Original Article

Evaluation of a deep learning supported remote diagnosis model for identification of diabetic retinopathy using wide-field Optomap

Terry Lee[1#^], Mingzhe Hu[2#], Qitong Gao[2], Joshua Amason[1], Durga Borkar[1], David D'Alessio[3], Michael Canos[3], Afreen Shariff[3], Miroslav Pajic[2], Majda Hadziahmetovic[1]

[1]Department of Ophthalmology, Duke University Medical Center, Durham, NC, USA; [2]Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA; [3]Department of Endocrinology, Duke University Medical Center, Durham, NC, USA

Contributions: (I) Conception and design: T Lee, M Hu, M Pajic, M Hadziahmetovic; (II) Administrative support: M Pajic, M Hadziahmetovic; (III) Provision of study materials or patients: D D'Alessio, M Canos, A Shariff; (IV) Collection and assembly of data: T Lee, M Hu, Q Gao, J Amason, D Borkar, M Hadziahmetovi; (V) Data analysis and interpretation: T Lee, M Hu, Q Gao, J Amason, D Borkar, M Hadziahmetovi; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Majda Hadziahmetovic, MD. Assistant Professor of Ophthalmology, Department of Ophthalmology, Duke University Medical Center, 2351 Erwin Rd., Durham, NC 27710, USA. Email: majda.hadziahmetovic@duke.edu.

Background: We test a deep learning (DL) supported remote diagnosis approach to detect diabetic retinopathy (DR) and other referable retinal pathologies using ultra-wide-field (UWF) Optomap.

Methods: Prospective, non-randomized study involving diabetic patients seen at endocrinology clinics. Non-expert imagers were trained to obtain non-dilated images using UWF *Primary*. Images were graded by two retina specialists and classified as DR or incidental retinal findings. Cohen's kappa was used to test the agreement between the remote diagnosis and the gold standard exam. A novel DL model was trained to identify the presence or absence of referable pathology, and sensitivity, specificity and area under the receiver operator characteristics curve (AUROC) were used to assess its performance.

Results: A total of 265 patients were enrolled, of which 241 patients were imaged (433 eyes). The mean age was 50±17 years, 45% of patients were female, 34% had a diagnosis of diabetes mellitus type 1, and 66% of type 2. The average Hemoglobin A1c was 8.8±2.3%, and 81% were on Insulin. Of the 433 images, 404 (93%) were gradable, 64 patients (27%) were referred to a retina specialist, and 46 (19%) were referred to comprehensive ophthalmologist for a referable retinal pathology on remote diagnosis. Cohen's kappa was 0.58, indicating moderate agreement. Our DL algorithm achieved an accuracy of 82.8% (95% CI: 80.3–85.2%), a sensitivity of 81.0% (95% CI: 78.5–83.6%), specificity of 73.5% (95% CI: 70.6–76.3%), and AUROC of 81.0% (95% CI: 78.5–83.6%).

Conclusions: UWF *Primary* can be used in the non-ophthalmology setting to screen for referable retinal pathology and can be successfully supported by an automated algorithm for image classification.

Keywords: Retina; screening; imaging; deep learning (DL); diabetic retinopathy (DR)

Received: 16 September 2021; Accepted: 10 February 2022; Published: 15 June 2022.

doi: 10.21037/aes-21-53

View this article at: https://dx.doi.org/10.21037/aes-21-53

^ ORCID: 0000-0002-3852-805X.

Introduction

Diabetic retinopathy (DR) is a leading cause of blindness that affects millions of people in the U.S. (1), with the prevalence of DR and vision-threatening retinopathy among patients with diabetes estimated to be 28–40% and 4–8%, respectively (2,3). In addition to the devastating consequences on those with the disease, the care of DR has been estimated to cost the U.S. healthcare system $500 million annually, placing a significant economic burden (4). Fortunately, early detection and treatment have shown to significantly improve clinical and cost-effectiveness (5,6).

As a result, much emphasis has been placed on developing effective and accessible ways of screening for DR and other retinal pathologies. Screening for DR has clear guidelines (7-9) and is becoming increasingly mandated in the U.S. Regardless, screening via traditional methods (e.g., dilated fundus exam by eye care specialists) is heavily dependent on resource availability in each community and thus frequently is not very successful (10-14). Teleophthalmology has attempted to address this unmet need, and while it can be very effective, there are yet several barriers to its widespread implementation, including cost and lack of universal standards (15,16). While there are many approaches to teleophthalmology, one type hereby referred to as remote diagnosis, has been proposed as a particularly effective form of screening (17). In contrast to other forms of teleophthalmology screening, remote diagnosis uses imaging devices permanently located at the point of service (e.g., primary care or endocrinology clinics) and operated by non-expert imagers (e.g., office clinical medical assistants, CMAs). Thus, it holds the potential to be more accessible and cost-effective.

