# Referral for disease-related visual impairment using retinal photograph-based deep learning: a proof-of-concept, model development study

*Yih-Chung Tham*, Ayesha Anees*, Liang Zhang, Jocelyn Hui Lin Goh, Tyler Hyungtaek Rim, Simon Nusinovici, Haslina Hamzah, Miao-Li Chee, Gabriel Tjio, Shaohua Li, Xinxing Xu, Rick Goh, Fangyao Tang, Carol Yim-Lui Cheung, Ya Xing Wang, Vinay Nangia, Jost B Jonas, Bamini Gopinath, Paul Mitchell, Rahat Husain, Ecosse Lamoureux, Charumathi Sabanayagam, Jie Jin Wang, Tin Aung, Yong Liu†, Tien Yin Wong†, Ching-Yu Cheng†*

## Summary

**Background** In current approaches to vision screening in the community, a simple and efficient process is needed to identify individuals who should be referred to tertiary eye care centres for vision loss related to eye diseases. The emergence of deep learning technology offers new opportunities to revolutionise this clinical referral pathway. We aimed to assess the performance of a newly developed deep learning algorithm for detection of disease-related visual impairment.

**Methods** In this proof-of-concept study, using retinal fundus images from 15 175 eyes with complete data related to best-corrected visual acuity or pinhole visual acuity from the Singapore Epidemiology of Eye Diseases Study, we first developed a single-modality deep learning algorithm based on retinal photographs alone for detection of any disease-related visual impairment (defined as eyes from patients with major eye diseases and best-corrected visual acuity of <20/40), and moderate or worse disease-related visual impairment (eyes with disease and best-corrected visual acuity of <20/60). After development of the algorithm, we tested it internally, using a new set of 3803 eyes from the Singapore Epidemiology of Eye Diseases Study. We then tested it externally using three population-based studies (the Beijing Eye study [6239 eyes], Central India Eye and Medical study [6526 eyes], and Blue Mountains Eye Study [2002 eyes]), and two clinical studies (the Chinese University of Hong Kong's Sight Threatening Diabetic Retinopathy study [971 eyes] and the Outram Polyclinic Study [1225 eyes]). The algorithm's performance in each dataset was assessed on the basis of the area under the receiver operating characteristic curve (AUC).

**Findings** In the internal test dataset, the AUC for detection of any disease-related visual impairment was 94·2% (95% CI 93·0–95·3; sensitivity 90·7% [87·0–93·6]; specificity 86·8% [85·6–87·9]). The AUC for moderate or worse disease-related visual impairment was 93·9% (95% CI 92·2–95·6; sensitivity 94·6% [89·6–97·6]; specificity 81·3% [80·0–82·5]). Across the five external test datasets (16 993 eyes), the algorithm achieved AUCs ranging between 86·6% (83·4–89·7; sensitivity 87·5% [80·7–92·5]; specificity 70·0% [66·7–73·1]) and 93·6% (92·4–94·8; sensitivity 87·8% [84·1–90·9]; specificity 87·1% [86·2–88·0]) for any disease-related visual impairment, and the AUCs for moderate or worse disease-related visual impairment ranged between 85·9% (81·8–90·1; sensitivity 84·7% [73·0–92·8]; specificity 74·4% [71·4–77·2]) and 93·5% (91·7–95·3; sensitivity 90·3% [84·2–94·6]; specificity 84·2% [83·2–85·1]).

**Interpretation** This proof-of-concept study shows the potential of a single-modality, function-focused tool in identifying visual impairment related to major eye diseases, providing more timely and pinpointed referral of patients with disease-related visual impairment from the community to tertiary eye hospitals.

**Funding** National Medical Research Council, Singapore.



## Introduction

Visual impairment is a major public health problem.[1] It is associated with reduced quality of life and increased risk of frailty and mortality.[2–4] Globally in 2020, an estimated 553 million people had a visual impairment and 43 million were blind.[5] 40% of visual impairment is related to refractive error (typically myopia) that requires the provision of spectacles in community settings;[6] however, the remaining 60% of cases cannot be corrected with spectacles and require assessment, diagnosis, treatment, and possibly surgery in eye-care settings led by ophthalmologists.[6] These 60% of people with visual impairment can be referred to as having disease-related visual impairment (ie, substantial loss of vision caused by an eye disease, and unrelated to refractive error).[1,6] The leading causes of disease-related visual impairment (eg, cataract, diabetic retinopathy, age-related macular degeneration, and glaucoma) are typically age related and are thus increasing in numbers globally. If detected early, these conditions can be treated, thus preventing or slowing development of vision loss.

*Lancet Digit Health* 2021; 3: e29–40

See **Comment** page e2

*Contributed equally

†Contributed equally

**Singapore Eye Research Institute, Singapore National Eye Centre, Singapore** (Y-C Tham PhD, L Zhang PhD, J H L Goh BEng, T H Rim MD, S Nusinovici PhD, H Hamzah BSc, M-L Chee BSc, R Husain MD, Prof E Lamoureux PhD, C Sabanayagam PhD, Prof T Aung MD, Y Liu PhD, Prof T Y Wong MD, Prof C-Y Cheng MD); **Duke-NUS Medical School, Singapore** (Y-C Tham, T H Rim, S Nusinovici, R Husain, Prof E Lamoureux, C Sabanayagam, Prof J J Wang PhD, Prof T Aung, Y Liu, Prof T Y Wong, Prof C-Y Cheng); **Institute of High Performance Computing, A*STAR, Singapore** (A Anees MSc, G Tjio MSc, S Li PhD, X Xu PhD, R Goh PhD, Y Liu); **School of Chemical and Biomedical Engineering, Division of Bioengineering, Nanyang Technological University, Singapore** (J H L Goh); **Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China** (F Tang PhD, C Y-L Cheung PhD); **Beijing Institute of Ophthalmology, Beijing Ophthalmology and Visual Science Key Lab, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China** (Y X Wang MD); **Suraj Eye Institute, Nagpur, India** (V Nangia MD); **Department of Ophthalmology, Medical Faculty Mannheim of the Ruprecht-Karis-University Heidelberg, Mannheim, Germany** (Prof J B Jonas MD); **Centre for Vision Research,**

**Department of Ophthalmology, The Westmead Institute for Medical Research, The University of Sydney, Westmead Hospital, Westmead, NSW, Australia** (Prof B Gopinath PhD, Prof P Mitchell MD); **Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore** (Prof T Aung, Prof T Y Wong, Prof C-Y Cheng)

Correspondence to: Prof Ching-Yu Cheng, Singapore Eye Research Institute, The Academia, Discovery Tower, Singapore 169856 **chingyu.cheng@duke-nus.edu.sg**

## Research in context

### Evidence before this study

We searched PubMed for articles published in English from database inception until Aug 14, 2020, about the use of deep learning and retinal imaging for community-based vision screening, using the search terms "deep learning (DL)", "visual impairment", "screening", "community", AND "retinal imaging". The studies identified suggest that previous work mainly focused on algorithms for detection of specific (or narrow-spectrum) eye diseases, such as diabetic retinopathy, age-related macular degeneration, and glaucoma, based on retinal photographs. Previous deep learning algorithms were designed to detect the presence of specific eye diseases without taking into account the visual function status (ie, whether there was substantial visual impairment), and thus might instead identify early diseases with normal vision (ie, early cataract, early age-related macular degeneration) that do not require immediate treatment, potentially causing over-referral tertiary hospitals. Furthermore, in the context of community screening and given that disease-related visual impairment might be attributed to multiple eye diseases, a so-called blanket-style algorithm that can broadly detect eyes with one or multiple diseases with substantial visual loss would be more efficient. Hence, a need exists for a new function-focused tool that enables more efficient identification and referral of only necessary cases (ie, people with eye disease and substantial visual impairment) to tertiary eye hospitals, to reduce the burden on finite health system resources.

### Added value of this study

In this study, we developed a single-modality, retinal photograph-based deep learning algorithm to detect disease-related visual impairment, using a total of 15 175 eyes from a multiethnic Asian population-based eye study. We also did independent validation of the algorithm using datasets of eyes from three other population-based studies and two clinic-based studies (total of 16 963 eyes), which generally showed that the algorithm had good performance. To our knowledge, this is the first study to show the use of a single-modality deep learning algorithm, using only a single macular-centred retinal photograph, for identification and referral of eyes with disease-related visual impairment. Based on saliency maps, we found that the regions probably used by the algorithm were congruent with pathological signs typically associated with major age-related eye diseases. Hence, the unique design of this algorithm enables it to potentially be used as an efficient automated referral tool in community screening.

