



OPEN

Epiretinal Membrane Detection at the Ophthalmologist Level using Deep Learning of Optical Coherence Tomography

Ying-Chih Lo^{1,2,3,4,18}, Keng-Hung Lin^{5,6,7,18}, Henry Bair⁸, Wayne Huey-Herng Sheu^{9,10,11,12}, Chi-Sen Chang¹³, Ying-Cheng Shen⁵ & Che-Lun Hung^{14,15,16,17} ✉

Purpose: Previous deep learning studies on optical coherence tomography (OCT) mainly focused on diabetic retinopathy and age-related macular degeneration. We proposed a deep learning model that can identify epiretinal membrane (ERM) in OCT with ophthalmologist-level performance. **Design:** Cross-sectional study. **Participants:** A total of 3,618 central fovea cross section OCT images from 1,475 eyes of 964 patients. **Methods:** We retrospectively collected 7,652 OCT images from 1,197 patients. From these images, 2,171 were normal and 1,447 were ERM OCT. A total of 3,141 OCT images was used as training dataset and 477 images as testing dataset. DL algorithm was used to train the interpretation model. **Diagnostic results** by four board-certified non-retinal specialized ophthalmologists on the testing dataset were compared with those generated by the DL model. **Main Outcome Measures:** We calculated for the derived DL model the following characteristics: sensitivity, specificity, F1 score and area under curve (AUC) of the receiver operating characteristic (ROC) curve. These were calculated according to the gold standard results which were parallel diagnoses of the retinal specialist. **Performance of the DL model** was finally compared with that of non-retinal specialized ophthalmologists. **Results:** Regarding the diagnosis of ERM in OCT images, the trained DL model had the following characteristics in performance: sensitivity: 98.7%, specificity: 98.0%, and F1 score: 0.945. The accuracy on the training dataset was 99.7% (95% CI: 99.4 - 99.9%), and for the testing dataset, diagnostic accuracy was 98.1% (95% CI: 96.5 - 99.1%). AUC of the ROC curve was 0.999. The DL model slightly outperformed the average non-retinal specialized ophthalmologists. **Conclusions:** An ophthalmologist-level DL model was built here to accurately identify ERM in OCT images. The performance of the model was slightly better than the average non-retinal specialized ophthalmologists. The derived model may play a role to assist clinicians to promote the efficiency and safety of healthcare in the future.

¹Division of Nephrology, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan.

²Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA. ³Harvard Medical School, Boston, Massachusetts, USA. ⁴Department of Data Science and Big Data Analytics, Providence University, Taichung, Taiwan.

⁵Department of Ophthalmology, Taichung Veterans General Hospital, Taichung, Taiwan. ⁶Central Taiwan University of Science and Technology, Taichung, Taiwan. ⁷Rong Hsing Research Center for Translational Medicine, National Chung-Hsing University, Taichung, Taiwan.

⁸Stanford University School of Medicine, Stanford, California, USA. ⁹Division of Endocrinology and Metabolism, Department of Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. ¹⁰College of Medicine, National Yang-Ming University, Taipei, Taiwan.

¹¹College of Medicine, National Defense Medical Center, Taipei, Taiwan. ¹²Institute of Biomedical Sciences, National Chung-Hsing University, Taichung, Taiwan. ¹³Division of Gastroenterology, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan.

¹⁴Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan. ¹⁵Department of Computer Science and Communication Engineering, Providence University, Taichung, Taiwan. ¹⁶Department of Computer Science and Information Engineering, Chang Gung University, Taoyuan, Taiwan.

¹⁷AI Innovation Research Center, Chang Gung University, Taoyuan, Taiwan. ¹⁸These authors contributed equally: Ying-Chih Lo and Keng-Hung Lin. ✉e-mail: chung@mail.cgu.edu.tw

Epi-retinal membrane. An epi-retinal membrane (ERM), also known as macular pucker or cellophane maculopathy, is a pathological fibrocellular tissue that forms on the inner surface of the retina. Clinical manifestations vary from asymptomatic cellophane-like films to fibrotic contractile membranes that result in blurred vision, monocular diplopia, micropsia, metamorphopsia, decreased visual acuity, and central vision loss^{1,2}. The exact pathogenic mechanisms remain determined. One hypothesis is that a separation of the vitreous membrane from the retina, or a posterior vitreous detachment, causes inflammation-mediated proliferation of retinal glial cells, fibrous astrocytes, hyalocytes, fibroblasts, myofibroblasts, and macrophages on the retinal surface^{3–5}. ERMs can be either idiopathic or secondary to retinal vascular diseases, ocular inflammatory diseases, and retinal tear or detachment^{6,7}.