The feasibility of such a model has recently been successfully tested in the detection of referable DR and age-related macular degeneration (AMD), using fundus photography and optical coherence tomography (OCT) camera located at the point of service such as endocrinology clinics and assisted living centers (17). This study showed near equivalent rates of detection compared to standard examination by a retinal specialist (17). Other imaging methods such as ultra-wide-field scanning laser ophthalmoscopy (Optomap), which are possibly better suited for imaging of the fundus through non-dilated pupils, have yet to be explored with a remote diagnosis model. However, these methods have significant costs associated with an expert grader required to review imaging findings and decide whether to refer the patient for further examination. Additionally, grading and information transfer are associated with a significant delay in decision making and patient scheduling. One way of bypassing the need for an expert grader is to train and implement a deep learning (DL) model to automate the grading. Previous studies have trained DL models using Optomap images to detect AMD (18), central retinal vein occlusion (19), macular holes (20), retinal detachment (21), and DR (22,23). However, none of these studies used Optomap images acquired in a prospective manner, in a remote diagnosis approach to detect referable retinal pathologies, including DR.

Thus, the purposes of the present study are twofold. First, we evaluate the feasibility of a remote diagnosis approach to screen for DR and other referable retinal pathology at two Duke endocrinology clinics using the UWF Primary imaging device by Optos. Then, we introduce a DL model using the acquired Optomap images and test its diagnostic precision compared to clinician evaluation of the same images as the reference standard. We present the following article in accordance with the STARD reporting checklist (available at https://aes.amegroups.com/article/view/10.21037/aes-21-53/rc).

Methods

This prospective, nonrandomized study was conducted in accordance with the tenets of the Declaration of Helsinki (as revised in 2013). The study was approved by Institutional Review Board of Duke University (IRB00012400) and informed consent was taken from all individual participants.

Patient population

A total of 265 patients were enrolled from 2 Duke University Heath System endocrinology clinics. All eligible patients over the age of 18 seen at these outpatient sites were invited to participate in this study. Pregnant or nursing women were excluded, as well as any patients who could not sit still for the duration of the imaging or tolerate the imaging. Patients were informed of risks and benefits prior to study involvement, and consent was obtained from all patients.

Study design

Remote diagnostic imaging of non-dilated pupils was performed in the endocrinology clinics by trained but non-

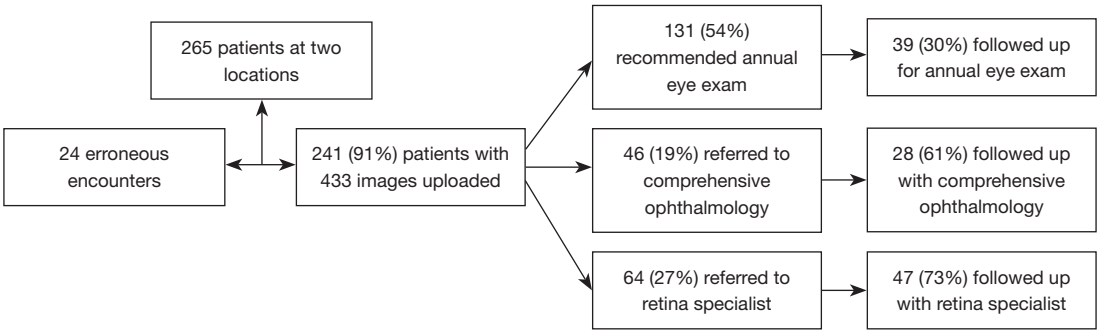

Figure 1 Flow chart of the remote diagnosis approach for retinal evaluation.

expert imagers provided by the clinics. Non-ophthalmologic health care professionals (e.g., clinical medical assistants, CMAs) were trained to use an FDA-approved noncontact, portable retinal imaging device (Optomap, Nikon) that takes up to 200° field-of-view retinal images. After the initial training, imagers were observed and coached periodically to ensure that the imaging remained at a high quality.

A large percentage of the screened patients referred for further evaluation followed up with either Duke University Health System retinal specialists or comprehensive ophthalmologists for standard examination, including a dilated fundus exam and any ancillary testing (e.g., color fundus photo, optical coherence tomography, fundus angiography) performed by expert photographers. Thus, each eye served as its own control for those patients. *Figure 1* depicts a flowchart of our study. At the conclusion of the study, the non-ophthalmologic health care professionals trained to use the Optomap imaging device completed a non-validated survey to assess the ease of using the device and acquiring images.

Remote diagnosis image grading

Two graders (M.H. and D.B.) graded the images by consensus, without any access to the patients' clinical information. All images were graded by the end of the work week during which they were acquired. Images were considered ungradable if no clear view of the macula was available (e.g., poor positioning, poor patient compliance, movement, droopy eyelids). Images were assessed for DR and its severity using the standardized ETDRS scale. By default, patients with DR on remote diagnosis imaging were referred to a retina specialist. Other findings were referred either to a retina specialist or comprehensive ophthalmology, depending on apparent severity. Any ungradable images were grouped with those with referable disease, as these eyes would have failed image screening. For accuracy assessment, the reference standard was the standard clinical examination findings from the same patient.

