

# Retinal Lesion Detection With Deep Learning Using Image Patches

Carson Lam,<sup>1,2</sup> Caroline Yu,<sup>3</sup> Laura Huang,<sup>4</sup> and Daniel Rubin<sup>1,4,5</sup>

<sup>1</sup>Department of Biomedical Data Science, Stanford University, Stanford, California, United States

<sup>2</sup>Department of Ophthalmology, Santa Clara Valley Medical Center, San Jose, California, United States

<sup>3</sup>Stanford University School of Medicine, Stanford, California, United States

<sup>4</sup>Department of Ophthalmology, Stanford University School of Medicine, Stanford, California, United States

<sup>5</sup>Department of Radiology, Stanford University School of Medicine, Stanford, California, United States

Correspondence: Daniel Rubin,  
Medical School Office Building  
(MSOB) Room X-335, MC 5464, 1265  
Welch Road, Stanford, CA 94305-  
5479, USA;  
rubin@stanford.edu.

Submitted: July 30, 2017

Accepted: December 28, 2017

Citation: Lam C, Yu C, Huang L, Rubin D. Retinal lesion detection with deep learning using image patches. *Invest Ophthalmol Vis Sci.* 2018;59:590-596. <https://doi.org/10.1167/iovs.17-22721>

**PURPOSE.** To develop an automated method of localizing and discerning multiple types of findings in retinal images using a limited set of training data without hard-coded feature extraction as a step toward generalizing these methods to rare disease detection in which a limited number of training data are available.

**METHODS.** Two ophthalmologists verified 243 retinal images, labeling important subsections of the image to generate 1324 image patches containing either hemorrhages, microaneurysms, exudates, retinal neovascularization, or normal-appearing structures from the Kaggle dataset. These image patches were used to train one standard convolutional neural network to predict the presence of these five classes. A sliding window method was used to generate probability maps across the entire image.

**RESULTS.** The method was validated on the eOphtha dataset of 148 whole retinal images for microaneurysms and 47 for exudates. A pixel-wise classification of the area under the curve of the receiver operating characteristic of 0.94 and 0.95, as well as a lesion-wise area under the precision recall curve of 0.86 and 0.64, was achieved for microaneurysms and exudates, respectively.

**CONCLUSIONS.** Regionally trained convolutional neural networks can generate lesion-specific probability maps able to detect and distinguish between subtle pathologic lesions with only a few hundred training examples per lesion.

Keywords: deep learning, detection, retina, machine learning, computer vision

Many diseases manifest in the retina that affect a large proportion of the population<sup>1,2</sup> and can lead to poor patient outcomes such as permanent vision loss if left untreated. The cost-effectiveness of regular retinal screenings has been well established,<sup>3</sup> but one of the major barriers to implementing more widespread screenings is the limited number of eye care practitioners who are trained in interpreting retinal images. Thus, there has been an active effort to create methods to automate screening of retinal images.

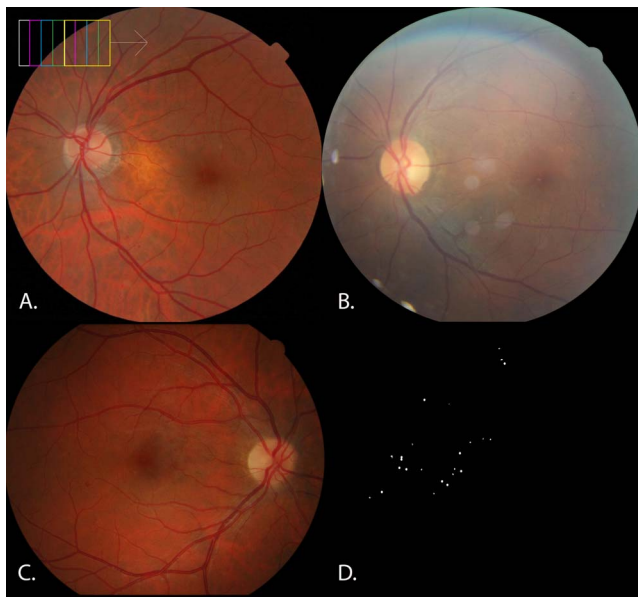
One approach to automating retinal screenings is deep convolutional neural networks (CNNs), which have become popular because of their ability to classify images with high sensitivity and specificity.<sup>4-7</sup> CNNs are a new but rapidly expanding machine-learning model used for object recognition in computer vision. The lure comes from the ability to learn from examples, or training datasets, rather than hard-coded rules. However, this strategy has several major limitations that affect its utility. First, CNNs often struggle in detecting the subtle pathologic lesions characteristic of early-stage disease as these pathologies that distinguish mild versus normal disease often reside in less than 1% of the total pixel volume.<sup>8,9</sup> Another limitation of deep learning is that it requires large training datasets, usually larger than tens of thousands of images. Especially for rare diseases, compiling such a database can be an arduous, if not

