



Deep learning for identification of peripheral retinal degeneration using ultra-wide-field fundus images: is it sufficient for clinical translation?

Tien-En Tan¹, Daniel Shu Wei Ting^{1,2}, Tien Yin Wong^{1,2}, Dawn A. Sim^{3,4,5}

¹Singapore Eye Research Institute, Singapore, Singapore National Eye Centre, Singapore; ²Duke-National University of Singapore Medical School, Singapore; ³Moorfields Eye Hospital, London, UK; ⁴National Institute for Health and Research Biomedical Centre, Moorfields Eye Hospital, London, UK; ⁵Institute of Ophthalmology, University College London, London, UK

Correspondence to: Dawn A. Sim, BSc (Hons), MBBS, FRCOphth, PhD. Consultant Ophthalmic Surgeon, Moorfields Eye Hospital, NHS Foundation Trust, 162 City Road, London EC1V 2PD, London, UK. Email: dawnsim@nhs.net.

Provenance and Peer Review: This article was commissioned by the editorial office, *Annals of Translational Medicine*. The article did not undergo external peer review.

Comment on: Li Z, Guo C, Nie D, *et al.* A deep learning system for identifying lattice degeneration and retinal breaks using ultra-widefield fundus images. *Ann Transl Med* 2019;7:618.

Submitted Jan 23, 2020. Accepted for publication Mar 04, 2020.

doi: 10.21037/atm.2020.03.142

View this article at: <http://dx.doi.org/10.21037/atm.2020.03.142>

In recent years, deep learning (DL) systems have emerged as powerful tools for automated analysis of medical images (1,2). In the field of ophthalmology, DL systems have similarly been successfully applied to a large number of ocular diseases using a variety of ocular imaging modalities (3). From standard colour fundus photography, DL systems have been able to accurately detect significant causes of blindness, such as referable or vision-threatening diabetic retinopathy (DR), age-related macular degeneration (AMD) and glaucoma (4-6). Using optical coherence tomography (OCT) scans, DL systems have also been able to accurately classify AMD and a number of other sight-threatening retinal diseases such as choroidal neovascularization, macular edema and central serous chorioretinopathy (7,8). As the technology continues to develop, clinically validated DL systems are likely to be integral to the future delivery of patient care; particularly in the field of teleophthalmology where DL systems are likely to be deployed, either as assistive or fully-automated decision-making triage and diagnostic tools (9).

In this study, Li *et al.* use ultra-widefield fundus (UWF) images, which provide a 200° panoramic image of the retina with a single image capture (compared to 45° in standard colour fundus photography). UWF imaging is particularly well-suited to use in teleophthalmology, as it has a rapid

acquisition time for the area of retina imaged, and allows better detection of peripheral retinal pathology. The clinical deployment of UWF imaging in the field of DR screening has demonstrated lower rates of ungradable images, reduced image evaluation time, and higher rates of pathology detection (10). DL systems have also been applied to UWF images. To date, these algorithms have been shown to be able to detect a number of sight-threatening ocular diseases; including proliferative DR, AMD, central and branch retinal vein occlusions, macular holes and rhegmatogenous retinal detachments (RRDs) (11-16).

In this study, Li *et al.* developed and validated a DL system for the detection of lattice degeneration and retinal breaks using UWF Optos images (Optos PLC, Dunfermline, Scotland), on a dataset of 5,606 images from two ophthalmic hospitals in China. All the images were manually classified by three experienced retinal specialists based on the presence or absence of lattice degeneration or retinal breaks [termed notable peripheral retinal lesions (NPRLs) by the authors], with adjudication by a fourth retinal specialist in cases of disagreement. After training and validation, 12 different DL models were tested on an independent test set of 750 images. These 12 models were derived from four different DL algorithms, each using three different methods for image preprocessing.

Analyzing the performance of these 12 different models showed that InceptionResNetV2 consistently outperformed the other algorithms [average area under the receiver operating characteristic curve (AUC) of 0.996], and that the optimal method of image preprocessing involved applying brightness shift, rotation and mirror flipping augmentation to approximately five times the original size. Their best DL model achieved remarkable performance, with an AUC of 0.999, sensitivity of 98.7%, and specificity of 99.1% for the detection of NPRLs. This DL model was also tested against, and outperformed, two general ophthalmologists in classification of UWF images in the test dataset. Furthermore, they examined heatmap visualizations of all 154 true-positive images, and found that 150 (97.4%) of these images showed accurate localization of the NPRLs in question.

