Deep Learning Predicts OCT Measures of Diabetic Macular Thickening From Color Fundus Photographs

Filippo Arcadu,[1] Fethallah Benmansour,[1] Andreas Maunz,[1] John Michon,[2] Zdenka Haskova,[2] Dana McClintock,[2] Anthony P. Adamis,[2] Jeffrey R. Willis,[2] and Marco Prunotto[1,3]

[1]Pharma Research and Early Development (pRED), Roche Innovation Center Basel, Basel, Switzerland
[2]Genentech, Inc., San Francisco, California, United States
[3]School of Pharmaceutical Sciences, University of Geneva, Switzerland

Correspondence: Jeffrey R. Willis, Genentech, Inc., 1 DNA Way, San Francisco, CA 94080, USA; willisj5@gene.com.

JRW and MP contributed equally to the work presented here and should therefore be regarded as equivalent co-senior authors.

Submitted: September 17, 2018
Accepted: January 24, 2019

Citation: Arcadu F, Benmansour F, Maunz A, et al. Deep learning predicts OCT measures of diabetic macular thickening from color fundus photographs. *Invest Ophthalmol Vis Sci*. 2019;60:852–857. https://doi.org/10.1167/iovs.18-25634

Purpose. To develop deep learning (DL) models for the automatic detection of optical coherence tomography (OCT) measures of diabetic macular thickening (MT) from color fundus photographs (CFPs).

Methods. Retrospective analysis on 17,997 CFPs and their associated OCT measurements from the phase 3 RIDE/RISE diabetic macular edema (DME) studies. DL with transfer-learning cascade was applied on CFPs to predict time-domain OCT (TD-OCT)–equivalent measures of MT, including central subfield thickness (CST) and central foveal thickness (CFT). MT was defined by using two OCT cutoff points: 250 μm and 400 μm. A DL regression model was developed to directly quantify the actual CFT and CST from CFPs.

Results. The best DL model was able to predict CST $\geq$ 250 μm and CFT $\geq$ 250 μm with an area under the curve (AUC) of 0.97 (95% confidence interval [CI], 0.89–1.00) and 0.91 (95% CI, 0.76–0.99), respectively. To predict CST $\geq$ 400 μm and CFT $\geq$ 400 μm, the best DL model had an AUC of 0.94 (95% CI, 0.82–1.00) and 0.96 (95% CI, 0.88–1.00), respectively. The best deep convolutional neural network regression model to quantify CST and CFT had an R^2 of 0.74 (95% CI, 0.49–0.91) and 0.54 (95% CI, 0.20–0.87), respectively. The performance of the DL models declined when the CFPs were of poor quality or contained laser scars.

Conclusions. DL is capable of predicting key quantitative TD-OCT measurements related to MT from CFPs. The DL models presented here could enhance the efficiency of DME diagnosis in tele-ophthalmology programs, promoting better visual outcomes. Future research is needed to validate DL algorithms for MT in the real-world.

Keywords: deep learning, diabetic macular edema, ocular imaging, tele-ophthalmology and public health ophthalmology

Diabetic macular edema (DME) is a condition in which the retina develops diabetic microangiopathy with subsequent accumulation of fluid in the macula.[1] It is a leading cause of vision impairment in people with diabetes, compromising their function and quality of life.[2–5] In 2017, approximately 425 million people worldwide had diabetes, and this number is estimated to grow to 629 million by 2045.[6] Adults with diabetes and DME also have a substantially higher risk of cardiovascular morbidity, mortality, and amputation risk than those without, creating a further public health hazard.[7,8] When individuals with DME are not treated in a timely fashion through intravitreal anti–vascular endothelial growth factor injections, they are at risk of irreversible vision loss.[9]

The current gold standard for DME diagnosis is based on optical coherence tomography (OCT) evaluation,[10,11] which is often not available for tele-ophthalmology screening in the real-world setting owing to the high cost and technical limitations.[12] An imaging modality that is widely accessible and commonly used for tele-ophthalmology is, instead, color fundus photography (CFP), which is less expensive and easier to operate. In the future, it is also likely that patients will be able to acquire their own CFPs via smartphone,[13] further expanding the horizon of screening possibilities provided by this imaging technique. The disadvantage of CFP with respect to OCT is that the presence of macular thickening (MT) cannot be easily identified with the human eye.[14–16] The modest correlation found between MT assessed on CFPs and MT assessed on OCT[16] may be related to the inability of the human eye to characterize CFP changes and novel patterns associated to MT at the CFP pixel level. For this reason, using CFP as a standalone tool for retinal screening may lead to an inaccurate diagnosis of DME.[14–16]

