-house using MATLAB 2015 (Mathworks,

Natick, MA) using the image processing toolbox. The code worked automatically through various steps to achieve a segmented image and then calculated physical measurement outcomes based on it, which reflected expert knowledge on the impact of diabetes in the eye. It was entirely automatic with no subjective steps. A summary of the main steps in the algorithm is provided here, with more detail provided in the Supplementary Section S1.

The original image first underwent noise removal via a Wiener filter, and subsequent sharpening filters were followed by techniques to equalize the image histogram. At this stage, the vascular network was refined using a seed-growing algorithm. Automated threshold levels were then used to define pixels representing vasculature, with alternate areas representing areas of nonperfusion. If these nonperfused areas matched predefined thresholds of size, they were defined as ischemic areas. Major vessels among the vascular network were located by starting with noise removal and an averaging filter and then pruning of smaller capillaries. A final segmented image was produced of capillaries, major vessels, ischemic areas, and FAZ.

Initial algorithms explored 12 clinically orientated measures of the earlier described segmented anatomic OCT-A features that were designed by a retinal expert to explore features that could be of pathophysiological importance. These are described in full in the Supplementary Section S1. We could potentially have submitted all these characteristics into our machine learning algorithms. However, limited numbers of patients available for training would have meant inappropriate measures would not necessarily be fully excluded, with a lower quality result. We opted therefore to include only those measures deemed to have the most clinical potential, including at least one measure of each anatomic segmented zone. We were guided by earlier feasibility studies that assessed differences in the outcome measures between patients who were DR compared with those who were normal. After reviewing results of these feasibility studies, five key outcome measures derived from the superficial capillary plexus were selected as the candidates for input to machine learning algorithms for the more challenging classification tasks and are described later.

FAZ Circularity ("Circ Max")

Representing the FAZ, this measure was shown to be significantly different between the two prior feasibility groups. Although the more fundamental value of perimeter length of FAZ was also significantly different, information on perimeter length is present within the circularity calculation and the value of FAZ circularity has been found to be of discriminatory value in previous publications.[9] It was calculated using the formula $4\pi A/P^2$ where A is the FAZ area and P is the FAZ perimeter in pixels. Other measures of FAZ including eccentricity and area were not significantly different in prior studies and were not included.

Area of Ischemic Zones Around FAZ ("Adjacent Area Sum")

The FAZ is easily recognizable as the large central area in images that is devoid of vascular structures. It is clinically known that diabetes may lead to increase in size of the FAZ but also understood that smaller areas of capillary loss may occur outside of this FAZ. The "adjacent area sum" measure represents the sum of all ischemic areas found around the FAZ in the entire image. Rather than direct measurement of the area of FAZ,[9,10] we found in preliminary studies that using morphologic techniques to measure the areas of capillary dropout not within, but surrounding the FAZ (Fig. 1), was more significant despite this precise measure not being previously reported. Although these ischemic areas lacking vasculature could be theoretically anywhere within the 3 × 3 area of the OCT-A, they were typically adjacent to the FAZ and this feature has been labeled as adjacent area sum.

Average Percentage of Skeletonized Capillary Vessels ("Ave Percent Sk")

A skeletonization process reduces all objects in any binary image to single-pixel wide lines, without changing the essential structure of the image.[11] Thus it extracts the center line while preserving the network topology. Our algorithm was novel in utilizing morphologic processing techniques to distinguish smaller and larger vessels, allowing for meaningful separate measures of smaller capillary and larger vessels that have not been previously studied. "Average percentage of skeletonized capillary vessels" expresses a measure of the extent of the smaller capillary vessels networking to cover the retina, excluding areas taken up by larger vessels and FAZ. This measure was significantly different between the clinical groups of early studies.

