HHS Public Access

Author manuscript

Br J Ophthalmol. Author manuscript; available in PMC 2021 August 01.

Published in final edited form as:

Br J Ophthalmol. 2021 August ; 105(8): 1133–1139. doi:10.1136/bjophthalmol-2020-316984.

Detection of Features Associated with Neovascular Age-Related Macular Degeneration in Ethnically Distinct Datasets by an Optical Coherence Tomography – Trained Deep Learning Algorithm

Tyler Hyungtaek Rim, MD, MBA[#1,2], **Aaron Yuntai Lee, MD MSCI**[#3], **Daniel Shu Wei Ting, MD, PhD**[1,2], **Kelvin Yi Chong Teo, MD**[1,2], **Bjorn Kaijun Betzler, MBBS**[4], **Zhen Ling Teo, MBBS, MRCS(Ed)**[1], **Tae Keun Yoo, MD**[5], **Geunyoung Lee**[6], **Youngnam Kim**[6], **Andrew Lin, MD**[7], **Seong Eun Kim, MD**[8], **Yih-Chung Tham, PhD**[1,2], **Sung Soo Kim, MD, PhD**[9], **Ching-Yu Cheng, MD, PhD**[1,2], **Tien Yin Wong, MD, PhD**[1,2], **Chui Ming Gemmy Cheung, MD, PhD**[1,2]

[1]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore

[2]Duke-NUS Medical School, National University of Singapore, Singapore

[3]Department of Ophthalmology, University of Washington School of Medicine, Seattle, Washington

[4]Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[5]Department of Ophthalmology, Aerospace Medical Center, Republic of Korea Air Force, Cheongju, South Korea

[6]Medi Whale Incorporated, Seoul, Korea

[7]New York University, New York

[8]Department of Ophthalmology, CHA Bundang Medical Center, CHA University, Seongnam, Korea

[9]Department of Ophthalmology, Severance Hospital, Institute of Vision Research, Yonsei University College of Medicine, Seoul, Korea

Corresponding author: Sung Soo Kim, MD, PhD, Department of Ophthalmology, Severance Hospital, Institute of Vision Research, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea, Tel: 82-10-2191-6313; Fax: 82-2-312-0541; semekim@yuhs.ac, Gemmy Chui Ming Gemmy, MD, PhD, Singapore Eye Research Institute, Singapore National Eye Centre, 11 Third Hospital Avenue, Singapore 168751, Phone: (65) 69881460, Fax: (65) 62263995, gemmy.cheung.c.m@snec.com.sg.

Author contributions statement

THR, AYL, SSK, CYC, TYW, and CMGC conceptualized the study.
THR, AYL, DSWT, KYCT, BKB, TKY, TZL, GL, YK, AL, YCT, and SEK reviewed the literature.
THR, AYL, SSK, CYC, TYW, and CMGC designed the study.
THR, AYL, GL, YK, and SSK collected the data.
THR, GL, and YK developed the algorithm.
THR, AYL, GL, YK, YCT, and AL analyzed data.
All authors contributed to data interpretation.
THR, KYCT, BKB, TKY, TZL, YCT, CYC, TYW, and CMGC drafted the manuscript.
YCT, DSWT, TYW, and CMGC did the critical revision.
All authors read and approved the final report.

Competing interests

THR was a scientific advisor to a start-up company called Medi-whale, Inc. GL and YNK are working in Medi Whale, Inc. TYW and DSWT hold a patent for a deep learning system developed for use in ophthalmology.

[#] These authors contributed equally to this work.

Abstract

Background: The ability of deep learning (DL) algorithms to identify eyes with neovascular age-related macular degeneration (nAMD) from optical coherence tomography (OCT) scans has been previously established. We herewith evaluate the ability of a DL model, showing excellent performance on a Korean dataset, to generalize onto an American dataset despite ethnic differences. In addition, expert graders were surveyed to verify if the DL model was appropriately identifying lesions indicative of nAMD on the OCT scans.

Methods: Model development dataset – 12,247 OCT scans from South Korea; External validation dataset – 91,509 OCT scans from Washington, USA. In both datasets, normal eyes or eyes with nAMD were included. After internal testing, the algorithm was sent to the University of Washington, USA, for external validation. Area under the receiver operating characteristic curves (AUC) and precision-recall curves (AUPRC) were calculated. For model explanation, saliency maps were generated using Guided GradCAM.

Results: On external validation, AUC and AUPRC remained high at 0.952 (0.942–0.962) and 0.891 (0.875–0.908) at the individual level. Saliency maps showed that in normal OCT scans, the fovea was the main area of interest; in nAMD OCT scans, the appropriate pathological features were areas of model interest. Survey of ten retina specialists confirmed this.

Conclusion: Our DL algorithm exhibited high performance for nAMD identification in a Korean population, and generalized well to an ethnically distinct, American population. The model correctly focused on the differences within the macular area to extract features associated with nAMD.

Introduction

Deep Learning (DL) is a novel artificial intelligence (AI) technology in which convolutional neural networks (CNN) are programmed to optimize a specific performance criterion using large datasets with a known outcome [1–3]. Robust algorithm performance has been reported in identifying candidate patients for corneal refractive surgery [4], classification of retinal diseases including diabetic retinopathy (DR) and age-related macular degeneration (AMD) [5–8], detection of glaucoma [9], and identifying retinopathy of prematurity [1011].

