Supplementary

Visualization method of ensemble DL scheme

To visualize the critical components in OCT images that are highly correlated with CFT and BCVA prediction, a popular occlusion test was used to interpret the results and increase model transparency. A blank 100×100 pixel box was systematically moved across every possible position in the image and the probabilities of the prediction were recorded. The highest drop in the probability represents the part of the OCT image that is most critical for accurate classification (shown as the red part in the Figure S1). Furthermore, whether the identified regions by occlusion test were the most clinically significant areas of predictive basis in DME eyes was further verified by our retinal specialists (YH, DC, and HY).

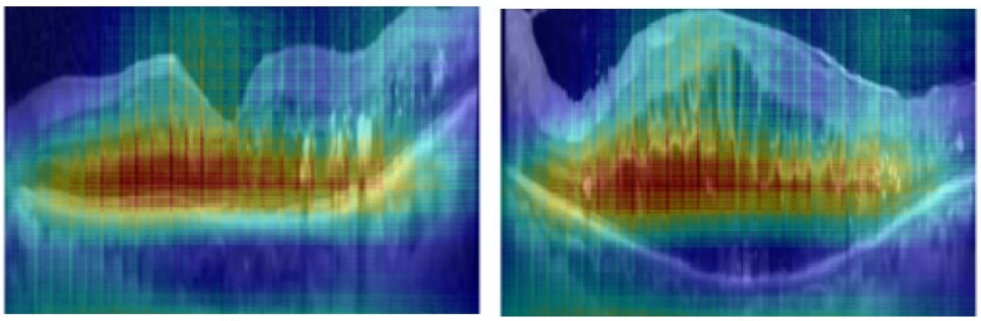

Figure S1 Occlusion test for visualization of ensemble deep learning scheme. Occlusion test successfully identified the predictive basis in the OCT images from different patterns of DME eyes. An occlusion map was generated by convolving an occluding kernel across the input image. The occlusion map is created after prediction by assigning the SoftMax probability of the correct label to each occluded area. The occlusion map could then be superimposed on the input image to represent the critical components in OCT images that showed highly correlation with the accurate prediction of CFT and BCVA in DME patients. The red part represents high correlation, while the blue part represents low correlation.

Table S1 Patient demographics

Variable	Training set	Validation set
No. of patients [female]	208 [143]	41 [22]
No. of eyes	304	59
Age, mean (SD), years	57.14 (13.90)	56.81 (13.96)
Preoperative CFT, mean (SD), μm	489.13 (214.37)	447.63 (186.36)
Postoperative CFT, mean (SD), μm	334.15 (137.53)	303.54 (92.47)
No.(percentage) of eyes responding in CFT	202 (66.45)	40 (67.80)
Preoperative BCVA, mean (SD)	0.79 (0.55)	0.57 (0.36)
Postoperative BCVA, mean (SD)	0.44 (0.41)	0.32 (0.28)
No.(percentage) of eyes responding in BCVA	200 (65.79)	37 (62.71)

No., number; SD, standard deviation; CFT, central foveal thickness; BCVA, the best-corrected visual acuity [in the logarithm of minimum angle of resolution (logMAR) unit].

http://dx.doi.org/10.21037/atm-20-1431

Table S2 The properties of applied CNNs

Networks	AlexNet	Vgg16	ResNet18
Depth	8	16	18
Parameters (millions)	61.0	138	11.7
Image input size	227-227-3	224-224-3	224-224-3

Depth means the largest number of fully-connected layers or sequential convolutional layers on a path from the input layer to output layer. Parameters were defined as the number of weights in the networks. The image input size means the required sizes of input images, in which 3 is the number of color channels and 227 or 224 is the number of pixels. VGG, visual geometry group.

http://dx.doi.org/10.21037/atm-20-1431