Mean Capillary Intensity ("Mean Cap Int")

This represented a measure of the amount of flow through small vessels by measuring the intensity of the small vessel/capillary pixels. As described earlier, our segmentation techniques allowed measurement of

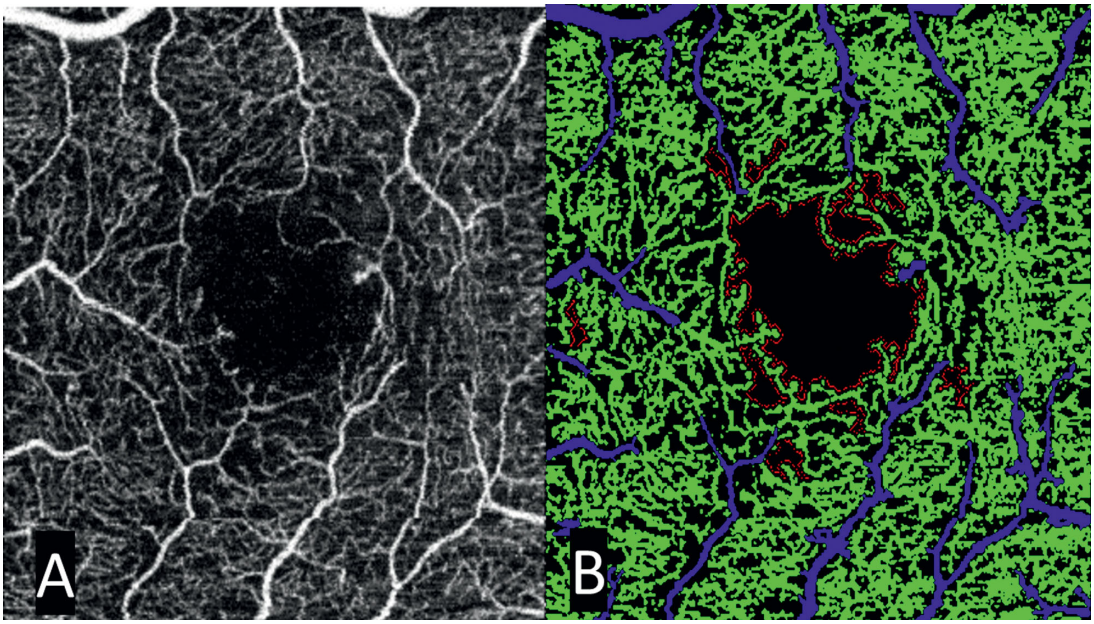

Figure 1. OCT-A 3 × 3 mm scan of the superficial capillary plexus of the left eye of a patient with known diabetic retinopathy. (A) Original scan, (B) segmentation by our algorithms. The large central area bounded by red is the FAZ. It is irregular and distorted with numerous spaces of capillary dropout surrounding it (areas bordered by red lines). The green lines represent capillaries (small vessels) and the blue lines those areas segmented as large vessels.

capillaries only, excluding larger vessels. This measure represented a distinct aspect of vessel pathophysiology to the feature of "average percentage of skeletonized capillary vessels" described earlier, and so both measures were adopted for further study.

Mean Vessel Intensity ("Mean Ves Int")

Despite not being shown to be significantly different in our small feasibility studies, this measure was still incorporated as the only discrete indicator of status of the larger vessels, to allow for some form of representation of all clinically recognizable anatomic features in resulting algorithms. It represents average intensity of the larger segmented vessels.

All eligible OCT-A scans were processed with the automated analysis software, and the key outcome measures described earlier were calculated. These image analysis outcome measures were used as potential predictor variables with the target outcome being the clinical cohort to which the patients belong: normals (N) or diabetic (DM). The DM cohort consisted of patients who were DR or DnR. In this article, we do not promote or explore the accuracy of individual stages of segmentation or analysis of the processing algorithms but focus on assessing and reporting the utility of our final outcome measures to provide data on which to differentiate diabetic status.

We conducted an initial exploratory analysis including distributions of the potential predictor variables and summary statistics of predictor variables in each of the clinical cohorts. As an extension to the exploratory analyses, we derived a decision tree multiclass classifier to illustrate the interaction and importance of the predictor variables to the clinical groups as a whole. We used R (R Foundation for Statistical Computing, Vienna, Austria) for these analyses.[12]

Rather than using a three-class classifier to separate a patient into one of three clinical groups, we deemed it more clinically appropriate to build separate binary classifiers for two distinct, clinically useful classifications. In the first scenario, we assess the capacity to distinguish any patients with diabetes, who may or may not have visible retinopathy, from normal patients. In the second, we assess the potential of algorithms to distinguish those diabetic patients with clinically significant retinopathy (requiring hospital care) from those diabetic patients with no retinopathy. The binary classifiers we assessed were Naïve Bayes,[13] Decision Tree,[13] Logistic Regression,[14] Random Forest,[15] and a gradient boosting classifier (XGBoost).[16]