To date, prior studies using DL systems for analysis of UWF images have focused mainly on posterior pole pathology (11-16). This is the first study to examine peripheral retinal pathologies, which are conceivably more difficult to detect even on clinical examination, and known to be more prone to image distortion (globe effect and pincushion distortion). It adds lattice degeneration or retinal breaks to the growing list of ocular diseases that can be potentially detected by DL systems, with potential future applications in screening and teleophthalmology platforms. Both lattice degeneration and retinal breaks increase the risk of RRD, a potentially blinding condition with an incidence of 6.3 to 17.9 per 100,000 population (17). Furthermore, RRD is highly associated with myopia—a growing global problem with a prevalence of 27% in 2010, and predicted to rise to 50% in 2050 (18). As such, an automated solution that enables early detection and treatment of risks factors of RRD will be of huge clinical and economic significance. In this study, the DL model developed performed the task of “automated detection of peripheral retinal pathology” with a high degree of accuracy. Additionally, it utilized a strong reference standard, relying on image classification by three experienced retinal specialists, with a fourth retinal specialist for adjudication. From a technical perspective, the number of different algorithms and image preprocessing techniques used in this study are valuable for moving the field of DL-based analysis of UWF images forward. The heatmap visualization analyses presented further lend credence to the strong performance of the DL system they developed. Heatmap visualizations are often performed in DL-based studies to show a few representative images, and demonstrate qualitatively that the DL system is detecting

the actual lesions of interest. However, by taking the extra step to quantify the percentage of all their true-positive cases with accurate heatmap localization (in a similar manner to studies by Kermany *et al.* and Keel *et al.*), the authors effectively demonstrate the robustness of their results, and further increase confidence in the performance of their system (19,20). Other studies on DL-based image analysis should strive to adopt a similar methodology for quantifying accurate heatmap visualization, which would help to further address the “black box” phenomenon that is a real barrier to clinical implementation of many DL systems.

This work represents a good “first step” in automating the detection of peripheral retinal features that places an individual at risk for RRD but other challenges lie ahead before clinical deployment of such an algorithm is practicable. Firstly, it is important to remember that this DL system was only tested on the primary dataset. DL systems should be tested on external “testing” datasets, ideally from a variety of different countries, populations and settings, in order to demonstrate that it is truly generalizable to other populations. For example, Bellemo *et al.* used a DL algorithm for DR screening developed in Singapore, and validated this in population-based DR screening program in Africa (21).

Second, studies on DL systems should endeavor to include more information on the datasets used for development and validation. Besides basic demographic data like age and sex, other relevant information in this context would include refractive status, lens status, and whether or not these patients were symptomatic for floaters or photopsia. Providing detailed information on the datasets used is crucial for clinicians and readers to determine if the DL system developed is likely to be applicable to their own patients.

Third, assessment of image quality is another important factor to consider in the translation of DL systems to clinical practice. In this study, as with many others, poor quality images were screened out manually by human graders—a poignant health economic consideration for clinical deployment. This DL system was also therefore only exposed to and tested on images of good quality. If it were to be clinically deployed, there would have to be an accompanying method of grading or screening images for quality. Human grading of quality before DL system analysis would be time-consuming and resource intensive, and reduce the efficiency of the system. While use of UWF images may reduce the number of ungradable images

compared to standard fundus photography, there is still likely to be a substantial number of these poor-quality images; 10.7% of the UWF images initially screened for this study were deemed to be of poor quality and excluded. Image quality is of particular concern when attempting to screen for peripheral retinal pathology, as any areas of peripheral retina that are obscured by the eyelids or lash artifacts would likely impact the sensitivity of the system negatively. One possible approach to tackling this problem is to first develop a DL algorithm for automated screening of image quality (22,23). This should be applied to all images as a “first cut”, before any subsequent downstream image analysis. In a real-world screening or tele-ophthalmology platform, such ungradable images would either be re-acquired, or referred on to ophthalmology services, to screen for media opacity, cataracts, or other pathology.