### Implications of all the available evidences

Real-world validation is required to further determine the accuracy and clinical utility of this algorithm. However, with further validation, our algorithm could be deployed as an automated tool for identifying referable disease-related visual impairment in the community, allowing potentially more timely and accurate referral of such cases to tertiary eye centres.

To address the burden of visual impairment, WHO recommends annual vision screening for individuals aged 60 years and older,[7] and some high-income countries have already implemented annual vision screening for older people in the community.[8–13] Nevertheless, strategies and models for simple and efficient screening and referral remain key challenges for sustainable implementation of these screening programmes. Most traditional models of vision screening in the community rely on a visual acuity chart test with a pinhole to identify people who should be referred for disease-related visual impairment.[14–16] However, such models have inherent limitations. For instance, although the pinhole visual acuity test is commonly thought to be a quick measure of vision level and can be done by a nurse or layperson, it has poor reliability and repeatability,[17,18] with inconsistencies in documentation of results among testers.[19] Importantly, suboptimal pinhole visual acuity results are only an estimate to indicate potential disease-related visual impairment but do not definitively confirm the presence of eye disease, thus often resulting in a high proportion of unnecessary false positive referrals.[18,20,21] A second model used in some countries is to add retinal photography[22,23] or other methods such as air-puff tonometry (for intraocular pressure measurement),[14] to identify eye disease. In particular, adding retinal photography offers the advantage of ascertaining the presence of disease by providing direct visualisation of the fundus. However, assessment of retinal photographs requires highly trained personnel (ie, optometrists or graders), which restricts feasibility and sustainability of current screening programmes with retinal photography.[14,24,25]

Taken together, pathways in current screening approaches to identify and refer disease-related visual impairment are not efficient. Furthermore, in community or national screening programmes, the pitfalls of over-referring patients with mild or early disease (especially those with normal vision), who might not require immediate treatment in the short term, need to be avoided. Additionally, over-referral from community screenings might further overwhelm the tertiary health system. Hence, a need exists for a new function-focused tool that enables more efficient identification and referral of only necessary cases (ie, people with major eye disease and substantial visual impairment) to tertiary eye hospitals to reduce the burden on health systems. To address this gap, we designed and tested a novel single-modality deep learning algorithm, using retinal photography alone, to identify referable disease-related visual impairment.

## Methods

### Study design and population

In this proof-of-concept study, we developed and tested (internally and externally) a deep learning algorithm using over 30 000 retinal photographs. First, we used clinical data and retinal photographs from the Singapore Epidemiology of Eye Diseases (SEED) study.[26–28] Once we had developed and internally tested our algorithm, we further validated it using five external independent datasets. We used three population-based studies: the Beijing Eye study (BES),[29] Central India Eye and Medical study (CIEMS),[30] and Blue Mountains Eye Study (BMES);[31] and two clinical studies: the Chinese University of Hong Kong's Sight Threatening Diabetic Retinopathy study (CUHK-STDR)[32] and the Outram Polyclinic Study (OPS).[33] Notably, the BES, CIEMS, and BMES used best-corrected visual acuity measurements based on subjective refraction done by qualified optometrists;[27–31] whereas the CUHK-STDR and OPS used pinhole visual acuity measurements. Details of these studies are further described in the appendix (pp 23–24).

Y-CT and HH extracted data and images for SEED, JBJ extracted data and images for BES and CIEMS, BG extracted data and images for BMES, FT extracted data and images for CUHK-STDR, and HH extracted data and images for OPS. Across the development and testing datasets, study participants with incomplete or missing data related to best-corrected visual acuity or pinhole visual acuity were excluded. We only used macular-centred retinal photographs in this study (appendix p 10). In instances where multiple photographs were available for the same eye, only one photograph with the best quality was selected. Retinal photographs with artefacts due to eye movement, blinking, or extremely small pupil (thus restricted view of the fundus) were also excluded. Image quality was assessed by manual inspection by the respective study representatives. Final adjudication was done by Y-CT and JHLG. Further details on image exclusion are in the appendix (pp 1, 24).

Different retinal cameras were used in each study; however, the manufacturer overlapped in some studies. In SEED, BES, and BMES the retinal cameras used were manufactured by Canon (Tokyo, Japan), in CUHK-STDR and OPS they were manufactured by Topcon (Tokyo, Japan), and in CIEMS they were manufactured by ZEISS (Oberkochen, Germany). Details of the resolution and field of view of each retinal camera is in the appendix (p 24).

For the development and internal test dataset, we randomly distributed (8:2) the SEED dataset (N=9747; 18 978 eyes) into a development dataset (n=7793; 15 175 eyes) and an internal test dataset (n=1954; 3803 eyes). To prevent overfitting of the model, we divided the datasets at the participant level to ensure no overlapping of data from the same individual across the development dataset and internal test dataset. We followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines.[34] Participants' written informed consent was obtained in each study. All included studies adhered to the tenets of the Declaration of Helsinki and had respective local ethical committee approval. We obtained permission from the principal investigator of each study to use the data.

### Definition of disease-related visual impairment

We defined any disease-related visual impairment as eyes from patients with major eye disease with best-corrected visual acuity or pinhole visual acuity of worse than 20/40 (ie, logarithm of the Minimum Angle of Resolution [logMAR] measurement >0·30). In each study, ground truth for presence of disease-related visual impairment were ascertained by trained ophthalmologists in each study by assessing clinical records and retinal photos. Final adjudication was done by a senior consultant in each study. Mild disease-related visual impairment was defined as eyes from patients with eye disease with best-corrected visual acuity or pinhole visual acuity between 20/40 and 20/60 (ie, logMAR 0·30–0·48). Moderate or worse disease-related visual impairment was defined as eyes from patients with eye disease with best-corrected visual acuity or pinhole visual acuity worse than 20/60 (ie, logMAR >0·48).

See Online for appendix

### Models of algorithms

We used two different models of algorithms for predicting disease-related visual impairment and best-corrected visual acuity (appendix p 11). For our primary outcome of the prediction of the presence of visual impairment (ie, a classification task), we used a classification-based model. For the secondary outcome of predicting best-corrected visual acuity level, we used a regression-based model.

### Development of the deep learning algorithm

We used supervised deep learning methods for both the classification-based and regression-based models. The primary input to each of the deep learning models were the pre-processed macula-centred retinal photographs, and the relevant clinical labels (ie, visual impairment status). With these annotated data, we used a deep convolutional neural network, called the Residual Neural Network ResNet-50 architecture.[35] We used a ResNet-50, which had been pretrained on the ImageNet dataset.[36] These convolutional neural networks then further extracted features from the retinal photographs. We further applied extreme gradient boosting techniques (a scalable tree boosting system)[37] to do the classification tasks for prediction of disease-related visual impairment and regression tasks for prediction of best-corrected visual acuity. The final output of the classification model was the probability of the presence of disease-related visual impairment in each dataset. The final output of the regression model was the algorithm-predicted best-corrected visual acuity in each dataset. Further details of

