Supplementary Materials

Appendix 1

Module Development

Normalization of retinal photographs

In this study, we captured retinal photographs with Topcon TRC 50DX, Topcon TRC-NW400, Canon CR-1 40D and Canon CR-DGi. Retinal photographs with 8 different resolutions, 1728 x 1152, 1956 x 1934, 2160 x 1440, 2544 x 1696, 2848 x 2848, 3504 x 2336, 3696 x 2448, 3888 x 2592, were collected. Image normalization was performed using the preprocessing module to standardize inputs to similar conditions.

Data Balancing and Augmentation

Data balancing and data augmentation were applied on the fly. During training, the pre-trained ImageNet weights were used for initial weighting. Furthermore, these assessment modules were converted into TFLite models to reduce latency inference. For all tasks, input data was randomly augmented with (-0.3, 0.3) brightness adjustment, (-0.5, 0.5) contrast adjustment, (-0.5, 0.5) saturation adjustment, (-0.1, 0.1) hue adjustment, along with 60 degrees of random rotation, 20% random translation, 10% scaling and 5 degrees of shearing. All images were augmented channel-wise with means of (0.485, 0.456, 0.406) and standard deviations of (0.229, 0.224, 0.225).

Appendix 2

Comparison with Other Deep Learning Algorithms

Performances of deep learning (DL) algorithms are directly dependent on the ground truth. If the training dataset only contains high quality retinal photographs free from artifacts, the DL algorithm may exhibit perfect or near perfect results in laboratory settings that are not replicated in clinical settings. This may explain improved algorithm performance in some

previous studies compared to our algorithm. For instance, the model of Saha et al. reported accuracy, sensitivity, and specificity performances of 100% for their algorithm to distinguish image-quality[1]. However, apart from conventionally classifying retinal photographs into "accept" or "reject", an "ambiguous" category referring to retinal photographs with intergrader discrepancies were excluded from their study. As Coyner et al. also acknowledged, Saha et al. considered only a bimodal distribution that inadequately reflects image capture in a clinical setting[2]. This overly stringent classification of images may give a false impression that the algorithm excelled in laboratory settings that may not reflect the clinical settings. Although our training dataset included some retinal photographs of lower quality which resulted in a relatively less perfect performance, our pre-diagnosis module was trained using a more realistic dataset that was validated on unseen datasets from different populations which would offer greater reliability in a real-world implementation.

Appendix 3

Analysis of Misclassified Retinal Photographs

We also analyzed the numbers and reasons for misclassified cases. Examples of misclassified retinal photographs can be found in Figure S1-4.

In the image-quality assessment, 77 out of 222 (34.7%) misclassified gradable retinal photographs, and 4 out of 14 (28.6%) misclassified ungradable retinal photographs showed either or both characteristics: 1) approximately 25% of the peripheral area of the retina was unobservable due to artifacts, or 2) less than 25% of the peripheral area of the retina was unobservable due to artifacts, but the center region was affected by minor artifacts. It is understandable for the module to misclassify these retinal photographs as they were considered "borderline" with respect to our definition of gradability. In the laterality-of-the-eye assessment module, 31 out of 96 (32.3%) misclassified right-eye retinal photographs, and

12 out of 65 (18.5%) misclassified left-eye retinal photographs were considered “borderline”. This may suggest that the “borderline” image quality may also have affected the laterality-of-the-eye assessment.

Appendix 4 Supplementary figures, tables, and videos.

Figure S1. Examples of misclassified gradable retinal photographs. A gradable retinal photograph should fulfil both of the following criteria: 1) less than 25% of the peripheral area of the retina was unobservable due to artifacts, and 2) the center region of the retina was absent of significant artifacts. a-b) approximately 25% of the peripheral area of the retina was unobservable due to artifacts, c) less than 25% of the peripheral area of the retina was unobservable due to artifacts, but the center region was affected by minor artifacts, d) gradable retinal photographs with fundus abnormalities.