A possible solution for improving the accuracy of MT detection on CFPs is the use of deep learning (DL) and, in particular, deep convolutional neural networks (DCNNs).[17] In recent years, DL has gained an incredible momentum in the field of clinical ophthalmology and has opened up new possibilities for the automated detection of anomalies and grading of retinal diseases.[18–28] Compared to feature-based machine learning, DL has the advantage of offering an end-to-end solution between raw images and a selected outcome variable. Consequently, DL does not require the specification of known clinical features to construct a detection model, as DL learns directly from raw images without being limited by a priori assumptions on the information contained by the images themselves. With this aspect, and through attribution maps

(also called focus maps), DL may eventually unravel novel predictive biomarkers that are associated with a given outcome variable after thorough clinical validation.

Within this context, the primary objective of this study was to assess whether DL can automatically predict OCT-equivalent quantitative MT measures from CFPs. Specifically, we present results related to four different DCNN models: (1) two models to detect the presence of clinically significant MT, using the cutoff points on time-domain OCT (TD-OCT) of central foveal thickness (CFT) of 250 μm and 400 μm; and (2) two models to detect the presence of clinically significant MT, using the cutoff points on TD-OCT of central subfield thickness (CST) of 250 μm and 400 μm. The cutoff point of 250 μm measured with TD-OCT is traditionally used to discriminate normal subjects from those with abnormal MT.[29] The cutoff point of 400 μm is used by the National Institute of Health and Care Excellence (NICE) in the United Kingdom to identify cases of severe DME for which ranibizumab treatment should be initiated.[30] The secondary objective was to assess the robustness of a DL regression model to predict the exact value in micrometers of CFT and CST. The tertiary objective was to assess how DL results are affected by CFP image quality and the presence of focal/grid/panretinal laser scars. The crucial difference between our study and prior studies looking at DL algorithms to detect DME is that our study aimed to predict actual quantitative OCT measurements from CFPs rather than being explicitly instructed to identify the presence of other DME markers such as exudates.[18,19]

Methods

Dataset

The combined CFP dataset of two different clinical trials, RIDE[31] and RISE,[32] was used to address the scientific questions of this work. RIDE/RISE were identically designed, phase 3, multicenter, double-masked, 24-month, sham injection–controlled randomized studies to test the efficacy of ranibizumab injections for patients affected by DME.[31,32] The eligible participants had diabetes mellitus (type 1 or 2), a best corrected visual acuity (BCVA) of 20/40 to 20/320, and a central macular thickness $\geq$ 275 μm for, at least, one eye.[33] In addition, CFPs from the Kaggle Diabetic Retinopathy challenge were used.[34] The trials adhered to the tenets of the Declaration of Helsinki, were Health Insurance Portability and Accountability Act compliant, and protocols were approved by institutional review boards, ethics committees, or as applicable. Patients provided written informed consent for future medical research and analyses, based on results of the trial.

Outcome Variables for Deep Learning Modeling

The quantitative OCT measurements, CST and CFT, were selected for the modeling among all OCT measurements collected for the study-eye for each patient at each visit. CST is defined as the average thickness in the central 1-mm-diameter circle of the ETDRS circle,[35] whereas CFT is defined as the mean thickness measured at the point of intersection of the six radial scans.[35,36] As a result, CFT is a more variable measure than CST. All TD-OCT scans were acquired with the Zeiss Stratus device (Carl Zeiss Meditec, Jena, Germany), and the CFT and CST outputs were assessed by masked readers at the central reading center.

The values in micrometers of CST and CFT were used as outcome variable for the DL regression problem. Two clinically significant thresholds for CST/CFT, namely, 250 μm and 400 μm, were used as cutoffs to construct the outcome variables for the binary classification problems. The cutoff point of 250 μm measured with TD-OCT is traditionally used to discriminate normal subjects from those with abnormal MT.[29] The cutoff point of 400 μm is used by NICE in the United Kingdom to identify cases of severe DME for which ranibizumab treatment should be initiated.[30]

Table 1. Characteristics of the Color Fundus Datasets Used to Train the Deep Learning Models

Characteristics	RIDE/RISE
No. of patients with CST	725
No. of CFPs with CST	12,374
CST $\geq$ 250 μm, in CFPs	7,041 (56.9%)
CST $\geq$ 400 μm, in CFPs	3,269 (26.4%)
CST, mean ± SD, μm	318 ± 145
No. of patients with CFT	753
No. of CFPs with CFT	17,997
CFT $\geq$ 250 μm, in CFPs	8,166 (45.3%)
CFT $\geq$ 400 μm, in CFPs	4,378 (24.3%)
CFT, mean ± SD, μm	308 ± 164

CFT, central foveal thickness[39]; CST, central subfield thickness[38]; SD, standard deviation.