impossible, task. Thus, while CNNs represent a promising tool to assist in the screening and diagnosis of a number of diseases, there are still significant limitations that must be overcome.

One strategy that has been employed to address these limitations is the use of sliding windows (Fig. 1a). This refers to analyzing an image through a series of overlapping windows that are each focused on a zoomed-in subsection of the image, allowing for detection or segmentation of lesions that otherwise would not reliably be detected. Previous studies have successfully employed this strategy in analyzing medical and ophthalmologic images in various domains.<sup>10-13</sup>

This paper presents an optimized method for detecting the subtle retinal lesions of microaneurysms and exudates using a relatively small training set of only a few hundred images by using a combination of manually cropped image patches that comprise lesions and sliding windows. The purpose of this study is to demonstrate that with very few training examples, compared to what is currently cited in the deep learning literature, a researcher can flexibly train a CNN to recognize any number of findings, even rare findings, while overcoming the resolution problem of CNNs unable to learn microscopic findings by training on a more appropriate scale or feature size to image size ratio.





**FIGURE 1.** Schematic of sliding window method and an example of whole retinal fundus images. (A) Schematic of sliding window method, demonstrating how the algorithm scans across subsections of the whole image by moving in incremental steps. (B) Example image from the Kaggle dataset. (C) Example image from the eOphta dataset. (D) Binary mask from the eOphta image displayed in C. Black pixels in the binary mask correspond to negative pixels in the original, and white pixels in the binary mask correspond to positive pixels in the original such that when the mask is overlaid onto the original, the white pixels cover the microaneurysms or exudates almost exactly.

## METHODS

### Datasets

Two datasets were used in this project: the Kaggle retinopathy dataset<sup>14</sup> for training and validation and the eOphta dataset (TeleOphta)<sup>15</sup> for testing. Datasets consisted of color retinal images that varied in height and width between the low hundreds to low thousands of pixels. The Kaggle dataset (EyePacs LLC, San Jose, CA, USA) is a collection of 35,126 images of diabetic retinopathy (DR) with five class labels (normal, mild, moderate, severe, and proliferative) (Fig. 1b). These images vary significantly both in image quality and patient demographics and are sometimes mislabeled. Therefore, two ophthalmologists screened a subset of these images for both correct labeling and image quality. Images with disagreement were excluded. The eOphta dataset contains retinal fundus images derived from a consortium of French hospitals and consists of 47 images with exudates, 35 exudate-free images, 148 images with microaneurysms or other small red lesions, and 233 microaneurysm-free images (Fig. 1c). For the eOphta images, two ophthalmologists have labeled every pixel in the image as either belonging to exudate, microaneurysms, or neither (Fig. 1d). The pixel-level classifications are provided as binary masks that are the same dimensions as the original image.

### Extraction of Image Patches Comprising Lesions

A small subset of Kaggle images (243) that were correctly labeled and of sufficient quality were annotated independently by two ophthalmologists to isolate subsections of the image containing important clinical findings such as microaneurysms, exudates, camera artifacts, neovascularization, and so forth.