Deep learning dataset

A total of 433 3-channel Optomap images acquired during the study were used to train and test our DL model. Of all these images, 269 were labeled as normal according to remote diagnosis evaluation, and the remaining 164 were labeled as having some retinal pathology, whether DR or other pathologies. *Figure 2* shows 3 representative examples of Optomap images: ungradable, normal, and diagnosed with DR.

We adopted a 4:1 ratio for splitting all the images into training and testing sets. To increase the amount of data and improve the robustness of the model, the images in the training dataset underwent augmentation. The augmentation operations involved: (I) cropping 0 to 16 pixels from each side of the images randomly; (II) flipping the images horizontally or vertically; (III) rotating the images between negative 25 degrees to positive 25 degrees. The purposes of augmentation were to increase the number of training data to reduce overfitting and improve the generalizability of our model. The images in the testing set remain unchanged. *Figure 3* shows a flow chart of our proposed DL approach.

Development of the deep learning algorithm

A DL model was built from scratch using the Keras library in Python to identify the presence or absence of retinal pathologies on Optomap images, using a convolutional

Figure 2 Ultra-wide-field (UWF) Fundus images. (A) Ungradable image. (B) Normal retina. (C) Retina with identifiable pathology.

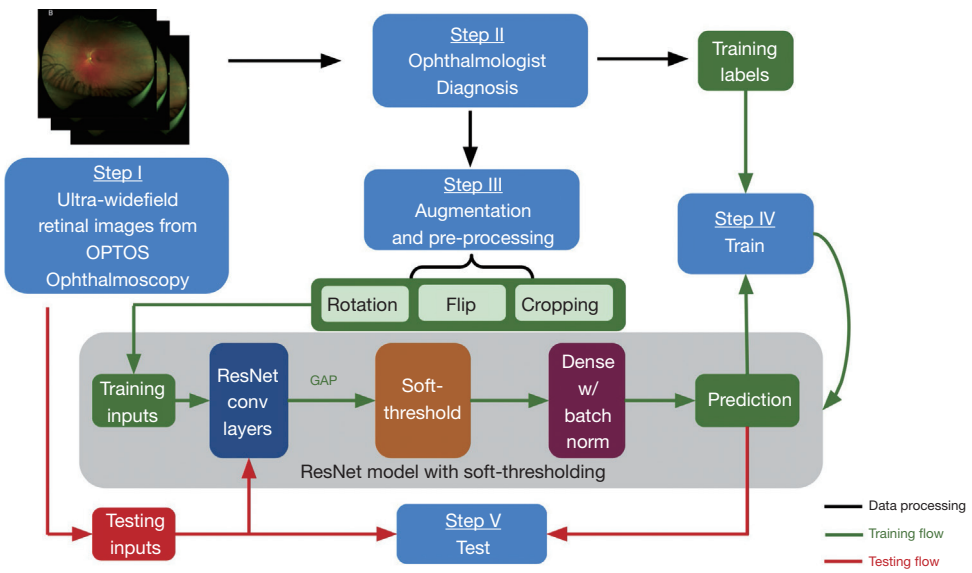

Figure 3 Overview of our proposed deep learning methodology. Optomap images acquired at the point of service (Step I) were labeled by ophthalmologists (Step II). After splitting the images into testing and training sets, data augmentation was conducted on the training dataset (Step III). Then, the training images and labels were used to train a ResNet model with a soft-thresholding layer (Step IV). Finally, we used the hold-out testing images to validate the performance of the deep learning (DL) model (Step V).

neural network (CNN) model built on top of ResNet's (24) architecture. We did not employ transfer learning since the standard well-established models are pre-trained on public datasets like ImageNet, which contains substantially different images from our UWF Optomap images. Besides, since large databases of ultrawide field images are not as readily available (compared to simple fundus photos, for example), there were not enough images to pre-train the model. Thus, we followed the classical architecture of ResNet to build a small model and additionally added soft-thresholding techniques to reduce the noise.

Specifically, ResNet is a CNN that uses shortcuts to bypass some layers to alleviate the vanishing gradient problem caused by increasing depths in deep neural networks. However, it falls short in analyzing images that are highly noisy or include substantial artifacts (24). As a result, the eyebrows and eyelids presented in the Optomap images could prevent ResNet models from classifying the images accurately. Instead, as proposed by Zhao *et al.* (24), we applied a trainable soft threshold to the features outputted from the global average pooling (GAP) layer. This helps the CNN to focus on the features that are most helpful for improving its performance during training as well as disregarding the noisy information introduced by

the artifacts from the input images. Similar techniques are widely used for other denoising methods (25,26). Specifically, as shown in *Figure 3*, soft-thresholding layer was employed after the GAP layer, followed by dense layers with rectified linear unit (27) as the activation function and batch normalization. The batch normalization standardized the inputs to the dense layer and improved the training speed and stability (28). Finally, the binary classification was implemented by the dense layer after global average pooling.