| | Development dataset (SEED) | Internal test dataset (SEED) | External test datasets | | | | |
|---|---|---|---|---|---|---|---|
| | | | BES | CIEMS | BMES | CUHK-STDR | OPS |
| Number of patients | 7793 | 1954 | 3307 | 3435 | 1034 | 504 | 939 |
| Number of eyes | 15175 | 3803 | 6239 | 6526 | 2002 | 971 | 1225 |
| Age, years | 58·8 (10·3) | 58·6 (10·1) | 64·2 (9·6) | 48·2 (13·4) | 76·6 (6·8) | 61·2 (13·4) | 61·6 (6·8) |
| Sex | | | | | | | |
| Female | 3950 (50·7%) | 997 (51·0%) | 1870 (56·5%) | 1854 (54·0%) | 597 (57·7%) | 246 (48·8%) | 616 (65·6%) |
| Male | 3843 (49·3%) | 957 (49·0%) | 1437 (43·5%) | 1581 (46·0%) | 437 (42·3%) | 258 (51·2%) | 323 (34·4%) |
| Any disease-related visual impairment (by eye) | 1351 (8·9%) | 333 (8·8%) | 384 (6·2%) | 662 (10·1%) | 170 (8·5%) | 134 (13·8%) | 147 (12·0%) |
| Mild disease-related visual impairment (by eye) | 710 (4·7%) | 186 (4·9%) | 240 (3·9%) | 149 (2·2%) | 57 (2·9%) | 74 (7·6%) | 102 (8·3%) |
| Moderate disease-related visual impairment (by eye) | 641 (4·2%) | 147 (3·9%) | 144 (2·3%) | 513 (7·9%) | 113 (5·6%) | 60 (6·2%) | 45 (3·7%) |
| Best-corrected visual acuity, logMAR unit | 0·10 (0·21) | 0·10 (0·21) | 0·07 (0·19) | 0·11 (0·28) | 0·11 (0·25) | .. | .. |
| Pinhole visual acuity, logMAR unit | .. | .. | .. | .. | .. | 0·17 (0·19) | 0·21 (0·14) |
| Spherical equivalent, dioptre | −0·30 (2·23) | −0·30 (2·33) | −0·08 (1·75) | −0·07 (1·57) | −0·23 (2·08) | .. | −0·08 (2·34) |

Data are n, mean (SD), or n (%). For n (%), the denominator is number of patients or number of eyes (by eye), as applicable. BES=Beijing Eye study. BMES=Blue Mountains Eye Study. CIEMS=Central India Eye and Medical study. CUHK-STDR=Chinese University of Hong Kong's Sight Threatening Diabetic Retinopathy study. logMAR= logarithm of the Minimum Angle of Resolution. OPS=Outram Polyclinic Study. SEED=Singapore Epidemiology of Eye Diseases.

*Table 1:* **Characteristics of the development and testing datasets**

model development including model pipeline, feature extraction, and pre-training are in the appendix (pp 24–27).

### Saliency maps

To understand which regions of the retinal photographs were most likely used by the neural network for prediction of visual impairment or normal vision (visual acuity ≥20/40 or ≤logMAR 0·30), we generated saliency maps using the GradCAM method.[38] These saliency maps are presented as coloured heatmaps, with hotter colours (ie, reds and oranges) indicative of regions with increased contributions towards the predicted output and colder colours (eg, blue and green) might be indicative of relatively less contributions towards the predicted output. Further details of saliency map generation are in the appendix (p 27).

### Statistical analysis

To assess the model's performance for binary classification of visual impairment (by eye), we used the measures of the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity. We selected the classification threshold on the basis of the Youden's index.[39] We calculated the 95% CIs for these performance measures using 2000 bootstrap replicates. To assess the model's performance for continuous predictions (ie, outputs from regression model), we used mean absolute error, mean difference, 95% limits of agreement, and intraclass correlation. For the estimation of intraclass correlation values, we used a two-way mixed effect intraclass correlation model (with average fixed raters assumption). Intraclass correlation values between 0·81 and 1·00 indicate almost perfect agreement, values between 0·61 and 0·80 indicate good agreement, and values between 0·41 and 0·60 indicate moderate agreement. Values less than 0·40 indicate poor-to-fair agreement.[40]

In a post-hoc analysis, we investigated the total number of false positive and false negative eyes in the external test datasets and we analysed the common causes of false negatives and positives in the internal test dataset.

We also did post-hoc analyses on various classification tasks in the test datasets. First, we did a subgroup analysis by patient age, in which we assessed the performance of classification algorithm among eyes of individuals aged 60 years and older in both internal and external test datasets. Second, we did a sensitivity analysis at the individual level, in which individuals were defined as having visual impairment when either eye was visually impaired. This analysis was done in both internal and external test datasets. Third, we investigated the algorithm's specific performances in detecting disease-related visual impairment due to cataract, diabetic retinopathy, and any maculopathy. These specific analyses were done only in the internal test dataset.

In another post-hoc analysis, using the internal test dataset, we assessed the algorithm's performance to differentiate between healthy eyes (control group) and eyes with disease with normal vision; and also between healthy eyes and eyes with disease with moderate or worse disease-related visual impairment.

### Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## Results

We developed the deep-learning algorithm using retinal fundus images from a development dataset of 7793 study participants (15 175 eyes) from the SEED study. We validated the performance of the algorithm using retinal images from an internal test dataset of 1954 participants (3803 eyes) from the SEED study, and five external test datasets including 3307 participants (6239 eyes) from the BES, 3435 participants (6526 eyes) from the CIEMS, 1034 participants (2002 eyes) from the BMES, 504 participants (971 eyes) from the CUHK-STDR, and 939 participants (1225 eyes) from the OPS. The mean age of participants in the development dataset was 58·8 years (SD 10·3), in the internal test dataset was 58·6 years (10·1), and in the five external test datasets ranged from 48·2 years (13·4) in the CIMES to 76·6 years (6·8) in the BMES (table 1). The ethnicity and race of participants varied by study, with participants in the SEED study and OPS, which were based in Asia, being Malays, Indian, and Chinese; participants in the BES and CUHK-STDR being Chinese; participants in the CIEMS being Indian; and participants in the BMES being White. Across all included datasets, the proportion of any disease-related visual impairment (by eye) ranged from 6·2% to 13·8%. Additional study participant demographics and characteristics are summarised in table 1.

We first examined the performance of the deep learning-based classification algorithm in detecting any and moderate or worse disease-related visual impairment. In the internal test dataset, the AUC for detection of any disease-related visual impairment was 94·2% (95% CI 93·0–95·3), with sensitivity of 90·7% (87·0–93·6) and specificity of 86·8% (85·6–87·9; table 2, figure 1). Across the three external test datasets from population-based studies that measured best-corrected visual acuity, the AUC for detection of any disease-related visual impairment ranged between 89·0% (95% CI 86·1–91·9; BMES) and 93·6% (92·4–94·8; BES). Sensitivity and specificity for detection for each of these three external datasets are shown in table 2. When applying the algorithm in the two external test datasets from clinical studies (CUHK-STDR and OPS), which defined visual impairment on the basis of pinhole visual acuity measurement, the algorithm achieved AUCs of 86·6% (83·4–89·7; CUHK-STDR) and 90·5% (88·2–92·8; OPS; table 2)

| | Detection of any disease-related visual impairment* | | | Detection of moderate or worse disease related visual impairment† | | |
|---|---|---|---|---|---|---|
| | AUC | Sensitivity | Specificity | AUC | Sensitivity | Specificity |
| **Internal dataset** | | | | | | |
| SEED‡ | 94·2% (93·0–95·3) | 90·7% (87·0–93·6) | 86·8% (85·6–87·9) | 93·9% (92·2–95·6) | 94·6% (89·6–97·6) | 81·3% (80·0–82·5) |
| **External datasets** | | | | | | |
| BES‡ | 93·6% (92·4–94·8) | 87·8% (84·1–90·9) | 87·1% (86·2–88·0) | 93·5% (91·7–95·3) | 90·3% (84·2–94·6) | 84·2% (83·2–85·1) |
| CIEMS‡ | 92·8% (91·7–94·0) | 82·3% (79·2–85·2) | 90·6% (89·8–91·3) | 89·8% (88·2–91·4) | 76·4% (72·5–80·0) | 89·1% (88·3–89·9) |
| BMES‡ | 89·0% (86·1–91·9) | 75·9% (68·7–82·1) | 89·7% (88·2–91·0) | 87·2% (83·5–91·0) | 79·6% (71·0–86·6) | 79·6% (77·7–81·4) |
| CUHK-STDR§ | 86·6% (83·4–89·7) | 87·5% (80·7–92·5) | 70·0% (66·7–73·1) | 85·9% (81·8–90·1) | 84·7% (73·0–92·8) | 74·4% (71·4–77·2) |
| OPS§ | 90·5% (88·2–92·8) | 82·3% (75·2–88·1) | 85·6% (83·4–87·7) | 87·6% (82·4–92·9) | 75·6% (60·5–87·1) | 85·8% (83·7–87·7) |

Data in parentheses are 95% CIs. AUC= Area under receiver operating characteristic curve. BES=Beijing Eye study. BMES=Blue Mountains Eye Study. CIEMS=Central India Eye and Medical study. CUHK-STDR=Chinese University of Hong Kong's Sight Threatening Diabetic Retinopathy study. logMAR=logarithm of the Minimum Angle of Resolution. OPS=Outram Polyclinic Study. SEED=Singapore Epidemiology of Eye Diseases. *Defined as best-corrected visual acuity >logMAR 0·30 (ie, <20/40). †Defined as best-corrected visual impairment >logMAR 0·48 (ie, <20/60). ‡Defined visual impairment on the basis of best-corrected visual acuity measured from subjective refraction. §Defined visual impairment based on visual acuity measured from pinhole measurement.