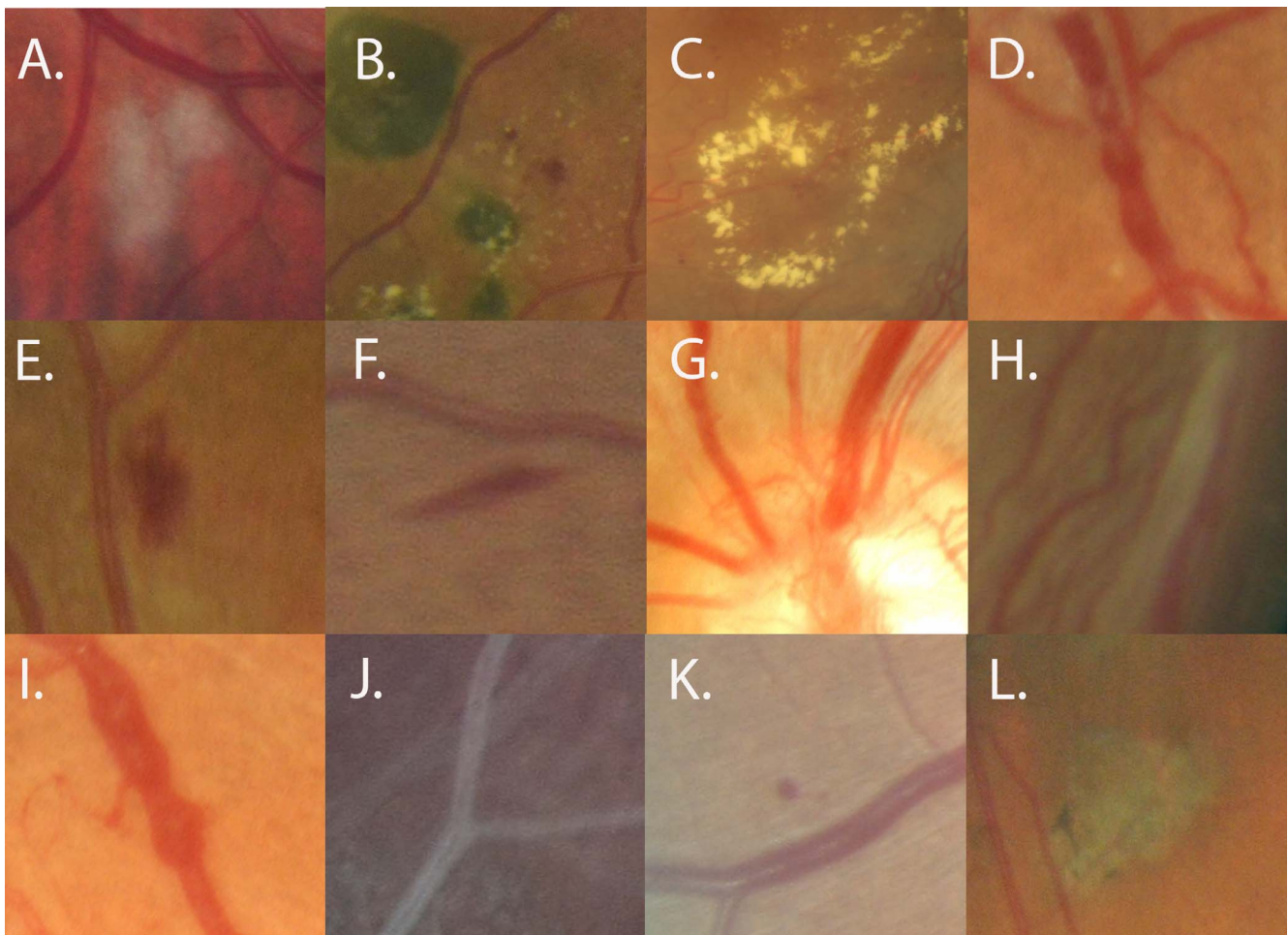
These subsections were made into image patches centered on the finding of interest; 1324 image patches mutually agreed upon for the accuracy of their clinical label were used. Of these, 1050 were included in the training set and 274 in the testing set. Image patches were extracted such that they were centered on the lesion of interest without constraints on patch size. Patches were extracted at whichever size and shape was appropriate to the lesion shape and scale. Normal patches were taken from the same image as the abnormal patches in regions that were free of abnormal lesions. Images varied significantly in size and shape, ranging from 25 to 1050 pixels in height and width. Overall, 609 images containing abnormal lesions and 576 images without abnormal lesions were taken from 243 whole images, and an additional 139 normal patches were randomly cropped from normal retinal images for a total of 1324 images. The dataset contained 260 images of microaneurysms, 128 images of dot-blot hemorrhages, 73 images of exudates, 33 images of cotton wool spots, and 31 images of retinal neovascularization. The other 84 abnormal images were distributed among various other abnormal lesions such as laser scars and sclerotic vessels (Fig. 2).