The images were divided into mini batches of size 200, and the model was trained for 5,000 epochs. The Adam optimizer (29) was chosen as the optimization algorithm, and the learning rate was 0.001. Early stopping—which monitored the validation loss—was set, and the minimum delta and patience were 0 and 2,000, respectively. This ensured that if the model's validation loss did not decrease within 2,000 epochs, the training would be stopped. The model was enabled to restore weights from the epoch that achieved minimum validation loss. Training accuracy curves, training loss curves, validation accuracy curves, and validation loss curves were recorded to fine-tune the parameters. This study's hardware specification was an Intel(R) Xeon(R) CPU @ 2.20GHz CPU, a Tesla P100 GPU, and 13GB memory. The code we used to train the CNN has been made open source on GitHub (28).

Statistical analysis

All statistical analyses were performed with open-source software (Python, version 3.7.6; Python Software Foundation, https://www.python.org). Descriptive statistics were calculated to assess the baseline demographics and medical comorbidities. Chi-square test was used to assess the association of categorical variables, two-sample *t*-tests were used to test differences in means, and logistic regressions were used to assess the relationship between predictor variables and a binary outcome. Cohen's kappa was used to test agreement between remote diagnosis and standard examination by eye specialists at a tertiary care center. Per standards set by Landis and Koch (30), a kappa of 0.00–0.20 indicated slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement. Ninety-five percent confidence intervals were derived by bootstrapping with 1,000 replications. To assess the performance of our DL model, we calculated sensitivity, specificity, and the area under the receiver operating characteristic curve (AUROC), along with each metric's respective 95% confidence intervals. The alpha level (type I error) was set at 0.05 for all statistical analyses.

Results

A total of 265 patients agreed to participate. However, due to various technical or logistical difficulties (e.g., patient positioning), 433 eyes of 241 patients were imaged. The mean age was 50±17 years, and 45% of patients were female. Of all patients, 56% identified as white, 31% as black or African American, 6% as Asian, 2% as multiracial, and 5% did not disclose or identify as any of the above. Thirty-four percent of the patients had a diagnosis of diabetes mellitus type 1, and the other 66% had a diagnosis of diabetes mellitus type 2. The average Hemoglobin A1c was 8.8%±2.3% (normal range, 5.7–6.4%), and 81% of the patients were on Insulin. *Table 1* lists baseline demographic information as well as the most common medical comorbidities for the overall cohort and subgroups stratified by type of diabetes (type 1 or type 2). While patients with Type 2 diabetes mellitus (T2DM) were more likely than Type 1 diabetic patients (T1DM) to have several comorbidities including hypertension (P<0.001), hyperlipidemia (P<0.001), and chronic kidney disease (P=0.027), among others, they had comparable rates of DR (P=0.741) and rates of referral to a retina specialist (P=0.725) or comprehensive ophthalmologist (P=0.657).

After remote evaluation of Optomap images by retina specialists, as expected, a large proportion of screened patients (54%) had no retinal pathology identified and were advised to follow-up annually for DR screening. Sixty-four patients (27%) received a referral to a retina specialist for further assessment and treatment, and 46 patients (19%) received a referral to a comprehensive ophthalmology for management of incidental findings. Of those patients referred for further ophthalmic attention, 52 had DR. Among these 52 patients, 32% had mild, 46% moderate, 10% severe non-proliferative DR, and 12% had proliferative DR. The most common incidental findings in eyes referred to a retinal specialist was AMD, and the most common incidental findings in those referred to a comprehensive ophthalmologist were choroidal nevi and pigmentary changes.

Using logistic regression modeling, we evaluated several factors that affected the likelihood of DR (*Table 2*) and retinal referral (*Table 3*) in remote diagnosis evaluation. Patients with greater HbA1c were more likely to have

Table 1 Demographics, comorbidities, and remote diagnosis evaluation for overall cohort and stratified by type of diabetes mellitus

Variables	All, n=265	T1DM, n=86[†]	T2DM, n=168[†]	P value
Demographics				
Age, mean ± SD, years	50.1±16.7	34.3±13.8	57.8±12.2	<0.001[a]
Sex, female	45%	45%	45%	0.979[b]
Race				<0.001[b]
Caucasian	56%	79%	45%	
African American	31%	17%	38%	
Hemoglobin A1c, mean ± SD	8.8±2.3	8.3±1.9	9.0±2.5	0.011[a]
Duration of diabetes, mean ± SD, years	3.6±3.4	3.2±3.4	3.8±3.3	0.163[a]
Insulin use	81%	100%	72%	<0.001[b]
Comorbidities				
Hypertension	51%	22%	65%	<0.001[b]
Obesity	46%	27%	58%	<0.001[b]
Hyperlipidemia	28%	7%	39%	<0.001[b]
OSA	13%	4%	19%	0.001[b]
GERD	13%	2%	19%	<0.001[b]
OA	12%	4%	17%	0.005[b]
Coronary artery disease	11%	5%	14%	0.035[b]
Anemia	11%	2%	14%	0.006[b]
Chronic kidney disease	9%	4%	13%	0.027[b]
Remote diagnosis evaluation				
Diabetic retinopathy	22%	24%	21%	0.741[b]
Referral to retina specialist	27%	25%	28%	0.725[b]
Referral to comprehensive ophthalmologist	19%	16%	19%	0.657[b]

[a], compared using two-sample *t*-test; [b], compared using chi-square test; [†], type of diabetes was unknown in 11 patients. GERD, gastroesophageal reflux disease; OA, osteoarthritis; OSA, obstructive sleep apnea; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

DR on remote evaluation [odds ratio (OR) 1.16 for each 1% increase in HbA1c, 95% CI: 1.02–1.33, P=0.029] as did patients on insulin (OR 3.62, 95% CI: 1.23–10.63, P=0.019). Patients with chronic kidney disease were also more likely to have DR (OR 2.96, 95% CI: 1.23–7.12, P=0.016). Obesity (OR 1.81, 95% CI: 1.01–3.24, P=0.045) and congestive heart failure (OR 7.54, 95% CI: 1.42–39.93, P=0.017) were associated with referral to a retinal specialist.

