

Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images

Cecilia S. Lee, MD, Doug M. Baughman, BS, Aaron Y. Lee, MD, MSCI

Purpose: The advent of electronic medical records (EMRs) with large electronic imaging databases along with advances in deep neural networks with machine learning has provided a unique opportunity to achieve milestones in automated image analysis. Optical coherence tomography is the most common imaging modality in ophthalmology and represents a dense and rich data set when combined with labels derived from the EMR. We sought to determine whether deep learning could be utilized to distinguish normal OCT images from images from patients with age-related macular degeneration (AMD).

Design: EMR and OCT database study.

Subjects: Normal and AMD patients who underwent macular OCT.

Methods: Automated extraction of an OCT database was performed and linked to clinical end points from the EMR. Optical coherence tomography scans of the macula were obtained by Heidelberg Spectralis, and each OCT scan was linked to EMR clinical end points extracted from EPIC. The central 11 images were selected from each OCT scan of 2 cohorts of patients: normal and AMD. Cross-validation was performed using a random subset of patients. Receiver operating characteristic (ROC) curves were constructed at an independent image level, macular OCT level, and patient level.

Main Outcome Measure: Area under the ROC curve.

Results: Of a recent extraction of 2.6 million OCT images linked to clinical data points from the EMR, 52 690 normal macular OCT images and 48 312 AMD macular OCT images were selected. A deep neural network was trained to categorize images as either normal or AMD. At the image level, we achieved an area under the ROC curve of 92.78% with an accuracy of 87.63%. At the macula level, we achieved an area under the ROC curve of 93.83% with an accuracy of 88.98%. At a patient level, we achieved an area under the ROC curve of 97.45% with an accuracy of 93.45%. Peak sensitivity and specificity with optimal cutoffs were 92.64% and 93.69%, respectively.

Conclusions: The deep learning technique achieves high accuracy and is effective as a new image classification technique. These findings have important implications in utilizing OCT in automated screening and the development of computer-aided diagnosis tools in the future. *Ophthalmology Retina* 2017;1:322-327 © 2016 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Optical coherence tomography has become the most commonly used imaging modality in ophthalmology, with 4.39 million, 4.93 million, and 5.35 million OCTs performed in 2012, 2013, and 2014, respectively, in the U.S. Medicare population.¹ Since its development in 1991,² a 70-fold increase in OCT use for diagnosing age-related macular degeneration (AMD) was reported between 2002 and 2009.³ Furthermore, since the development of antiangiogenic agents, OCT has become a critical tool for baseline retinal evaluation before initiation of therapy and monitoring therapeutic effect.^{4,5} This increase in the use of OCT, with images stored in large electronic databases, highlights the ever-increasing time and effort spent by providers interpreting images.

The key OCT findings in AMD, including drusen, retinal pigment epithelium changes, and subretinal and intraretinal

fluid,⁴ share some common OCT features that are distinctively different from a normal retina.⁶ Correct identification of these characteristics allows for precise management of neovascular AMD and guides the decision of whether intravitreal therapy with anti-vascular endothelial growth factor agents should be given or not.⁷⁻⁹ Computer-aided diagnosis (CAD) has the potential for allowing more efficient identification of pathologic OCT images and directing the attention of the clinician to regions of interest on the OCT images.

The concept of CAD is not novel and has been applied in radiology, a field where the increasing demand of imaging studies has begun to outpace the capacity of practicing radiologists.¹⁰ A number of CAD systems have been approved by the US Food and Drug Administration for lesion detection and volumetric

analysis in mammography, chest radiography, and chest computed tomography.¹¹

Traditional image analysis required the manual development of convolutional matrices applied to an image for edge detection and feature extraction. In addition, prior work on OCT image classification of diseases has relied on machine learning techniques such as principal component analysis, support vector machine, and random forest.^{12–14} However, there has recently been a revolutionary step forward in machine learning techniques with the advent of deep learning, where a many-layered neural network is trained to develop these convolutional matrices purely from training data.¹⁵ Specifically, the development of convolutional neural network layers allowed for significant gains in the ability to classify images and detect objects in a picture.^{16–18} Within ophthalmology, deep learning has been recently applied at a limited capacity to automated detection of diabetic retinopathy from fundus photographs, visual field perimetry in glaucoma patients, grading of nuclear cataracts, and segmentation of foveal microvasculature, each with promising initial findings.^{19–22}

Although deep learning has revolutionized the field of computer vision, its application is usually limited due to the lack of large training sets. Often several tens of thousands of examples are required before deep learning can be used effectively. With the ever-increasing use of OCT as an imaging modality in ophthalmology along with the use of codified structured clinical data in the electronic medical record (EMR), we sought to link 2 large data sets together to use as a training set for developing a deep learning algorithm to distinguish AMD from normal OCT images.