### Classification of Patches via CNN

The object of interest in these images occupied a proportion of the total image that mimics that of the ImageNet dataset. Typically, between 5% and 50% of pixels belonged to the object of interest. A single CNN using the GoogLeNet architecture with a  $128 \times 128 \times 3$  input with a five-class output was constructed: (1) normal, (2) microaneurysms, (3) dot-blot hemorrhages, (4) exudates or cotton wool spots, and (5) high-risk lesions such as neovascularization, venous beading, scarring, and so forth. A sample size of 33 was not enough to train a reliable classifier for cotton wool spots specifically. Due to the very small sample of cotton wool spots in our training set, cotton wool spots and exudates were combined into one class. This seemed more appropriate than placing the cotton wool spots in a separate class of its own or placing them in class 5 with the high-risk lesions. The CNN was trained on 1050 images and tested on 274 randomly selected images.<sup>10</sup> Training and inference were performed using the FirstAid Deep Learning repository<sup>16</sup> implemented in Tensorflow and two graphical processing units (Nvidia GeForce GTX 1080 ti; Nvidia Corp., Beaverton, OR, USA). Training data were augmented continuously during training by random zoom-in and zoom-out, from 50% to 200% of image size, with zero padding in cases of zoom-out. Random translation and rolling of the image were performed up to half the height and width of the image. Brightness and contrast were randomly adjusted by a factor of 0.5 each. Training was continued for 600 epochs, saving the model parameter state with lowest validation loss, learning rate of 0.001 and decay of 0.99, and 0.5 dropout rate implemented in the last fully connected or dense layer.

### Sliding Window Patch-Based Classification

The original retinal images were cropped to remove the black pixels at the image margins. The remaining image was up-sampled to  $2048 \times 2048 \times 3$  pixels in size. Pixels were normalized by subtracting the minimum pixel value and dividing by the maximum pixel value. Using a sliding window, the trained CNN was passed over the full scan to give a multiclass probability distribution across the image of the aforementioned pathologies. A  $128 \times 128 \times 3$  window was moved across full images with a slide of 32 pixels such that windows overlapped. Each window was the input for a forward pass through our trained CNN, producing a probability score within that patch of the image for each of the five classes



**FIGURE 2.** Examples of manually cropped image patches containing lesions of interest. (A) Cotton wool spot, (B) laser scars and exudate, (C) circinate ring, (D) venous beading and intraretinal vascular abnormality, (E) intraretinal hemorrhage, (F) intraretinal hemorrhage, (G) neovascularization of the disc, (H) fibrotic band, (I) venous beading, (J) sclerotic ghost vessel, (K) microaneurysm, and (L) laser scar.

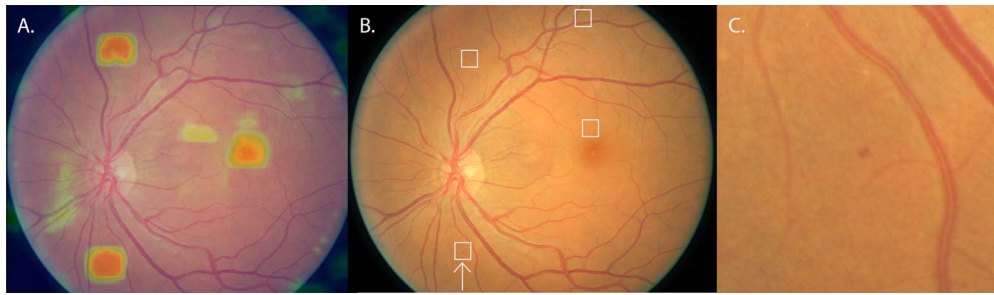
of normal and pathologic lesions. The result of this sliding window was a blanket of probability values over the entire image for each of the five classes, mostly flat but with smooth hills and valleys where the peaks correspond to high-probability pixels and warm colors in a heat map for the class of interest. These probability scores at each position in the image were used to construct heat maps representing the probability of pathology in local regions across the whole image, as depicted in Figures 3 and 4. End to end, this process took 1.1 minutes using a laptop with an Intel i7-6700HQ Processor.