Of the 433 images, 404 (93%) were deemed gradable. A total of 114 patients out of the 241 patients that were imaged during remote diagnoses followed up at Duke Eye Center for a standard examination by a retinal specialist (breakdown by type of referral shown in *Figure 1*). Cohen's kappa score for agreement between remote evaluation and standard examination was 0.58 (95% CI: 0.44–0.72), which indicated moderate agreement. There were 13 eyes in which remote diagnosis with Optomap images missed a retinal pathology and 11 eyes with misdiagnoses compared to standard evaluation. These findings are summarized in *Table 4*.

On the non-validated survey of CMAs to assess the feasibility of implementing remote imaging and diagnosis as a standard practice, the results indicated that generally, the camera was well received, with about 75% of those surveyed

Table 2 Logistic regression analyses to assess putative predictive variables for detection of diabetic retinopathy on remote diagnosis evaluation

Predictor	Odds ratio	95% CI	P value
Demographics			
Age	1.01	[0.99, 1.03]	0.429
Sex, male	1.13	[0.60, 2.09]	0.709
Race, Caucasian	0.73	[0.38, 1.43]	0.365
Hemoglobin A1c	1.16	[1.02, 1.33]	0.029
Duration of diabetes	0.94	[0.85, 1.05]	0.277
Insulin use	3.62	[1.23, 10.63]	0.019
Comorbidities			
Hypertension	1.24	[0.67, 2.30]	0.488
Hyperlipidemia	0.55	[0.26, 1.17]	0.118
Obesity	1.48	[0.80, 2.73]	0.216
OSA	0.66	[0.24, 1.82]	0.425
Coronary artery disease	1.16	[0.44, 3.07]	0.765
Chronic kidney disease	2.96	[1.23, 7.12]	0.016
GERD	0.63	[0.23, 1.74]	0.376
Congestive heart failure	1.46	[0.28, 7.77]	0.654
Hypothyroidism	2.10	[0.74, 5.98]	0.165

OSA, obstructive sleep apnea; GERD, gastroesophageal reflux disease.

expressing mastery with the camera. However, the results of our survey also indicated that about 50% of respondents had some difficulty correctly positioning patients to get the retina imaged as clearly as possible with the Optomap. Additionally, our survey results showed that, on average, it only required 3–6 minutes to image a patient's retina.

The accuracy of our DL model for identifying referable retinal pathology was 82.8% (95% CI: 80.3–85.2%), sensitivity was 81.0% (95% CI: 78.5–83.6%), and specificity was 73.5% (95% CI: 70.6–76.3%). *Figure 4* shows the receiver operating characteristic (ROC) curve of our model for detecting referable retinal pathology from Optomap. The AUROC was 81.0% (95% CI: 78.5–83.6%). *Table 5* summarizes additional performance metrics for our DL model.

Discussion

Burdened with multiple co-morbidities, diabetic patients seem to be very compliant on average with their primary care and endocrinology visits (31,32), making those clinics an ideal location to implement screening for DR. Herein, we demonstrate the feasibility of such a screening approach. We also demonstrate that the UWF camera *Primary* by Optos may be a good choice for a screening device at the point of service. To facilitate timely image interpretation and information transfer, we developed and made open-source an automated DL-based image interpretation model with a binary classification outcome (Yes/No retinal pathology). With a larger number of images, it is expected that the accuracy of the proposed DL model could be further improved. However, it is important to emphasize that, while previous studies have developed DL models to detect or stage DR and other retinal pathologies, this is the first computational model for automated analysis for identification of any referable pathology using Optomap images acquired in a prospective manner at multiple endocrinology clinics.

Retina screening in this setting has multiple advantages: improved accessibility, patient capture and triage, and clinical and cost-effectiveness. In some forms of teleophthalmology screening, color fundus photographs

Table 3 Logistic regression analyses to assess putative predictive variables for referral to a retina specialist after remote diagnosis evaluation

Predictor	Odds ratio	95% CI	P value
Demographics			
Age	1.02	[1.00, 1.04]	0.056
Sex, male	1.10	[0.62, 1.96]	0.748
Race, Caucasian	0.55	[0.29, 1.02]	0.059
Hemoglobin A1c	1.20	[1.05, 1.37]	0.006
Duration of diabetes	1.04	[0.95, 1.13]	0.379
Insulin use	2.38	[1.00, 5.64]	0.050
Comorbidities			
Hypertension	1.65	[0.92, 2.95]	0.092
Hyperlipidemia	0.94	[0.49, 1.79]	0.848
Obesity	1.81	[1.01, 3.24]	0.045
OSA	0.80	[0.33, 1.95]	0.619
Coronary artery disease	1.68	[0.70, 4.01]	0.246
Chronic kidney disease	2.20	[0.92, 5.24]	0.076
GERD	1.33	[0.59, 2.99]	0.491
Congestive heart failure	7.54	[1.42, 39.93]	0.017
Hypothyroidism	2.09	[0.76, 5.74]	0.154

OSA, obstructive sleep apnea; GERD, gastroesophageal reflux disease.

