tvst

Deep-Learning–Based Pre-Diagnosis Assessment Module for Retinal Photographs: A Multicenter Study

Vincent Yuen[1,*], Anran Ran[1,*], Jian Shi[1], Kaiser Sham[1], Dawei Yang[1], Victor T. T. Chan[1], Raymond Chan[1], Jason C. Yam[1,2], Clement C. Tham[1,2], Gareth J. McKay[3], Michael A. Williams[4], Leopold Schmetterer[5–11], Ching-Yu Cheng[5,6], Vincent Mok[12], Christopher L. Chen[13], Tien Y. Wong[5,6], and Carol Y. Cheung[1]

[1] Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong
[2] Hong Kong Eye Hospital, Hong Kong
[3] Center for Public Health, Royal Victoria Hospital, Queen's University Belfast, Belfast, UK
[4] Center for Medical Education, Royal Victoria Hospital, Queen's University Belfast, Belfast, UK
[5] Singapore Eye Research Institute, Singapore National Eye Center, Singapore
[6] Ophthalmology and Visual Sciences Academic Clinical Programme, Duke-NUS Medical School, Singapore
[7] SERI-NTU Advanced Ocular Engineering (STANCE) Program, Nanyang Technological University, Singapore
[8] School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore
[9] Department of Clinical Pharmacology, Medical University of Vienna, Vienna, Austria
[10] Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria
[11] Institute of Molecular and Clinical Ophthalmology, Basel, Switzerland
[12] Gerald Choa Neuroscience Center, Therese Pei Fong Chow Research Center for Prevention of Dementia, Lui Che Woo Institute of Innovative Medicine, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong
[13] Memory, Aging and Cognition Center, Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Purpose: Artificial intelligence (AI) deep learning (DL) has been shown to have significant potential for eye disease detection and screening on retinal photographs in different clinical settings, particular in primary care. However, an automated pre-diagnosis image assessment is essential to streamline the application of the developed AI-DL algorithms. In this study, we developed and validated a DL-based pre-diagnosis assessment module for retinal photographs, targeting image quality (gradable vs. ungradable), field of view (macula-centered vs. optic-disc-centered), and laterality of the eye (right vs. left).

Methods: A total of 21,348 retinal photographs from 1914 subjects from various clinical settings in Hong Kong, Singapore, and the United Kingdom were used for training, internal validation, and external testing for the DL module, developed by two DL-based algorithms (EfficientNet-B0 and MobileNet-V2).

Results: For image-quality assessment, the pre-diagnosis module achieved area under the receiver operating characteristic curve (AUROC) values of 0.975, 0.999, and 0.987 in the internal validation dataset and the two external testing datasets, respectively. For field-of-view assessment, the module had an AUROC value of 1.000 in all of the datasets. For laterality-of-the-eye assessment, the module had AUROC values of 1.000, 0.999, and 0.985 in the internal validation dataset and the two external testing datasets, respectively.

Conclusions: Our study showed that this three-in-one DL module for assessing image quality, field of view, and laterality of the eye of retinal photographs achieved excellent performance and generalizability across different centers and ethnicities.

Translational Relevance: The proposed DL-based pre-diagnosis module realized accurate and automated assessments of image quality, field of view, and laterality of

1

the eye of retinal photographs, which could be further integrated into AI-based models to improve operational flow for enhancing disease screening and diagnosis.

Introduction

Artificial intelligence (AI) deep learning (DL) algorithms developed for retinal photographs have shown highly accurate detection and diagnosis of major eye diseases (e.g., diabetic retinopathy,[1–6] age-related macular degeneration,[7–9] glaucoma[10–12]), measurement of retinal vessel caliber,[13] retinal vessel segmentation,[14,15] and even estimation of cardiovascular risk factors.[16–18] As such, integrating DL algorithms into real-time clinical workflow is a priority to realize the significant potential of AI for clinical diagnosis and disease risk stratification.[3,19–21] However, although individual DL algorithms have shown promising results in laboratory and research settings, the performance of many of these DL algorithms in real-world clinical settings requires further evaluation.[18]

A major challenge is that retinal photographs captured from real-world clinical settings may have lower quality than the retinal photographs carefully curated and used specifically in the developments of DL algorithms,[3,22–24] Thus, the performance of such DL algorithms is less reliable when applied clinically.[3,24] For example, Abràmoff et al.[3] reported that, although their DL algorithm achieved sensitivity of 97% in a retrospective dataset under a laboratory setting, the performance dropped to 87.2% in a prospective study conducted in a primary-care setting. In another prospective study conducted by Beede et al.,[24] about 21% of retinal photographs were unsuitable for DL-based diabetic retinopathy screening because of low image quality. These were likely due to the exclusion of low-quality retinal photographs when training DL algorithms for eye disease diagnosis.[6,25–28] As such, application of these algorithms in real-world clinical settings would require the exclusion of retinal photographs of low image quality to prevent deterioration of diagnostic performance.[29–32] In addition to image-quality assessment, DL can further provide other useful information such as field of view and laterality of the eye before disease diagnosis by subsequent DL processing algorithms. For example, DL algorithms developed for optic disc diseases (e.g., papilledema, glaucoma) should focus on optic-disc-centered retinal photographs, as the algorithm may work less well for macula-centered retinal photographs.[12,33,34] Indeed, previous studies have developed algorithms for pre-diagnosis assessments with satisfactory results, but these were largely trained and tested on datasets from a single study population,[29,35–40] whereas others developed from multiple cohorts have not undergone external testing by independent datasets.[35–38]