The incidence of ERM is 1.1% per eye-year⁸, with estimated prevalence as high as 28.9% (population-dependent)⁹. ERMs occur at higher rates in the elderly population (>65 years of age). Thus, the number of people afflicted likely increases with expanding aging populations.

ERMs are diagnosed based on clinical examination historically. In comparison, the more recently developed optical coherence tomography (OCT) has greater sensitivity¹⁰, and becoming the mainstay for guiding ERM diagnosis and treatment^{11,12}. Spectral domain OCT is a noncontact, noninvasive imaging technique based on the spectral analysis of interference patterns of back-scattered light to form two- and three-dimensional views of living retinal tissues^{13,14}. Depending on the severity of the ERM, its management involves either conservative observation or surgical intervention to peel the membrane away from the retina^{15,16}. If left untreated, ERM may eventually lead to blurred vision and metamorphopsia, impairing the life quality and self-care capability of patients. OCT now plays a vital role in visualizing ERMs, determining the appropriate timing and procedures for their management, as well as the prediction of postoperative outcomes¹⁷.

Computer-aided diagnosis for ocular diseases. Despite the diagnostic advantage of OCT on ocular diseases, interpretation of images is a time-consuming procedure for ophthalmologists. To accelerate the diagnostic process, several studies on ocular images were made to automate the interpretation workflow using various computer vision approaches^{18,19}. Even though, there is still a lot of limitation for the conventional handcrafted feature approach to hinder the widely adoption of computer-aided diagnosis in the clinical settings.

Deep learning in medical imaging. Deep learning (DL) is an algorithm in machine learning. It utilizes statistical and computational methodology to allow the computer to perform intelligent tasks in a data-driven manner. In recent years, due to the rapid growth of data volume and computational capacity, DL approaches have made great advancements in many fields, such as computer vision, voice recognition and nature language processing. The surprising improvement over conventional approaches has positioned DL in the mainstream technique in implementing applications of the artificial intelligence.

Due to the huge success of DL in the field of computer vision, several researchers attempted to apply the technique to medical imaging. For example, Gulshan *et al.* built an automated interpretation model for images of the retinal fundus. It detects referable diabetic retinopathy (RDR) with excellent performance (area under the receiver operating curve, AUC = 0.99)²⁰. Its performance is well comparable with the assessment of ophthalmologists. Ting *et al.* later developed a DL system that can identify disorders like RDR, glaucoma and age-related macular degeneration (AMD) in a multiethnic population²¹. Poplin *et al.* also established a DL model that predicts common cardiovascular risk factors and the occurrence of 5-year major adverse cardiovascular events (MACE)²². Their results supported the usefulness of the DL model in detecting image characteristics perceived by human observers, as well as those more subtle abnormalities human observers do not perceive.

Regarding optical coherence tomography (OCT), DL has been used to discriminate images between age-related macular degeneration and normal retina²³. Kermany *et al.* built a DL model that detects choroidal neovascularization, diabetic macular edema, and drusen OCT images²⁴. The occlusion map further allows the DL model in assisting diagnostic decisions according to manifestations of certain features recognized as deterministic abnormality by domain experts. In addition to image classification, DL was also used to solve segmentation problem for intraretinal fluid in OCT images²⁵.

Aim of the study. DL has been used for the detection of several ocular diseases (such as RDR), but only few studies focus on the ERM identification. Sonobe *et al.*²⁶ confirmed DL model outperform support vector machine (SVM) in the task of ERM detection on 3D-OCT images. However, the performance on routine OCT images was not investigated. In addition, Lu *et al.*²⁷ built a DL model to detect ERM, macular hole, cytooid macular edema and serous macular detachment. The accuracy was non-inferior to domain experts but the model interpretability was not elucidated. Due to ERM is a common manifestation of OCT abnormality (especially in the elderly population), it should be fully studied and regarded as a fundamental building block in developing an OCT interpretation decision support system. The present study is aimed to determine the value of DL in model detection of ERM in retinal OCT images with more comprehensive evaluation.

Materials and Methods

This study was approved by the Institutional Review Board of Taichung Veterans General Hospital (CE18178B) with waiver of informed consent from study participants and adhered to the tenets of Declaration of Helsinki. All collected OCT images received de-identification before further processing.