*Table 2:* **Performance of the classification algorithm in detection of any and moderate or worse disease-related visual impairment**

The AUC for detection of moderate or worse disease-related visual impairment in the internal test dataset was 93·9% (95% CI 92·2–95·6), with sensitivity of 94·6% (89·6–97·6) and specificity of 81·3% (80·0–82·5; table 2, figure 1). When tested across the external test datasets, the algorithm showed the highest AUC in the BES dataset for detection of moderate or worse disease-related visual impairment (93·5% [91·7–95·3]), followed by CIEMS (89·8% [88·2–91·4]), OPS (87·6% [82·4–92·9]), BMES (87·2% [83·5–91·0]), and CUHK-STDR (85·9% [81·8–90·1]). When assessing mild disease-related visual impairment, the AUC in the internal test dataset was 92·7% (95% CI 90·9–94·6), with sensitivity of 87·6% (82·0–92·0) and specificity of 87·6% (86·5–88·7). Across the external test datasets, the AUCs ranged between 81·9% (77·2–86·6; CUHK-STDR) and 92·5% (90·8–94·2; BES; appendix p 2).

At a specificity level of 80%, the sensitivity for detecting any disease-related visual impairment was 94·0% (95% CI 90·9–96·3) in the internal test dataset, followed by 91·7% (88·4–94·2) in the BES, 88·8% (86·2–91·1) in the CIEMS, 86·4% (79·8–91·5) in the OPS, 80·0% (73·2–85·7) in the BMES, and 72·8% (64·5–80·1) in the CUHK-STDR (table 3). For detection of moderate or worse (table 3) and mild (appendix p 3) disease-related visual impairment, the sensitivity levels (at 80% of specificity) were largely similar to those for any disease-related visual impairment albeit slightly lower across all test datasets.

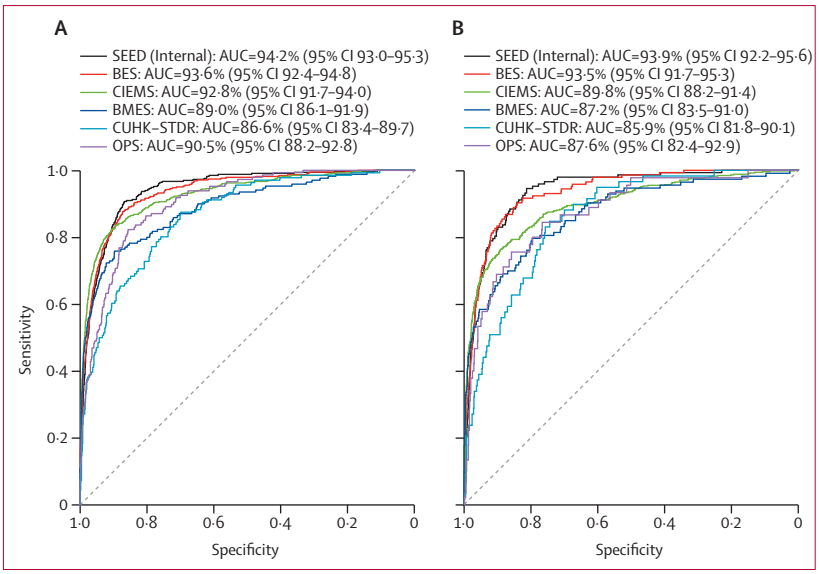


***Figure 1:*** **Performance of the classification algorithm for detection of any (A) and moderate or worse (B) disease-related visual impairment**

AUC=area under the receiver operating characteristic curve. BES=Beijing Eye study. BMES=Blue Mountains Eye Study. CIEMS=Central India Eye and Medical study. CUHK-STDR=Chinese University of Hong Kong's Sight Threatening Diabetic Retinopathy study. OPS=Outram Polyclinic Study. SEED=Singapore Epidemiology of Eye Diseases.

| | Detection of any disease-related visual impairment* | | | Detection of moderate or worse disease related visual impairment† | | |
|---|---|---|---|---|---|---|
| | At 70% specificity | At 80% specificity | At 90% specificity | At 70% specificity | At 80% specificity | At 90% specificity |
| **Internal dataset** | | | | | | |
| SEED‡ | 96·7% (94·2–98·3) | 94·0% (90·9–96·3) | 83·8% (79·4–87·6) | 98·0% (94·2–99·6) | 94·6% (89·6–97·6) | 82·3% (75·2–88·1) |
| **External dataset** | | | | | | |
| BES‡ | 95·1% (92·4–97·0) | 91·7% (88·4–94·2) | 82·8% (78·7–86·5) | 94·4% (89·3–97·6) | 91·7% (85·9–95·6) | 83·3% (76·2–89·0) |
| CIEMS‡ | 92·6% (90·3–94·5) | 88·8% (86·2–91·1) | 82·5% (79·4–85·3) | 88·9% (85·8–91·5) | 83·0% (79·5–86·2) | 74·9% (70·9–78·6) |
| BMES‡ | 87·1% (81·1–91·7) | 80·0% (73·2–85·7) | 73·5% (66·2–80·0) | 84·1% (76·0–90·3) | 77·9% (69·1–85·1) | 65·5% (56·0–74·2) |
| CUHK-STDR§ | 86·8% (79·9–92·0) | 72·8% (64·5–80·1) | 60·3% (51·6–68·6) | 86·4% (75·0–94·0) | 67·8% (54·4–79·4) | 50·8% (37·5–64·1) |
| OPS§ | 92·5% (87·0–96·2) | 86·4% (79·8–91·5) | 69·4% (61·3–76·7) | 86·7% (73·2–94·9) | 77·8% (62·9–88·8) | 68·9% (53·4–81·8) |

BES=Beijing Eye study. BMES=Blue Mountains Eye Study. CIEMS=Central India Eye and Medical study. CUHK-STDR=Chinese University of Hong Kong's Sight Threatening Diabetic Retinopathy study. logMAR=logarithm of the Minimum Angle of Resolution. OPS=Outram Polyclinic Study. SEED=Singapore Epidemiology of Eye Diseases. *Defined as best-corrected visual acuity >logMAR 0·30 (ie, <20/40). †Defined as best-corrected visual impairment >logMAR 0·48 (ie, <20/60). ‡Defined visual impairment on the basis of best-corrected visual acuity measured from subjective refraction. §Defined visual impairment on the basis of visual acuity measured from pinhole measurement.

***Table 3:*** **Sensitivity of the algorithm for the detection of any and moderate or worse disease-related visual impairment at different specificity levels**

Next, we examined the agreement in best-corrected visual acuity measurements between algorithm-predicted values (ie, outputs from the regression model) and actual best-corrected visual acuity level (ie, measured by subjective refraction; appendix p 4). In the internal test dataset, the mean absolute error was 0·076 (in logMAR unit), which is almost equivalent to a 4-letter (0·02 logMAR unit represents one letter) difference on a logMAR visual acuity chart. The mean difference was 0·01, with 95% limits of agreement of –0·221 to 0·240. The intraclass correlation coefficient value was 0·82 (95% CI 0·80 to 0·83), indicating almost perfect agreement between the two measurements. Across the relevant external test datasets, the mean absolute error ranged between 0·097 (BES) and 0·150 (CIEMS), and the intraclass correlation ranged between 0·62 (0·58 to 0·65; BMES) and 0·72 (0·71 to 0·73; BES), indicating good agreement between two measurements across all relevant external test datasets.

We used saliency maps to provide insights into the regions in the fundus image that could have influenced the algorithm's predictions. For prediction of non-visually impaired eyes, the saliency map consistently denoted the foveal area as the region used by the algorithm (figure 2). Additionally, the saliency maps highlighted regions identified by the algorithm when predicting disease-related visual impairment (figure 3). The highlighted regions were congruent with pathological features typically present in diabetic retinopathy, cataract, age-related macular degeneration, and severe glaucoma (figure 3).