The Naïve Bayes classifier provided a baseline against which to assess the predictive accuracy of the other classifiers, predictive accuracy markedly below that of the Naïve Bayes classifier may be an indication of overfitting, and therefore less reliable for inference about the importance of the various

Table 1. Summary Statistics by Diabetic Status of all Patients Included in Study Analyses

Variable	Statistic	Normals, N (n = 49)	Diabetic no Retinopathy, DnR (n = 50)	Diabetic with Retinopathy, DR (n = 53)
Sex		Male 32, Female 17	Male 36, Female 14	Male 43, Female 10
Age	Mean (SD)	57.14 (13.56)	61.06 (12.77)	58.38 (13.06)
	Median	57.0	61.5	60
BCVA	Mean (SD)	85 (19.66)	96.36 (7.39)	86.17 (13.43)
	Median	95	95	90
Adjacent area sum	Mean (SD)	3117.12 (3662.07)	3613.92 (3458.93)	12427.91 (6899.22)
	Median	1953.0	1978.5	10959
Circ max	Mean (SD)	0.32 (0.16)	0.24 (0.10)	0.18 (0.088)
	Median	0.29	0.22	0.18
Ave percent sk	Mean (SD)	12.89 (0.44)	13.14 (0.53)	12.58 (0.41)
	Median	12.85	13.03	12.52
Mean ves int	Mean (SD)	180.65 (6.43)	181.38 (6.04)	179.28 (7.45)
	Median	182.0	183.5	181.0
Mean cap int	Mean (SD)	98.69 (6.23)	94.22 (5.41)	93.47 (6.03)
	Median	99	95	94

Image analysis metrics shown are area of ischemic zones around FAZ (adjacent area sum), FAZ circularity (circ max), skeletonized capillary[11] percentage area (ave percent sk), mean vessel intensity (mean ves int), and mean capillary intensity (mean cap int).

predictor variables. ROC curves for the classifiers were evaluated by performing a leave-one-out (LOO) cross-validation. The ROC curve is essentially a plot of sensitivity against specificity for the classifier. Due to both the Random Forest and XGBoost algorithms having a nondeterministic element to them, for each cross-validation fold we ran these algorithms 100 times, each with a different random initialization, and used the mean predicted probability across these 100 runs as our estimate for the left-out data point. Predictive accuracy for the classifiers was assessed using the AUC of the ROC curve. We used the AUC package in R to calculate AUC values of the ROC curves. Confidence intervals for the AUC values were estimated using the pROC package in R. We presented and explored in further detail those classifiers demonstrating the highest AUC values.

Results

We assessed 182 eyes for this study, of which 30 were excluded due to significant artefacts or poor quality: 12 in the DR group and 9 in both DnR and normal groups. A total of 152 eyes from 152 patients were therefore used in the analyses. There were 49 eyes in the normal control cohort of patients, 50 eyes in the cohort of diabetic eyes with no retinopathy, and 53 eyes in the cohort with diabetic retinopathy. The summary of demographic data for each cohort are reported in Table 1.

Exploratory Data Analysis

From initial histograms of variables, we were able to infer which of them help differentiate between the different diabetic statuses (see Supplementary Section S2). In particular, age did not appear to differentiate between the three diabetic populations. As a known diabetic risk factor, confirmation that the sample populations had been matched sufficiently for age is provided by Table 1, which gives summary statistics for each of the variables by diabetic status. The differences in age are not statistically significant ($P = 0.07$ for two-sided, two-sample Wilcoxon test of normals vs. DnR; $P = 0.33$ for two-sided, two-sample Wilcoxon test of normals vs. DR). As the different populations are age matched and we are interested in identifying purely ophthalmic imaging factors for diabetic status, we do not include age as a predictor in the classifier models we will build. Likewise, we consider the lower BCVA values in the normal population to likely be due to nondiabetic patients with suspected vision issues being referred for assessment and a priori do not consider BCVA to be a useful diagnostic or causal factor of diabetic status for our study. Consequently, we do not include BCVA as a predictor variable in our classifier models. As BCVA does not display strong correlations within