Datasets. We retrospectively collected OCT images from patients in the Taichung Veterans General Hospital between January 2010 and April, 2018. OCT studies were conducted according to recommendations of board-certified ophthalmologists based on clinical indications. The OCT images were obtained with spectral-domain OCT (Spectralis; Heidelberg Engineering, Heidelberg, Germany) and the raw image data were stored in a centralized workstation. In total, we collected 7,652 central fovea cross section OCT images from 1,197

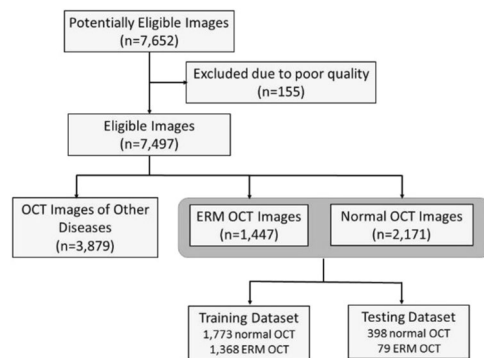


Figure 1. Optical coherence tomography image dataset used for the detection of epiretinal membrane. Flowchart of handling optical coherence tomography (OCT) images, showing data collection and the separation of training and testing datasets. The training dataset was used to train and validate the deep learning model.

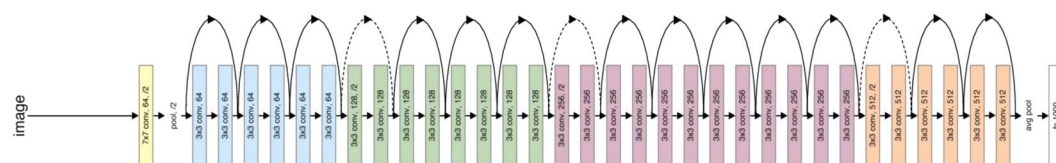


Figure 2. Schematic architecture of residual network (ResNet). ResNet was composed of stacking with multiple residual block. Shortcut connections between layers were added to facilitate the training process. Currently, a number of versions of ResNet are available (such as ResNet-50, ResNet-101 and ResNet-152). In this study, we adopted ResNet-101 for modeling.

patients. Duplicated and poor quality images were first excluded. Each OCT image was classified as normal, ERM or other ocular disease by a senior retinal specialist (with > 18 years of experience). After remove OCT images of other ocular diseases, a total of 3,618 central fovea cross section OCT images from 1,475 eyes of 964 patients were left. Normal ($n = 2,171$) and ERM ($n = 1,447$) OCT images were subsequently selected for analysis. Data were randomly split into either training dataset ($n = 3,141$) for training (and validation), or testing dataset ($n = 477$) for final evaluation of model performance to compare with ophthalmologists (see Fig. 1), and testing dataset is kept aside which is not included in the training dataset. We randomly choose 80% of the training dataset to be the actual training set and the remaining 20% to be the validation set. In order to facilitate the training process, we split the training and testing dataset in a way to let the training dataset have a more balanced class distribution (normal vs. ERM). On the other hand, we created a testing dataset with small proportion of ERM cases, that is similar to the real world OCT images data distribution. Therefore, the evaluation performance would be more likely to reflect that in the real world.

Data preprocessing and labeling. First, the retinal specialists used a well-known open source tool, LabImg²⁸, to annotate the images as ERM or normal. In OCT images, the characteristic morphology of ERM was localized around the central fovea. All labeled images were verified by two experienced retinal specialists. The images with disagreement by the specialists were not included in the experimental dataset. Meanwhile, confusion matrix is usually used to observe the result of classification of a trained model on the training dataset after completing training process. We then performed confusion matrix to verify in case of mislabeling images to affect the classification accuracy. In this study, no images are mislabeled from confusion matrix.