Among the 333 eyes with visual impairment due to major eye diseases in the internal dataset, the algorithm correctly identified 303 eyes (ie, sensitivity of 90·7% [95% CI 87·0–93·6]). However, there were 30 false negative cases (ie, falsely classified by the algorithm as non-visual impairment); of which, 25 eyes had mild visual impairment, and five had moderate or worse visual impairment. The most common eye conditions that had false negative classification were mild cataract (14 [47%] of 30 eyes), age-related macular degeneration (five [17%]), epiretinal membrane (three [10%]), early lamellar hole (three [10%]), and diabetic macular oedema (diabetic retinopathy; two [7%]; appendix pp 5, 12–14).

In the internal test dataset, the algorithm identified 459 false positive eyes. Among these eyes, most had some form of irregularity albeit mild or sparing the foveal region (appendix pp 6, 15–18). The most common characteristics were hazy fundus photo due to the presence of cataract (272 [59%] of 459 eyes); drusen or pigmentary changes of the retinal pigment epithelium (51 [11%]), diabetic retinopathy (45 [10%]), tessellated fundus or extensive peripapillary atrophy (38 [8%]), and epiretinal membrane (25 [5%]). Normal appearance fundus was observed in eight (2%) of 459 retinal photographs in the internal dataset (appendix p 18).

In a post-hoc subgroup analysis that further assessed the performance of classification algorithm in the eyes of individuals aged 60 years and older, the AUC for detection of any disease-related visual impairment was 88·9% (95% CI 86·9–90·9; sensitivity of 83·1% [78·3–87·2]; specificity 80·9% [78·6–83·0]) in the internal test dataset.

Across all external test datasets, the AUC ranged from 82·4% (78·2–86·6; CUHK-STDR) to 91·2% (89·6–92·8; BES; appendix p 7). When assessing the detection of moderate or worse disease-related visual impairment, the AUC in the internal test dataset was 88·3% (95% CI 85·5–91·2; sensitivity 78·7% [70·6–85·5]; specificity 83·2% [81·1–85·5]). Across the external datasets, the AUC ranged from 81·3% (75·7–86·8; CUHK-STDR) to 90·3% (87·5–93·1; BES; appendix p 7).

In another post-hoc sensitivity analysis, we explored the algorithm's performance at an individual level. For the detection of any disease-related visual impairment, the sensitivity was the highest in the CUHK-STDR dataset at 92·1% (95% CI 85·0–96·5), followed by 91·4% (87·3–94·5) in the internal test dataset, 89·1% (85·2–92·3) in the BES, 85·6% (82·1–88·7) in the CIEMS, 84·9% (77·5–90·7) in the OPS, and 76·9% (68·7–83·9) in the BMES (appendix p 8). For the detection of moderate or worse disease-related visual impairment, the sensitivity was generally higher than for detecting any disease-related visual impairment, ranging from 76·9% (95% CI 60·7–88·9; OPS) to 93·5% (87·6–97·2; internal test dataset; appendix p 8). We found a slight improvement in sensitivity compared with the main analysis, with sensitivity improved by 0·7% to 4·6%, but specificity reduced by 1·2% to 8·7%.

We also further assessed the algorithm's specific performances in detecting disease-related visual impairment due to cataract, diabetic retinopathy, and any maculopathy. Across these three subgroups of causes of visual impairment, the algorithm's performance was generally similar, with AUCs ranging between 91·9% (95% CI 88·8–95·0; maculopathy) to 96·7% (94·9–98·5; diabetic retinopathy) for any disease-related visual impairment, and 91·1% (81·3–100; diabetic retinopathy) to 95·0% (92·6–97·4; cataract) for moderate or worse disease-related visual impairment (appendix p 9).

In another post-hoc subgroup and sensitivity analysis, the AUC for the algorithm's performance in differentiating between healthy eyes and eyes with disease and normal vision was 68·2% (95% CI 65·7–70·6); while differentiation between healthy eyes and eyes with disease and moderate or worse visual impairment, the AUC was 95·9% (94·4–97·4; data not shown)—ie, slightly higher than the AUC of 93·9% for moderate visual impairment in main analysis.

## Discussion

We developed a novel, single-modality, deep learning algorithm to detect disease-related visual impairment using only retinal photographs. This proof-of-concept, function-focused model could be deployed as a simple, automated, and comprehensive tool to identify referable disease-related visual impairment in the community, allowing more timely and pinpointed referral of such cases to tertiary eye centres.

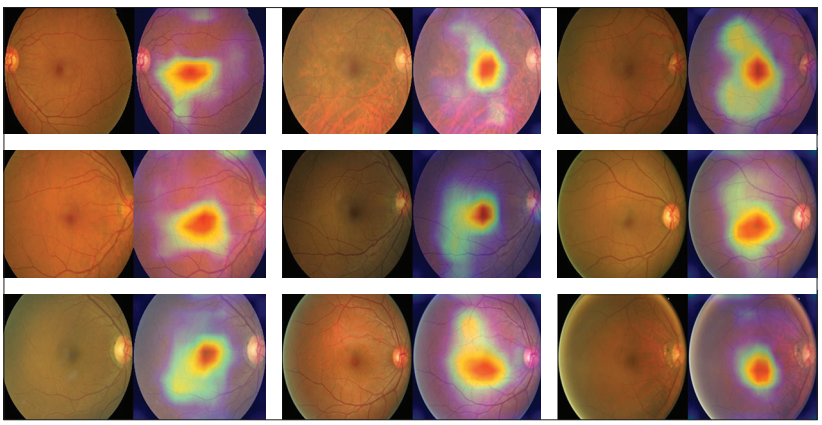

***Figure 2:*** **Saliency maps denoting the foveal area as the region used by the deep learning algorithm for specification of non-visually impaired eyes**
Images are from the SEED internal test dataset (Canon retinal camera). In these heatmaps, hotter areas (reds and oranges) are indicative of regions with increased contributions towards the predicted output, and colder regions (blues and greens) might be indicative of relatively less contributions. In these nine different images (from nine different eyes) the algorithm consistently picks up the foveal region across different eyes.

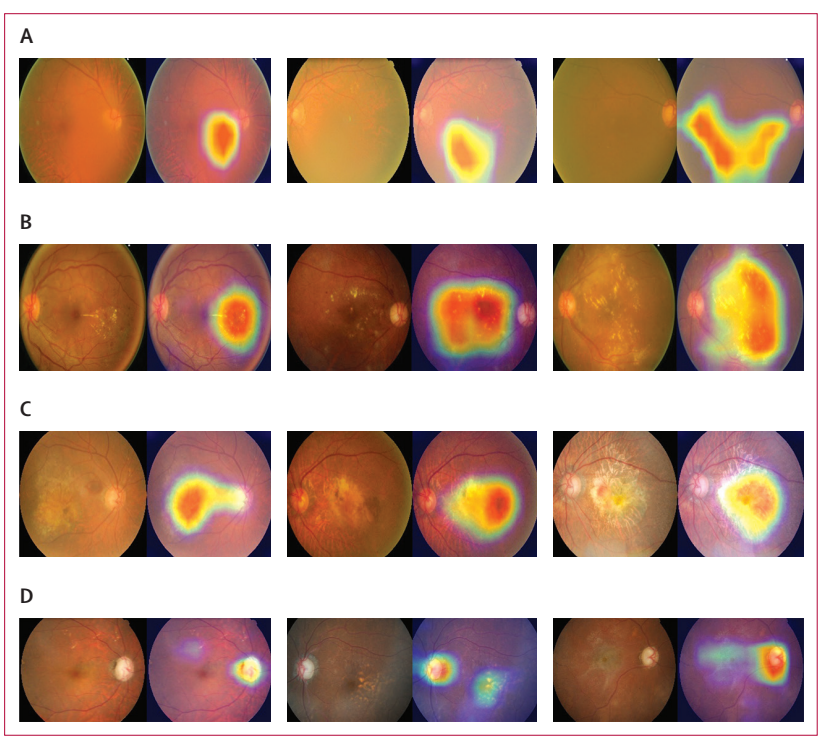


***Figure 3:*** **Saliency maps highlighting regions that the algorithm focuses on when predicting the disease-related visual impairments of cataract (A), diabetic retinopathy (B), age-related macular degeneration (C), and glaucoma (D)**
Images are from the SEED internal test dataset (Canon retinal camera). In these heatmaps, hotter areas (ie, reds and oranges) are indicative of regions with increased contributions towards the predicted output and colder regions (blues and greens) might be indicative of relatively less contribution. For each disease subgroup, each set of three images (from three different eyes) consistently shows the same region or feature is highlighted by the algorithm.