Methods

This study was approved by the Institutional Review Board of the University of Washington (UW) and adhered to the tenets of the Declaration of Helsinki and the Health Insurance Portability and Accountability Act.

OCT and Electronic Medical Record Extraction

Macular OCT scans from 2006 to 2016 were extracted using an automated extraction tool from the Heidelberg Spectralis (Heidelberg Engineering, Heidelberg, Germany) imaging database. Each macular scan was obtained using a 61-line raster macula scan, and every image of each macular OCT was extracted. The images were then linked by patient medical record number and dates to the clinical data stored in EPIC. Specifically, all clinical diagnoses and the dates of every clinical encounter, macular laser procedure, and intravitreal injection were extracted from the EPIC Clarity tables.

Patient and Image Selection

A normal patient was defined as having no retinal International Classification of Diseases, 9th Revision (ICD-9) diagnosis and better than 20/30 vision in both eyes during the entirety of their recorded clinical history at UW. An AMD patient was defined as having an ICD-9 diagnosis of AMD (codes 362.50, 362.51, and 362.52) by a retina specialist, at least 1 intravitreal injection in either eye, and worse than 20/30 vision in the better-seeing eye.

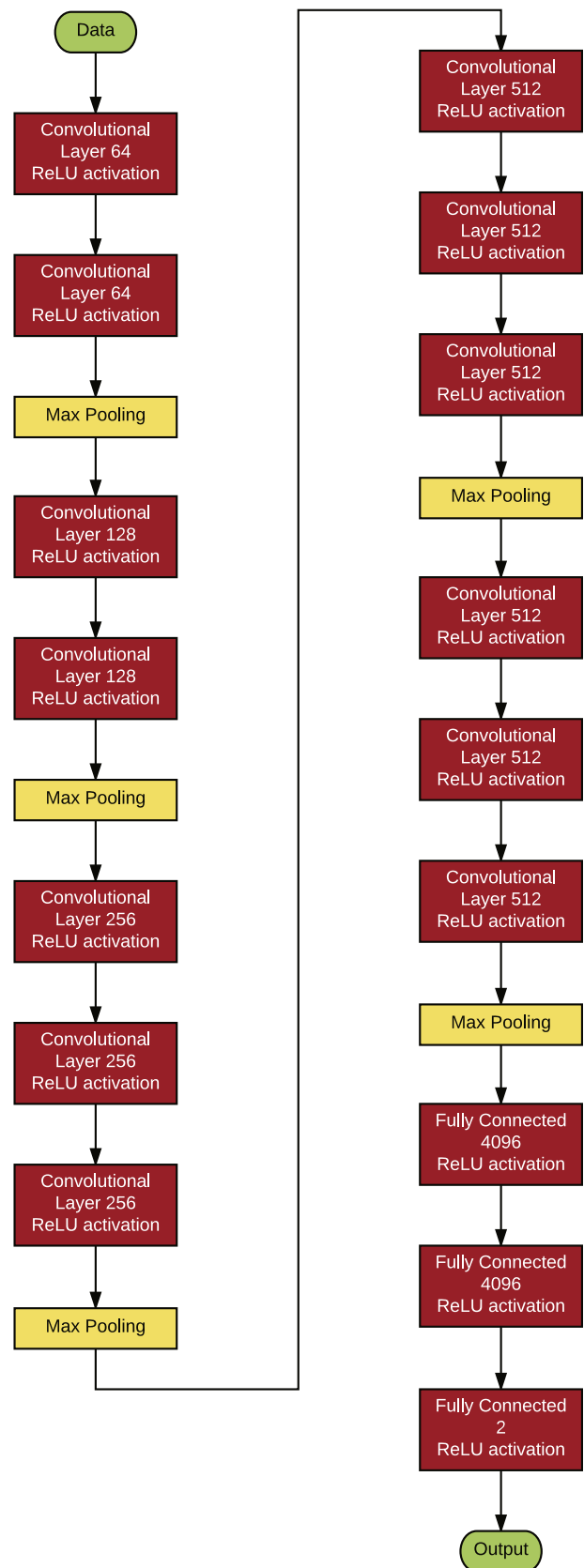


Figure 1. Schematic of the deep learning model used. A total of 21 layers with rectified linear unit (ReLU) activations were used.