After creating a robust DL algorithm, the next challenge is allowing clinicians, regulators and ultimately patients to understand what the computer ‘sees’ and how it ‘thinks’ when undertaking the designated task – essentially to present its reasons for disease classification and diagnosis. In image analysis, this is where model explanation techniques are often applied. Model explanation techniques typically utilize saliency maps (also known as heatmaps) to highlight areas which the CNN deems critical in making decisions. The strength at which an area is highlighted is pitched according to its weightage in the decision-making process. There are several visualization methods to do this, for example, occlusion testing, integrated gradients, soft attention and class activation mapping [12]. Using these methods, researchers often analyze heatmaps post model development to ascertain that the decision-making process of CNN is based on clinically relevant, and not random, features.

In our study, we use the Guided GradCAM technique, a form of class activation mapping useful for identifying discriminative regions and widely used for classification [13]. Guided GradCAM saliency maps have been shown to provide both high-resolution and class-discriminative visualizations of the retinal images [1415].

In addition to model explanation, external validation studies are important to illustrate the ability of the DL algorithm to generalize. With regard to OCT scans, generalization across populations is difficult because variations in ethnicity, age, and sex are known to significantly alter the appearance of retinal structures, even in the absence of disease [16]. Despite this, generalizability across populations and clinical settings is a key step to be fulfilled if real-world application is desired.

The use of OCT – trained DL algorithms to clinically classify AMD has been pursued by a handful of studies in recent years [17–22]. In 2019, Bhatia et al was the first, and at this time only, to externally validate their DL algorithm on OCT scans acquired from independent centres in foreign populations [22]. However, model explanation and visualization techniques were not included. In this manuscript, we report the performance of our DL algorithm on a local South Korean dataset and on an external American dataset. To the best of our knowledge, our present study adds to current literature by being the first to implement visualization techniques on both internal and external validation datasets. We will examine the areas of model interest highlighted by the saliency maps, correlating them with the clinically important pathological features of neovascular AMD (nAMD). In addition, expert graders were surveyed to verify if the DL model was appropriately identifying lesions indicative of nAMD on the OCT scans.

Methods

Ethics Statement

This retrospective cross-sectional study was approved by the institutional ethics committee at Yonsei University (4–2019–0442) and by the Institutional Review Board of the University of Washington (UW; STUDY00010512), and adhered to the tenets of the Declaration of Helsinki. The ethics committee waived written informed consent given de-identification of image and clinical data.

Participants and Image database

OCT images with relevant accompanying clinical data were obtained from Severance Hospital, Yonsei University, South Korea between 2005 and 2017. All macular OCT scan images were extracted using an in-house developed automated extraction tool from the Heidelberg Spectralis (Heidelberg Engineering, Heidelberg, Germany) imaging database.[18] Each raster macular OCT scan was extracted as individual OCT images. We included the 5 images centred on the fovea of both conventional and enhanced depth imaging (EDI) OCT scans. Macular OCT scans with B-scan widths between 6 mm and 9 mm were included. Multiple OCT scans from the same patient taken at different visits were included. The clinical information of participants was extracted from in-house electronic medical records and the physician’s order communication system housed in the clinical data repository and

data warehousing in Severance Hospital. This clinical information was linked to OCT images via patient medical record number and visit dates.[23] All clinical diagnoses and the dates of every clinical encounter and intravitreal injection were extracted from the clinical data repository.

Reference Standard for OCT images

The following methodology describes how OCT scans were classified into 'normal' and 'nAMD' before model development for use as the ground truth. 'Normal' OCT images were handpicked by a retina specialist (THR) from an image pool of participants who 1) Underwent a baseline OCT scan and 2) Not indicated for further OCT examinations during the follow-up period (till 2017). This was done because this pool represents participants expected to be free of macular pathologies that would require repeat OCT scans. OCT images of poor quality and steep curvature due to high myopia were excluded.

'nAMD' OCT images were selected by the same retina specialist (THR) from pooled OCT images extracted from patients who 1) Received the Korean Classification of Disease diagnosis code of 'nAMD' and 2) Received an intravitreal injection of anti-vascular endothelial growth factor (VEGF) in either eye at the same visit.

From this pool, images clearly indicative of nAMD were selected. This was defined as the presence intraretinal fluid/ cyst (IRC), subretinal fluid (SRF), fluid below the retinal pigment epithelium (RPE), detachments (PED), hard exudates, or subretinal fibrous scar tissue [24]. Other retinal pathologies such as diabetic macular oedema and myopic choroidal neovascularization (CNV) were excluded based on OCT images. Images from the same patient taken at different visits that fulfilled these criteria were included. Poor quality images and images of retinal areas without any typical nAMD features (SRF, IRC, PED) were excluded.

Internal Validation and External Validation

A total of 12,247 OCT images in the internal dataset were divided into a training set to develop our models and a test set reserved to evaluate model performance. This division was performed participant-wise, meaning that images taken from 1972/ 2465 participants (80%) were randomly sorted into the training set, and images taken from 493/ 2465 participants were randomly sorted into the test set (Supplementary Table 1). OCT images from the same patients have similar morphological characteristics. If these images from the same patients are distributed in both the training and test sets, it will result in overfitting and potentially inflate the performance. Hence, in our study design, the training set and the test set should be separated based on individual-level, and not image-level. For external validation, a dataset of macular OCT images from the UW was used [18]. The OCT images of the external dataset were not transferred to Severance Hospital in South Korea, in accordance with privacy protection guidelines. Instead, the developed DL algorithm was sent over to the UW, USA, for external validation. The external validation dataset included 91 509 OCT images (48 543 nAMD; 42 966 normal) captured from 1440 participants (319 nAMD; 1121 normal) of different ethnic groups in different clinical setting in the United States.