Given this new algorithm only requires a single macular-centred retinal photograph as input, without the need for a skilled expert to assess the photographs or multiple tests, it could potentially offer an efficient option

for identifying referable cases when screening in the community. When deployed alongside visual acuity tests, this algorithm could be used to better support national screening programmes in the efficient identification of individuals who should be referred, while catering for large screening volumes in the community. A possible scenario could be that when an individual does not pass the preliminary visual acuity test, a subsequent retinal photograph could be taken and automatically assessed using the algorithm, and the algorithm's output would indicate whether the individual needs further referral. This new algorithm would also be useful when applied in low-resource communities (ie, that do not have trained eye-care personnel) with a large number of individuals to screen, such as in rural areas or low-income and middle-income countries. However, due to the increasing accessibility of retinal photography in primary care settings and community screening programmes for diabetic retinopathy in Singapore and Asia,[1,2] during which ophthalmologists or optometrists are not readily onsite, this new algorithm might also be used as an add-on in these existing settings with minimal additional cost because retinal photographs are already a routine procedure. In these situations, our deep learning algorithm could also help to expand the capability of current screening programmes for diabetic retinopathy by additionally identifying disease-related visual impairment in the same setting. This implementation could be especially relevant for diabetes care settings because visual impairment due to age-related eye diseases are also more prevalent among people with diabetes.[41–44] Finally, this algorithm could potentially be used in optometry practices alongside subjective refraction as a complementary test. In situations in which a patient's vision cannot be improved with use of spectacles, an automatic assessment using a retinal photograph can be done using this algorithm, which will help to further ascertain the need for referral to tertiary eye care centre.

Previous deep learning studies in ophthalmology mostly focused on developing algorithms for detection of specific (or narrow-spectrum) eye diseases, such as diabetic retinopathy, age-related macular degeneration, and glaucoma, on the basis of retinal photographs.[41,45–49] However, these previous studies did not take into account the visual function status (ie, whether substantial visual loss was present that further justifies referral decision). The application of algorithms such as those in these previous studies does not align well with the actual scenario and demand in the community, where the priority lies in identifying and referring individuals with eye diseases concomitant with substantial visual loss versus just detection of eye diseases. Furthermore, in the context of community screening and given that disease-related visual impairment might be attributed to multiple eye diseases, an algorithm that can more broadly identify individuals with eye disease (of many types) with visual impairment who should be referred to tertiary eye care would be more useful than previous single purpose-designed algorithms. Applying multiple algorithms for different eye diseases in a real-world community scenario would be less efficient and more cumbersome for deployment. Furthermore, the function-focused and so-called blanket-style approach of this deep learning algorithm might make it less likely to miss eye diseases with visual impairment in community screening than current disease-specific algorithms.

This study is unique because of the type of data we used for the development and testing of the algorithm. To design and develop a relevant tool to be deployed in community settings, the algorithm must learn from population-representative data that reflect the patterns of visual impairment in the general population. For this reason, we curated development data from the SEED study, which is a well established population-based study comprising people from the three main ethnic groups in Asia—Malays, Indian, and Chinese.[26] Additionally, we did external validation tests of our algorithm using three other landmark population-based studies (BES, CIEMS, BMES), and two other clinical studies (CUHK-STDR, OPS). Via this external validation, we found that our algorithm was robust and we gained further insight into the generalisability of our algorithm. For instance, across the three external population-based datasets, the performance of the algorithm was highest in BES but slightly lower in BMES, which predominantly included White participants. The improved performance observed in the BES dataset might be explained by the fact that the same retinal camera type was used in BES as was used in the SEED. And the lower performance in the BMES dataset than in the other datasets might be because the eyes of White people have lighter retinal pigmentation (thus different fundus appearance) than do the eyes of Asian people.[41] Furthermore, the performance of the algorithm on the two external clinical studies was lower than on the three external population-based studies. This reduced performance is probably due to the use of pinhole visual acuity to define visual impairment in the clinical studies. The pinhole visual acuity measurement is known to have worse repeatability and validity than best-corrected visual acuity measured from subjective refraction, which was the standard method used in the three external population-based studies.[15,21] The overall findings from the external test datasets provided useful insights into the generalisability of the algorithm.

In a subgroup analysis that further assessed the performance of classification algorithm among eyes of individuals aged 60 years and older, we observed similar results as the original main analyses. This finding indicates that the algorithm would perform similarly well on this older age group that has increased risk for disease-related visual impairment. However, in another sensitivity analysis that explored the algorithm's performance at an individual level (ie, an individual would be defined as having visual impairment when

either eye has visual impairment), we observed largely similar results as in the main analyses, with slight improvement in sensitivity by 0·7% to 4·6%, but with a slight reduction in specificity values. We also further assessed the algorithm's specific performances in detecting disease-related visual impairment due to cataract, diabetic retinopathy, and any maculopathy. Across these three subgroups of causes of visual impairment, the algorithm's performance was generally similar for any disease-related visual impairment, and for moderate or worse disease-related visual impairment.

In the internal test dataset, there were also 2965 eyes without disease and with normal vision (best-corrected visual acuity ≥20/40) and 505 eyes with early stages of disease but with normal vision (best-corrected visual acuity <20/40). In our additional assessment of the algorithm's performance using health eye controls, we found that the AUC for the differentiation between healthy eyes and eyes with disease with moderate or worse vision impairment was slightly higher than was observed for moderate or worse visual impairment in the main analysis. This observation was not unexpected, given that the algorithm was originally designed as a function-focused tool to detect eyes with disease with substantial visual loss impairment, and not to merely detect cases of early features of eye disease with normal vision.

To further explore the algorithm's decision-making process, we presented saliency maps that highlighted regions in the retinal photographs that the algorithm focused on to make predictions of whether eyes had normal or impaired vision. For instance, among eyes with normal vision, the saliency maps consistently highlighted the foveal region as the area mainly used by the algorithm to make predictions for outputs denoting so-called normal eyes. This finding corresponds with the current knowledge that the fovea is the retinal area responsible for central vision and its integrity is highly correlated with best-corrected visual acuity level.[50] However, among eyes with disease-related visual impairment, the regions highlighted in the saliency maps are mostly congruent with pathological signs present in major age-related eye diseases, such as cataract, diabetic retinopathy, age-related macular degeneration, and glaucoma. Interestingly, in instances in which two different pathological features were present in the same eye (SEED internal dataset), the algorithm identified both features while assigning different contributory weights to both. For example, regions with epiretinal membrane and glaucomatous optic disc in the same eye were both highlighted in the saliency map (appendix p 19). On closer inspection, the optic disc region with features of advanced glaucomatous damage was in fact highlighted with greater intensity than the epiretinal membrane region, suggesting that the algorithm deciphered advanced glaucoma as the main contributing cause of visual impairment in this patient, as was indeed the case.[51] A similar case was identified (SEED internal dataset), with concomitant presence of advanced glaucoma and drusen (appendix p 19). Overall, the saliency maps showed encouraging evidence that the algorithm might identify clinically relevant features associated with disease-related visual impairment. Building on these findings, another so-called layer of the deep learning algorithm could be built for further subclassification of the cause or causes of disease-related visual impairment should visual impairment be detected. Such a feature would further enhance the algorithm's overall capability and would be even more clinically informative than the current algorithm. We plan to further explore this area in the next phase of this project.