Table 1. Baseline Characteristics of Patients Who Had Macular OCT

	Normal	AMD
Images, no.	48 312	52 690
Macular volumes, no.	4392	6364
Patients, no.	1259	347
Age (yrs), mean (SD)	56.96 (14.76)	70.1 (17.58)
Gender, no. (%)		
Male	21 615 (44.74)	25 069 (47.58)
Female	26 697 (55.26)	27 621 (52.42)
Eye, no. (%)		
Right	24 167 (50.02)	26 246 (49.81)
Left	24 145 (49.98)	26 444 (50.19)
VA (logMAR), mean (SD)	0.00057 (0.0648)	0.3839 (0.4999)

AMD = age-related macular degeneration; SD = standard deviation; VA = visual acuity.

Patients with other macular pathology by ICD-9 code were excluded. These parameters were chosen a priori to ensure that macular pathology was most likely present in both eyes in the AMD patients and absent in both eyes in the normal patients. Consecutive images of patients meeting these criteria were included, and no images were excluded due to image quality. Labels from the EMR were then linked to the OCT macular images, and the data were stripped of all protected health identifiers.

As most of the macular pathology is concentrated in the foveal region, the decision was made a priori to select the central 11 images from each macular OCT set, and each image was then treated independently, labeled as either normal or AMD. The images were histogram equalized and the resolution downsampled to 192×124 pixels due to limitations of memory. The image set was then divided into 2 sets, with 20% of the patients in each group placed into the validation set; the rest were used for training. Care was taken to ensure that the validation set and the training set contained images from a mutually exclusive group of patients (ie, no single patient contributed images to both the training and validation sets). The order of images was then randomized in the training set.

Deep Learning Classification Model

A modified version of the VGG16 convolutional neural network²³ was used as the deep learning model for classification (Fig 1). Weights were initialized using the Xavier algorithm.²⁴ Training was then performed using multiple iterations, each with a batch size of 100 images, with a starting learning rate of 0.001 with stochastic gradient descent optimization. At each iteration, the loss of the model was recorded; at every 500 iterations, the performance of the neural network was assessed using cross-validation with the validation set. The training was stopped when the loss of the model decreased and the accuracy of the validation set decreased.

An occlusion test¹⁷ was performed to identify the areas contributing most to the neural network's assigning the category of AMD. A blank 20×20 -pixel box was systematically moved across every possible position in the image and the probabilities were recorded. The highest drop in the probability represents the region of interest that contributed the highest importance to the deep learning algorithm.

Caffe (<http://caffe.berkeleyvision.org/>) and Python (<http://www.python.org>) were used to perform deep learning. All training occurred using the NVIDIA Pascal Titan X graphics processing unit with NVIDIA cuda (version 8.0) and cu-dnn (version 5.5.1) libraries (<http://www.nvidia.com>). Macular OCT-level analysis was performed by averaging the probabilities of the images obtained from the same macular OCT. Patient-level analysis was performed by averaging the probabilities of the images obtained from the same patient. Receiver operating characteristic (ROC) curves were constructed using the probability output from the deep learning model. Statistics were performed using R (<http://www.r-project.org>).

Results

We successfully extracted 2.6 million OCT images of 43 328 macular OCT scans from 9285 patients. After linking the macular OCT scans to the EMR, 48 312 images from 4392 normal OCT scans and 52 690 images from 4790 AMD OCT scans were selected. A total of 80 839 images (41 074 from AMD, 39 765