Color Fundus Photographs

In RIDE/RISE, CFPs were obtained at each patient's screening visit and at months 3, 6, 12, 18, and 24. While stereoscopic seven-field photographs were captured by using cameras with a 35° setting at each visit, this study only used the fovea-centered field two photographs (FC-CFPs). The CFP dataset contains 12,374 FC-CFPs from 725 patients with associated CST measurements and 17,997 CFPs from 753 patients with associated CFT. Among these images, 56.9% ($n = 7041/12,374$) have CST $\geq$ 250 μm, 26.4% ($n = 3269/12,374$) have CST $\geq$ 400 μm, 45.3% ($n = 8166/17,997$) have CFT $\geq$ 250 μm, and 24.3% ($n = 4378/17,997$) have CFT $\geq$ 400 μm (Table 1).

Photographs were evaluated at the University of Wisconsin Fundus Photograph Reading Center (Madison, WI, USA) by trained evaluators masked to both treatment assignment and images from previous visits. These evaluators, in turn, annotated various features of the CFPs, including the presence of scars due to either focal/grid photocoagulation (FGPC) or scatter photocoagulation (SCATPC). CFPs were also annotated for their general quality (QC) based on previously described criteria by Gulshan et al.[19] Specifically, the quality of each CFP was evaluated by two instructed readers supervised by a board-certified retina specialist. The quality assessment of the CFPs was based on focus, illumination, image field definition, and presence of artifacts. In case of disagreement between the two readers, the retinal specialist served as the final adjudicator. Our study subsequently used these annotations and QC measures to conduct a sensitivity analysis of the DL performance, based on the presence of laser scars and the level of CFP image quality.

Deep Learning Algorithms

The Inception-V3[37] architecture was used to address both the binary classification and the regression tasks. This architecture offers a very good tradeoff in terms of depth (313 layers) versus number of parameters (~23 million). That is, it is very deep, as needed to better learn from images, while having a relatively small amount of parameters, which helps in preventing overfitting. The Inception-V3 models were trained by using a *transfer-learning cascade*. Transfer learning[38] is a robust and efficient technique, in which training does not start from