Minimisation of false negative misclassifications is essential to avoid missing patients that should be referred to tertiary eye centres who typically warrant timely treatment. In this regard, we investigated the causes of false negative classifications in our internal test dataset. Among the false-negative misclassifications (n=30), most were cases of mild visual impairment (n=25), but the algorithm missed five eyes with moderate or worse visual impairment, which were attributed to age-related macular degeneration, cataract, early lamellar hole, diabetic macular oedema, and other reasons. Most of these missed cases of moderate or worse disease-related visual impairment had subtle features of disease that are also difficult for less experienced ophthalmologists and graders to identify on the basis of retinal photographs alone. Additional refinement and training of the algorithm involving more of these cases would probably help to further improve the algorithm's performance. Nevertheless, the current observations indicate that the algorithm is less likely to miss more severe visual impairment cases.

Reduction of false positive results is also an important consideration in community screening programmes to avoid unnecessary referrals. However, notably, in our internal test dataset, among the false positive classifications (n=459) identified by the algorithm, only eight (2%) were eyes with normal vision. Most eyes that gave false positive results actually presented with some features indicative of disease or irregularities on the retinal photographs even though they had relatively normal best-corrected visual acuity. These features included hazy fundus due to presence of cataract, drusen of the retinal pigment epithelium, which are early markers of age-related macular degeneration, signs of diabetic retinopathy, and epiretinal membrane. Additionally, patients with stable diabetic retinopathy who had scars from laser eye treatment were also identified as having any disease-related vision impairment by the algorithm (from the SEED internal dataset: ten of 45 eyes with diabetic retinopathy, areas with laser scars were highlighted by saliency maps; three examples are shown in the appendix [p 20]). On closer inspection of these cataract cases with slit lamp photographs, we observed

that most did not have media opacities affecting the visual axes (appendix p 21). Likewise, in eyes that gave false positive results with diabetic retinopathy and epiretinal membrane, the foveal region was largely spared (from SEED internal dataset; appendix pp 16–17). These observations collectively explain the relatively normal best-corrected visual acuity in these cases, despite the presence of disease. Overall, most of these false positive cases would not be deemed entirely to be incorrect referrals, and would probably benefit from further clinical assessment by eye care professionals. However, we also identified some true misclassifications involving tessellated fundus (so-called thin-looking fundus) with normal best-corrected visual acuity, which was especially evident in tessellated fundus concomitant with extensive peripapillary atrophy (from SEED internal dataset; appendix p 16). Saliency maps also showed that the algorithm probably interpreted the peripapillary atrophy region as an irregular feature responsible for disease-related visual impairment (appendix p 22). This observation provides a useful indication that further data training involving cases of extensive peripapillary atrophy (but with normal best-corrected visual acuity) is required to minimise the algorithm's false positive rates. This is especially important for Asian patients, given that extensive peripapillary atrophy is more prevalent in the eyes of Asian people than those of other ethnicities and races.[52,53]

Our study had several limitations. First, our deep learning algorithm was trained using a single Asian study. Subsequently, slightly lower performance was noted in a dataset of eyes from White individuals (the BMES), indicating that further refinement of the algorithm is needed by including more diverse training datasets to further improve the generalisability of the algorithm. Second, notably, the ground truth of visual impairment was defined on the basis of best-corrected visual acuity measurements, which are highly dependent on participants' response during measurement. Thus, misclassifications of eyes as a result of this subjective measurement error cannot be entirely ruled out. Third, in this proof-of-concept study, to determine the best performance of the algorithm, we used Youden's index, which provides a threshold with balanced maximisation of sensitivity and specificity.[54] However, for eventual real-world deployment, other contextual factors should also be taken into account when determining the classification threshold. These considerations include, the deployment site (ie, whether in community or clinical settings), local regulation for health technology deployment (where a minimum level of specificity and sensitivity must be achieved before roll out is granted), and the deployment model (ie, whether as a replacement or complementary tool). Finally, potential selection bias cannot be entirely ruled out in our study because the examination setting, image types, and qualities assessed in this proof-of-concept study might differ from those in eventual deployment sites. Further testing of the algorithm in real-world community settings in future studies and further assessment of its feasibility, efficiency, and cost-effectiveness (compared with existing models) is essential.

In summary, we developed and tested a new single-modality, retinal photograph-based deep learning algorithm for detection of disease-related visual impairment that warrants referral to tertiary eye care. This new tool could potentially help to improve the provision of more timely and accurate referral of patients with disease-related visual impairment from community settings to tertiary eye hospitals.

#### Contributors

Y-CT, AA, YL, TYW, and C-YC conceptualised the study. Y-CT, AA, LZ, JJW, YL, TYW, and C-YC designed the study. Y-CT, JHLG, THR, SN, HH, M-LC, FT, CY-LC, YXW, VN, JBJ, BG, PM, RH, EL, CS, TA, TYW, and C-YC collected the data. AA, GT, SL, XX, RG, and YL developed the algorithm. Y-CT, AA, LZ, JHLG, THR, SN, and M-LC analysed the data. Y-CT, AA, LZ, JHLG, YL, TYW, and C-YC drafted the manuscript. JHLG, THR, SN, HH, and M-LC confirm they had access to the raw datasets of the SEED study. YXW and JBJ confirm they had access to the raw datasets of the BES. VN and JBJ confirm they had access to the raw datasets of the CIEMS. HH, TYW, BG, and PM confirm they had access to the raw datasets of the BMES. FT and CY-LC confirm they had access to the raw datasets of the CUHK-STDR. JHLG, HH, and TA confirm they had access to the raw data sets of the OPS. Y-CT, AA, LZ, SL, XX, YL, and C-YC accessed and verified each dataset during the course of the study. All authors approved the final manuscript.

#### Declaration of interests

THR was a scientific advisor to a start-up company called Medi Whale; he received stock as a part of the standard compensation package. THR also reports personal fees from Allergan and Novartis, and patents pending for Cardiovascular disease diagnosis assistant method and apparatus (10–2018–0166720(KR), 10–2018–0166721(KR), 10–2018–0166722(KR), and PCT/KR2018/016388); diagnosis assistance system (10–2018–0157559(KR), 10–2018–0157560(KR), and 10–2018–0157561(KR)); diagnosis technology using artificial intelligence (62/694,901(US) and 62/776,345(US)); method for controlling portable fundus camera and diagnosing disease using the portable fundus camera (62/715,729(US)); method for predicting cardio-cerebrovascular disease using eye image (10–2017–0175865(KR). TYW is a consultant and a member of the advisory board for Allergan, Bayer, Boehringer Ingelheim, Genentech, Merck, Novartis, Oxurion (formerly ThromboGenics), Roche, and Samsung Bioepis and cofounder of the start-up companies Plano Pte and EyRiS. TYW also has a patent issued for Deep Learning System for Retinal Diseases (PCT/SG2018/050363, Singapore and 10201901218S [provisional] Singapore). All other authors declare no competing interests.

#### Data sharing

The algorithm mentioned in the current study is currently undergoing filing for licensing (Singhealth Intellectual Property Filing Reference number: TEC-18-97) and thus the relevant code cannot be shared. Data in this study cannot be shared publicly due to regulations of local ethical committees (SingHealth Centralised Institutional Review Board, 2018/2717 and 2018/2570). Data might be made available to researchers who meet the criteria (to be provided once all data are available) for access to confidential data and upon Institutional Review Board's approval; requests for access can be made to Prof Ching-Yu Cheng (chingyu.cheng@duke-nus.edu.sg).

#### Acknowledgments

We thank Cong Ling Teo and Kai Hui Koh for their assistance in the preparation and compilation of the figures. Y-CT is supported by a grant (NMRC/MOH-TA18nov-0002) from the National Medical Research Council, Singapore, which was used to fund this study. C-YC is supported by a grant (NMRC/CSA-SI/0012/2017) from the National Medical Research Council, Singapore. These funders had no role in study design, data collection, data analysis, decision to publish, or preparation of the manuscript.

**References**

1 Bourne RRA, Flaxman SR, Braithwaite T, et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet Glob Health* 2017; **5:** e888–97.

2 Lamoureux EL, Chong EW, Thumboo J, et al. Vision impairment, ocular conditions, and vision-specific function: the Singapore Malay Eye Study. *Ophthalmology* 2008; **115:** 1973–81.

3 Siantar RG, Cheng CY, Gemmy Cheung CM, et al. Impact of visual impairment and eye diseases on mortality: the Singapore Malay Eye Study (SiMES). *Sci Rep* 2015; **5:** 16304.

4 Dai W, Tham YC, Chee ML, et al. Falls and recurrent falls among adults in a multi-ethnic Asian population: the Singapore epidemiology of eye diseases study. *Sci Rep* 2018; **8:** 7575.