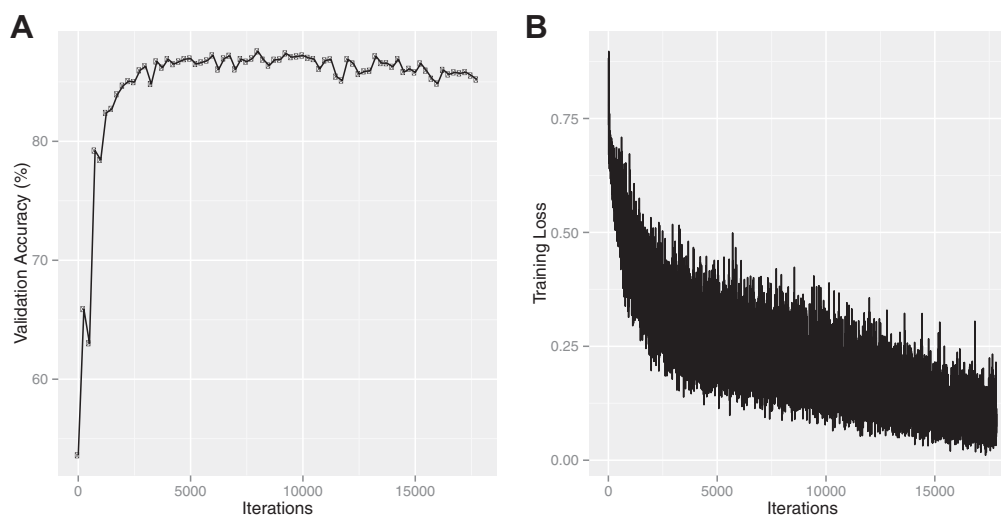


Figure 2. Learning curve of the training of the neural network with (A) accuracy and (B) loss. Each iteration represents 100 images being trained on the neural network.

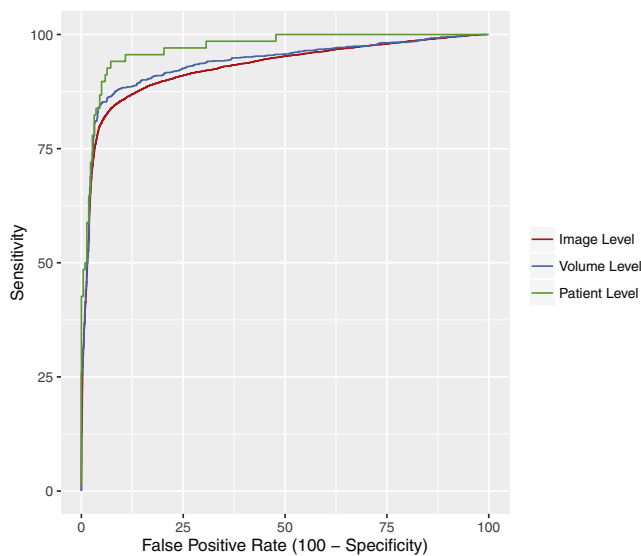


Figure 3. Receiver operator curves of 3 levels of classification. Image-level classification was performed by considering each image independently. Macula-level classification was performed by averaging the probabilities of all the images in a single macular volume. Patient-level classification was performed by averaging the probabilities of all the images belonging to a single patient.

from normal) were used for training and 20 163 images (11 616 from AMD, 8547 from normal) were used for validation (Table 1).

After 8000 iterations of training the deep learning model, the training was stopped due to overfitting occurring after that point (Fig 2). ROC curves are shown at the image level, OCT macular level, and patient level in Figure 3. The average time to evaluate a single image after training was complete was 4.97 ms.

At the level of each individual image, we achieved an accuracy of 87.63% with a sensitivity of 84.63% and a specificity of 91.54%.

After constructing an ROC curve, the peak sensitivity and specificity with optimal cutoffs were 87.08% and 87.05%, respectively. The area under the ROC curve (AUROC) was 92.77%.

By grouping the images in the same OCT scan and averaging the probabilities from each image, we achieved an accuracy of 88.98% with a sensitivity of 85.41% and a specificity of 93.82%. After constructing the ROC curve, the peak sensitivity and specificity with optimal cutoffs were 88.63% and 87.77%, respectively. The AUROC was 93.82%.

By averaging the probabilities from each image from the same patient, we achieved an accuracy of 93.45% with a sensitivity of 83.82% and a specificity of 96.40%. After constructing the ROC curve, the peak sensitivity and specificity with optimal cutoffs were 92.64% and 93.69%, respectively. The AUROC was 97.46%.

Example images from the occlusion test, shown in Figure 4, demonstrate that the neural network was successfully able to identify pathologic regions on the OCT. These areas represent the area in each image most critical to the trained network in categorizing the image as AMD.