Image Preprocessing, Model Development, and Model Explanation

Images were preprocessed to remove noise and to enhance contrast. Original images in the acquired datasets came in two sizes – 768 X 496 pixels (9 mm scan) and 512 X 496 pixels (6 mm scan). Blank space was added to the right and left margins to create identical image sizes of 768 × 496 pixels. Finally, we resized all OCT images from 768 X 496 pixels to 192 X 124 pixels (one fourth) while maintaining the width: height ratio. Data augmentation was performed. Our proposed DL model, based on the VGG-16 neural network architecture [25], was developed, trained and evaluated in TensorFlow (http://tensorflow.org). Supplementary Figure 1 provides further detail regarding image preprocessing, augmentation and model architecture. Concerning model explanation, saliency maps created with Guided GradCAM [1526] were superimposed over input images of the test set to facilitate understanding of how our model classified OCT images of nAMD and normal eyes.

Survey of Retina Specialists

Ten retina specialists independently accessed one hundred twenty OCT scans with overlaid saliency maps to determine if areas of high model interest corresponded to lesions indicative of nAMD (PED, SRF, and IRC). The 10 specialists are practising experts in the field from Japan, South Korea, and Singapore, and experienced in interpreting OCT images on a regular basis. To create a questionnaire, OCT scans with single pathology (e.g. PED), and with two or more pathologies were selected. For three types of findings of PED, SRF, and IRC, 10 questions (total 30) were made, and 10 questions were made for mixed OCT scans. Each question contains three representative images, and each retinal specialist was asked to answer a total of 40 questions. Each specialist was asked to evaluate how well the generated heatmaps (which represent areas of model interest) reflect actual lesions characteristic of nAMD. The Likert scale was used with a 5-point scale that ranges from “strongly agree (5)” to “(1).” Supplementary Document 1 provides details and a summary of the results.

Statistical Analysis

Python packages including NumPy, SciPy, matplotlib, scikit-learn were used to process and analyze the data[27]. Performance was evaluated using the test set of 2485 OCT scans (1478 nAMD; 1007 normal) from Severance Hospital, Yonsei University, South Korea (Figure 1), while the external validation dataset from the UW, USA included 91 509 OCT images (48 543 nAMD; 42 966 normal). The statistical methods used for analysis include – area under the ROC curve (AUC), area under the PR curve (AUPRC), sensitivities and specificities with 95% Confidential Intervals (CIs). Analysis was conducted both at the individual level and image level.

Results

The characteristics of our study population are described in Supplementary Table 1. It was observed that normal participants (mean [SD] age of 54.9 [6.2] years) were younger than nAMD patients (mean [SD] age of 70.2 [10.0] years). The learning curve of training of the neural network with accuracy and loss is provided in Supplementary Fig 1.

We first evaluated performance of the model in internal test set (Supplementary Fig 2). The DL algorithm performed well, predicting the presence of nAMD with AUC of 0.999 (95% CI, 0.998–0.999) at the individual level. Correspondingly, the AUPRC was 0.999 (95% CI, 0.999 – 0.999) at the individual level. Figure 1 describes the ROC curves and precision-recall curves. of external validation. A total of 91 509 OCT images (48 543 nAMD and 42 966 normal OCT scans) from 1440 participants (319 nAMD patients and 1121 normal participants) were included for external validation. The performance was AUC of 0.901 (95% CI, 0.899–0.903) at each image level and AUC of 0.952 (95% CI, 0.942–0.962) at each individual level. AUPRCs were 0.928 (95% CI, 0.926–0.929) at each image level and 0.891 (95% CI, 0.875–0.908) at each individual level.

Representative images from generated saliency maps in the internal validation set are provided in Fig 2. The highlighted areas represent parts of the image which the trained CNN deemed to be most critical in the classification task. Fig 2A is an example of one OCT scan classified as ‘normal’; its overlying heatmap reflects that model focus on the fovea and its surroundings. This was consistently observed in ‘normal’ scans, with more instances provided in Supplementary Fig 3. Similarly, Fig 2B is an example of an OCT scan classified as ‘nAMD’, with green highlights reflecting high model interest in components of nAMD (IRC, SRF, PED). Figures 3 – 5 each show close-up representative examples of SRF, PED and IRC respectively. More examples are included in Supplementary Figures 4 – 6 respectively. These clinically relevant areas were identified by CNN without feature engineering. Fig 3 shows that the network paid attention to the inner margin of SRF and on retinal pigment epithelial mottling.

A similar pattern in the saliency maps was observed in the external validation set from the UW. Details are provided in Supplementary Fig 7. Saliency maps also showed high model attention on the fovea in normal OCT images while three components of IRC, SRF, and PED were areas of focus in nAMD OCT images. Supplementary Fig 8 shows the false positive and false negative cases with saliency maps in the UW dataset. Most of false-negative cases (ground truth ‘nAMD’; predicted as ‘normal’) shows small PED or IRC, and mottled RPE. The false-positive cases (ground truth ‘normal’; predicted as ‘nAMD’) shows that the raised portion near the prelaminar optic nerve, or the posterior hyaloid membranes; or slight focal disruption of the ellipsoid zone/interdigitation zone were highlighted in saliency maps.