Model training, validation and testing. AlexNet, the state-of-the-art convolutional neural network (CNN); in designing newer network architecture is to go deeper into the data with more layers in the model. The conventional AlexNet has only 5 convolutional layers, other networks like VGG network²⁹ or GoogleNet (also code-named as Inception_v1)³⁰ have more layers (like 19 or 22). He *et al.* proposed a residual learning framework, called ResNet³¹, and they obtained a remarkably successful outcome in the ILSVRC 2015 competition. The key idea of ResNet (Fig. 2) is in its modeling the residual of the intermediate output, instead of the intermediate output (like in the traditional models). ResNet is able to train extremely deep networks with stochastic gradient descent (SGD) through the use of residual modules. It is also able to train a network with large amount of layers while keeping low complexity (compared with VGGNet) and it has achieved with a particular dataset a top-5 error rate of 3.57%, a performance level better than human. Currently, a number of versions of ResNet are available, with the more popular ones being ResNet-50, ResNet-101 and ResNet-152. In this study, we adopted ResNet-101 for modeling. In total, we used 3,141 OCT images for model training. Among the training datasets, 20% was used as the validation data to guide the tuning of the network hyperparameters.

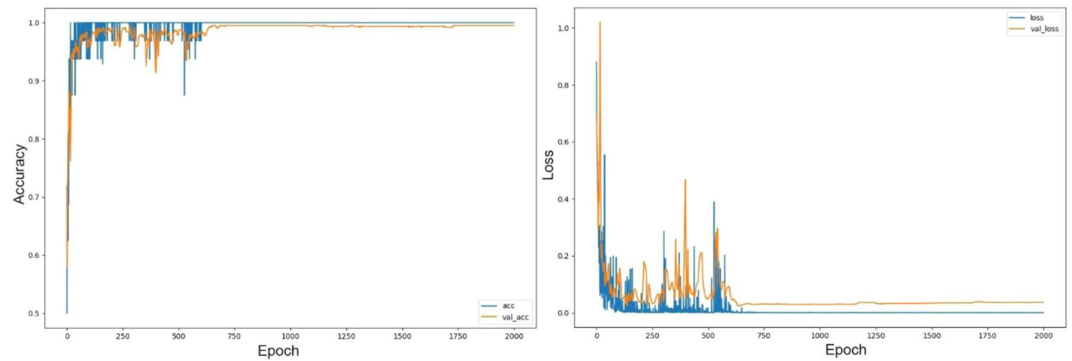


Figure 3. Learning curve of the derived deep learning model. The blue one is the result for the training dataset, while the orange one indicate that for the validation dataset. (Left panel: accuracy, Right panel: loss).

The framework used to train our models is Python 3.6.4 + Keras 2.2.4 on a workstation equipped with Intel Core i7-6850K, 128 GB ram and NVIDIA GTX 1080Ti graphic card. The parameters utilized in the training were the following: learning rate, 0.0001; batch size, 32; epoch, 2000, and optimizer, Adaptive Moment Estimation (Adam).

Statistics on testing dataset. In order to evaluate the performance of the derived model, we first calculated the area under curve (AUC) of the receiver operating characteristic (ROC) curve for the model prediction in an unseen testing dataset. Next, we determined the following as evaluation metrics for the final model: the accuracy on the training data and the accuracy, sensitivity, specificity and F1 score on the testing data. Cohen's kappa index was used to measure the inter-rater agreement of the four ophthalmologists on the testing dataset. Confusion matrix were also generated to investigate the detail of the misinterpretation. All statistical analyses were performed using R Statistics software (v3.4.1).

Model performance compared with clinicians. To evaluate the usefulness of the DL model in the clinical settings, four board-certified non-retinal specialized ophthalmologists of different clinical experiences were asked to interpret the unseen testing dataset which was used for the final model evaluation. Statistics with sensitivity and specificity were used to evaluate the performance of human expert on the task of OCT ERM identification. The performance of the ophthalmologists was finally compared with the DL model to validate its usefulness in the real world.

Model visualization. To gain deeper understanding on the logic of DL model, some methods were proposed to make the prediction result more explainable. Gradient-weighted class activation mapping (Grad-CAM)³² is a well-known approach to produce a coarse localization map highlighting the important regions of the image that the machine learned to identify the classes. In our study, this approach was implemented before the last fully-connected layer of ResNet.

Results

Finally, 3,141 OCT images were used for model training and 20% ($n = 628$) of them were validation dataset. During the training process, the accuracy and loss metrics were monitored and plotted as learning curves. Figure 3 shows that the model converged after 700 epochs and the training continued until 2,000 iterations. No obvious model overfitting was found. The prediction accuracy on training data was 99.7% (95% confidence interval: 99.4 - 99.9%). Due to the limitation of memory capacity of GPU device, the batch size we used is 32. Therefore, the issue of mini-batch gradient descent led to the spikes of loss values at the early stage before 700 epochs shown in Fig. 3.