5 GBD 2019 Blindness and Vision Impairment Collaborators. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. *Lancet Glob Health* 2020; published online Dec 1. https://doi.org/10.1016/S2214-109X(20)30425-3.

6 Bourne RR, Stevens GA, White RA, et al. Causes of vision loss worldwide, 1990-2010: a systematic analysis. *Lancet Glob Health* 2013; **1:** e339–49.

7 WHO. Integrated care for older people: guidelines on community-level interventions to manage declines in intrinsic capacity. Geneva: World Health Organization, 2017.

8 Arun CS, Al-Bermani A, Stannard K, Taylor R. Long-term impact of retinal screening on significant diabetes-related visual impairment in the working age population. *Diabet Med* 2009; **26:** 489–92.

9 Lin S, Gupta B, James N, Ling RH. Visual impairment certification due to diabetic retinopathy in north and eastern Devon. *Acta Ophthalmol* 2017; **95:** e756–62.

10 Green C, Goodfellow J, Kubie J. Eye care in the elderly. *Aust Fam Physician* 2014; **43:** 447–50.

11 Feder RS, Olsen TW, Prum BE Jr, et al. Comprehensive adult medical eye evaluation preferred practice pattern(®) guidelines. *Ophthalmology* 2016; **123:** 209–36.

12 Clinical Practice Guideline Expert Committee. Canadian Ophthalmological Society evidence-based clinical practice guidelines for the periodic eye examination in adults in Canada. *Can J Ophthalmol* 2007; **42:** 39–45.

13 Singapore Ministry of Health. More seniors to benefit from functional screening. Sept 19, 2018. https://www.moh.gov.sg/news-highlights/details/more-seniors-to-benefit-from-functional-screening (accessed Nov 4, 2020).

14 Brinks M, Zaback T, Park DW, Joan R, Cramer SK, Chiang MF. Community-based vision health screening with on-site definitive exams: Design and outcomes. *Cogent Med* 2018; **5:** 1560641.

15 Sun JK, Aiello LP, Cavallerano JD, et al. Visual acuity testing using autorefraction or pinhole occluder compared with a manual protocol refraction in individuals with diabetes. *Ophthalmology* 2011; **118:** 537–42.

16 van Nispen R, van der Aa H, Timmermans F, et al. Reducing avoidable visual impairment in elderly home healthcare patients by basic ophthalmologic screening. *Acta Ophthalmol* 2019; **97:** 401–08.

17 Kaiser PK. Prospective evaluation of visual acuity assessment: a comparison of snellen versus ETDRS charts in clinical practice (an AOS thesis). *Trans Am Ophthalmol Soc* 2009; **107:** 311–24.

18 Falkenstein IA, Cochran DE, Azen SP, et al. Comparison of visual acuity in macular degeneration patients measured with snellen and early treatment diabetic retinopathy study charts. *Ophthalmology* 2008; **115:** 319–23.

19 Nurse C. Preschool vision screening tests administered by nurse screeners compared with lay screeners in the vision in preschoolers study. *Invest Ophthalmol Vis Sci* 2005; **46:** 2639–48.

20 Lovie-Kitchin JE. Validity and reliability of visual acuity measurements. *Ophthalmic Physiol Opt* 1988; **8:** 363–70.

21 Eagan SM, Jacobs RJ, Demers-Turco PL. Study of luminance effects on pinhole test results for visually impaired patients. *Optom Vis Sci* 1999; **76:** 50–58.

22 Chou R, Dana T, Bougatsos C, Grusing S, Blazina I. Screening for impaired visual acuity in older adults: updated evidence report and systematic review for the US preventive services task force. *JAMA* 2016; **315:** 915–33.

23 Tapp RJ, Svoboda J, Fredericks B, Jackson AJ, Taylor HR. Retinal photography screening programs to prevent vision loss from diabetic retinopathy in rural and urban Australia: a review. *Ophthalmic Epidemiol* 2015; **22:** 52–59.

24 Cicinelli MV, Marmamula S, Khanna RC. Comprehensive eye care—issues, challenges, and way forward. *Indian J Ophthalmol* 2020; **68:** 316–23.

25 Quigley HA, Park CK, Tracey PA, Pollack IP. Community screening for eye disease by laypersons: the Hoffberger program. *Am J Ophthalmol* 2002; **133:** 386–92.

26 Wong TY, Tham YC, Sabanayagam C, Cheng CY. Patterns and risk factor profiles of visual loss in a multiethnic Asian population: the Singapore Epidemiology of eye diseases study. *Am J Ophthalmol* 2019; **206:** 48–73.

27 Lavanya R, Jeganathan VS, Zheng Y, et al. Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians. *Ophthalmic Epidemiol* 2009; **16:** 325–36.

28 Foong AW, Saw SM, Loo JL, et al. Rationale and methodology for a population-based study of eye diseases in Malay people: the Singapore Malay eye study (SiMES). *Ophthalmic Epidemiol* 2007; **14:** 25–35.

29 Jonas JB, Xu L, Wang YX. The Beijing eye study. *Acta Ophthalmol* 2009; **87:** 247–61.

30 Nangia V, Jonas JB, Sinha A, Matin A, Kulkarni M. Refractive error in central India: the Central India Eye and Medical Study. *Ophthalmology* 2010; **117:** 693–99.

31 Attebo K, Mitchell P, Smith W. Visual acuity and the causes of visual loss in Australia. The Blue Mountains Eye Study. *Ophthalmology* 1996; **103:** 357–64.

32 Sun Z, Tang F, Wong R, et al. OCT angiography metrics predict progression of diabetic retinopathy and development of diabetic macular edema: a prospective study. *Ophthalmology* 2019; **126:** 1675–84.

33 Tun TA, Chua J, Shi Y, et al. Association of iris surface features with iris parameters assessed by swept-source optical coherence tomography in Asian eyes. *Br J Ophthalmol* 2016; **100:** 1682–85.

34 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; **162:** 55–63.

35 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA; June 27–30, 2016.

36 Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 2012; **2012:** 1097–105.

37 Chen T, Guestrin C. Xgboost: a scalable tree boosting system. 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining; San Francisco, CA, USA; Aug 13–17, 2016.

38 Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. *Proc IEEE Int Conf Comput Vis* 2017; **2017:** 618–26.

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

40 Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973; **33:** 613–19.

41 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:** 2211–23.

42 Tan AG, Kifley A, Tham YC, et al. Six-year incidence of and risk factors for cataract surgery in a multi-ethnic Asian population: the Singapore epidemiology of eye diseases study. *Ophthalmology* 2018; **125:** 1844–53.

43 Klein BE, Klein R, Wang Q, Moss SE. Older-onset diabetes and lens opacities. The Beaver Dam Eye study. *Ophthalmic Epidemiol* 1995; **2:** 49–55.

44 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.

45 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:** 2402–10.

46 Keel S, Li Z, Scheetz J, et al. Development and validation of a deep-learning algorithm for the detection of neovascular age-related macular degeneration from colour fundus photographs. *Clin Exp Ophthalmol* 2019; **47:** 1009–18.

47 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:** 565–75.

48 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:** 1410–20.

49 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:** 1353–60.

50 Provis JM, Dubis AM, Maddess T, Carroll J. Adaptation of the central retina for high acuity vision: cones, the fovea and the avascular zone. *Prog Retin Eye Res* 2013; **35:** 63–81.

51 Baskaran M, Foo RC, Cheng CY, et al. The prevalence and types of glaucoma in an urban Chinese population: the Singapore Chinese eye study. *JAMA Ophthalmol* 2015; **133:** 874–80.

52 Chang L, Pan CW, Ohno-Matsui K, et al. Myopia-related fundus changes in Singapore adults with high myopia. *Am J Ophthalmol* 2013; **155:** 991–99.

53 Koh V, Tan C, Tan PT, et al. Myopic Maculopathy and Optic Disc Changes in Highly Myopic young Asian eyes and impact on visual acuity. *Am J Ophthalmol* 2016; **164:** 69–79.

54 Yu M, Tham Y-C, Rim TH, Ting DSW, Wong TY, Cheng C-Y. Reporting on deep learning algorithms in health care. *Lancet Digit Health* 2019; **1:** e328–29.