Table 4 Disagreements between remote diagnosis and standard evaluation by a retinal specialist

False-positive findings by remote diagnosis[†] (n=11)	Missed by remote diagnosis[‡] (n=13)
Mild to moderate NPDR (n=8)	PDR (n=1)
Venous tortuosity (n=1)	Severe NPDR (n=2)
Macular lesions (n=1)	Mild to moderate NPDR (n=6)
Possible retinoschisis in the periphery (n=1)	Cystoid macular edema (n=1)
	Dry AMD (n=3)

[†], indicates retinal findings noted on remote diagnosis evaluation but were not found on standard evaluation by a retinal specialist; [‡], indicates retinal findings that were missed on remote diagnosis evaluation but were diagnosed on standard evaluation by a retinal specialist. NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy; AMD, age-related macular degeneration.

(CFP) are taken by expert photographers in mobile vans or dedicated brick-and-mortar imaging centers, then transmitted to grading experts. These approaches often suffer from high costs, requiring dedicated vehicles or office space, as well as photographers and graders. While CFP is the current standard in teleophthalmology care of retinal disease (7,9,33), we employed an UWF Optomap camera. Despite images being taken on non-dilated pupils by non-expert imagers, we found a relatively low percentage of ungradable images (15%) in our study. Our rate of ungradable images was higher than the rate of 6.1% previously published (11) which might be secondary to patient demographic factors (e.g., age, difficult positioning), different UWF imaging devices (Primary *vs.* Daytona), imager skill, or applied protocol (e.g., not to disturb clinic flow, we did not encourage our imagers to retake

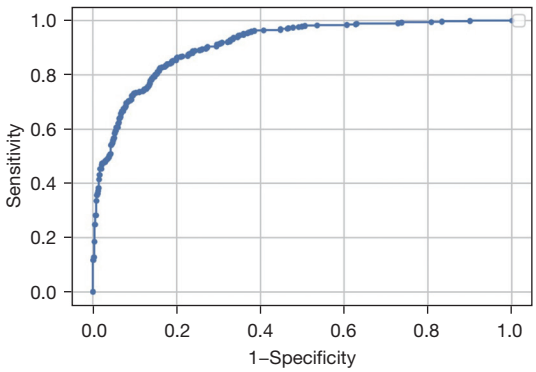

Figure 4 Receiver operating characteristic (ROC) curve of our model for referrable retinal pathology detection from Optomap images.

Table 5 Deep learning model performance metrics

Metric	Performance
Accuracy	82.8% (95% CI: 80.3–85.2%)
False negative rate	19.0% (95% CI: 16.4–21.5%)
Sensitivity	81.0% (95% CI: 78.5–83.6%)
Specificity	73.5% (95% CI: 70.6–76.3%)
Positive predictive value	81.4% (95% CI: 79.0–84.0%)
F1 score	81.4% (95% CI: 78.9–83.9%)
AUROC	81.0% (95% CI: 78.5–83.6%)

AUROC, area under the receiver operating characteristics curve.

substandard images).

Our study demonstrated that in an environment with relatively high disease prevalence, 46% of screened patients required further ophthalmology attention. Out of all referred patients, 22% had some stage of DR and by consensus were referred to a retina specialist, 5% had retina-referable incidental finings (most commonly age-related macular degeneration, data not shown), and 19% had incidental findings referable to comprehensive ophthalmology (most commonly choroidal nevi and pigmentary changes, data not shown). We found that various factors were associated with an increased likelihood of being referred to a retina specialist. We found that higher HbA1c, chronic kidney disease, and obesity increased that likelihood. While this study was conducted in a population of relatively high disease prevalence (endocrinology clinic patient population), if this approach is used in a population of low disease prevalence, we would set different thresholds of various demographic and clinical characteristics (e.g., HbA1c levels) to potentially determine whom to screen with imaging. This would enrich the pre-test probability of disease in the screened population, thereby increasing the post-test probability, minimizing the false positive rate for referrals, and making the program more cost-effective and available to wider range of healthcare providers.

To potentially decrease the time and cost associated with image grading and information transfer, we developed a DL model to identify the presence of retinal pathology from acquired images. Our model achieved a sensitivity of 81%, specificity of 73%, and an AUROC of 81%. The combination of remote diagnosis with an accurate DL model would, in theory, bypass the need for expert graders to determine referable retinal pathologies and further improve the cost-effectiveness of screening, effectively facilitating the delivery of the grand promise of teleophthalmology.