### Statistical Analysis

To evaluate the efficacy of the patch-trained CNN used as a sliding window across whole images, the eOphta dataset of images was selected because it contained a large collection of publicly annotated retinal images and pixel-level information in the form of binary masks for microaneurysms and exudates, but not for any other lesion types. As a measure of saliency, or the ability to highlight important regions, pixel-level sensitivity and specificity across varying thresholds were calculated as one measure of performance. This is the same metric used in binary classification tasks, only performed at the pixel level. Pixels belonging to microaneurysms or exudates according to the ground truth mask, which were predicted by the model to

have a probability score above threshold or below threshold, were counted as true positive or false negative, respectively. Pixels not belonging to microaneurysms or exudates according to the ground truth mask, which were predicted by the model to have a probability score above threshold or below threshold, were counted as false positive or true negative, respectively. The threshold refers to the probability threshold at which a pixel was classified as positive for the lesion of interest. Various thresholds were used between 0.0 and 1.0 to generate an area under the curve of the receiver operating characteristic (AUC-ROC) for every image in the eOphta test set with respect to microaneurysms and exudates; the two lesions are presented separately.

Another metric calculated was precision and recall, as defined by positive predictive value compared to sensitivity in a detection task. Although not explicitly trained to serve as detection algorithm, the multiclass probability map was adapted into a detection map by thresholding the probability values and turning the probability map into a binary mask, then considering each connected component as a declaration that a lesion lies in this region. Connected components are groups of positive pixels touching each other either by being adjacent or through unbroken chains of other adjacent positive pixels. Detection is typically measured by bounding box overlap.<sup>17</sup> The eOphta dataset does not come with ground truth bounding boxes such as in traditional detection



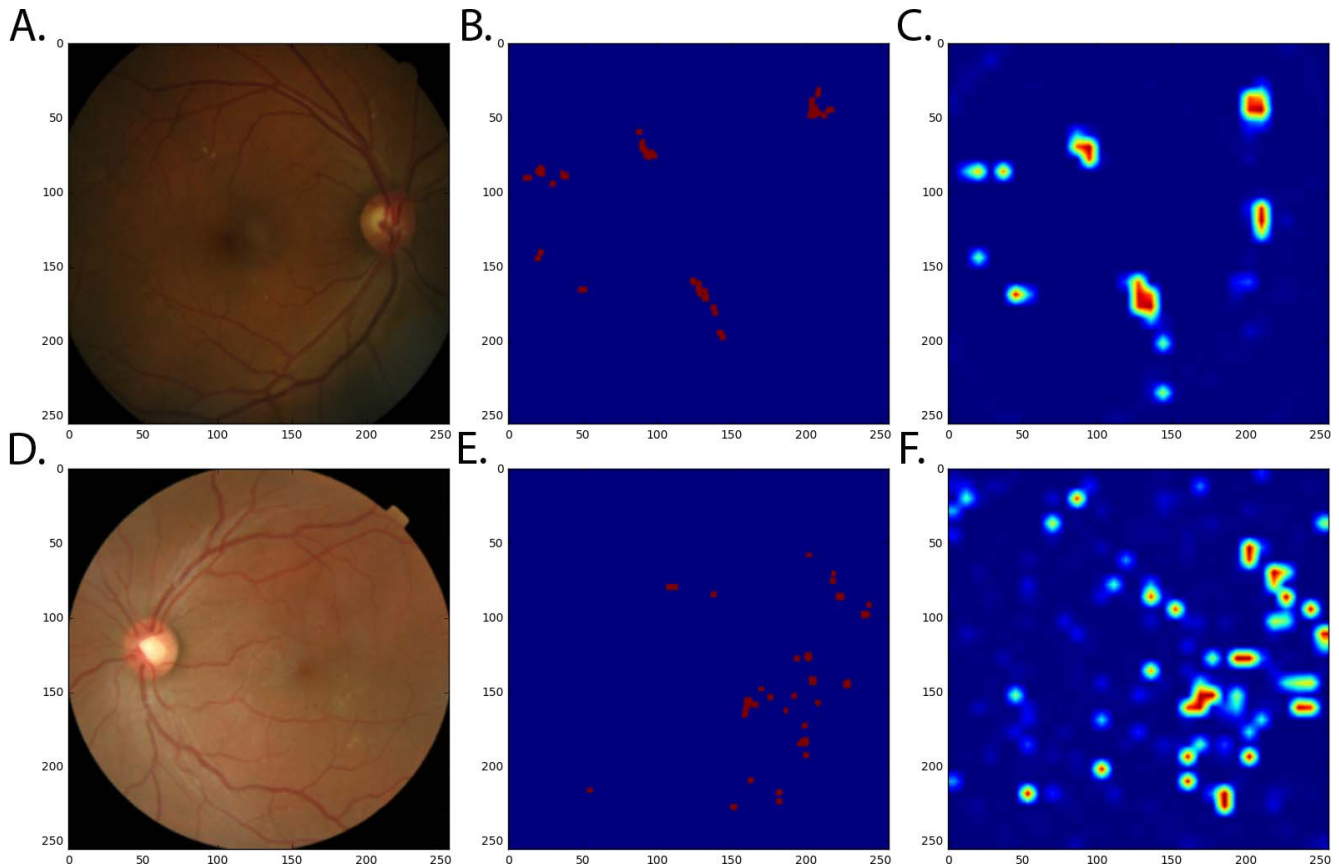
**FIGURE 3.** Probability heat maps highlight important regions. (A) Microaneurysm heat map over original image. Colors closer to red on the spectrum denote high probability. Green denotes probability near 0.5, and blue denotes low probability. (B) Original image with bounding boxes around four ophthalmologist-verified microaneurysms. The algorithm has missed the microaneurysm in the upper right of the image. The white arrow points to the zoomed-in image on the right. (C) Zoomed-in image of the arrowed box in the lower left of the middle image (B) have a microaneurysm in the center.