Our survey of ten retina specialists showed strong agreement that the features that the DL algorithm was focusing on are the features of PED (mean Likert scale, 4.8 ± 0.3), and SRF (mean Likert scale, 4.5 ± 0.4). However, retina specialists gave relatively lower agreement on IRC (mean Likert scale, 3.6 ± 0.6), and were neutral (mean Likert scale, 3.1 ± 0.4) on the images that included more than two types of pathologies. (Supplementary Document 1).

Discussion

Our study adds to current literature demonstrating that a DL algorithm can effectively classify nAMD from normal eyes using OCT images, doing so with an extremely high performance (AUC of 0.999 at the image and patient-level). The novelty of our study stems from successful reproduction of the results in an ethnically distinct population. Moreover,

saliency maps demonstrated that our DL model focused on pathological features associated with nAMD, in agreement with what clinicians deem important for nAMD diagnosis. This adds transparency and interpretability to the DL algorithm.

We confirmed that our model, trained on a South Korean (predominantly Asian) dataset, performed well (AUC of 0.952 at the patient level) upon external validation on an American (predominantly Caucasian) dataset. This reflects the generalizability of our algorithm across geographic areas and ethnic differences. About OCT scans, generalization across populations is difficult because variations in ethnicity, age, and sex are known to significantly alter the appearance of retinal structures, even in the absence of disease [16]. Furthermore, methodological and statistical considerations such as biased sampling, image quality, unknown systemic error and lack of adequate sample size may confound the results. Concerning AMD subtypes, polypoidal choroidal vasculopathy is a common subtype of AMD in Asian populations, while choroidal neovascularization secondary to AMD is the more common subtype in Western populations [28]. The ability of the algorithm to perform well in a completely different ethnic population suggests that the OCT features that the DL model was extracting and using for decision-making are generalizable regardless of the subtype of AMD.

We believe that the generalizability of our algorithm was, in part, enhanced by the common use of the Heidelberg Spectralis (Heidelberg Engineering, Heidelberg, Germany) imaging system in both centres. The Heidelberg Spectralis is a favoured platform for the diagnosis and monitoring of macular diseases amongst retina specialists and is involved in many OCT-based DL studies [18–22]. With similar imaging conditions, this provided a favourable environment with reduced ‘noise’, contributing to the good performance of our algorithm across two different institutions.

Moving on to model explanation, saliency maps demonstrated that our DL model focused on pathological features associated with nAMD, in agreement with what clinicians deem important for nAMD diagnosis. This is seen on both internal validation (Figures 2–5; Supplementary Figures 3–6) and external validation (Supplementary Figure 6). This verification allows us to safely suggest that the DL model is making classifications based on pathological features rather than systemic errors that cannot be detected. A common systemic error that paradoxically results in the correct classifying patients with nAMD is poor-quality OCT images secondary to cataracts in the elderly. Since nAMD are more common in the elderly, there is a high chance that a patient with significant cataract would also have nAMD. Hence while elderly patients with significant cataract are correctly classified as patients with nAMD, it may not be based on the pathomorphological features of nAMD but instead, due to poor image quality. In contrast, saliency maps in our study showed that the areas of interest identified by our model are associated with clinical diagnoses of SRF, IRC, and PED [24].

A majority of the surveyed retina specialists (Supplementary Document 1) agreed that the developed DL algorithm was appropriately classifying images based on the clinical features of nAMD, specifically PED and SRF. However, when IRC and several pathologies were mixed, saliency maps could not highlight the lesions enough to be deemed satisfactory by

the retina specialists. This is because, as seen in the figures, PED and SRF are well highlighted, whereas in cases of mixed or IRC, saliency maps highlighted only small portions of these lesions. Also, one notable fact is the false positive cases shown in the 9 mm OCT scan in Supplementary Fig 8. The highlighted area shown in the false positive is the raised portion near the prelaminar optic nerve. Therefore, we can assume that DL infers the presence of nAMD from a structure that rises upward on the OCT scans. In our explorative analysis, the performance from 6 mm images was greater than that from 9 mm images (AUCs, 0.999 in 6 mm vs. 0.989 in 9 mm in internal test set) because the optic disc is not shown in 6 mm images. All of the external sets (UW dataset) were 9mm images, which may make performance somewhat lower. Finally, the performance of the DL algorithm for detecting macular lesions such as nAMD can be lowered when the OCT field contains the optic disc (e.g. 9 mm). Our study has several limitations. First, OCT scans with nAMD features were manually selected by a retina specialist, reducing the likelihood of including OCT images of other diseases. Second, our DL algorithm was trained on a very specific purified set of central 5-B-scan images with extremes in phenotypes and age was not matched. The performance of this model may be degraded in the real-world setting when confounded by other ocular pathology or in extrafoveal lesions. Third, this model may not be applicable to images acquired from other OCT platforms. Ideally, future studies would include images from different OCT manufacturers for further generalizability. In addition, our DL algorithm does not diagnose or detect nAMD in a real-world setting with many retinal diseases. This excellent performance was obtained in special controlled data including only normal and nAMD. Therefore, clinical applications could not be discussed based on current results. The features that the DL algorithm is focused on are not specific to nAMD, and the distinction between other retinal diseases such as dry AMD and diabetic retinopathy should be further investigated.