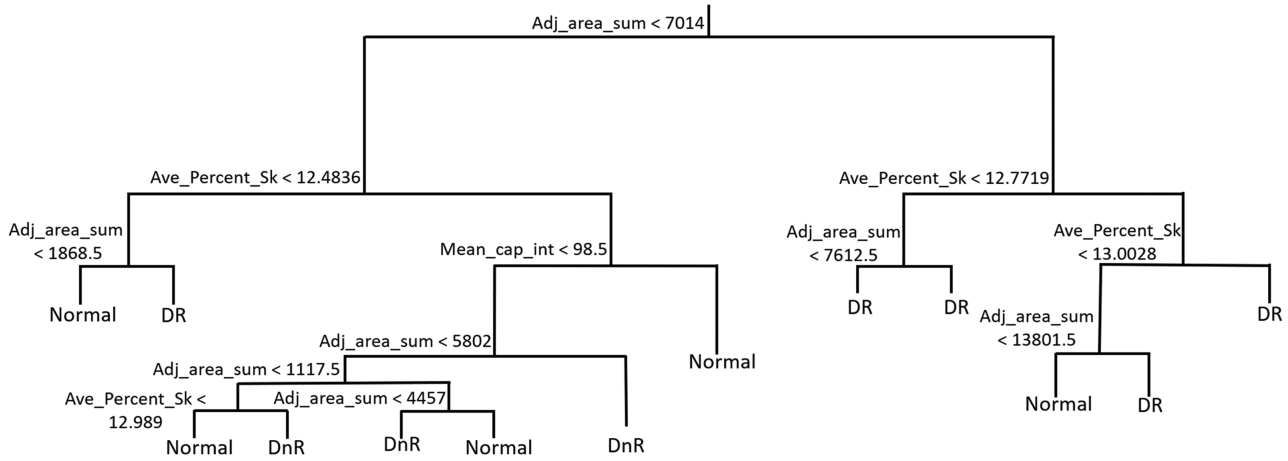

Figure 2. Multiclass decision tree. At the root of each branch is the variable that the algorithm uses to decide the split at that branch. The labels at the terminal nodes are those in which the algorithm predicts to be the most probable diabetic status. Where we see the same terminal label at both nodes of a binary split, this means that the variable has detected different distributions despite the label being the same.

any of the other variables (see Supplementary Section S2), this should not induce any bias in the resultant classifiers.

An initial decision tree built on all the observations (all three sample populations) using the R tree package is shown in Figure 2. This classification technique serves as a useful visual representation of the importance of the various image analysis outcomes in distinguishing between the different clinical scenarios. It is distinct from the Random Forest algorithm used subsequently as a binary classifier. It is evident that the decision tree uses "adjacent area sum" at the root node, primarily distinguishing DR cases (43 out of 53 of the DR cases) from the other sample populations. "Skeletonization percentage" and "mean cap int" are then also used to distinguish primarily between normals and those with diabetes.

Binary Classifiers and Predictive Accuracy

The results from using the various binary classifiers for each of the two clinical classifications of all diabetic patients from nondiabetic (normal), or those with significant diabetic retinopathy from those without significant retinopathy (DR vs. DnR), are presented in Table 2. Estimates of the AUC values of the ROC curves, along with 95% confidence intervals, for the Naïve Bayes, Decision Tree, Logistic Regression, Random Forest, and XGBoost classifiers are provided for each scenario. To further aid comparison of the different classifiers, we have listed estimates (along with 95% confidence intervals) of the specificity for each classifier at 90% sensitivity. Further details on those classifiers demonstrating the highest AUC values are provided later.

Table 2. AUC Values from Analyses of Different Binary Classifiers

Classifier	AUC All Diabetic vs. Normal	AUC DR vs. DnR	Specificity All Diabetic vs. Normal	Specificity DR vs. DnR
Naïve Bayes	0.78 (0.70, 0.86)	0.90 (0.83, 0.95)	0.43 (0.29, 0.59)	0.68 (0.40, 0.88)
Decision Tree	0.61 (0.52, 0.70)	0.83 (0.75, 0.91)	0.12 (0.04, 0.24)	0.46 (0.23, 0.78)
Logistic Regression	0.79 (0.71, 0.87)	0.91 (0.85, 0.96)	0.47 (0.31, 0.63)	0.72 (0.56, 0.88)
Random Forest	0.80 (0.73, 0.87)	0.86 (0.79, 0.93)	0.49 (0.31, 0.69)	0.66 (0.34, 0.82)
Gradient Boosting	0.75 (0.67, 0.83)	0.84 (0.76, 0.92)	0.41 (0.12, 0.59)	0.62 (0.22, 0.80)

Values for AUC for each of the various forms of machine learning and statistical analysis that have been applied are given with respect to the classification tasks in the first two columns; N versus DM individuals and DR versus DnR. A perfect classifier has an AUC value of 1. In the rightmost two columns, specificity estimates of each of the binary classifiers are provided for a fixed 90% value of sensitivity. Values for the 95% confidence intervals are shown in brackets.