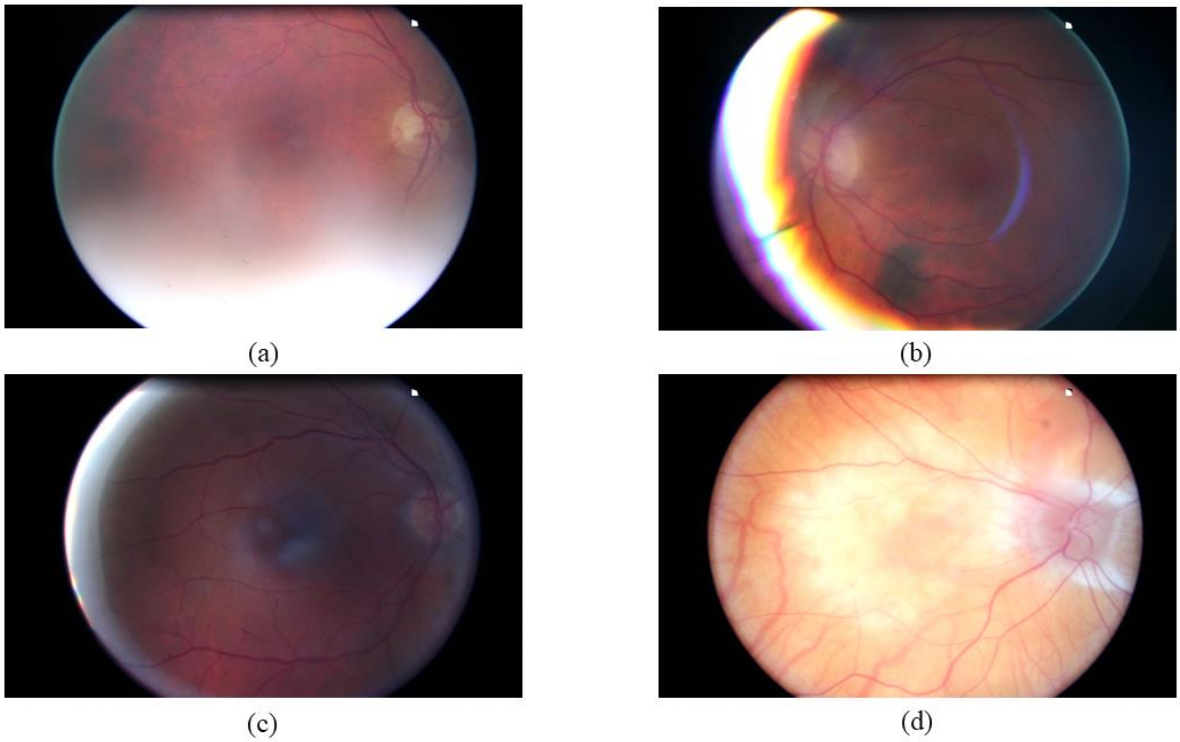

Figure S2. Examples of misclassified ungradable retinal photographs. A gradable retinal photograph should fulfil both of the following criteria: 1) less than 25% of the peripheral area of the retina was unobservable due to artifacts, and 2) the center region of the retina was absent of significant artifacts. a) about 25% of the peripheral area of the retina was unobservable due to artifact, b) less than 25% of the peripheral area of the retina was unobservable due to artifacts, but the center region was affected by minor artifacts, c-d) blurry retinal photographs.