Our remote diagnosis approach using Optomap images had a moderate agreement with a standard clinical evaluation by a retinal specialist or comprehensive ophthalmologist, yielding a Cohen's kappa of 0.58. This was lower than our previous work (17) using an imaging device combining CFP and OCT, which achieved almost perfect agreement (Cohen's kappa >0.80). The reason may be that many patients concurrently underwent the screening and standard examination in the prior study. Anyhow, this finding elucidates the need for further investigations on what might be the most appropriate imagining modality in a teleophthalmology setting. Multiple components are important for the efficient logistics behind successful DR telemedicine program, and the screening system is the critical one. The screening unit besides providing good image quality on the non-mydriatic pupil, has to also be cost-effective, of adequate size, and easy to use with a short time for image acquisition.

Our study has some limitations. The screened population was homogenous, consisting of diabetic patients seen at an endocrinology clinic. As such, the baseline prevalence of DR was relatively high. The performance of this approach in a population with relatively low disease prevalence (e.g., primary care clinics) needs to be validated. Additionally, we did not assess for retinal lesion distribution relative to the standard ETDRS photographic field. This might be needed to better understand the retinal area of interest for capture and whether using UWF retinal imaging *vs.* a narrow field camera combined with OCT would be more appropriate in this setting. Another limitation of this study that likely

influenced Cohen's agreement was that not all screened patients had standard examination following the screening. We believe that this is not always feasible and, in essence, defeats the purpose of a screening program, especially when thousands of patients get screened. Nonetheless, increasing the number of these patients would likely increase our study's agreement outcome. Furthermore, while our DL model was trained simply to binarily classify Optomap images as the presence or absence of retinal pathology, future studies with larger datasets could develop models that more precisely classify retinal pathologies. For example, multilabel convolutional neural network (34) can be used to classify images as age-related macular degeneration, hypertensive retinopathy, or DR (and its stages).

Herein we demonstrate the feasibility of the remote diagnosis approach using Optos device *Primary*, combined with an automated DL model to identify referable retinal pathology. Emerging new imaging technologies with simplified interfaces that can be used in non-ophthalmic clinics by non-expert imagers, in conjunction with automated image interpretation, make remote diagnosis very attractive to take the lead in early diagnosis and monitoring of retinal pathology.

Acknowledgments

We thank Jim Dorogi and Larry Paul (Optos, Nikon) for providing the screening units and training, Duke South and South Durham Endocrinology CMAs for beautiful image capture, and Maria Maga and Joey Giattino for helping with the project logistics.
Funding: None.

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at https://aes.amegroups.com/article/view/10.21037/aes-21-53/rc

Data Sharing Statement: Available at https://aes.amegroups.com/article/view/10.21037/aes-21-53/dss

Peer Review File: Available at https://aes.amegroups.com/article/view/10.21037/aes-21-53/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://aes.amegroups.com/article/view/10.21037/aes-21-53/coif). AS reports consulting for BMS for unrelated work to the current manuscript or study. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by Institutional Review Board of Duke University (IRB00012400) and informed consent was taken from all individual participants.

References

1. Fong DS, Aiello LP, Ferris FL 3rd, et al. Diabetic retinopathy. Diabetes Care 2004;27:2540-53.
2. Zhang X, Saaddine JB, Chou CF, et al. Prevalence of diabetic retinopathy in the United States, 2005-2008. JAMA 2010;304:649-56.
3. Kempen JH, O'Colmain BJ, Leske MC, et al. The prevalence of diabetic retinopathy among adults in the United States. Arch Ophthalmol 2004;122:552-63.
4. Rein DB, Zhang P, Wirth KE, et al. The economic burden of major adult visual disorders in the United States. Arch Ophthalmol 2006;124:1754-60.
5. Javitt JC, Aiello LP, Chiang Y, et al. Preventive eye care in people with diabetes is cost-saving to the federal government: Implications for health-care reform. Diabetes Care 1994;17:909-17.
6. Javitt JC, Aiello LP. Cost-effectiveness of detecting and treating diabetic retinopathy. Ann Intern Med 1996;124:164-9.
7. Farley TF, Mandava N, Prall FR, et al. Accuracy of primary care clinicians in screening for diabetic retinopathy using single-image retinal photography. Ann Fam Med 2008;6:428-34.
8. Ogunyemi O, George S, Patty L, et al. Teleretinal