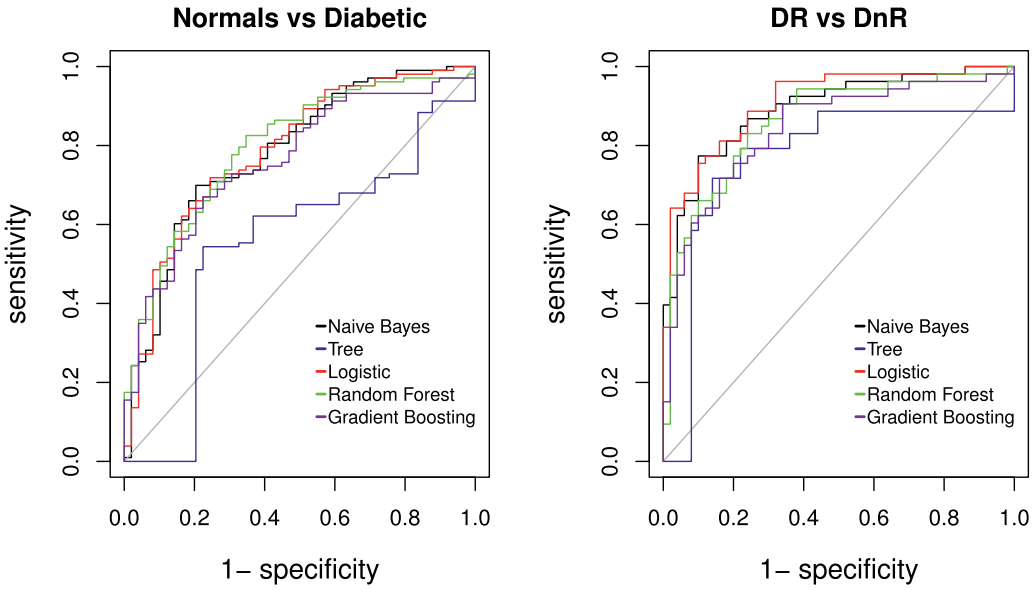

Figure 3. The left image represents the ROC curves for all classifiers for N versus DM. The ROC curves shows how the true-positive rate (sensitivity) increases with increase in the false-positive rate (1-specificity). The ROC curve for a near perfect classifier would start from the origin and rise steeply into the top-left quadrant. The AUC measures how close the ROC curve is to that of a perfect classifier. A perfect classifier has an AUC value of 1. The right image similarly represents the ROC curves for all classifiers for DR versus DnR.

Diabetic versus Normal: Random Forest

The Random Forest algorithm provided the highest predictive accuracy, as measured by the AUC value of 0.8, and is therefore presented in detail. The Random Forest ROC curve estimated from the LOO cross-validation analysis is shown in the left image of Figure 3 in green, along with the ROC curves (for comparison) for all the other classifiers. The importance of each variable, as measured by the values of mean decrease in accuracy and mean decrease in Gini coefficient, when the Random Forest algorithm was applied to the full dataset, is shown in Table 3. The decrease in accuracy measure indicates the reduction in classifier accuracy when we explicitly break any potential link between the predictor variable and the outcome variable. The Gini coefficient is a standard measure of how unequal a probability distribution is, and in this instance indicates

Table 3. Variable Importance Measures for the Random Forest Classifier when Differentiating N from DM Individuals

Variable	Mean Decrease in Accuracy	Mean Decrease in Gini Coefficient
Adjacent area sum	21.11 (21.059, 21.15)	18.45 (18.44, 18.47)
Circ max	10.21 (10.17, 10.26)	15.27 (15.25, 15.28)
Ave percent sk	6.44 (6.40, 6.48)	11.01 (11.00, 11.017)
Mean ves int	2.00 (1.96, 2.045)	7.69 (7.68, 7.70)
Mean cap int	12.81 (12.76, 12.85)	12.99 (12.97, 13.00)