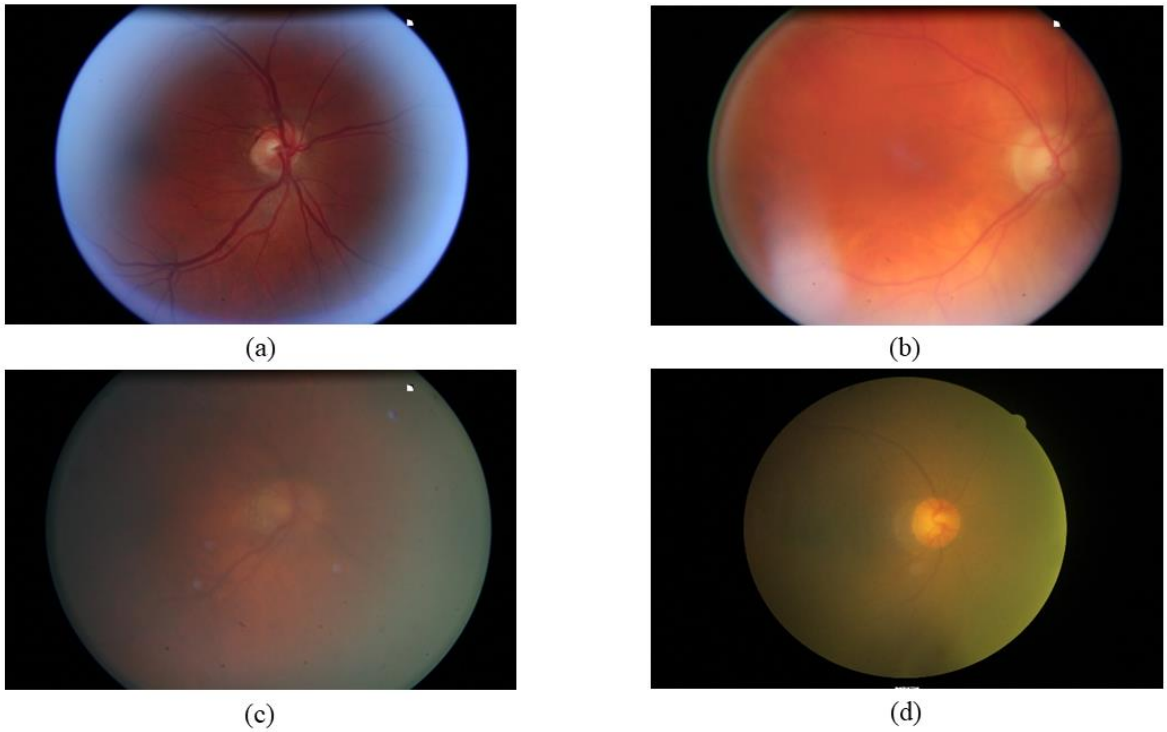

Figure S3. Examples of misclassified right-eye retinal photographs. a-b) misclassified right-eye retinal photographs with image-quality that was not considered borderline, c) misclassified right-eye retinal photograph with approximately 25% of the peripheral area of the retina was unobservable due to artifacts, d) misclassified right-eye retinal photograph with less than 25% of the peripheral area of the retina was unobservable due to artifacts, but the center region was affected by minor artifacts.