To address these gaps, we developed a three-in-one, DL-based pre-diagnosis module for retinal photographs to assess three tasks (i.e., image quality, field of view, and laterality of the eye) simultaneously. We further tested the module in two external unseen datasets to evaluate the generalizability. This module offers a useful prior application for a range of AI-DL algorithms for eye disease detection and other related tasks.

Materials and Methods

This study adhered to the tenets of the Declaration of Helsinki, and the protocols were approved by the institutional review board of the Chinese University of Hong Kong (CUHK), Hong Kong; the Hospital Authority Kowloon Central Cluster and New Territories East Cluster, Hong Kong; the National Healthcare Group Domain-Specific Review Board, Singapore; and Queen's University Belfast, United Kingdom. Informed consent was waived based on the retrospective design of the study, anonymized dataset of retinal photographs, minimal risk, and confidentiality protections.

Data Collection

A total of 21,348 retinal photographs from 1914 patients, composed of primary and external datasets, were used in this study. Table 1 summarizes the number of retinal photographs, number of eyes, number of subjects, image format, image size, and retinal camera used, as well as age, gender, ethnicity, and pharmacological pupil dilation of the subjects. For training and internal validation (primary dataset), we collected and used retrospective datasets from the Study of Novel Retinal Imaging Biomarkers for Cognitive Decline and the CUHK healthy volunteer cohort recruited from the CUHK Eye Center, Hong Kong (Internal-1), as well as a cohort study recruited from the memory clinic at the National University Hospital, Singapore (Internal-2).[41,42] For external testing, we used unseen datasets

Table 1. Summary of Internal and External Datasets

Dataset	Retinal Photographs, n	Eyes, n	Subjects, n	Image Format	Image Size (pixels)	Retinal Camera	Age (yr), Mean (SD)	Gender, Female, n (%)	Ethnicity, n (%)	Pharmacologic Pupil Dilation?
Internal-1	1205	340	173	.tif	3696 × 2448	Topcon TRC-50DX Mydriatic, 50°	69.3 (8.7)	109 (63.0)	Chinese, 173 (100)	Yes
Internal-2	13,217	1299	693	.jpg	2160 × 1440, 2848 × 2848, 3504 × 2336, 3888 × 2592	Canon CR-1 with EOS 40D, Non-Mydriatic, 45°	74.8 (7.8)	404 (58.3)	Chinese, 564 (81.4); Malay, 69 (9.96); Indian, 46 (6.64); other, 14 (2.02)	Yes
External-1	2385	1021	514	.jpg, .tif	1956 × 1934	Topcon TRC-NW400 Non-Mydriatic, 45°	40.2 (6.2)	336 (65.4)	Chinese, 509 (99.0); other, 5 (0.973)	No
External-2	4541	1036	534	.jpg	1728 × 1152, 2544 × 1696, 3504 × 2336	Canon CR-DGi Non-Mydriatic, 45°	78.0 (7.4)	326 (61.0)	United Kingdom, 534 (100)	Yes
Total	21,348	3696	1914	—	—	—	—	—	—	—
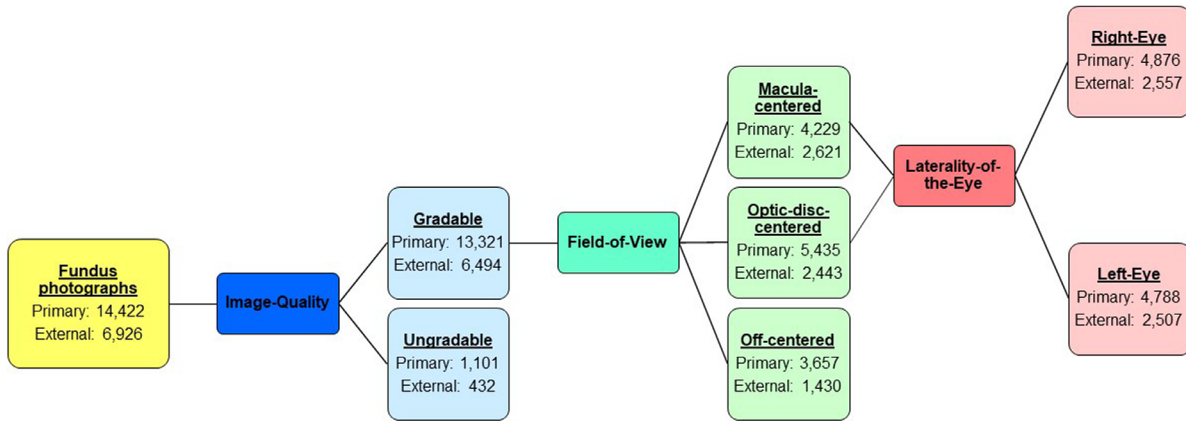
Figure 1. Ground-truth labeling and the data distribution of the primary and external datasets.