When DL model was applied on an unseen testing dataset ($n = 477$), the accuracy was 98.1% (95% confidence interval: 96.5 - 99.1%). Sensitivity, specificity and F1 score on the testing data were 98.7%, 98.0% and 0.945, respectively. ROC curve of the model (AUC: 0.999) are shown in Fig. 4 together with the results of evaluation by four ophthalmologists. The close-up view (Fig. 4B) shows the DL model performed slightly better than the average of the participated ophthalmologists (pink symbol). During the error analysis, we found the DL model was more likely to result in false positive and false negative error with OCT images from myopia patients. Besides, after reviewing the false positive cases, we also identify some cases with suspicious early manifestation of ERM, indicating the derived model is quite sensitive in ERM detection. Table 1 showed the inter-rater agreement between the ophthalmologists and DL model and the confusion matrices of the clinicians' interpretation on the testing dataset were provided in Table 2. During reviewing the disagreed images between the four ophthalmologists, we found majority of the disagreement occur in OCT images with only subtle ERM change. However, there are still few apparent misinterpretation by the clinicians noted.

Figure 5A,B shows examples of normal and ERM OCT images with Grad-CAM visualization effect overlaid. Regions highlighted with warmer colors represent those areas more important for the final class determination. The ERM region of interest (ROI) was captured precisely and results are compatible with judgement of the retinal specialist.

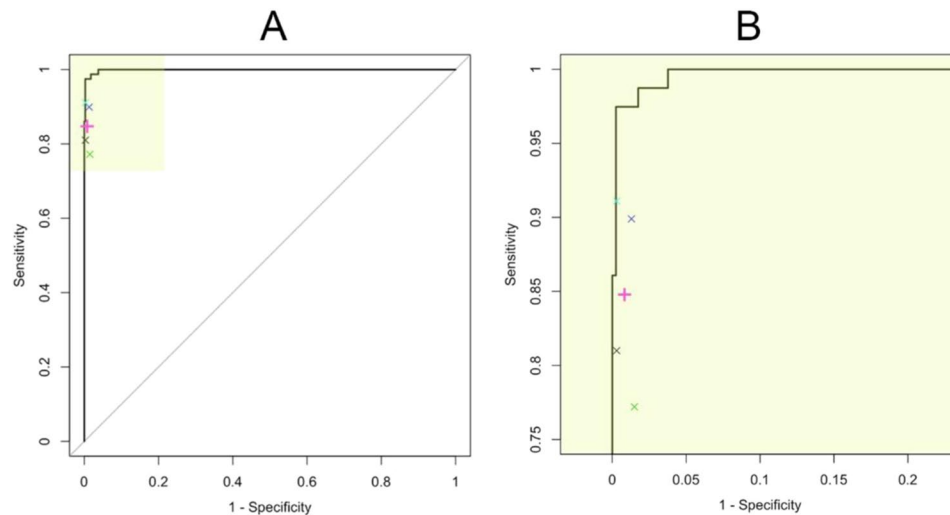


Figure 4. Receiver operating characteristic (ROC) curve for the identification of epiretinal membrane in the testing dataset. Evaluation results of four ophthalmologists are plotted with their average performance (pink symbol). (Panel A: original ROC curve; Panel B: close-up view of the high-lighted area in panel A).

	Clinician 1	Clinician 2	Clinician 3	Clinician 4	DL model**
Clinician 1	1.00				
Clinician 2	0.86	1.00			
Clinician 3	0.88	0.88	1.00		
Clinician 4	0.78	0.77	0.78	1.00	
DL model	0.87	0.87	0.92	0.79	1.00

Table 1. Inter-rater agreement* for clinicians and deep learning model. *Measurement with Cohen's kappa index. **DL: Deep learning.

	Actual (+)	Actual (-)
Clinician 1		
Predict (+)	71	5
Predict (-)	8	393
Clinician 2		
Predict (+)	64	1
Predict (-)	15	397
Clinician 3		
Predict (+)	72	1
Predict (-)	7	397
Clinician 4		
Predict (+)	61	6
Predict (-)	18	392

Table 2. Confusion matrix of the clinicians.

Discussion

Beginning with the proposal of Krizhevsky *et al.* modifications on conventional architecture of CNN were made with wider use of multiple graphics processing units (GPU) to accelerate the computational operations, DL has greatly improved results in the field of computer vision³³. In 2016, Gulshan *et al.* of Google successfully developed a DL model that can detect RDR in retinal fundus color photography with ophthalmologist-level performance²⁰. Other studies applied DL in different medical images, such as on skin, pathology, chest X-ray and electrocardiography^{34–37}. Increasing evidence has indicated the potential and feasibility of utilizing DL in the interpretation of medical images.