datasets, but each connected component can be considered to be one instance of a microaneurysm or exudate. Images typically contain several instances of each lesion per image. Centroids were defined as the center of mass of these connected components. Using a threshold distance of one-tenth of the dimensions of the image, true positives were defined as ground truth centroids falling within a threshold distance of a pixel predicted to be positive or the predicted centroid falling within the threshold distance of a true-positive pixel. Predicted connected components failing to meet this criteria were considered false positives, and ground truth connected components failing to meet this criteria were

considered false negatives. True negatives were undefined in detection tasks, and thus they were evaluated with precision recall curves.

**RESULTS**

The five CNN models that were tested included AlexNet, VGG16, GoogLeNet, ResNet, and Inception-v3. Ultimate CNN selection was based on accuracy, calculated as total correctly classified patches divided by total number of all patches in a held-out test set of 274 randomly selected patches from the



**FIGURE 4.** Comparison of probability map with ground truth. The heat map is generated from the probability map resulting from the output of the CNN at each location on the image. Since the training set is relatively balanced, the detector is highly sensitive but poor at generating tight bounds around the lesions. High-probability regions are in red, equivocal regions are highlighted in green, low probability regions in blue. (A-C) Microaneurysms, (D-F) exudates. (A, D) Original image (B, E) binary mask ground truth, (C, F) generated probability heat map.

**TABLE.** Accuracy of Image Patch Detection Across Model Architectures

Model	Five-Class Accuracy, %*	Binary Accuracy, %
AlexNet	74	79
VGG16	86	90
GoogLeNet	95	98
ResNet	92	95
Inception-v3	96	98

Data were calculated from a test set of 274 randomly selected patches. The binary classification tasks are normal and abnormal.

\* The five-class classification task includes normal, microaneurysm, intraretinal hemorrhage, exudate or cotton wool spot, and neovascularization or vitreous hemorrhage or intraretinal microvascular abnormality or venous beading or chorioretinal scars or sclerotic vessels.

Kaggle dataset (Table). A patch classification accuracy of 98% and AUC-ROC of 0.99 was achieved on a held-out test set of 274 randomly selected image patches in which the majority classifier accuracy would have been 58%. Similar results were obtained using GoogLeNet-v1, Inception-v3, and ResNet, and GoogleLeNet-v1 was chosen for the sliding window for its efficient parameter-to-performance ratio (Table).