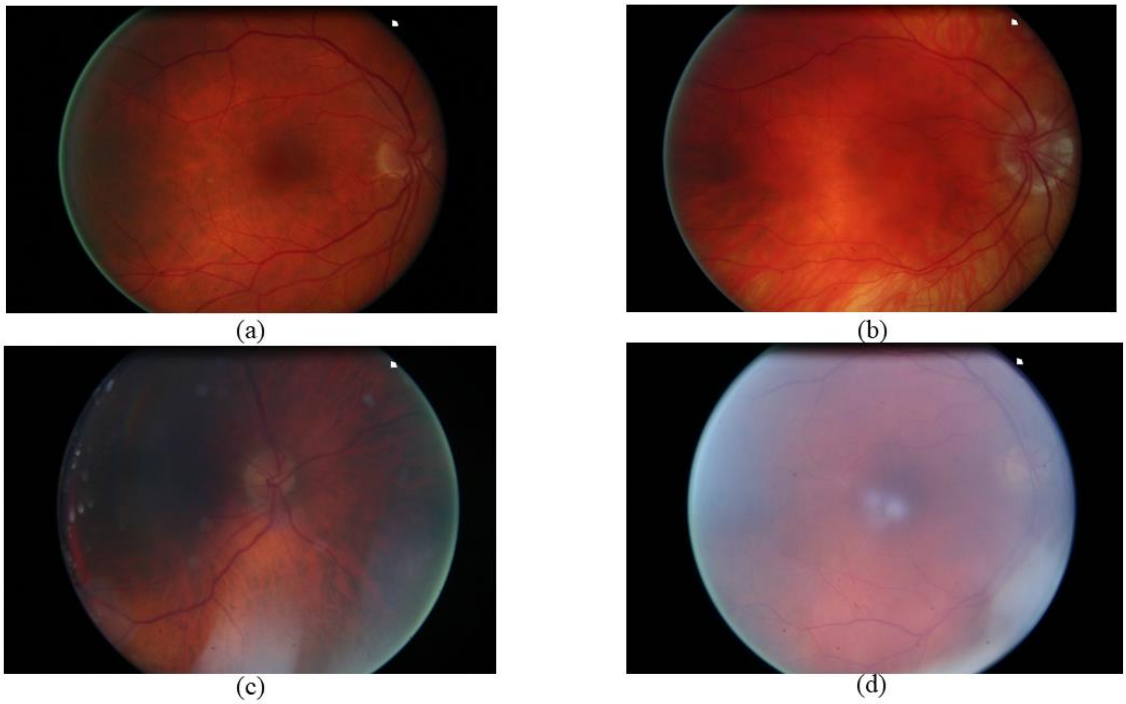

Figure S4. Examples of misclassified left-eye retinal photographs. a-b) misclassified left-eye retinal photographs with image-quality that was not considered borderline, c) misclassified left-eye retinal photograph with approximately 25% of the peripheral area of the retina was unobservable due to artifacts, d) misclassified left-eye retinal photograph with less than 25% of the peripheral area of the retina was unobservable due to artifacts, but the center region was affected by minor artifacts.