In this study, we implemented a DL model that outperformed non-retinal specialized ophthalmologists in ERM identification. Our application can help to accelerate the process and lower the cost of ERM diagnosis. It is especially useful for regions with limited access to retinal specialist due to various reasons (such as economic issues or medical resource allocation). Further and timely referral to retinal specialist can be allocated to those whose abnormality has been detected by the DL model.

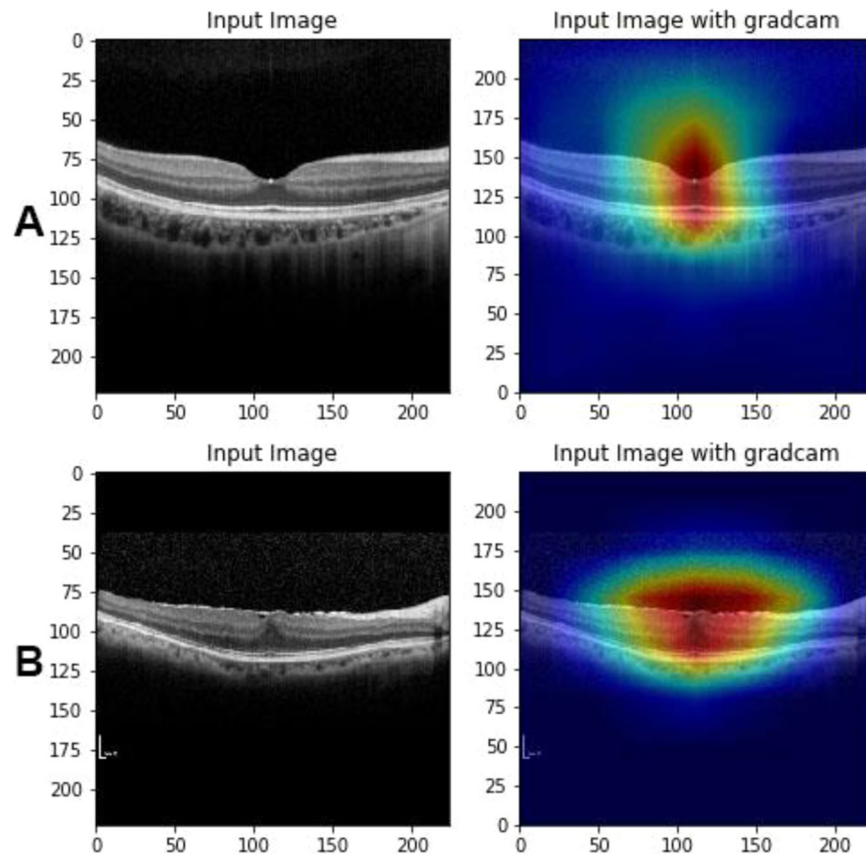


Figure 5. Exemplary OCT Images of normal and the epiretinal membrane (ERM) in patients. Important area for pattern recognition is highlighted with gradient-weighted class activation mapping shown on the right panels. (Panel A: normal OCT, Panel B: ERM OCT).

As for the diagnosis of ocular diseases, non-mydriatic fundus photography is a convenient tool of examination due to its non-requirement of pupil dilatation, and hence widely used to screen for diabetic retinopathy. Its drawback is not able to detect subtle abnormalities. Therefore, the OCT remains the gold-standard diagnostic tool for many retinal diseases. In previous studies, DL has been used to interpret and identify choroidal neovascularization (CNV), diabetic macular edema (DME), and drusen OCT images³⁸. Sonobe *et al.* also confirm the superiority of DL model over SVM in ERM detection with 3D-OCT images. However, the 3D-OCT images were not supported by all the OCT imaging machine and the generalizability to routine OCT images were not investigated in their study. An OCT image DL classification model with competitive performance with domain experts were developed by Lu *et al.* However, the interpretability of the model was not elucidated. In our study, DL model showed no inferiority compared with the ophthalmologists, supporting the potential use of DL in OCT interpretation. Grad-CAM visualization confirm the validity and the robustness of the derived ERM detection DL model. ERM has not been fully studied yet but it is a prevalent disease among the elderly and is also a common finding in OCT images. An DL model for ERM identification could be an essential component in an automatic and comprehensive interpretation model for OCT. In this study, we have developed a DL model that can distinguish between ERM and normal OCT with ophthalmologist-level accuracy. We believed the established model can further improve the applicability of DL model in the highly versatile clinical settings when combined with previous developed models (like that by Kermany *et al.*) in analyzing OCT images³⁸. The derived DL model may be used in the clinical settings to shorten the time period from examination to the diagnosis and increase the efficiency and efficacy of our healthcare. In addition, when the automatic DL model combine into the clinical workflow, it can also help the clinicians to avoid the occurrence of the medical error and misdiagnosis. Therefore, the derived model may also potentially play a role as a clinical decision support system to promote the patient safety in the future. In the critical period of the healthcare burden overloading, such as the COVID-19 pandemic, the DL based automatic model may also assist the clinicians to decrease the healthcare workload and prevent the healthcare providers from burnout.