Finally, to really have effective translation of DL systems into clinical practice, studies utilizing DL systems for automated image analysis should be designed with the end in mind. As the authors rightly point out, their study did not classify lattice degeneration and retinal breaks independently, which can make clinical application challenging. In clinical practice, lattice degeneration and retinal breaks are lesions with vastly different implications and risks for development of RRD. Consequently, the urgency of ophthalmic assessment and treatment required for the two are quite different. If these two types of lesions were independently classified and the DL system was trained to independently detect them, this would make the system much more effective in clinical practice. Also, while it is important to acknowledge that the DL system may be technically very accurate, and capable of outperforming general ophthalmologists in terms of UWF image classification, the real question in clinical translation is whether it is truly comparable to the “gold standard” evaluation.

DL-based analysis of UWF images has potential to play an important role in the future delivery of teleophthalmology screening platforms. Li et al have developed a novel DL system that is able to detect peripheral retinal pathology such as lattice degeneration and retinal breaks on UWF images, with performance surpassing that of human readers on UWF images. This adds to the growing list of ocular diseases that can potentially be detected from UWF images using DL systems. Not only is significant work required to translate this to meaningful clinical application, but the medicolegal

territories of algorithms “missing retinal breaks” remains uncharted. Future work in this area should focus on tackling these real-world challenges, and should be designed with the end goal of scalable and economically viable clinical translation in mind.

Acknowledgments

Funding: None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
2. Ker J, Wang L, Rao J, et al. Deep Learning Applications in Medical Image Analysis. *IEEE Access* 2018;6:9375-89.
3. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019;103:167-75.
4. 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.
5. Li Z, He Y, Keel S, et al. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology* 2018;125:1199-206.

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 Ophthalmol* 2017;135:1170-6.
7. 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.
8. 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.
9. Korot E, Wood E, Weiner A, et al. A renaissance of teleophthalmology through artificial intelligence. *Eye (Lond)* 2019;33:861-3.
10. Silva PS, Cavallerano JD, Tolls D, et al. Potential efficiency benefits of nonmydriatic ultrawide field retinal imaging in an ocular telehealth diabetic retinopathy program. *Diabetes Care* 2014;37:50-5.
11. Nagasawa T, Tabuchi H, Masumoto H, et al. Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naïve proliferative diabetic retinopathy. *Int Ophthalmol* 2019;39:2153-9.
12. Nagasato D, Tabuchi H, Ohsugi H, et al. Deep Neural Network-Based Method for Detecting Central Retinal Vein Occlusion Using Ultrawide-Field Fundus Ophthalmoscopy. *J Ophthalmol* 2018;2018:1875431.
13. Nagasato D, Tabuchi H, Ohsugi H, et al. Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion. *Int J Ophthalmol* 2019;12:94-9.
14. Matsuba S, Tabuchi H, Ohsugi H, et al. Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning, a machine-learning technology, for detecting age-related macular degeneration. *Int Ophthalmol* 2019;39:1269-75.
15. Ohsugi H, Tabuchi H, Enno H, et al. Accuracy of deep learning, a machine-learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. *Sci Rep* 2017;7:9425.
16. Nagasawa T, Tabuchi H, Masumoto H, et al. Accuracy of deep learning, a machine learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting idiopathic macular holes. *PeerJ* 2018;6:e5696.
17. Mitry D, Charteris DG, Fleck BW, et al. The epidemiology of rhegmatogenous retinal detachment: geographical variation and clinical associations. *Br J Ophthalmol* 2010;94:678-84.
18. Holden BA, Wilson DA, Jong M, et al. Myopia: a growing global problem with sight-threatening complications. *Community Eye Health* 2015;28:35.
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.
20. Keel S, Wu J, Lee PY, et al. Visualizing Deep Learning Models for the Detection of Referable Diabetic Retinopathy and Glaucoma. *JAMA Ophthalmol* 2019;137:288-92.
21. Bellemo V, Lim ZW, Lim G, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *Lancet Digit Health* 2019;1:e35-44.
22. Abràmoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018;1:39.
23. Niemeijer M, Abràmoff MD, van Ginneken B. Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. *Med Image Anal* 2006;10:888-98.

Cite this article as: Tan TE, Ting DSW, Wong TY, Sim DA. Deep learning for identification of peripheral retinal degeneration using ultra-wide-field fundus images: is it sufficient for clinical translation? *Ann Transl Med* 2020;8(10):611. doi: 10.21037/atm.2020.03.142