The table shows the mean decrease in accuracy estimated from 2000 runs of the Random Forest algorithm applied to the full dataset. The more the accuracy of the Random Forest classifier decreases when breaking the link between the predictor variable and the outcome variable, the more important that predictor variable. Variables with a large mean decrease in accuracy are therefore more important for classification of the data. The Gini importance measures the mean decrease of node impurity by splits of a given variable. If the variable is useful, it tends to split mixed labeled nodes into pure single class nodes. The table shows the average across the 2000 runs of the algorithm of the mean decrease in Gini coefficient for each predictor variable. In all cases the numbers in brackets denote the range corresponding to ± 1.96 standard errors around the average. Thus we can see that area of ischemic zones around FAZ (adjacent area sum) and FAZ circularity (circ max) are the most important variables.

Table 4. Table of Coefficient Values for Logistic Regression of Image Analysis Measures Differentiating DR versus DnR

Variable	Effect	Standard Error	*P* Value
Adjacent area sum	2.56	0.69	0.00021***
Circ max	−0.36	0.36	0.31
Ave percent sk	−1.65	0.48	0.00052***
Mean ves int	0.085	0.45	0.85
Mean cap int	−0.11	0.45	0.81

It is evident from this table that significant effect is seen from the variables of area of ischemic zones around FAZ (adjacent area sum) and percentage skeletonized area (ave percent sk).

***P Value < 0.001.

how pure, with respect to case mix, the child nodes of a tree are after splitting using the predictor variable. Mean decrease in accuracy and mean decrease in Gini coefficient values are obtained by averaging over all the trees in the Random Forest ensemble, and were calculated using the randomForest package in R. We ran the Random Forest classifier process 2000 times, each with a different random initialization, to average over the nondeterministic element of the Random Forest algorithm. The table shows the averages of the mean decrease in accuracy and mean decrease in Gini coefficient across these 2000 runs. The numbers in brackets in Table 3 denote the range corresponding to ± 1.96 standard errors around the average, with the standard error estimated from the sample standard deviation across the 2000 runs of the Random Forest algorithm. From these figures we would conclude that "adjacent area sum" and "circ max" are the most important predictors in distinguishing normals from diabetics.

DR versus DnR: Logistic Regression

The Logistic Regression algorithm provided the highest predictive accuracy, as measured by the AUC value of 0.91, and is therefore presented in detail. The ROC curve estimated from the LOO cross-validation analysis is shown in the right image of Figure 3 in red, along with the ROC curves (for comparison) for all the other classifiers. Parameter estimates for the logistic regression model comparing DR versus DnR are shown in Table 4. We have used standardized versions of each of the predictor variables (centered to zero mean and scaled to unit variance), to enable easy direct comparison of the contribution of each variable to the probability of a patient having diabetic retinopathy. It is apparent that the most important predictors appear to be adjacent area sum and ave percent sk.

Discussion

The dramatic advances in ophthalmic imaging of recent years have fueled equally intense research into optimizing use of these images for diagnosis of systemic as well as ocular disease. In this study, we combined novel and expert-led image processing outcomes with modern statistical and machine learning techniques to demonstrate feasibility of classifying the diabetic status of patients.

There have been a number of excellent studies that have highlighted the potential power of OCT-A in discriminating diabetic status.[17–21] In particular, pioneering work by Durbin et al.[7] demonstrated the ability of measuring retinal microvasculature to distinguish diabetic status in patients. In their albeit small cohort they demonstrated highest area under the ROC curve (i.e., AUC) of 0.893 for measures of the vessel density in the superficial capillary network. FAZ area gave a much lower AUC of 0.472. More recently, Eladawi et al.[8] developed a comprehensive computer-aided diagnosis system that segmented superficial and deep retinal maps and determined density and caliber of the blood vessels, and size of the FAZ for both the superficial and deep retinal maps, which they used to train and test a support vector machine classifier. Their system demonstrated an overall accuracy (i.e., AUC) of 94.3%. Both these articles represent significant steps forward in the potential use of OCT-A information for assessing diabetic status.

However, both these systems addressed the single task of distinguishing patients with diabetic retinopathy from normal patients. For our work we addressed what we consider to be two more clinically useful but challenging tasks. First, we aimed to differentiate normal patients from any patients with systemic diabetes. Second, we elected to differentiate diabetic patients with retinopathy from patients who were diabetic but without retinopathy. Our challenges were greater still as we did not have patients with any significant macular edema in our study, which often artificially disrupts OCT-A imaging.