from the parental cohort of the Hong Kong Children Eye Study, Hong Kong[43] (External-1) and from the Study of the Prevalence of Age-Related Macular Degeneration in Alzheimer's Disease, the Queen's University Belfast, United Kingdom (External-2).[44]

Ground Truth Labeling

Each retinal photograph was labeled for three classifications: (1) image quality, (2) field of view, and (3) laterality of the eye. Image quality was assessed first. Only gradable retinal photographs underwent subsequent field-of-view and laterality-of-the-eye assessments. Three trained human graders (graders 1, 2, and 3) labeled the retinal photographs as "ground-truth" at the CUHK Ophthalmic Reading Center. Grader 1 was an ophthalmologist with more than 5 years in ocular imaging, and graders 2 and 3 were well-trained medical students with basic knowledge in retinal fundus examination. All graders underwent a reliability test for the image-quality assessment using a separate dataset consisting of 157 gradable and 42 ungradable retinal photographs, respectively. The inter-grader reliability was high, with Cohen's kappa coefficients ranging from 0.868 to 0.925. Grader 1, as the senior grader, made final decisions on retinal photographs for which grader 2 and grader 3 could not make decisions (e.g., retinal photographs with borderline quality).

We used the following definitions for the ground-truth labeling, and Figure 1 shows the process and numbers for each category. For the image-quality assessment, each retinal photograph was classified as gradable or ungradable. A gradable retinal photograph had to fulfill both of the following criteria: (1) less than 25% peripheral area of the retina was unobservable due to artifacts, including the presence of foreign objects, out-of-focus imaging, blurring, and extreme

illumination conditions[45]; and (2) the center region of the retina had no significant artifacts that would affect analysis. The center region was defined as the circular region of radius (in pixels) with the largest integer not greater than one-tenth the width of the image at the center of the image (the green circle in Figs. 2a and 2b), whereas the remaining area of the retinal photograph was considered the peripheral area. Figure 2a also shows examples of gradable and ungradable retinal photographs. For the field-of-view assessment, all gradable retinal photographs were further classified as (1) macula-centered, a retinal photograph with the fovea in the center region; (2) optic-disc-centered, a retinal photograph with the optic disc in the center region; or (3) off-centered, a retinal photograph with neither the fovea nor the optic disc in the center region. Examples of macula-centered, optic-disc-centered, and off-centered retinal photographs can be found in Figure 2b. For the laterality-of-the-eye assessment, all gradable macula-centered and optic-disc-centered retinal photographs were further classified as (1) right-eye, a retinal photograph with the optic disc on the right of the macula; or (2) left-eye, a retinal photograph with the optic disc on the left of the macula.