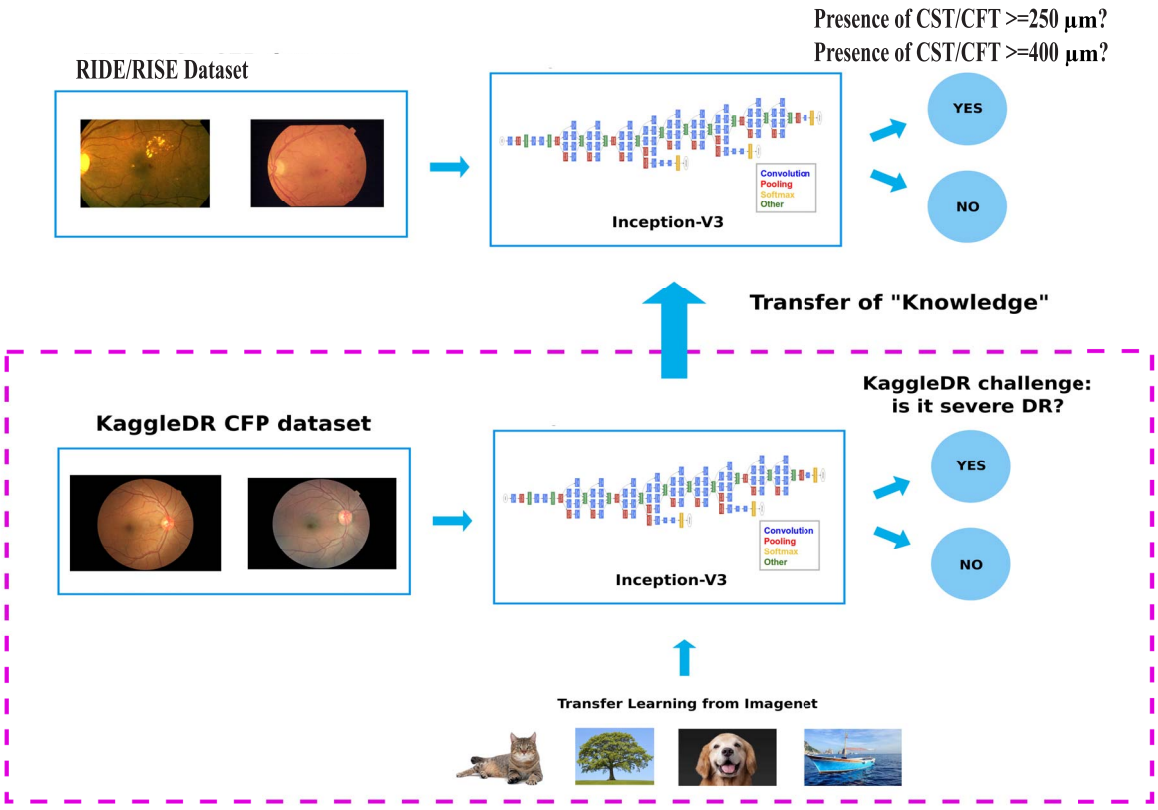

FIGURE 1. Transfer-learning cascade to learn robust DL models with a relatively small imaging dataset. The target DL model is the one on top, to be trained on the RIDE/RISE CFP dataset to detect the presence of either CST or CFT $\geq$ 250 or $\geq$ 400 µm. The training does not start from scratch, but rather from a model trained on the CFP dataset of the Kaggle Diabetic Retinopathy Detection challenge (KaggleDR) to identify CFPs with signs of severe diabetic retinopathy. The training of the KaggleDR model started in turn from a model trained to address the ImageNet challenge.[34]

scratch, but rather from a "warmed-up" model (i.e., a model trained on another dataset to address a different question). In our study, the starting point was a model trained on the Kaggle Diabetic Retinopathy challenge[34] CFP dataset, whose training, in turn, started from a model trained on the ImageNet challenge[39] dataset of natural images. The transfer-learning cascade allows to build well-performing DL models when dealing with a relatively small dataset, such as that used in this study (Fig. 1). More specifically, transfer learning involves replacing and training the softmax layer of the architecture in case of binary classification and the linear layer of the architecture in case of regression while keeping all the other layers frozen. This is followed by fine-tuning[38] of the weights throughout the network with the exception of a few initial layers close to the input. CFPs were resized to 299 × 299 pixels and normalized to [−1, 1], as required by the Inception-V3.[39] No additional preprocessing was performed on the images. The following parameters were used for training: 10 epochs for transfer-learning, 50 epochs for fine-tuning, and Adam optimizer with values of learning rate ranging in $[10^{-5}, 10^{-2}]$.

Evaluation of the DL Models

The CFP dataset was split as follows: 80% for training, 10% for testing (i.e., selection of the model), and 10% for validation (i.e., hold-out set). The split into these three sets changes for each target outcome variable.

Since the datasets contain multiple CFP acquisitions for the same patient at each visit across multiple visits, the random split occurred at the patient level and not at the level of individual CFPs. This prevents the allocation of CFPs from the same patient inside more than one set among training, testing, and validation.

The metrics to evaluate the model are computed on the validation sets, keeping just one CFP per patient visit in case multiple acquisitions are available for a certain visit. The area under the receiver operator characteristic curve (AUC) is used to assess the performance of the DL models for classifications, whereas the R^2 value is used to benchmark the DL regression model. Additionally, 95% confidence intervals (CIs) were computed with bootstrapping for every AUC and R^2 value. Sensitivity and specificity computed at Youden's operating point[40] were also reported for the best DL classification models. Since the dataset splitting changes for each outcome variable and is constrained with respect to the patient ID, the number of CFPs used to validate the different DL models also changes.

DL models are "black boxes" by construction, because the features used for prediction are learnt internally and not engineered beforehand. To gain insight into the inner workings of the DL models, we constructed attribution maps[41] created by means of guided back-propagation. These maps display the image locations that the DL model focused on to make its decision about the presence of MT. Maps were originally gray-scale images and were segmented with a threshold of 0.5 to provide cleaner and more salient pictures.