Conclusion

An ophthalmologist-level DL model has been developed here to accurately identify epiretinal membrane in OCT images. Due to the high prevalence disorders of epiretinal membrane, our model could form an essential component in automatic interpretation system for OCT images. The derived DL model may assist the clinicians to promote the efficiency and safety of healthcare in the future.

Received: 27 February 2020; Accepted: 4 May 2020;

Published online: 21 May 2020

References

- Mandal, N., Kofod, M. & Vorum, H. *et al.* Proteomic analysis of human vitreous associated with idiopathic epiretinal membrane. *Acta Ophthalmol.* **91**(4), e333–4 (2013).
- Ghazi-Nouri, S. M., Tranos, P. G. & Rubin, G. S. *et al.* Visual function and quality of life following vitrectomy and epiretinal membrane peel surgery. *Br. J. Ophthalmol.* **90**(5), 559–62 (2006).
- Gupta, P., Yee, K. M. & Garcia, P. *et al.* Vitreoschisis in macular diseases. *Br. J. Ophthalmol.* **95**(3), 376–80 (2011).
- Sebag, J. Anomalous posterior vitreous detachment: a unifying concept in vitreo-retinal disease. *Graefes Arch. Clin. Exp. Ophthalmol.* **242**(8), 690–8 (2004).
- Joshi, M., Agrawal, S. & Christoforidis, J. B. Inflammatory mechanisms of idiopathic epiretinal membrane formation. *Mediators Inflamm.* **2013**, 192582 (2013).
- de Bustros, S., Thompson, J. T. & Michels, R. G. *et al.* Vitrectomy for idiopathic epiretinal membranes causing macular pucker. *Br. J. Ophthalmol.* **72**(9), 692–5 (1988).
- Appiah, A. P. & Hirose, T. Secondary causes of premacular fibrosis. *Ophthalmology.* **96**(3), 389–92 (1989).
- Fraser-Bell, S., Guzowski, M. & Rochtchina, E. *et al.* Five-year cumulative incidence and progression of epiretinal membranes: the Blue Mountains Eye Study. *Ophthalmology.* **110**(1), 34–40 (2003).
- Ng, C. H., Cheung, N. & Wang, J. J. *et al.* Prevalence and risk factors for epiretinal membranes in a multi-ethnic United States population. *Ophthalmology.* **118**(4), 694–9 (2011).
- Do, D. V., Cho, M. & Nguyen, Q. D. *et al.* The impact of optical coherence tomography on surgical decision making in epiretinal membrane and vitreomacular traction. *Trans. Am. Ophthalmol. Soc.* **104**, 161–6 (2006).
- Koizumi, H., Spaide, R. F. & Fisher, Y. L. *et al.* Three-dimensional evaluation of vitreomacular traction and epiretinal membrane using spectral-domain optical coherence tomography. *Am. J. Ophthalmol.* **145**(3), 509–17 (2008).
- Goldberg, R. A., Waheed, N. K. & Duker, J. S. Optical coherence tomography in the preoperative and postoperative management of macular hole and epiretinal membrane. *Br. J. Ophthalmol.* **98**(Suppl 2), ii20–3 (2014).
- Wojtkowski, M., Leitgeb, R. & Kowalczyk, A. *et al.* *In vivo* human retinal imaging by Fourier domain optical coherence tomography. *J. Biomed. Opt.* **7**(3), 457–63 (2002).
- Yaqoob, Z., Wu, J. & Yang, C. Spectral domain optical coherence tomography: a better OCT imaging strategy. *Biotechniques.* **39**(6 Suppl), S6–13 (2005).
- Folk, J. C., Adelman, R. A. & Flaxel, C. J. *et al.* Idiopathic Epiretinal Membrane and Vitreomacular Traction Preferred Practice Pattern(R) Guidelines. *Ophthalmology.* **123**(1), P152–81 (2016).
- Dawson, S. R., Shunmugam, M. & Williamson, T. H. Visual acuity outcomes following surgery for idiopathic epiretinal membrane: an analysis of data from 2001 to 2011. *Eye.* **28**(2), 219–24 (2014).
- Stevenson, W., Prospero Ponce, C. M. & Agarwal, D. R. *et al.* Epiretinal membrane: optical coherence tomography-based diagnosis and classification. *Clin. Ophthalmol.* **10**, 527–34 (2016).