Our work is also distinct in that we compare multiple learning algorithms on new anatomic OCT-A features. We describe novel outcome measures such as ischemic areas adjacent to the fovea and separate metrics for small and larger vessels. We believe our strategies have led to algorithms with greater external validity for potential future real-world usage.

Finally, our classification of three distinct stages of eye disease has led to insights to vascular changes associated with diabetic retinopathy progression. In terms of differentiating diabetic patients with retinopa-

thy from those without retinopathy, skeletonization percentage and "adjacent area sum" were found to be the most informative variables. However, a different metric, the "mean capillary intensity," only helps differentiate diabetic populations from normal patients. Further studies would be needed to confirm any interpretation, but it may be that a reduced flow to capillaries (reduced capillary intensity) occurs early on in changing from normal to diabetic. It would then be in later stages of diabetes that the complete loss of such small vessels (skeletonization percentage) becomes an important distinguishing factor.

There are several limitations to this study that further research should address. First, our model involves a variety of automated steps including vessel segmentation into larger vessels and more fine capillaries, before applying a variety of image processing algorithms aiming to produce expert-led, disease-specific metrics. In this article, we provide evidence only for the resultant algorithm when used in its entirety to produce outcomes that can be combined for discrimination of diabetic status and cannot claim utility of individual components.

We used relatively low numbers of patients considering involvement of machine learning techniques. Rather than separating out a single validation set at the beginning of the study, we instead made maximum use of our dataset by using LOO cross-validation. Such cross-validation can test an algorithm's ability to predict new data that was not used in estimating it, preventing problems of overfitting or selection bias.[22] The levels of accuracy despite low numbers reflect the power of the image analysis outcome measures chosen, as well as the appropriate use of advanced statistical techniques.

As a retrospective study, there are inevitable potential sources of bias, such as in the clinical groups, which largely came from different hospital referral streams. Some images were excluded from the study if they were not deemed of suitable quality by a researcher, representing a subjective step in the otherwise automated algorithms. For the algorithm to be of eventual clinical use, we will need to ensure any exclusion process is automated.

Despite these weaknesses, our study demonstrates the clinical potential of our system of processing and analysis of OCT-A images. This is especially evident for distinguishing diabetic patients with retinopathy that requires hospital care from those without retinopathy (AUC of 0.91). Demonstration of ability to diagnose diabetic patients (including with no visible retinopathy) from those without diabetes has an AUC of 0.80. This discriminative ability, however, has arguably greater potential clinical implications but would need to be improved, particularly the false-positive rate, as evidenced by values for specificity given in Table 2. The algorithms presented in this article are not proposed as screening tools in their current form, but improvements as described later may lead to algorithms for clinical use.

We expect to improve the diagnostic accuracy through a variety of means. First, newer imaging technologies provide higher quality images of broader areas of the retina. We plan to use greater levels of information from these OCT-As such as by using deep or intermediate capillary plexi. Previous articles[8] suggest that level of information can be improved by accessing these deep capillary plexi, and we also plan to use new imaging scanners presenting larger areas for analysis. Additional patient metadata such as recent specific symptoms and details of race and sex might further increase the power of the classification algorithms and provide more assurance on external validity. We plan to adapt our own algorithms to manage any artefactual lesions more effectively and incorporate this into the automated analysis. With such improvements and subsequent prospective validation, we hope to optimize use of noninvasive OCT-A scans to automatically determine clinically important groups with regard to diabetic status without any clinician intervention, even when there are no discernible photographic or clinical signs of diabetes. Future studies will also investigate the potential of these OCT-A algorithms to determine diabetic microvascular disease in other parts of the body such as in renal tissue.

Conclusions

The importance of our study is that it is distinct in demonstrating the potential for combining expert-derived and novel OCT-A features, with a range of learning algorithms to address challenging and clinically relevant discrimination tasks addressing diabetic status. Further work will aim to develop this potential to produce clinically useful diagnostic tools for general and nonspecialist practices to improve screening for diabetes and diabetic eye complications.