RESULTS

Binary Classifications

The best DL model was able to predict CST $\geq$ 250 µm and CST $\geq$ 400 µm with an AUC of 0.97 (95% CI, 0.89–1.00; sensitivity =

Table 2. Performance of the Deep Learning Models for Binary Classification of Macular Thickening for the Estimation of Central Subfield Thickness on Color Fundus Photographs

	Central Subfield Thickness, µm					
	≥250			≥400		
Data Subset	AUC	95% CI	*N*	AUC	95% CI	*N*
All	0.86	0.81–0.90	307	0.84	0.79–0.88	342
No SCATPC	0.89	0.84–0.93	194	0.88	0.83–0.92	207
No FGPC	0.94	0.85–1.0	36	0.92	0.84–0.99	47
No laser	0.92	0.81–1.0	31	0.92	0.81–1.0	39
QC filtering	0.89	0.85–0.93	232	0.94	0.82–1.0	28
QC filtering + no SCATPC	0.86	0.73–0.95	42	0.94	0.82–1.0	28
QC filtering + no FGPC	0.97	0.89–1.0	28	0.94	0.82–1.0	28
QC filtering + no laser	0.97	0.89–1.0	24	0.94	0.82–1.0	28

N, number of CFPs available for validation; no FGPC, no focal/grid photocoagulation; no laser, no scatter photocoagulation and no focal/grid photocoagulation; no SCATPC, no scatter photocoagulation; QC filtering, filtered out all CFPs of low quality.

87.5%; specificity = 96.4%; $N = 28$ CFPs) and of 0.94 (95% CI, 0.82–1.00; sensitivity = 99.0%; specificity = 94.4%; $N = 28$ CFPs), respectively (Table 2).

To predict CFT $\geq$ 250 µm and CFT $\geq$ 400 µm, the best DL model had an AUC of 0.91 (95% CI, 0.76–0.99; sensitivity = 80.0%; specificity = 85.0%; $N = 41$ CFPs) and of 0.97 (95% CI, 0.88–1.00; sensitivity = 90.0%; specificity = 94.0%; $N = 45$ CFPs), respectively (Table 3).

Sensitivity analyses showed that the performance of the DL models increased when the models were trained on CFPs of high quality and without laser scars. For instance, when training a DL model to detect the presence of MT of CST $\geq$ 400 µm, the AUC based on the overall validation set was 0.84, and increased to 0.92 when CFPs with signs of laser were filtered out. Moreover, for the same case, the AUC improved to 0.94 when the DL model was further trained on high-quality images only (Table 2).

Regressions

The best DCNN regression model to quantify CST and CFT had an R^2 of 0.74 (95% CI, 0.49–0.91; $N = 24$ CFPs) and 0.54 (95% CI, 0.20–0.87; $N = 24$ CFPs), respectively (Table 4).

Once again, sensitivity analysis demonstrated that the performance of the DCNN regression model increased when trained on CFPs without laser scars and of high quality. That is, when training a DL model to quantify the exact CST value from CFPs, the R^2 based on the overall validation set was 0.57, and this increased to 0.65 and 0.73 when the study filtered out images of poor quality and with laser scars, respectively.

Table 3. Performance of the Deep Learning Models for Binary Classification of Macular Thickening for the Estimation of Central Foveal Thickness on Color Fundus Photographs

	Central Foveal Thickness, µm					
	≥250			≥400		
Data Subset	AUC	95% CI	*N*	AUC	95% CI	*N*
All	0.80	0.76–0.84	479	0.87	0.82–0.91	463
No SCATPC	0.80	0.74–0.86	251	0.85	0.80–0.90	362
No FGPC	0.88	0.74–0.97	55	0.96	0.93–0.98	251
No laser	0.86	0.72–0.98	40	0.95	0.91–0.98	205
QC filtering	0.83	0.78–0.88	294	0.88	0.84–0.92	293
QC filtering + no SCATPC	0.82	0.75–0.87	176	0.87	0.81–0.92	239
QC filtering + no FGPC	0.91	0.80–0.99	41	0.97	0.89–1.0	45
QC filtering + no laser	0.91	0.76–0.99	30	0.96	0.88–1.0	37

The regression of CFT is characterized by lower performance with respect to the regression of CST and the performance drop is consistent for all the sensitivity analyses conditions (i.e., filtering out laser scars and poor-quality images). These results can be explained by the fact that the performance of a DL regression is generally more sensitive to the stability of the endpoint and the CFT is a less reliable endpoint compared to CST. The CST is defined as an average thickness over an area,[34] whereas the CFT is defined as a 1-point measure,[35] making CFT a measurement more subject to noise by definition. Consequently, while the choice of outcome variable as either CST or CFT does not significantly affect the outcome of a classification task, it does make a difference for the associated regression problem.