In conclusion, our proposed DL algorithm shows high performance for the differentiation of nAMD from normal OCT images in a Korean population, and generalized well to an ethnically distinct, American population. The model correctly focused on the macula and its surroundings to extract features associated with nAMD, specifically SRF and PED that were confirmed to be clinically relevant. Our present study adds to current literature by implementing visualization techniques on both internal and external validation datasets, improving the interpretability and transparency of DL algorithm.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

We thank Dr. Seo Hee Kim, Dr. Seonghee Choi, Dr. Yong Joon Kim from Yonsei University, Dr. Hong Kyu Kim form Dankook University Hospital, Dr. Jong Yun Yang from Siloam Eye Hospital, Dr. Kyouho Lee from Yonseiwoori eye clinic, and Prof. Yasuo Yanagi from Asahikawa Medical University for their feedback of survey.

Funding

This work was supported by the Agency for Science, Technology and Research of Singapore (A19D1b0095), the National Medical Research Council of Singapore (NMRC/OFLCG/004a/2018; NMRC/CIRG/1488/2018) and the National Medical Research Council of US (K23EY029246).

Data availability statement

Data cannot be shared publicly due to the violation of patient privacy and lack of informed consent for data sharing. Data are available from the Yonsei University, Department of Ophthalmology (contact Prof. Sung Soo Kim, semekim@yuhs.ac) for researchers who meet the criteria for access to confidential data.

References

1. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: The technical and clinical considerations. Prog Retin Eye Res 2019.
2. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. British Journal of Ophthalmology 2019;103:167–75.
3. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, et al. Artificial intelligence in retina. Prog Retin Eye Res 2018;67:1–29. [PubMed: 30076935]
4. Yoo TK, Ryu IH, Lee G, et al. Adopting machine learning to automatically identify candidate patients for corneal refractive surgery. npj Digital Medicine 2019;2:59. [PubMed: 31304405]
5. 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. [PubMed: 29653860]
6. Burlina PM, Joshi N, Pekala M, et al. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. JAMA ophthalmology 2017;135:1170–6. [PubMed: 28973096]
7. Burlina PM, Joshi N, Pacheco KD, et al. Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration. JAMA ophthalmology 2018;136:1359–66. [PubMed: 30242349]
8. Ting DSW, Cheung CY-L, 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. [PubMed: 29234807]
9. Hood DC, De Moraes CG. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. Ophthalmology 2018;125:1207–8. [PubMed: 30032794]
10. Brown JM, Campbell JP, Beers A, et al. Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks. JAMA ophthalmology 2018;136:803. [PubMed: 29801159]
11. Ting DSW, Wu W-C, Toth C. Deep learning for retinopathy of prematurity screening. Br J Ophthalmol 2018:bjophthalmol-2018–313290.
12. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: The technical and clinical considerations. Prog Retin Eye Res 2019;72:100759. [PubMed: 31048019]
13. Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016:2921–9.
14. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering 2018;2:158.
15. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision 2019;128:336–59.
16. Kashani AH, Zimmer-Galler IE, Shah SM, et al. Retinal thickness analysis by race, gender, and age using Stratus OCT. Am J Ophthalmol 2010;149:496–502.e1. [PubMed: 20042179]
17. Treder M, Lauermann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. Graefes Arch Clin Exp Ophthalmol 2018;256:259–65. [PubMed: 29159541]
18. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus Age-related Macular Degeneration. Ophthalmol Retina 2017;1:322–7. [PubMed: 30693348]

19. Kermany DS, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 2018;172:1122–31 e9. [PubMed: 29474911]
20. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018;24:1342–50. [PubMed: 30104768]
21. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning. Ophthalmology 2018;125:549–58. [PubMed: 29224926]
22. Bhatia KK, Graham MS, Terry L, et al. DISEASE CLASSIFICATION OF MACULAR OPTICAL COHERENCE TOMOGRAPHY SCANS USING DEEP LEARNING SOFTWARE: Validation on Independent, Multicenter Data. Retina 2019.
23. Chang BC, Kim NH, Kim YA, et al. Ubiquitous-severance hospital project: implementation and results. Healthc Inform Res 2010;16:60–4. [PubMed: 21818425]
24. Mitchell P, Liew G, Gopinath B, et al. Age-related macular degeneration. Lancet 2018;392:1147–59. [PubMed: 30303083]
25. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 2014.
26. Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention. International conference on machine learning, 2015:2048–57.
27. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. Journal of machine learning research 2011;12:2825–30.
28. Wong CW, Yanagi Y, Lee WK, et al. Age-related macular degeneration and polypoidal choroidal vasculopathy in Asians. Prog Retin Eye Res 2016;53:107–39. [PubMed: 27094371]

Synopsis

Deep learning shows high performance for the differentiation of neovascular age-related macular degeneration from normal OCT images and generalized well to an ethnically distinct. The model correctly focused on the macula and clinically relevant features.