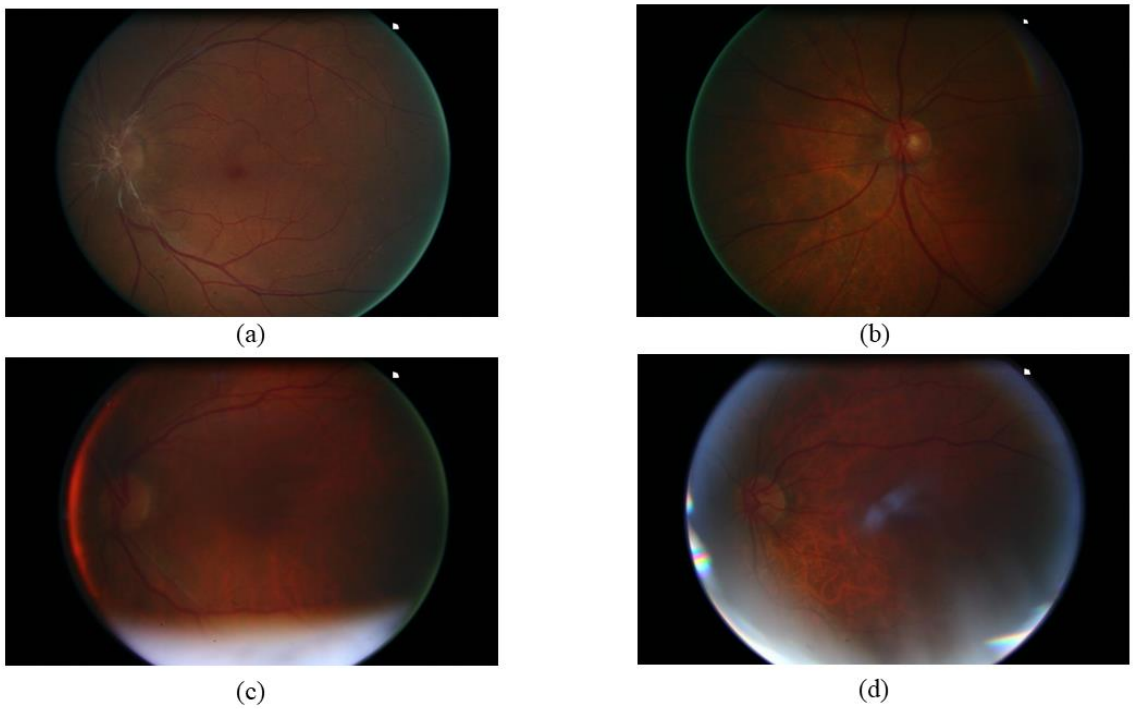

Figure S5. Potential real-life implementation of retinal fundus imaging in clinical settings.

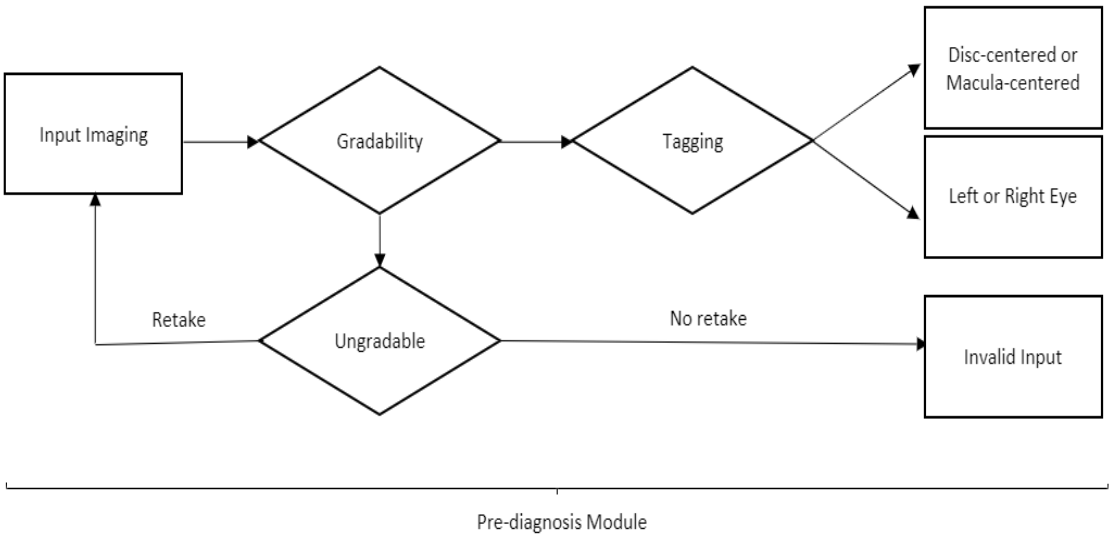

Table S1 Comparison of the image-quality assessment.

Study	Year	Mixed dataset	Testing dataset	Accuracy	Sensitivity	Specificity	AUROC
Tennakoon et al.[3]	2016	No	Internal validation	98.27%	99.12%	97.46%	/
Mahapatra et al.[4]	2016	No	Internal validation	97.90%	98.20%	97.80%	/
Yu et al.[5]	2017	No	Internal validation	95.42%	96.63%	93.10%	0.954
Saha et al.[1]	2017	No	Internal validation	100%	100%	100%	/
Zago et al.[6]	2018	No	Internal validation	87.73% - 97.10%	75.46% - 95.65%	98.55% - 100%	0.974 - 0.998
		No	External testing	94.00% - 98.55%	92.00% - 97.10%	96.00% - 100%	0.986 - 0.999
Coyner et al.[2]	2019	No	Internal validation	/	/	/	0.953 - 0.965
		No	External testing	/	93.90%	83.60%	0.965
Chalakkal et al.[7]	2019	No	Internal validation	97.50%	98.40%	95.20%	/
			External testing	85.00% - 99.80%	85.00% - 98.90%	91.20% - 98.10%	/
Zapata et al.[8]	2020	No	Internal validation	97.40%	98.30%	96.60%	0.989
Rim et al[9*]	2020	Yes	Internal validation	99.00%	/	/	/