Examples of Attribution Maps

The attribution maps corresponding to the DCNN models for classification are shown in Figure 2. Specifically, the maps of the DCNN models for MT with CFT $\geq$ 250 µm (Figs. 2a, 2b) and CFT $\geq$ 400 µm (Figs. 2c, 2d) have signals located inside the macula or close by, focusing on hemorrhages, exudates, and vessel contours.

Table 4. Performance of the Deep Learning Models for Regression of Macular Thickening on Color Fundus Photographs

	Central Subfield Thickness			Central Foveal Thickness		
Data Subset	R^2	95% CI	*N*	R^2	95% CI	*N*
All	0.57	0.48–0.64	307	0.42	0.34–0.49	459
No SCATPC	0.65	0.55–0.73	194	0.43	0.32–0.52	276
No FGPC	0.73	0.45–0.91	36	0.51	0.40–0.63	43
No laser	0.73	0.46–0.87	31	0.52	0.39–0.67	39
QC filtering	0.65	0.56–0.72	232	0.47	0.38–0.55	300
QC filtering + no SCATPC	0.68	0.60–0.76	157	0.49	0.43–0.55	187
QC filtering + no FGPC	0.73	0.47–0.89	28	0.52	0.25–0.81	31
QC filtering + no laser	0.74	0.49–0.91	17	0.54	0.20–0.87	24

N, number of CFPs used for validation; R^2, coefficient of determination.

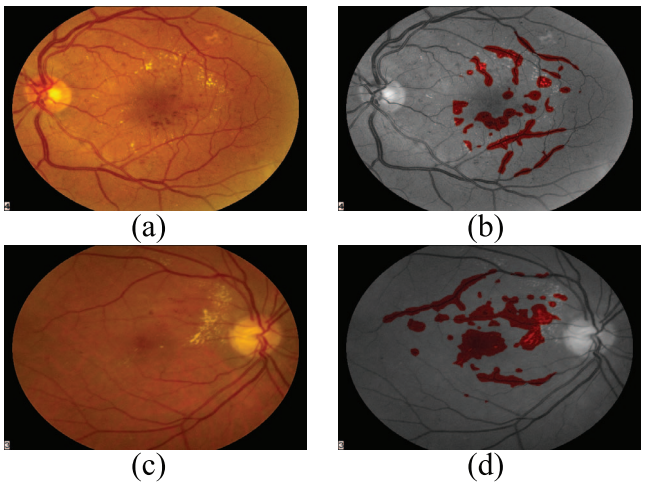

FIGURE 2. Examples of attribution maps or "hot spots" on CFPs generated by the DL classification models. (**a**, **b**) Example of map created by the DL model to detect MT with CFT > 250 μm. (**c**, **d**) Example of map created by the DL model to detect MT with CFT > 400 μm.

DISCUSSION

DL is capable of predicting quantitative OCT-equivalent measures of MT from CFPs. To our knowledge, this is the first time that DL has been shown to accurately accomplish such a challenging task in the field of ophthalmic imaging (i.e., reproducing three-dimensional clinical measurements from two-dimensional clinical images). This finding, together with what has been shown in previous studies,[18–28] underlines the value of DL in enhancing ophthalmic disease surveillance through an automated approach.

Our study showed that DL models can accurately identify which CFPs are associated with a clinically significant level of MT. Specifically, our DL models showed that they can identify CFPs with a CST $\geq$ 250 μm and $\geq$ 400 μm with an AUC of 0.97 and 0.91, respectively. Additionally, our DL models demonstrated that they can identify CFPs with a CFT $\geq$ 250 μm and $\geq$ 400 μm with an AUC of 0.91 and 0.97, respectively. Moreover, we also demonstrated the feasibility of predicting the value of CST in micrometers from CFPs ($R^2 = 0.74$; 95% CI, 0.49–0.91). These results are significantly better than those seen in previous studies, where a modest correlation of $R^2 = 0.37$ between center-point retinal thickness measured by OCT and Early Treatment Diabetic Retinopathy Study interpretation of MT at the center of the macula performed by a retina specialist on CFPs has been reported.[16] However, it is important to note that the performance of the models was not as high when the models were trained on CFPs that included images that were of poor quality or had laser scars.

Our study also showed that it is possible to construct high-performing DL models by using a relatively small dataset (in this case, the combined CFP dataset of two clinical trials, RIDE[31] and RISE[32]). This was made feasible by the use of a transfer-learning cascade approach, where the training of the DL models started after a warm-up learning phase on the Kaggle Diabetic Retinopathy[34] CFP dataset, which in turn started from another warm-up learning phase that occurred on the ImageNet[39] dataset. Furthermore, our sensitivity analyses showed the relative impact of image quality and laser scars on the final performance of the DL binary classification and regression models.