Discussion

The ever-increasing use of digital imaging and EMRs provides opportunities to create deep and rich data sets for analysis. Our study demonstrates that a deep learning neural network was effective at distinguishing AMD from normal OCT images, and its accuracy was even higher when aggregate probabilities at the OCT macular scan and patient level were combined. The increase in the AUROC mainly occurred with an increase in sensitivity (Fig 3), with the inclusion of more images when aggregated. This most likely occurred because AMD infrequently affects the entire macula and normal-appearing OCT images may be mixed in with a macular OCT scan obtained from an AMD patient. Another possible explanation is that the etiology of the choroidal neovascular membrane (CNVM) may be

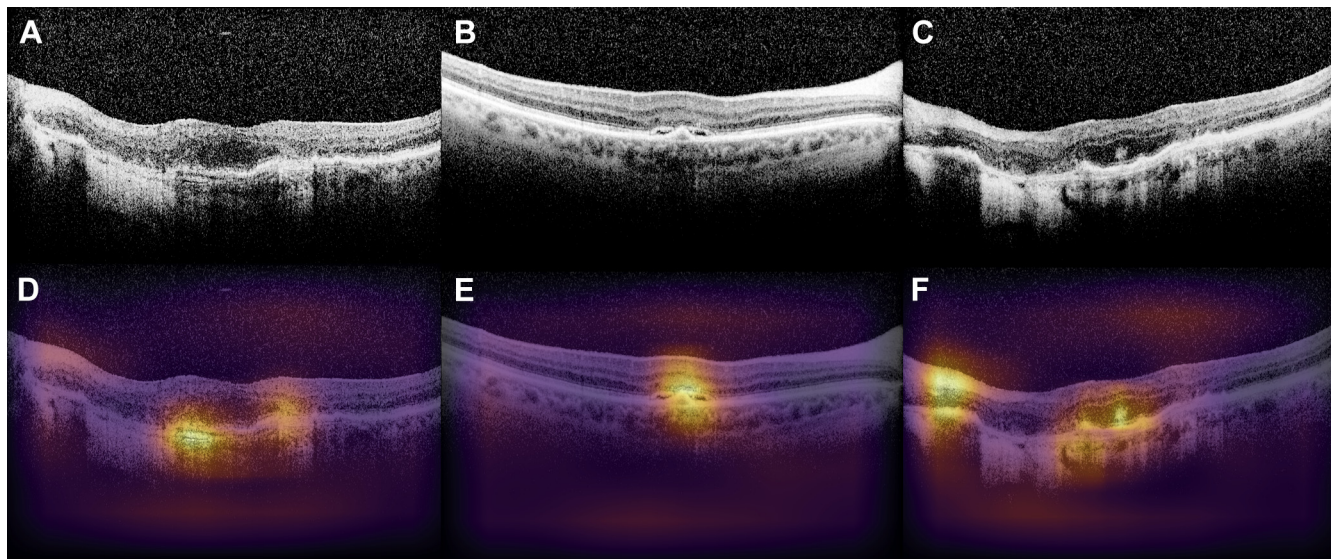


Figure 4. Examples of identification of pathology by the deep learning algorithm. Optical coherence tomography images showing age-related macular degeneration (AMD) pathology (A, B, C) are used as input images, and hotspots (D, E, F) are identified using an occlusion test from the deep learning algorithm. The intensity of the color is determined by the drop in the probability of being labeled AMD when occluded.

incorrect in the EMR. For example, our images labeled as AMD may have included myopic CNVM, which has localized pathology with relatively normal OCTs outside the CNVM area. Thus, adding additional images may explain the increased sensitivity.

Only limited applications of deep learning models exist in ophthalmology. Abramoff et al²⁰ reported sensitivity of 96.8% and specificity of 87.0% in detecting referable diabetic retinopathy in 874 patients using a deep learning model, but no details of the algorithm were included. Asaoka et al¹⁹ used a deep learning method to differentiate the visual fields of 27 patients with preperimetric open-angle glaucoma from those of 65 controls. An AUROC of 92.6% was achieved with that study's feed-forward network classifier, but the algorithm used only 4 layers of neurons. In contrast, our study's deep learning algorithm included 21 layers of neurons with state-of-the-art convolutional neural networking layers. Other, smaller studies that have applied automated OCT classification algorithms did not integrate deep learning strategies and were limited by small sample sizes, ranging from 32 to 384.^{12,19,20}

Our deep learning algorithm is a novel application to OCT classification in ophthalmology. To our best knowledge, the use of training and validation images from large EMR extraction has never been shown. To verify how the deep learning algorithm categorized the images as AMD, we performed an occlusion test where we systematically occluded every location in the image with a blank 20×20 -pixel area. The deep learning neural network successfully identified key areas of interest on the OCT image, which corresponded to the areas of pathology (Fig 4).