- Mookiah, M. R., Acharya, U. R. & Koh, J. E. *et al.* Automated diagnosis of Age-related Macular Degeneration using greyscale features from digital fundus images. *Comput. Biol. Med.* **53**, 55–64 (2014).
- Balakrishnan, U., Venkatachalapathy, K. & Marimuthu, G. S. A Hybrid PSO-DEFS Based Feature Selection for the Identification of Diabetic Retinopathy. *Curr. Diabetes Rev.* **11**(3), 182–90 (2015).
- 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.* **316**(22), 2402–10 (2016).
- Ting, D. S. W., Cheung, C. Y. & 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.* **318**(22), 2211–23 (2017).
- Poplin, R., Varadarajan, A. V. & Blumer, K. *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Engineering.* **2**(3), 158–64 (2018).
- Lee, C. S., Baughman, D. M. & Lee, A. Y. Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images. *Ophthalmol. Retina.* **1**(4), 322–7 (2017).
- Kermany, D. S., Goldbaum, M. & Cai, W. *et al.* Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell.* **172**(5), 1122–31 e9 (2018).
- Lee, C. S., Tyring, A. J. & Deruyter, N. P. *et al.* Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed. Opt. Express* **8**(7), 3440–8 (2017).
- Sonobe, T., Tabuchi, H. & Ohsugi, H. *et al.* Comparison between support vector machine and deep learning, machine-learning technologies for detecting epiretinal membrane using 3D-OCT. *Int. Ophthalmol.* **39**(8), 1871–1877 (2019).
- Lu, W., Tong, Y. & Yu, Y. *et al.* Deep Learning-Based Automated Classification of Multi-Categorical Abnormalities From Optical Coherence Tomography Images. *Transl. Vis. Sci. Technol.* **7**, 41 (2018).
- Tzatalin. Labelling Free Software. MIT License. 2015.
- Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv e-prints2014; v. 1409.
- Szegedy, C., Wei, L., Yangqing, J. *et al.* Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)2015.
- He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. ArXiv e-prints2015; v. 1512.
- Selvaraju, R. R., Cogswell, M., Das, A. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. ArXiv e-prints2016; v. 1610.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. Lake Tahoe, Nevada: Curran Associates Inc., 2012.
- Esteva, A., Kuprel, B. & Novoa, R. A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* **542**(7639), 115–8 (2017).
- Ehteshami Bejnordi, B., Veta, M. & Johannes van Diest, P. *et al.* Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA.* **318**(22), 2199–210 (2017).
- Rajpurkar, P., Irvin, J., Zhu, K. *et al.* CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. ArXiv e-prints2017; v. 1711.
- Rajpurkar, P., Hannun, A. Y., Haghpanahi, M. *et al.* Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. ArXiv e-prints2017; v. 1707.
- Kermany, D. S., Goldbaum, M. & Cai, W. *et al.* Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell.* **172**(5), 1122–31 e9 (2018).

Acknowledgements

We highly appreciate the participated ophthalmologists CC Chou, LC Wei, YS Cheng and YC Wu for their assistance in providing a baseline interpretation as the benchmark for evaluating the DL model. Supported by Taichung Veterans General Hospital (grant TCVGH- 1070105D), MOST108-2218-E-126-003, and MOST108-2221-E-010-013-MY3.

Author contributions

Ying-Chih Lo, Keng-Hung Lin, and Che-Lun Hung wrote the main manuscript text and Che-Lun Hung designed the algorithm and experiments. Ying-Chih Lo and Keng-Hung Lin collected the data and verify the data. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.-L.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020