A major limitation of the study was that training was based on data from clinical trials and may not be generalizable to the overall population with diabetes. Additionally, our results may not be generalizable to macular edema secondary to other causes such as exudative age-related macular degeneration, retinal vein occlusion, or central serous chorioretinopathy. The generalizability of these results to current standards may also be limited given that the study was based on TD-OCT images, which is not the current standard. Critics may also argue that our DL model is not truly detecting MT, but rather retinal phenotypes, such as ETDRS diabetic retinopathy (DR) severity and hard exudates, that are typically correlated with MT. While it is true that MT is correlated with such phenotypes, our analysis showed that our DL model detects the presence of MT irrespective of the presence of DR severity or the presence of hard exudates (Supplemental Figs. S1, S2). Future analyses on a larger validation set will be important to ascertain the correlation between our DL models, DRSS, and the presence of hard exudates. Moreover, our work has to be considered a pilot and a proof-of-feasibility study, and future research will be needed to validate these DL models against data from other clinical trials and from the real world. Through such efforts, it will be possible to create more accurate and unbiased DL models capable of predicting OCT measures in a clinical setup.

In summary, DL is capable of automatically predicting OCT-equivalent measures of MT from CFPs and could significantly benefit tele-ophthalmology screening programs. This could contribute to earlier diagnoses of abnormal MT, timely referral to specialists, faster recruitment of patients into clinical trials, and enhanced visual/health outcomes among individuals with diabetes.

Acknowledgments

Supported by Genentech, Inc. (San Francisco, CA, USA), a member of the Roche Group. Genentech, Inc. supported and contributed to all aspects of the study, including the study design, analyses, data interpretation, report writing, and decision to submit the manuscript for publication.

Disclosure: **F. Arcadu**, Roche (E); **F. Benmansour**, Roche (E); **A. Maunz**, Roche (E); **J. Michon**, Genentech (C); **Z. Haskova**, Genentech (E); **D. McClintock**, Genentech (I); **A.P. Adamis**, Genentech (E); **J.R. Willis**, Genentech (E); **M. Prunotto**, Roche (E)

References

1. Bhagat N, Grigorian RA, Tutela A, Zarbin MA. Diabetic macular edema: pathogenesis and treatment. *Surv Ophthalmol*. 2009;54:1–32.
2. Zhang X, Saaddine JB, Chou CF, et al. Prevalence of diabetic retinopathy in the United States, 2005–2008. *JAMA*. 2010;304:649–656.
3. Fenwick EK, Xie J, Ratcliffe J, et al. The impact of diabetic retinopathy and diabetic macular edema on health-related quality of life in type 1 and type 2 diabetes. *Invest Ophthalmol Vis Sci*. 2012;53:677–684.
4. Hariprasad SM, Mieler WF, Grassi M, Green JL, Jager RD, Miller L. Vision-related quality of life in patients with diabetic macular oedema. *Br J Ophthalmol*. 2008;92:89–92.
5. Bressler NM, Varma R, Mitchell P, et al. Effect of ranibizumab on the decision to drive and vision function relevant to driving in patients with diabetic macular edema: report from RESTORE, RIDE, and RISE trials. *JAMA Ophthalmol*. 2016;134:160–166.
6. International Diabetes Federation. IDF Diabetes Atlas - 8th edition. http://www.diabetesatlas.org/. 2017. Accessed January 19, 2018.
7. Leveziel N, Ragot S, Gand E, et al.; DIAB2NEPHROGENE Study Group. Association between diabetic macular edema and

cardiovascular events in type 2 diabetes patients: a multicenter observational study. *Medicine (Baltimore)*. 2015;94:e1220.