The application of occlusion testing provides insight into the trained deep learning model and which features were most important in distinguishing AMD images from normal images. In Figure 4C, occlusion testing did not show high-intensity dependence in the area of nasal high choroidal transmission, suggesting that the classifier was not using this as an important feature. One possible explanation is that in distinguishing normal from AMD OCT images, the classifier already achieved very low loss (Fig 2B) using the identified features and never discovered further improvements. Further studies could be performed where a classifier is specifically trained to distinguish specific AMD OCT features such as drusen, subretinal fluid, pigment epithelial detachments, and high choroidal transmission from each other.

Our study has several limitations. We included only images from patients who met our study criteria, and the neural network was only trained on these images. However, we included consecutive real-world images and did not exclude images with poor quality. In addition, this model was trained using images from a single academic center, and the external generalizability is unknown. Future studies would include expanding the number of diagnoses, using all images from a macular OCT scan, including images from different OCT manufacturers, and validation on OCT scans from other institutions.

In the future, our approach may be used to develop deep learning models that can have a number of wide-reaching applications. First, the models can be applied

to various retinal or choroidal pathologies in which OCT evaluations are essential, including diabetic retinopathy or retinal vein occlusions. Second, in future studies using deep learning, automated macular OCT classification could be used as a screening tool for retinal pathology when the cost of OCT machines decreases. This automated classification could be added to the majority of OCT machines being used in clinical practice without the need of integrating a graphics processing unit, as the inference step of deep learning is computationally inexpensive compared with training and can be run on standard computers. The automated classification feature will likely be beneficial in a large screening model such as that in an AMD screening system. Finally, deep learning models can identify concerning macular OCT images and efficiently display them to the clinician to aid in the diagnosis and treatment of macular pathology, much like CAD models found in radiology.

In conclusion, we demonstrate the ability of a deep learning model to distinguish AMD from normal OCT images and show encouraging results for the first application of deep learning to OCT images. Future follow-up studies will include widening the number of diseases and showing external validity of the model using images from other institutions.

Acknowledgments. The authors thank Dr Michael Boland for his assistance in identifying EPIC Clarity database values, the UW eScience Institute for their infrastructure support, and especially NVIDIA Corporation for donating the hardware graphics processing unit.

References

- Centers for Medicare & Medicaid Services. Medicare Provider Utilization and Payment Data: Physician and Other Supplier. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Physician-and-Other-Supplier.html>. Accessed May 20, 2016.
- Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science*. 1991;254:1178–1181.
- Stein JD, Hanrahan BW, Comer GM, Sloan FA. Diffusion of technologies for the care of older adults with exudative age-related macular degeneration. *Am J Ophthalmol*. 2013;155:688–696.e2.
- Keane PA, Patel PJ, Liakopoulos S, Heussen FM, Sadda SR, Tufail A. Evaluation of age-related macular degeneration with optical coherence tomography. *Surv Ophthalmol*. 2012;57:389–414.
- Ilginis T, Clarke J, Patel PJ. Ophthalmic imaging. *Br Med Bull*. 2014;111:77–88.
- Sayanagi K, Sharma S, Yamamoto T, Kaiser PK. Comparison of spectral-domain versus time-domain optical coherence tomography in management of age-related macular degeneration with ranibizumab. *Ophthalmology*. 2009;116:947–955.
- Keane PA, Liakopoulos S, Jivrajka RV, et al. Evaluation of optical coherence tomography retinal thickness parameters for use in clinical trials for neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2009;50:3378–3385.
- Reznicek L, Muhr J, Ulbig M, et al. Visual acuity and central retinal thickness: fulfilment of retreatment criteria for recurrent neovascular AMD in routine clinical care. *Br J Ophthalmol*. 2014;98:1333–1337.