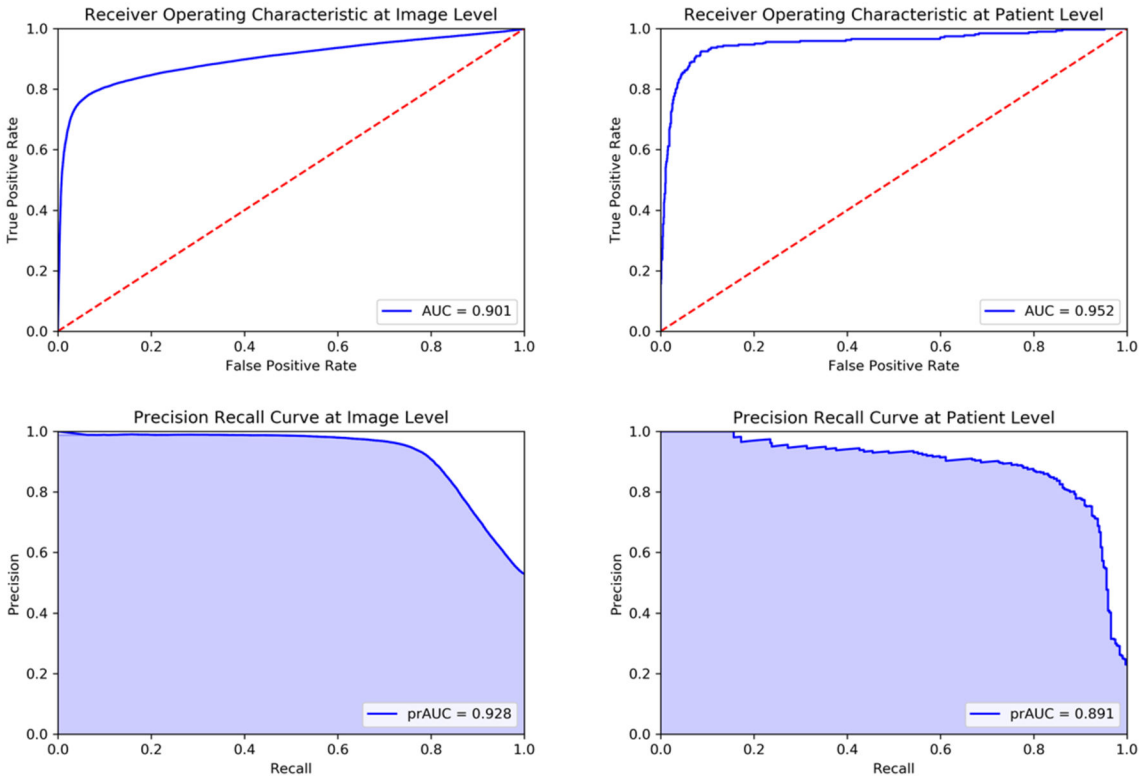

Figure 1.

External test using data at the University of Washington. AUC, area under the receiver operating characteristic curve; AUPRC, area under the precision–recall curve.

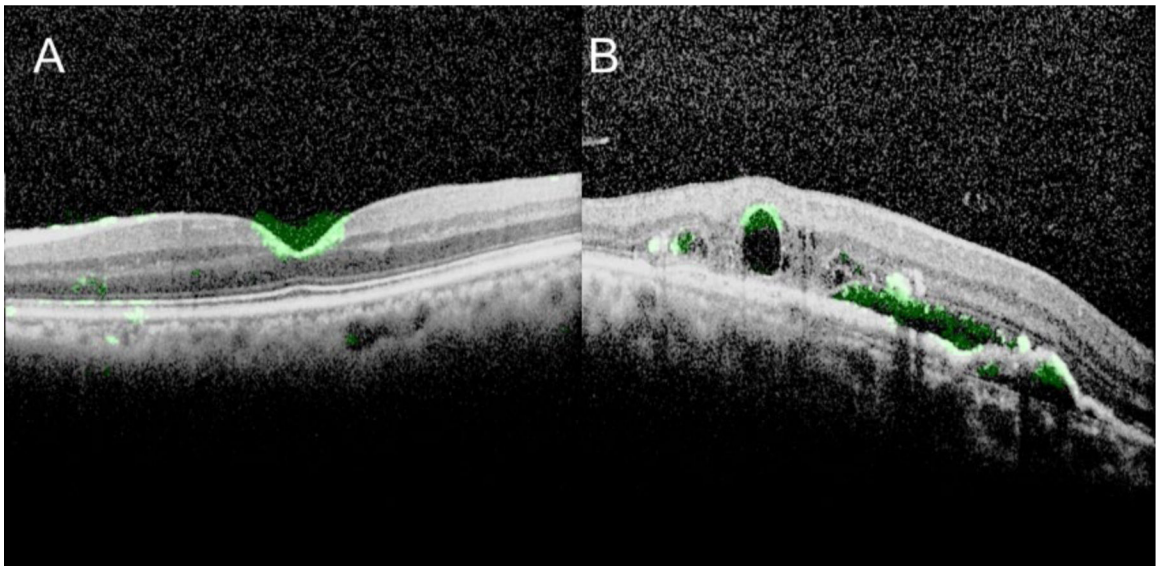

Figure 2. Representative OCT images of normal and neovascular age-related macular degeneration with saliency map.