screening for diabetic retinopathy in six Los Angeles urban safety-net clinics: final study results. AMIA Annu Symp Proc 2013;2013:1082-8.
9. Chedid EH, Golden QR, Jager RD. Operational challenges in delivery of a charity care program for diabetic retinopathy screening in an urban setting. Perm J 2013;17:21-5.
10. Wong TY, Sun J, Kawasaki R, et al. Guidelines on Diabetic Eye Care: The International Council of Ophthalmology Recommendations for Screening, Follow-up, Referral, and Treatment Based on Resource Settings. Ophthalmology 2018;125:1608-22.
11. Silva PS, Horton MB, Clary D, et al. Identification of Diabetic Retinopathy and Ungradable Image Rate with Ultrawide Field Imaging in a National Teleophthalmology Program. Ophthalmology 2016;123:1360-7.
12. Ouyang Y, Heussen FM, Keane PA, et al. The retinal disease screening study: prospective comparison of nonmydriatic fundus photography and optical coherence tomography for detection of retinal irregularities. Invest Ophthalmol Vis Sci 2013;54:1460-8.
13. Early Treatment Diabetic Retinopathy Study Research Group. Grading Diabetic Retinopathy from Stereoscopic Color Fundus Photographs—An Extension of the Modified Airlie House Classification. Ophthalmology 1991;98:786-806.
14. Davis MD, Gangnon RE, Lee LY, et al. The Age-Related Eye Disease Study severity scale for age-related macular degeneration: AREDS Report No. 17. Arch Ophthalmol 2005;123:1484-98.
15. Horton MB, Silva PS, Cavallerano JD, et al. Clinical Components of Telemedicine Programs for Diabetic Retinopathy. Curr Diab Rep 2016;16:129.
16. Horton MB, Silva PS, Cavallerano JD, et al. Operational Components of Telemedicine Programs for Diabetic Retinopathy. Curr Diab Rep 2016;16:128.
17. Hadziahmetovic M, Nicholas P, Jindal S, et al. Evaluation of a Remote Diagnosis Imaging Model vs Dilated Eye Examination in Referable Macular Degeneration. JAMA Ophthalmol 2019;137:802-8.
18. Matsuba S, Tabuchi H, Ohsugi H, et al. Accuracy of ultra-wide-field fundus ophthalmoscopy-assisted deep learning, a machine-learning technology, for detecting age-related macular degeneration. Int Ophthalmol 2019;39:1269-75.
19. Nagasato D, Tabuchi H, Ohsugi H, et al. Deep Neural Network-Based Method for Detecting Central Retinal Vein Occlusion Using Ultrawide-Field Fundus Ophthalmoscopy. J Ophthalmol 2018;2018:1875431.
20. Nagasawa T, Tabuchi H, Masumoto H, et al. Accuracy of deep learning, a machine learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting idiopathic macular holes. PeerJ 2018;6:e5696.
21. Ohsugi H, Tabuchi H, Enno H, et al. Accuracy of deep learning, a machine-learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. Sci Rep 2017;7:9425.
22. Nagasawa T, Tabuchi H, Masumoto H, et al. Accuracy of Diabetic Retinopathy Staging with a Deep Convolutional Neural Network Using Ultra-Wide-Field Fundus Ophthalmoscopy and Optical Coherence Tomography Angiography. J Ophthalmol 2021;2021:6651175.
23. Tang F, Luenam P, Ran AR, et al. Detection of Diabetic Retinopathy from Ultra-Widefield Scanning Laser Ophthalmoscope Images: A Multicenter Deep Learning Analysis. Ophthalmol Retina 2021;5:1097-106.
24. Zhao M, Zhong S, Fu X, et al. Deep Residual Shrinkage Networks for Fault Diagnosis. IEEE Trans Ind Informatics 2020;16:4681-90.
25. Isogawa K, Ida T, Shiodera T, et al. Deep shrinkage convolutional neural network for adaptive noise reduction. IEEE Signal Process Lett 2018;25:224–8.
26. Wu S, Zhu Q, Xie Y. Evaluation of various speckle reduction filters on medical ultrasound images. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:1148-51.
27. Arora R, Basu A, Mianjy P, et al. Understanding deep neural networks with rectified linear units. 6th Int Conf Learn Represent ICLR 2018 - Conf Track Proc 2018.
28. Hu M. Deep learning approach for automated detection of retinal pathology from ultra-widefield retinal images 2021. Available online: https://github.com/MingzheHu-Duke/Retinopathy-Detection-from-UWF-Retinal-Images
29. Kingma DP, Ba JL. Adam: A method for stochastic optimization. 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc 2015:1-15.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
31. Rhee MK, Slocum W, Ziemer DC, et al. Patient adherence improves glycemic control. Diabetes Educ 2005;31:240-50.
32. Karter AJ, Parker MM, Moffet HH, et al. Missed appointments and poor glycemic control: an opportunity to identify high-risk diabetic patients. Med Care 2004;42:110-5.
33. Boucher MC, Desroches G, Garcia-Salinas R, et al. Teleophthalmology screening for diabetic retinopathy

through mobile imaging units within Canada. Can J Ophthalmol 2008;43:658-68.
34. Wei Y, Xia W, Lin M, et al. HCP: A Flexible CNN Framework for Multi-label Image Classification. IEEE Trans Pattern Anal Mach Intell 2016;38:1901-7.

doi: 10.21037/aes-21-53
Cite this article as: Lee T, Hu M, Gao Q, Amason J, Borkar D, D'Alessio D, Canos M, Shariff A, Pajic M, Hadziahmetovic M. Evaluation of a deep learning supported remote diagnosis model for identification of diabetic retinopathy using wide-field Optomap. Ann Eye Sci 2022;7:11.