8. Xie J, Ikram MK, Cotch MF, et al. Association of diabetic macular edema and proliferative diabetic retinopathy with cardiovascular disease: a systematic review and meta-analysis. *JAMA Ophthalmol*. 2017;135:586–593.
9. American Academy of Ophthalmology. *Preferred Practice Pattern® Guidelines: Diabetic Retinopathy*. San Francisco, CA: American Academy of Ophthalmology; 2008.
10. Baskin D. Optical coherence tomography in diabetic macular edema. *Curr Opin Ophthalmol*. 2010;21:172–177.
11. Goebel W, Kretzchmar-Gross T. Retinal thickness in diabetic retinopathy: a study using optical coherence tomography (OCT). *Retina*. 2002;22:759–767.
12. Rathi S, Tsui E, Mehta N, Zahid S, Schuman JS. The current state of teleophthalmology in the United States. *Ophthalmology*. 2017;124:1729–1734.
13. Nazari Khanamiri H, Nakatsuka A, El-Annan J. Smartphone fundus photography. *J Vis Exp*. 2017;(125).
14. Surendran TS, Raman R. Teleophthalmology in diabetic retinopathy. *J Diabetes Sci Technol*. 2014;8:262–266.
15. Mackenzie S, Schmermer C, Charnley A, et al. SDOCT imaging to identify macular pathology in patients diagnosed with diabetic maculopathy by a digital photographic retinal screening programme. *PLoS One*. 2011;6:e14811.
16. Davis MD, Bressler SB, Aiello LP, et al.; Diabetic Retinopathy Clinical Research Network Study Group. Comparison of time-domain OCT and fundus photographic assessments of retinal thickening in eyes with diabetic macular edema. *Invest Ophthalmol Vis Sci*. 2008;49:1745–1752.
17. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems 25*. Available at: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf. Accessed May 2, 2017.
18. Abramoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57:5200–5206.
19. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;315:2402–2410.
20. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124:962–969.
21. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318:2211–2223.
22. Burlina PM, Joshi N, Pekala M, et al. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol*. 2017;135:1170–1176.
23. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. *Ophthalmology*. 2018;125:549–558.
24. Prahs P, Radeck V, Mayer C, et al. OCT-based deep learning algorithm for the evaluation of treatment indication with antivascular endothelial growth factor medications. *Graefes Arch Clin Exp Ophthalmol*. 2018;256:91–98.
25. Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. *Ophthalmol Retina*. 2017;1:322–327.
26. Chen X, Xu Y, Wong DWK, et al. Glaucoma detection based on deep convolutional neural network. *Conf Proc IEEE Eng Med Biol Soc*. 2015;2015:715–718.
27. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*. 2016;123:1974–1980.
28. Muhammad H, Fuchs TJ, De Cuir N, et al. Hybrid deep learning on single widefield optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma*. 2017;26:1086–1094.
29. Yau JW, Rogers SL, Kawasaki R, et al.; Meta-Analysis for Eye Disease (META-EYE) Study Group. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*. 2012;35:556–564.
30. Ding J, Wong TY. Current epidemiology of diabetic retinopathy and diabetic macular edema. *Curr Diab Rep*. 2012;12:346–354.
31. RIDE: A Study of Ranibizumab Injection in Subjects With Clinically Significant Macular Edema (ME) With Center Involvement Secondary to Diabetes Mellitus. Registered on ClinicalTrials.gov as NCT00473382. Available at: https://clinicaltrials.gov/ct2/show/NCT00473382. Accessed May 2, 2017.
32. RISE: A Study of Ranibizumab Injection in Subjects With Clinically Significant Macular Edema (ME) With Center Involvement Secondary to Diabetes Mellitus. Registered on ClinicalTrials.gov as NCT00473330. Available at: https://clinicaltrials.gov/ct2/show/NCT00473330. Accessed May 2, 2017.
33. Nguyen QD, Brown DM, Marcus DM, et al.; RISE and RIDE Research Group. Ranibizumab for diabetic macular edema: results from 2 phase III randomized trials: RISE and RIDE. *Ophthalmology*. 2012;119:789–801.
34. Kaggle, Diabetic Retinopathy Detection, sponsored by the California Healthcare Foundation, dataset provided by EyePacs. 2017. Available at: https://www.kaggle.com/c/diabetic-retinopathy-detection. Accessed January 3, 2018.
35. Huang J, Liu X, Wu Z, et al. Macular thickness measurements in normal eyes with time domain and fourier domain optical coherence tomography. *Retina*. 2009;29:980–987.
36. Chan A, Duker JS, Tony HK, et al. Normal macular thickness measurements in healthy eyes using stratus optical coherence tomography. *Arch Ophthalmol*. 2006;124:193–198.
37. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception architecture for computer vision. *arXiv:1512.00567v3*. 2015.
38. Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? *arXiv:1411.1792v1*. 2014.
39. Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis*. 2015;115:211–252.
40. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–35.
41. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. *arXiv:1703.01365v2*. 2017.

Investigative Ophthalmology & Visual Science