Development of the Pre-Diagnosis Module

The pre-diagnosis module consists of one pre-processing model, with three additional models for classification tasks. As different retinal cameras captured retinal photographs with different image resolutions and degrees of view, image pre-processing was first performed to normalize the inputs to similar conditions. More details of image pre-processing, data balancing, and augmentation can be found in
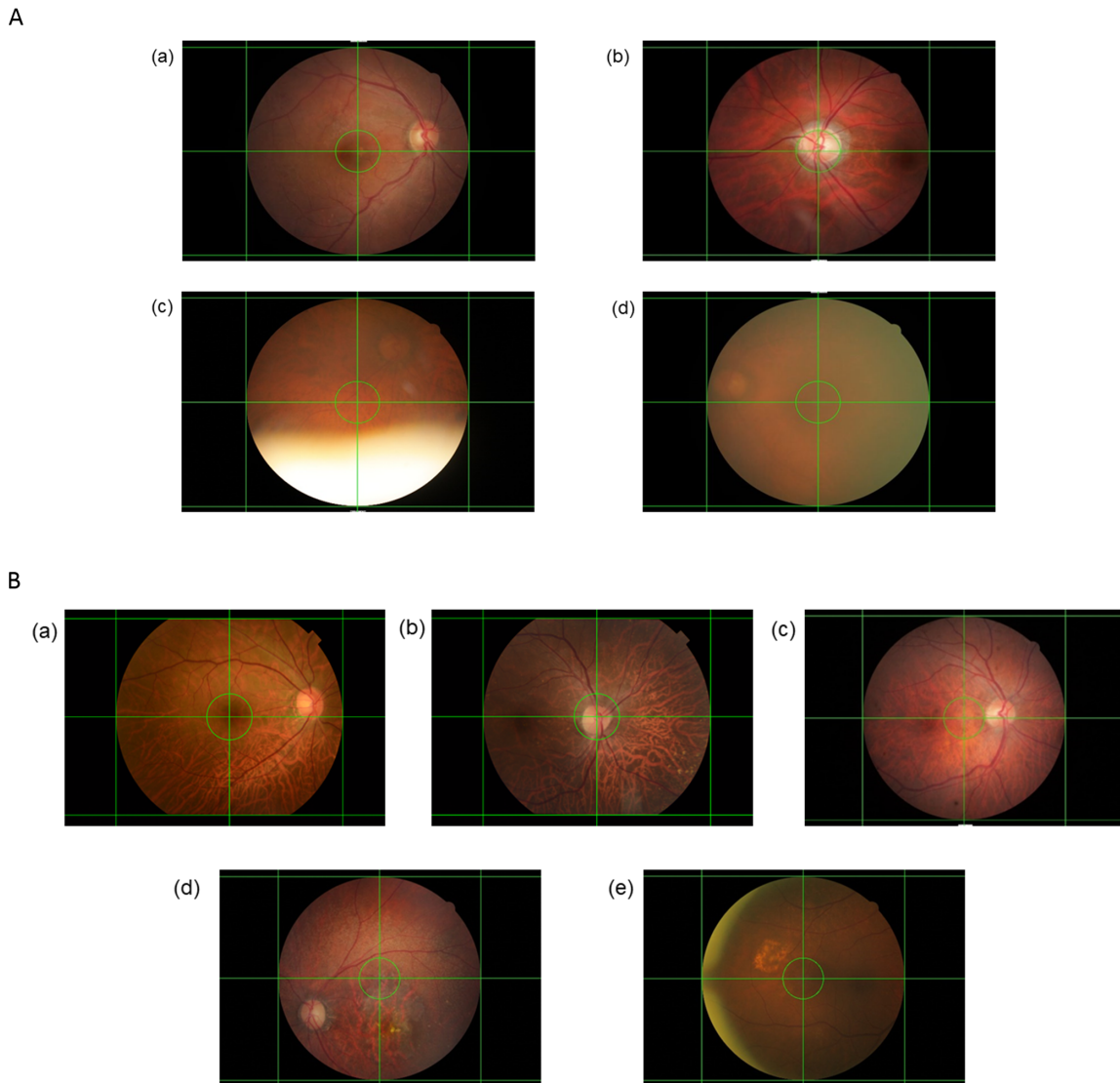
Figure 2. (A) Examples of retinal photographs with different levels of image quality. A gradable retinal photograph had to fulfill both of the following criteria: (1) less than 25% of the peripheral area of the retina was unobservable due to artifacts, and (2) the center region of the retina had no significant artifacts. (a, b) Gradable and absence of any artifacts; (c) ungradable as more than 25% of the peripheral area of the retina was unobservable due to the presence of eyelid; (d) ungradable due to presence of significant artifact in the center region. (B) Example of retinal photographs with different fields of view. (a) Macula-centered; (b) optic-disc-centered; (c–e) off-centered. Horizontal and vertical auxiliary lines were added to locate the center region of the retinal photograph, which was bounded by a *green circle*. The center region of a retinal photograph is defined as the circular region of radius (in pixels) with the largest integer not greater than one-tenth the width of the image at the center of the image. A *green circle* bounded the center region of the retinal photograph.

Supplementary Appendix S1. As tasks differ and models tend to learn different features, we used EfficientNet-B0[46] for the image-quality and the field-of-view tasks and MobileNetV2[47] for the laterality-of-the-eye task, aiming to make use of advantages in different architectures.

The primary dataset included 14,422 retinal photographs, 90% and 10% of which were used for training and internal validation, respectively. All of the retinal photographs were used to train the image-quality assessment model. After excluding ungrad-able retinal photographs and off-centered retinal photographs, only gradable macula-centered and optic-disc-centered retinal photographs were used to train the DL algorithms for field-of-view and laterality-of-the-eye assessments. The external datasets, including 2385 retinal photographs from External-1 and 4541 retinal photographs from External-2, respectively, were additionally used to test the pre-diagnosis module.