Fig 2A shows an example of a normal OCT image with the saliency map. The attention is located at the concave part of the inner retina region of the fovea. This image is from enhanced depth imaging (EDI) OCT with the B-scan width of 6 mm. The trained network appeared to value inner retina at the foveal depression in normal eyes to make the classification. For the majority of normal OCT images in the test set, similar attention to the inner retina at the foveal depression was observed (Supplementary Fig 3). In Fig 2B, three components of nAMD, intraretinal fluid or cyst (IRC), subretinal fluid (SRF), and retinal pigment epithelial detachments (PED) were noted in the heatmap.

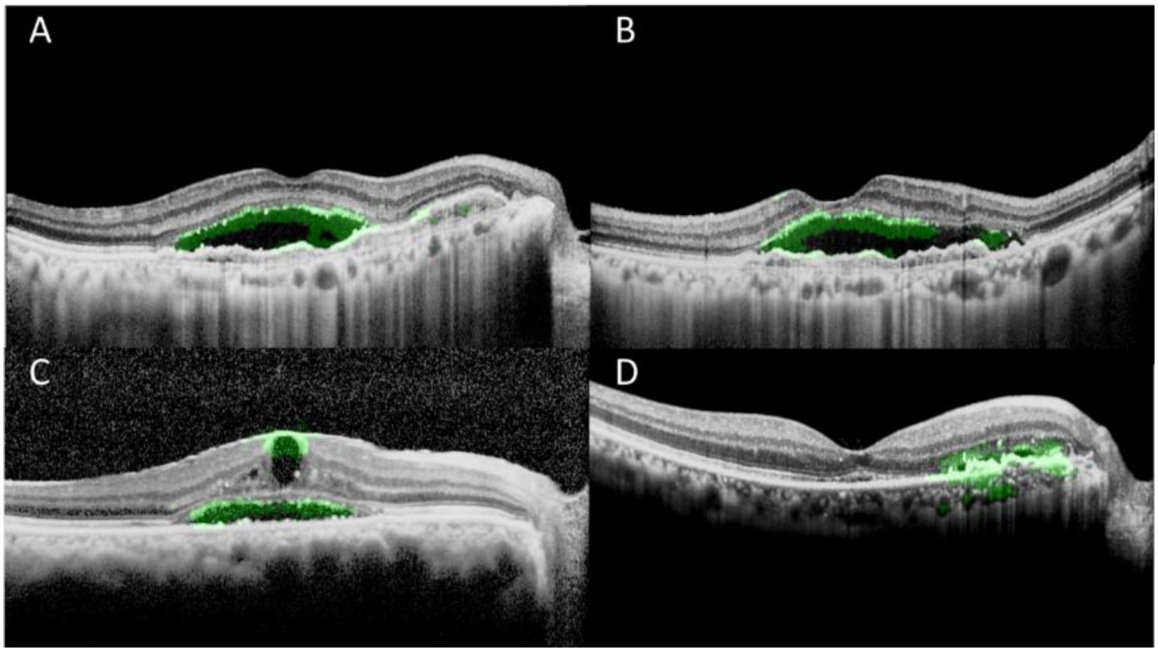

Figure 3. Representative examples of subretinal fluid OCT images with saliency map.

Fig 3A and 3B were from the same patient at different levels of the B-scan. They show that the network located attention on the inner margin of subretinal fluid (SRF), and some on retinal pigment epithelial mottling. Both intraretinal cyst and SRF were noted in Fig 3C, and both pigment epithelial detachment and a small amount of SRF were noted in Fig 3D. More exemplary images of SRF are provided in Supplementary Fig 4.