			External testing	99.20% - 100%	/	/	/
Our study	2020	Yes	Internal validation	92.50%	92.10%	98.30%	0.975
			External testing	95.10% - 99.30%	95.00% - 99.30%	96.40% - 100%	0.987-0.999

* The study of Rim et al. performed both field-of view and laterality-of-the-eye assessments concurrently.

Table S2 Comparison of the field-of-view assessment.

Study	Year	Mixed dataset	Testing dataset	Accuracy	Sensitivity	Specificity	AUROC
Bellemo et al.[10]	2018	Yes	Internal validation	97.20%	97.60%	96.70%	0.990
			External testing	93.80% - 99.40%	90.10% - 99.00%	97.10% - 99.60%	0.955 - 0.993
Rim et al.[9*]	2020	Yes	Internal validation	99.00%	/	/	/
			External testing	99.20% - 100%	/	/	/
Our study	2020	Yes	Internal validation	100%	100%	100%	1.000
			External testing	100%	100%	100%	1.000

* The study of Rim et al. performed both field-of-view and laterality-of-the-eye assessments concurrently.

Table S3 Comparison of the laterality-of-the-eye assessment.

Study	Year	Mixed dataset	Testing dataset	Accuracy	Sensitivity	Specificity	AUC
Roy et al.[11]	2016	No	Internal validation	94.00%	/	/	0.990
Raju et al.[12]	2017	No	Internal validation	93.30%	/	/	/
Jang et al.[13]	2018	No	Internal validation	98.98%	/	/	/
Bellemo et al.[10]	2018	Yes	Internal validation	95.70%	95.90%	95.50%	0.978
			External testing	93.80% - 96.70%	94.30% - 96.70%	91.90% - 96.80%	0.956 - 0.976
Lai et al.[14]	2019	No	Internal validation & external testing	93.14% - 98.47%	/	/	/
Liu et al.[15]	2019	No	Internal validation	99.13%	99.18%	99.08%	/
			External testing	99.02%	99.00%	99.10%	0.995
Zapata et al.[8]	2020	No	Internal validation	97.40%	98.30%	96.60%	0.989
Our study	2020	Yes	Internal validation	100%	100%	100%	1.000

			External testing	94.80% - 99.70%	94.00% - 99.70%	95.80% - 99.70%	0.985 – 0.999

Table S4 Numbers and possible explanations of misclassified retinal photographs in image-quality, field-of-view and laterality-of-the-eye assessments.

Assessment	Category	Number of misclassified photographs	Possible Explanations
Image-quality assessment	Gradable but predicted ungradable	222	34.7% of misclassified retinal photographs were perceived as "borderline". It is understandable for the module to misclassify these retinal photographs.
	Ungradable but predicted gradable	14	28.6% of misclassified retinal photographs were perceived as "borderline". It is understandable for the module to misclassify these retinal photographs.
Field-of-view assessment	Optic disc-centered but predicted macula-centered	0	There were no misclassified retinal photographs in this category.
	Macula-centered but predicted optic disc-centered	0	There were no misclassified retinal photographs in this category.
Laterality-of-the-eye assessment	Right-eye but predicted left-eye	96	The image-quality of 32.3% of misclassified retinal photographs was perceived as "borderline". The "borderline" image quality may also have affected the laterality-of-the-eye assessment.
	Left-eye but predicted right-eye	65	The image-quality of 18.5% of misclassified retinal photographs was perceived as "borderline". The "borderline" image quality may also have affected the laterality-of-the-eye assessment.