We further proposed a cloud-based web application, integrating the whole process of data pre-processing, data analysis, and data output (Fig. 3). The application
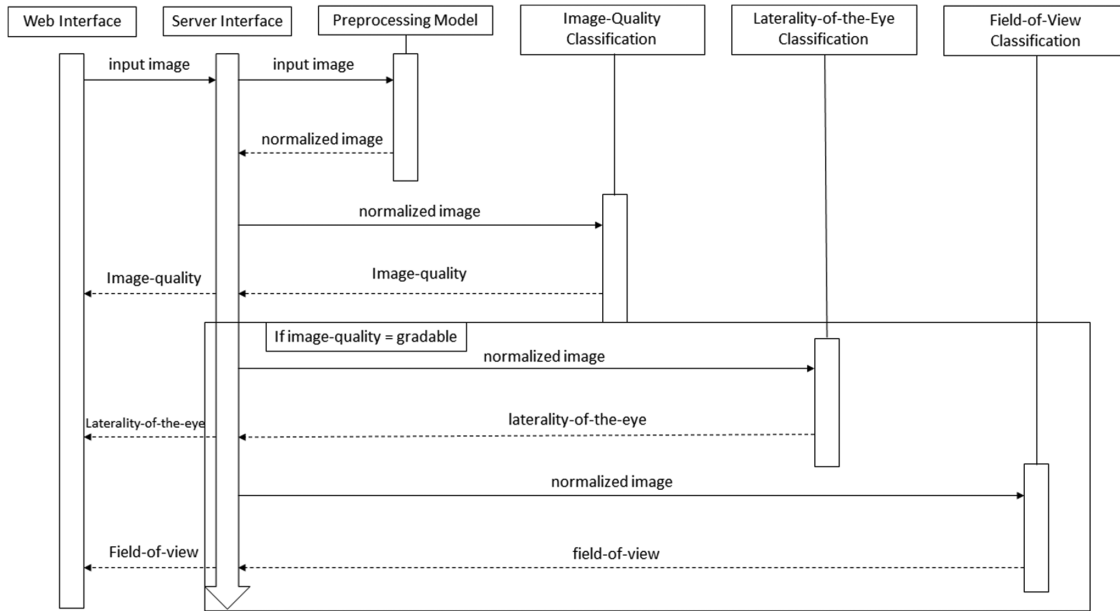
Figure 3. Sequence diagram of the pre-diagnosis module.

was composed under the service-oriented architecture protocol, which facilitates ease of maintenance. From the user perspective, no additional operation was required apart from uploading retinal photographs for pre-diagnosis assessment. Two videos, *batch.mkv* and *single.mkv* demonstrating the use of this application can be found in Supplementary Appendix S4.

Statistical Analysis

The statistical analyses were performed using RStudio 1.1.463. We calculated the area under the receiver operating characteristic curve (AUROC) with a 95% confidence interval (CI) for each outcome. Sensitivity, specificity, and accuracy were also reported at

a threshold with the highest Yuden's index (Yuden's index = sensitivity + specificity – 1) to evaluate the discriminative performance.[48] Misclassification analysis was also conducted based on this threshold.

Results

Table 2 shows the AUROC values, sensitivities, specificities, and accuracies of the pre-diagnosis module in the internal validation and two external testing datasets. In the image-quality assessment, the pre-diagnosis module achieved AUROC values of 0.975 (95% CI, 0.956–0.995), 0.999 (95% CI,

Table 2. Performance of the Pre-Diagnosis Module

Dataset	AUROC (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Accuracy, % (95% CI)
Image quality				
Internal validation	0.975 (0.956–0.995)	92.1 (88.2–95.5)	98.3 (91.5–100)	92.5 (89.0–95.6)
External-1	0.999 (0.999–1.000)	99.3 (98.9–99.7)	100 (100–100)	99.3 (99.0–99.6)
External-2	0.987 (0.981–0.993)	95.0 (92.4–96.9)	96.4 (93.7–98.7)	95.1 (92.9–96.8)
Field of view				
Internal validation	1.000 (1.000–1.000)	100 (100–100)	100 (100–100)	100 (100–100)
External-1	1.000 (1.000–1.000)	100 (100–100)	100 (100–100)	100 (100–100)
External-2	1.000 (1.000–1.000)	100 (99.8–100)	100 (99.9–100)	100 (99.9–100)
Eye laterality				
Internal validation	1.000 (1.000–1.000)	100 (100–100)	100 (100–100)	100 (100–100)
External-1	0.999 (0.998–1.000)	99.7 (99.4–100)	99.7 (99.4–100)	99.7 (99.5–99.9)
External-2	0.985 (0.982–0.989)	94.0 (91.7–96.2)	95.8 (93.3–97.7)	94.8 (94.1–95.6)

0.999–1.000), and 0.987 (95% CI, 0.981–0.993), respectively. The sensitivities, specificities, and accuracies were all higher than 92%. In the field-of-view assessment, the pre-diagnosis module achieved AUROC values of 1.000 (95% CI, 1.000–1.000) in all testing datasets. The sensitivities, specificities, and accuracies were all 100%. In the laterality-of-the-eye assessment, the pre-diagnosis module achieved AUROC values of 1.000 (95% CI, 1.000–1.000), 0.999 (95% CI, 0.998–1), and 0.985 (95% CI, 0.982–0.989), respectively. The sensitivities, specificities, and accuracies were all higher than 94%. More comparisons of this study and previous studies can be found in Supplementary Appendices S2 and S4 (Supplementary Tables S1–S3). The numbers and reasons for the misclassified cases are reported in Supplementary Appendices S3 and S4 (Supplementary Table S4). Examples of misclassified retinal photographs are shown in Supplementary Appendix S4 (Supplementary Figs. S1–S4).