Acknowledgments

Disclosure: **T.M. Aslam,** Bayer (F), Novartis (F), Laboratories Thea (F), Oraya (F), Bausch and Lomb (F); **D.C. Hoyle,** None; **V. Puri,** None; **G. Bento,** None

References

1. Bresnick GH, Condit R, Syrjala S, Palta M, Groo A, Korth K. Abnormalities of the foveal avascular zone in diabetic retinopathy. *Arch Ophthalmol*. 1984;102:1286–1293.
2. Lopez-Saez MP, Ordoqui E, Tornero P, et al. Fluorescein-induced allergic reaction. *Ann Allergy Asthma Immunol*. 1998;81:428–430.
3. Eladawi N, Elmogy M, Helmy O, et al. Automatic blood vessels segmentation based on different retinal maps from OCTA scans. *Comput Biol Med*. 2017;89:150–161.
4. Ting DSW, Tan GSW, Agrawal R, et al. Optical coherence tomographic angiography in type 2 diabetes and diabetic retinopathy. *JAMA Ophthalmol*. 2017;135:306–312.
5. Cao D, Yang D, Huang Z, et al. Optical coherence tomography angiography discerns preclinical diabetic retinopathy in eyes of patients with type 2 diabetes without clinical diabetic retinopathy. *Acta Diabetol*. 2018;55:469–477.
6. Nesper PL, Roberts PK, Onishi AC, et al. Quantifying microvascular abnormalities with increasing severity of diabetic retinopathy using optical coherence tomography angiography. *Invest Ophthalmol Vis Sci*. 2017;58:BIO307–BIO315.
7. Durbin MK, An L, Shemonski ND, et al. Quantification of retinal microvascular density in optical coherence tomographic angiography images in diabetic retinopathy. *JAMA Ophthalmol*. 2017;135:370–376.
8. Eladawi N, Elmogy M, Khalifa F, et al. Early diabetic retinopathy diagnosis based on local retinal blood vessel analysis in optical coherence tomography angiography (OCTA) images. *Med Phys*. 2018;45:4582–4599.
9. Kwon J, Choi J, Shin JW, Lee J, Kook MS. Glaucoma diagnostic capabilities of foveal avascular zone parameters using optical coherence tomography angiography according to visual field defect location. *J Glaucoma*. 2017;26:1120–1129.
10. Kwon J, Choi J, Shin JW, Lee J, Kook MS. An optical coherence tomography angiography study of the relationship between foveal avascular zone size and retinal vessel density. *Invest Ophthalmol Vis Sci*. 2018;59:4143–4153.
11. Pedinielli A, Bonnin S, Sanharawi ME, et al. Three different optical coherence tomography angiography measurement methods for assessing capillary density changes in diabetic retinopathy. *Ophthalmic Surg Lasers Imaging Retina*. 2017;48:378–384.
12. R_Core_Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
13. Mitchell T. *Machine learning*. Boston, MA: McGraw-Hill; 1997.
14. Agresti A. *Categorical data analysis*. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2013.
15. Breiman L. Random forests. *Machine Learning*. 2001;45:5–32.
16. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA; Association for Computing Machinery; 2016;785–794.
17. Niestrata-Ortiz M, Fichna P, Stankiewicz W, Stopa M. Enlargement of the foveal avascular zone detected by optical coherence tomography angiography in diabetic children without diabetic retinopathy. *Graefes Arch Clin Exp Ophthalmol*. 2019;257:689–697.
18. Li L, Almansoob S, Zhang P, Zhou YD, Tan Y, Gao L. Quantitative analysis of retinal and choroid capillary ischaemia using optical coherence tomography angiography in type 2 diabetes. *Acta Ophthalmol*. 2019;97:240–246.
19. Or C, Das R, Despotovic I, et al. Combined multimodal analysis of peripheral retinal and macular circulation in diabetic retinopathy (COPRA Study). *Ophthalmol Retina*. 2019;3:580–588
20. Arya M, Rebhun CB, Alibhai AY, et al. Parafoveal retinal vessel density assessment by optical coherence tomography angiography in healthy eyes. *Ophthalmic Surg Lasers Imaging Retina*. 2018;49:S5–S17.
21. Alibhai AY, De Pretto LR, Moult EM, et al. Quantification of retinal capillary nonperfusion in diabetics using wide-field optical coherence tomography angiography. *Retina*. 2020;40:412–420, doi:10.1097/IAE.0000000000002403. [Epub ahead of print].
22. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res*. 2010;11:2079–2107.