Videos

The two videos, *single* and *batch*, demonstrate the use of the cloud-based pre-diagnosis module. We can upload retinal photographs to the module, and then receive image-quality (gradable vs. ungradable), field-of-view (macula-centered vs. optic disc-centered), and laterality-of-the-eye (right vs. left) assessments. The module can analyze retinal photographs separately, as shown in *single*, or together, as shown in *batch*. In the video, we uploaded two ungradable retinal photographs and four gradable retinal photographs, including one macula-centered right-eye retinal photograph, one optic disc-centered right-eye retinal photograph, one macula-centered left-eye retinal photograph, and one optic disc-centered left eye retinal photograph. As we can see, the module successfully assessed all the retinal photographs with 100% accuracy.

References:

1. Saha S, Fernando B, Cuadros J, Xiao D, Kanagasingam Y. Deep Learning for Automated Quality Assessment of Color Fundus Images in Diabetic Retinopathy Screening. ArXiv. 2017;abs/1703.02511.
2. Coyner AS, Swan R, Campbell JP, et al. Automated Fundus Image Quality Assessment in Retinopathy of Prematurity Using Deep Convolutional Neural Networks. *Ophthalmol Retina.* 2019;3(5):444-450.
3. Tennakoon RB, Mahapatra D, Roy PK, Sedai S, Garnavi R. Image Quality Classification for DR Screening Using Convolutional Neural Networks. 2016.
4. Mahapatra D, Roy PK, Sedai S, Garnavi R. Retinal Image Quality Classification Using Saliency Maps and CNNs. 2016; Cham
5. u H, Agurto C, Barriga S, Nemeth SC, Soliz P, Zamora G. Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening. Paper presented at: 2012 IEEE Southwest Symposium on Image Analysis and Interpretation; 22-24 April 2012, 2012.
6. Zago GT, Andreão RV, Dorizzi B, Teatini Salles EO. Retinal image quality assessment using deep learning. *Computers in Biology and Medicine.* 2018;103:64-70.
7. Chalakkal RJ, Abdulla WH, Thulaseedharan SS. Quality and content analysis of fundus images using deep learning. *Comput Biol Med.* 2019;108:317-331.
8. Zapata MA, Royo-Fibla D, Font O, et al. Artificial Intelligence to Identify Retinal Fundus Images, Quality Validation, Laterality Evaluation, Macular Degeneration, and Suspected Glaucoma. *Clin Ophthalmol.* 2020;14:419-429.
9. Rim TH, Soh ZD, Tham Y-C, et al. Deep Learning for Automated Sorting of Retinal Photographs. *Ophthalmol Retina.* 2020.
10. Bellemo V, Yip MYT, Xie Y, et al. Artificial Intelligence Using Deep Learning in Classifying Side of the Eyes and Width of Field for Retinal Fundus Photographs. 2019; Cham.
11. Roy PK, Chakravorty R, Sedai S, Mahapatra D, Garnavi R. Automatic Eye Type Detection in Retinal Fundus Image Using Fusion of Transfer Learning and Anatomical Features. Paper presented at: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA); 30 Nov.-2 Dec. 2016, 2016.
12. Raju M, Pagidimarri V, Barreto R, Kadam A, Kasivajjala V, Aswath A. Development of a Deep Learning Algorithm for Automatic Diagnosis of Diabetic Retinopathy. *Stud Health Technol Inform.* 2017;245:559-563.
13. Jang Y, Son J, Park KH, Park SJ, Jung K-H. Laterality Classification of Fundus Images Using Interpretable Deep Neural Network. *Journal of Digital Imaging.* 2018;31(6):923-928.
14. Lai X, Li X, Qian R, Ding D, Wu J, Xu J. Four Models for Automatic Recognition of Left and Right Eye in Fundus Images. 2019; Cham.
15. Liu C, Han X, Li Z, et al. A self-adaptive deep learning method for automated eye laterality detection based on color fundus photography. *PLOS ONE* 2019;14:e0222025.