Discussion

To the best of our knowledge, we are the first to develop and test a three-in-one DL-based pre-diagnosis module for retinal photographs assessment. In either internal validation or external testing, the pre-diagnosis module provided highly reliable assessments of image quality, field of view, and laterality of the eye. The three assessments of this pre-diagnosis module offer significant value in clinical practice. The image quality of retinal photographs in real-world clinical settings should be assessed, as those of inferior quality diminish the application of DL algorithms. Dodge et al.[22] reported that existing DL algorithms were particularly susceptible to blurring and noise. Yip et al.[23] reported that phakic lens status and cataract, with resultant impact on media opacity and the image quality, would reduce the specificity of DL algorithms. Therefore, it is necessary for a pre-diagnosis module to screen retinal photographs of poor image quality for subsequent algorithm analyses.[29–31] In clinical settings, it is important to label the laterality of the eye to allow monitoring the pathologies of each eye independently and correlate them with the other eye of the same subject for clinical decisions.[49] Traditionally, the laterality of the eye can be denoted by examination sequence (first right eye and then left eye) or manual classification. However, the former method lacks flexibility whereas the latter method risks error and is time consuming.[50,51] In contrast to the laterality-of-the-eye assessment, the field-of-view assessment by DL algorithm is not frequently explored. To the best of our knowl-

edge, Bellemo et al.[52] and Rim et al.[53] are the only two studies that have utilized a DL approach for field-of-view assessment. The field-of-view assessment can act as a triage to refer retinal photographs to suitable algorithms for further diseases detection. For example, optic-disc-centered retinal photographs can be referred to algorithms that differentiate papilledema or glaucomatous optic neuropathy, whereas macula-centered retinal photographs can be referred to algorithms that identify age-related macular degeneration.[12,33,52] It is noted that AI algorithms may confuse the bright appearance of the optic disc with other bright lesions, such as exudates, which may otherwise affect diagnostic accuracy of the AI algorisms for disease classification.[54] Therefore, an automated identification of optic-disc-centered retinal photographs may be useful.

Previous research efforts have attempted to develop DL algorithms to assess image quality, field of view, and laterality of the eye, many of which have provided satisfactory results. However, there are certain limitations in these studies. First, many studies have investigated individual assessments of image quality, field of view, and laterality of the eye[12,29,35,36,38,39,49,51,55–58] but have not incorporated them into a single complete module. Developing individual DL algorithms for each assessment can prove effectiveness but does not align with the workflow of clinical practice—a prerequisite for successful clinical implementation.[20] It is also not practically feasible for users without technical AI or software expertise to switch between DL algorithms.[21] Second, some DL algorithms have been trained on datasets consisting of only a single cohort,[29,35–40] which increases the risk of selection bias and reduces the DL model generalizability. For example, the models of Chalakkal et al.[29] were trained on a dataset from a single cohort, and demonstrated significant variation (>10%) in accuracy, sensitivity, and specificity when tested on other datasets. Third, some algorithms were only evaluated by internal validation datasets,[35–38] without testing by external independent datasets. External datasets with variable and diversified settings such as brands and models of retinal cameras, image formats and sizes, provides additional testing to AI algorithms to improve their robustness.[59]

Our three-in-one module has several strengths that address the above limitations. First, to the best of our knowledge, our study is the first to integrate assessments of image quality, field of view, and laterality of the eye into a single pre-diagnosis module. Second, our study addresses the variability present in real-world clinical settings by combining three different cohorts in the training dataset. Apart from datasets from Hong Kong, a predominantly Chinese society, the

training dataset also included subjects obtained from Singapore, an ethnically diverse society of Chinese, Malays, and Indians. In addition, subjects' demographics, brands and models of retinal cameras, image formats, image sizes, and prevalence of artifacts also differ between these different cohorts. With the diversity of the cohorts used in the training dataset, our module has achieved an AUROC value of 1.000 in the field-of-view assessment even when validated against an external dataset of a predominantly Caucasian population from the United Kingdom. In the image-quality and laterality-of-the-eye assessments, the variations in AUROC values observed were very low (less than 0.024).

Despite a reliable performance, a limitation of our study was the exclusion of off-centered retinal photographs from the primary training dataset for assessments of the field of view and laterality of the eye. Although gradable off-centered retinal photographs could be used to assess laterality of the eye, the eventual purpose of our proposed three-in-one, DL-based pre-diagnosis module is for subsequent eye disease detection based on macula-centered and optic-disc-centered retinal photographs. Therefore, we excluded all off-centered retinal photographs and designed the DL algorithm for laterality-of-the-eye assessment from macula-centered and optic-disc-centered retinal photographs only. Another potential limitation was that we used only retinal photographs with 45° and 50° for training and testing in this study. Further research is warranted to train and test models to assess retinal photographs with different fields of view (i.e., different degrees).