9. Pron G. Optical coherence tomography monitoring strategies for A-VEGF-treated age-related macular degeneration: an evidence-based analysis. *Ont Health Technol Assess Ser*. 2014;14(10):1–64. [online]. <http://www.hqontario.ca/evidence/publications-and-ohtac-recommendations/ontario-health-technology-assessment-series/OCT-monitoring-strategies>.
10. Smieliauskas F, MacMahon H, Salgia R, Shih Y-CT. Geographic variation in radiologist capacity and widespread implementation of lung cancer CT screening. *J Med Screen*. 2014;21:207–215.
11. Van Ginneken B, Schaefer-Prokop CM, Prokop M. Computer-aided diagnosis: how to move from the laboratory to the clinic. *Radiology*. 2011;261:719–732.
12. Lemaître G, Rastgoo M, Massich J, et al. Classification of SD-OCT volumes using local binary patterns: experimental validation for DME detection. *J Ophthalmol*. 2016;2016:1–14.
13. Srinivasan PP, Kim LA, Mettu PS, et al. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed Opt Express*. 2014;5:3568–3577.
14. Liu Y-Y, Chen M, Ishikawa H, Wollstein G, Schuman JS, Rehg JM. Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Med Image Anal*. 2011;15:748–759.
15. Arel I, Rose DC, Karnowski TP. Deep machine learning - a new frontier in artificial intelligence research [research frontier]. *IEEE Comput Intell M*. 2010;5:13–18.
16. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*; 2012:1097–1105. <http://papers.nips.cc/paper/4824-imagenet-classification-w>. Accessed October 20, 2016.
17. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision*. 2014:818–833. http://link.springer.com/chapter/10.1007/978-3-319-10590-1_53. Accessed October 20, 2016.
18. Kavukcuoglu K, Sermanet P, Boureau Y-L, Gregor K, Mathieu M, Cun YL. Learning convolutional feature hierarchies for visual recognition. In: *Advances in Neural Information Processing Systems*; 2010:1090–1098. <http://papers.nips.cc/paper/4133-learning-convolutional-feature-hierarchies-for-visual-recognition>. Accessed October 20, 2016.
19. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*. 2016;123:1974–1980.
20. Abramoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57:5200–5206.
21. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Trans Biomed Eng*. 2015;62:2693–2701.
22. Prentašić P, Heisler M, Mammo Z, et al. Segmentation of the foveal microvasculature using deep learning networks. *J Biomed Opt*. 2016;21:075008.
23. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv:1409.1556. 2014. <http://arxiv.org/abs/1409.1556>. Accessed October 20, 2016.
24. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Aistats*. 9. 2010: 249–256. Accessed October 20, 2016 http://www.jmlr.org/proceedings/papers/v9/glorot10a/glorot10a.pdf?hc_location=ufi.

Footnotes and Financial Disclosures

Originally received: October 21, 2016.

Accepted: December 16, 2016.

Available online: February 13, 2017. Manuscript no. ORET_2016_140.

Department of Ophthalmology, University of Washington School of Medicine, Seattle, Washington.

Financial Disclosure(s):

Financial support: The authors have made the following disclosures:

C.S.L.: Grants — National Eye Institute (grant no. K23EY02492), Latham Vision Science Innovation, Research to Prevent Blindness, Inc.

D.M.B.: Grant — Latham Vision Science Innovation.

A.Y.L.: Grant — Research to Prevent Blindness, Inc.

The sponsors and funding organizations had no role in the design or conduct of this research.

Conflict of interest: No conflicting relationship exists for any author.

Author Contributions:

Research design: A.Y. Lee

Data analysis and/or interpretation: C.S. Lee, Baughman, A.Y. Lee

Data acquisition and/or research execution: A.Y. Lee

Manuscript preparation: C.S. Lee, Baughman, A.Y. Lee

Abbreviations and Acronyms:

AMD = age-related macular degeneration; **AUROC** = area under the receiver operating characteristic curve; **CAD** = computer-aided diagnosis; **CNVM** = choroidal neovascular membrane; **EMR** = electronic medical record; **ICD-9** = International Classification of Diseases, 9th Revision; **ROC** = receiver operating characteristic; **UW** = University of Washington.

Correspondence:

Aaron Y. Lee, MD, MSCI, Department of Ophthalmology, University of Washington, Box 359608, 325 Ninth Avenue, Seattle, WA 98104. E-mail: leeay@uw.edu.