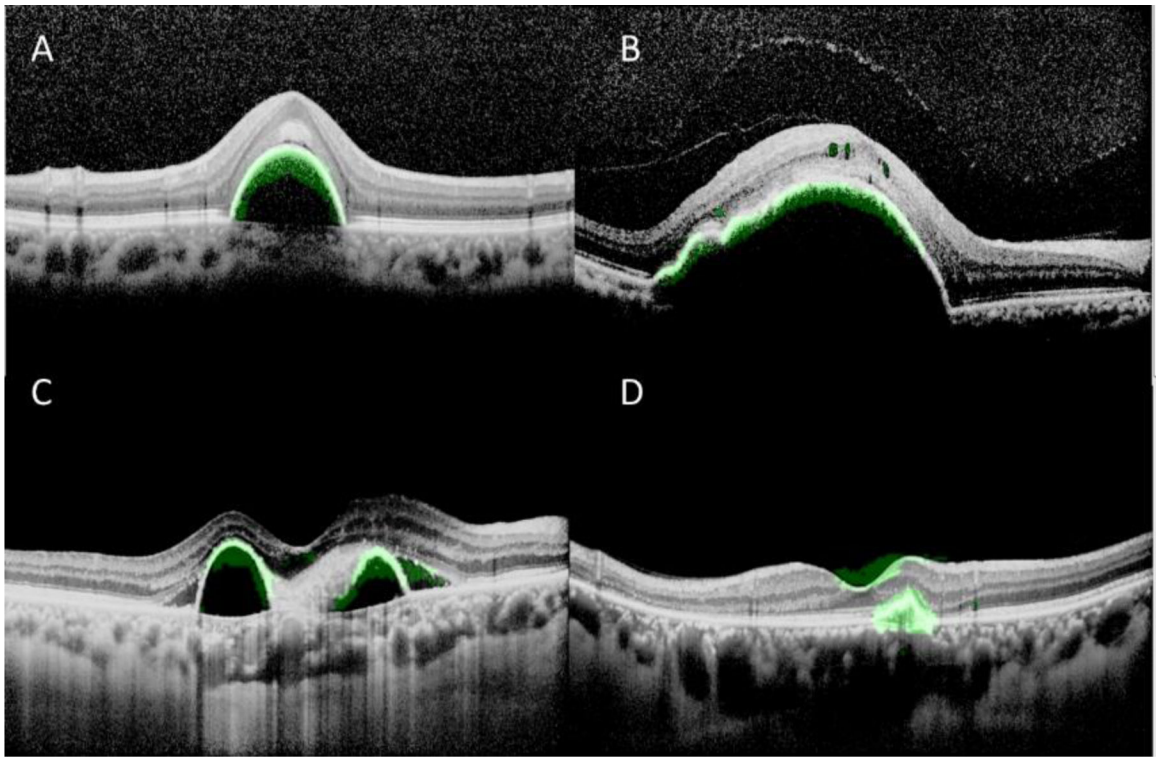

Figure 4. Representative examples of large pigmented epithelial detachment OCT images with saliency map.

It was found that the inner margin of the retinal pigment epithelial detachments (PEDs) were identified regardless of the size of the PED (small lesion in Fig 4A and large lesion in Fig 4B) or a number of the PED (multiple PEDs in Fig 4C). In Fig 4D, there is a small PED with a relatively normal fovea contour without elevation. Both the fovea area and the PED were identified, and the algorithm still correctly predicted this image as a nAMD. More exemplary images of PED are provided in Supplementary Fig 5.

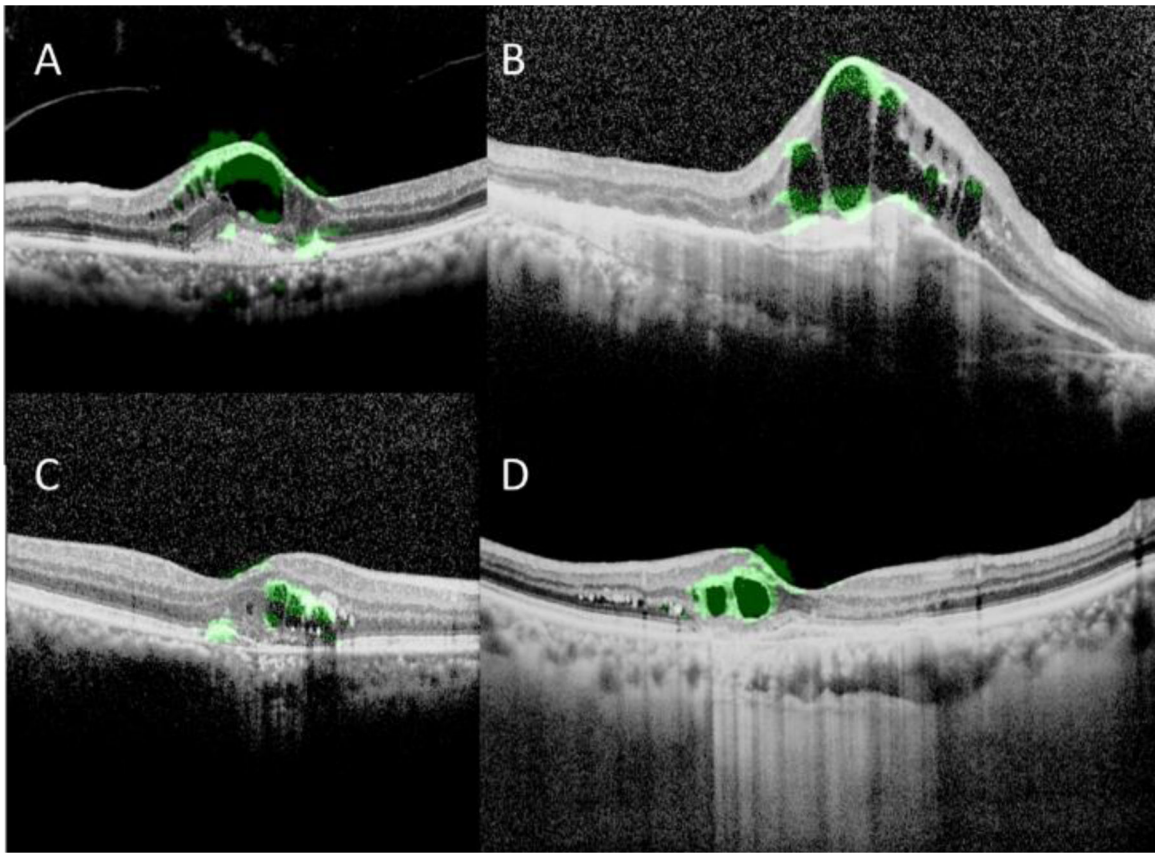

Figure 5. Representative examples of intraretinal cyst OCT images with saliency map.

Fig 5A and 5B show intraretinal fluid or cysts (IRCs) of a relatively large size and Fig 5C shows IRCs of a relatively smaller size. Despite the differences in IRC, attention is located at the inner and outer margin of IRC. In Fig 5D, the entire IRC is noted in the saliency map. More exemplary images of IRC are provided in Supplementary Fig 6.