The three-in-one pre-diagnosis module has two potential implementations in clinical practice. First, this module can ensure the gradability of retinal photographs by providing an immediate onsite assessment of image quality (Supplementary Fig. S5). This can allow retaking of retinal photographs, if necessary, of subjects within the same visit and also reduce the expertise required in collecting retinal photographs. In addition, the automatic identification of field of view and laterality of the eye can also minimize mislabeling errors and provide more information to facilitate diagnosis. Second, this module can accommodate algorithm use for disease diagnosis by pre-screening and removing retinal photographs with low image quality before algorithm-based disease classification or other tasks are performed, such as blood vessel segmentation, retinal vessel caliber estimation, and cardiovascular risk factor prediction. This DL-based pre-diagnosis screening module will potentially facilitate a more widespread use of AI algorithms in clinical practice to assist low-cost disease screening and lead to improved disease prevention.[60]

Conclusions

In conclusion, the proposed three-in-one pre-diagnosis module achieved reliable and outstanding performance with good generalizability to identify the image quality, field of view, and laterality of the eye of retinal photographs. The outcomes from this DL-based pre-diagnosis module demonstrated the utility for the screening and removal of low-quality retinal photographs to improve the accuracy and efficiency of image acquisition, which will potentially facilitate the use of AI-based algorithms for further disease diagnosis in clinical settings.

Acknowledgments

References

1. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124:962–969.
2. 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.
3. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1:39.
4. Heydon P, Egan C, Bolter L, et al. Prospective evaluation of an artificial intelligence-enabled

algorithm for automated diabetic retinopathy screening of 30 000 patients. *Br J Ophthalmol*. 2020;105:723–728.

5. Li Z, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. *Diabetes Care*. 2018;41:2509–2516.

6. 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;316:2402–2410.

7. Grassmann F, Mengelkamp J, Brandl C, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology*. 2018;125:1410–1420.

8. Peng Y, Dharssi S, Chen Q, et al. DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. *Ophthalmology*. 2019;126:565–575.

9. Burlina PM, Joshi N, Pacheco KD, Freund DE, Kong J, Bressler NM. Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration. *JAMA Ophthalmol*. 2018;136:1359–1366.

10. Christopher M, Belghith A, Bowd C, et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci Rep*. 2018;8:16685.

11. Keel S, Wu J, Lee PY, Scheetz J, He M. Visualizing deep learning models for the detection of referable diabetic retinopathy and glaucoma. *JAMA Ophthalmol*. 2019;137:288–292.

12. Liu H, Li L, Wormstone IM, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. *JAMA Ophthalmol*. 2019;137:1353–1360.

13. Cheung CY, Xu D, Cheng C-Y, et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat Biomed Eng*. 2020;5:498–508.

14. Welikala RA, Fraz MM, Habib MM, et al. Automated quantification of retinal vessel morphometry in the UK biobank cohort. In: *IPTA 2017: International Conference on Image Processing Theory, Tools & Applications*. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2017:1–6.

15. Tapp RJ, Owen CG, Barman SA, et al. Associations of retinal microvascular diameters and tortuosity with blood pressure and arterial stiffness: United Kingdom Biobank. *Hypertension*. 2019;74:1383–1390.

16. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2:158–164.

17. Rim TH, Lee G, Kim Y, et al. Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. *Lancet Digit Health*. 2020;2:e526–e536.

18. Sharafi SM, Sylvestre JP, Chevrefils C, et al. Vascular retinal biomarkers improves the detection of the likely cerebral amyloid status from hyperspectral retinal images. *Alzheimers Dement (N Y)*. 2019;5:610–617.

19. Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. *JAMA Intern Med*. 2019;179:293–294.

20. Maddox TM, Rumsfeld JS, Payne PRO. Questions for artificial intelligence in health care. *JAMA*. 2019;321:31–32.

21. Gunasekeran DV, Wong TY. Artificial intelligence in ophthalmology in 2020: a technology on the cusp for translation and implementation. *Asia Pac J Ophthalmol (Phila)*. 2020;9:61–66.

22. Dodge SF, Karam LJ. Understanding how image quality affects deep neural networks. In: *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX 2016)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2016;1–6.

23. Yip MYT, Lim G, Lim ZW, et al. Technical and imaging factors influencing performance of deep learning systems for diabetic retinopathy. *NPJ Digit Med*. 2020;3:40.

24. Beede E, Baylor E, Hersch F, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery; 2020;1–12.

25. Phene S, Dunn RC, Hammel N, et al. Deep learning and glaucoma specialists: the relative importance of optic disc features to predict glaucoma referral in fundus photographs. *Ophthalmology*. 2019;126:1627–1639.

26. Shah P, Mishra DK, Shanmugam MP, Doshi B, Jayaraj H, Ramanjulu R. Validation of deep convolutional neural network-based algorithm for detection of diabetic retinopathy - artificial intelligence versus clinician for screening. *Indian J Ophthalmol*. 2020;68:398–405.

27. Li F, Liu Z, Chen H, Jiang M, Zhang X, Wu Z. Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm. *Transl Vis Sci Technol*. 2019;8:4.

28. Asaoka R, Tanito M, Shibata N, et al. Validation of a deep learning model to screen for glaucoma using images from different fundus cameras and data augmentation. *Ophthalmol Glaucoma*. 2019;2:224–231.

29. Chalakkal RJ, Abdulla WH, Thulaseedharan SS. Quality and content analysis of fundus images using deep learning. *Comput Biol Med*. 2019;108:317–331.

30. Karnowski TP, Aykac D, Giancardo L, et al. Automatic detection of retina disease: robustness to image quality and localization of anatomy structure. *Conf Proc IEEE Eng Med Biol Soc*. 2011;2011:5959–5964.

31. Patton N, Aslam TM, MacGillivray T, et al. Retinal image analysis: concepts, applications and potential. *Prog Retin Eye Res*. 2006;25:99–127.

32. Dai L, Wu L, Li H, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat Commun*. 2021;12:3242.

33. Milea D, Najjar RP, Jiang Z, et al. Artificial intelligence to detect papilledema from ocular fundus photographs. *N Engl J Med*. 2020;382:1687–1695.

34. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018;125:1199–1206.

35. Tennakoon RB, Mahapatra D, Roy PK, Sedai S, Garnavi R. Image quality classification for DR screening using convolutional neural networks. In: *Proceedings of the 2016 Ophthalmic Medical Image Analysis International Workshop*. Iowa City, IA: University of Iowa; 2016.

36. Mahapatra D, Roy PK, Sedai S, Garnavi R. *Retinal Image Quality Classification Using Saliency Maps and CNNs*. Cham: Springer International Publishing; 2016;172–179.

37. FengLi Y, Jing S, Annan L, Jun C, Cheng W, Jiang L. Image quality classification for DR screening using deep learning. *Conf Proc IEEE Eng Med Biol Soc*. 2017;2017:664–667.

38. 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.

39. 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;3:444–450.

40. 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.

41. Liu S, Ong YT, Hilal S, et al. The association between retinal neuronal layer and brain structure is disrupted in patients with cognitive impairment and Alzheimer's disease. *J Alzheimers Dis*. 2016;54:585–595.

42. Hilal S, Chai YL, Ikram MK, et al. Markers of cardiac dysfunction in cognitive impairment and dementia. *Medicine (Baltimore)*. 2015;94: e297.

43. Yam JC, Tang SM, Kam KW, et al. High prevalence of myopia in children and their parents in Hong Kong Chinese population: the Hong Kong Children Eye Study [online ahead of print]. *Acta Ophthalmol*. 2020, https://doi.org/10.1111/aos.14350.

44. Williams MA, Silvestri V, Craig D, Passmore AP, Silvestri G. The prevalence of age-related macular degeneration in Alzheimer's disease. *J Alzheimers Dis*. 2014;42:909–914.

45. Paquit VC, Karnowski TP, Aykac D, Li Y, Tobin KW, Chaum E. Detecting flash artifacts in fundus imagery. *Annu Int Conf IEEE Eng Med Biol Soc*. 2012;2012:1442–1445.

46. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. *ArXiv*. 2019;abs/1905.11946.

47. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L. MobileNetV2: inverted residuals and linear bottlenecks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2018:4510–4520.

48. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–35.

49. 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*. Cham: Springer International Publishing; 2019:507–517.

50. 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.

51. 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. In: *2016 International Conference on Digital Image Computing: Techniques and*

Applications (DICTA). Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2016:1–7.

52. 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*. Cham: Springer International Publishing; 2019:309–315.

53. Rim TH, Soh ZD, Tham Y-C, et al. Deep learning for automated sorting of retinal photographs. *Ophthalmol Retina*. 2020;4:793–800.

54. Asiri N, Hussain M, Al Adel F, Alzaidi N. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: a survey. *Artif Intell Med*. 2019;99:101701.

55. Yu H, Agurto C, Barriga S, Nemeth SC, Soliz P, Zamora G. Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening. In: *2012 IEEE Southwest Symposium on Image Analysis and Interpretation*. Piscataway, NJ: Institute of Electrical and Electronics Engineers; 2012:125–128.

56. Zago GT, Andreão RV, Dorizzi B, Teatini Salles EO. Retinal image quality assessment using deep learning. *Comput Biol Med*. 2018;103:64–70.

57. 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.

58. Jang Y, Son J, Park KH, Park SJ, Jung K-H. Laterality classification of fundus images using interpretable deep neural network. *J Digit Imaging*. 2018;31:923–928.

59. Goh JHL, Lim ZW, Fang X, et al. Artificial intelligence for cataract detection and management. *Asia Pac J Ophthalmol (Phila)*. 2020;9:88–95.

60. Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *JAMA*. 2016;316:2366–2367.