The outputs of our sliding window CNN were multichannel probability maps for each image, where each channel represented a different lesion (Figs. 3, 4). Since the training data consisted of patches centered on the lesion of interest, the higher probability elevations in the probability map tended to cover both the lesion of interest and the surrounding pixels. Although with smaller strides, the higher probabilities bound the lesion of interest with tighter margins, the increase in computational time does not change the accuracy of lesion detection in a 1024 × 1024 image when the stride is less than 32 pixels.

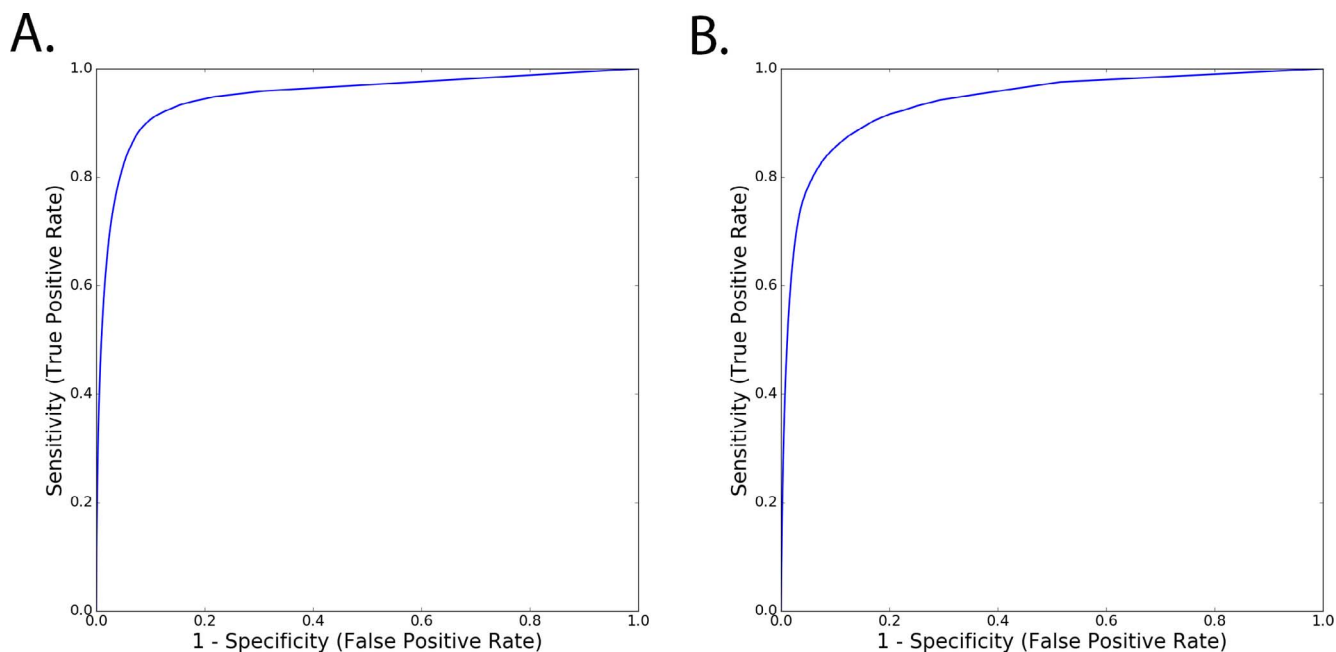
Figure 5 plots the ROC curves across all eOphta test images at the pixel level for microaneurysm and exudates separately. The average probability value for abnormal classes at each

pixel was 0.03 with a standard deviation of 0.018, meaning that the CNN was fairly confident of its prediction in the sea of blue in the background of the image. The islands of orange and red that represent high-probability regions tend to cover the ground truth lesions and extend well into the surrounding pixels of lesions. This characteristic partially explains the high true-positive rate and false-positive rate. In addition, the large proportion of abnormal training examples biased the detector to a high sensitivity for artifacts such as black specks on the camera or artefactual white reflections, despite the inclusion of such artifacts labeled as normal in the training set.

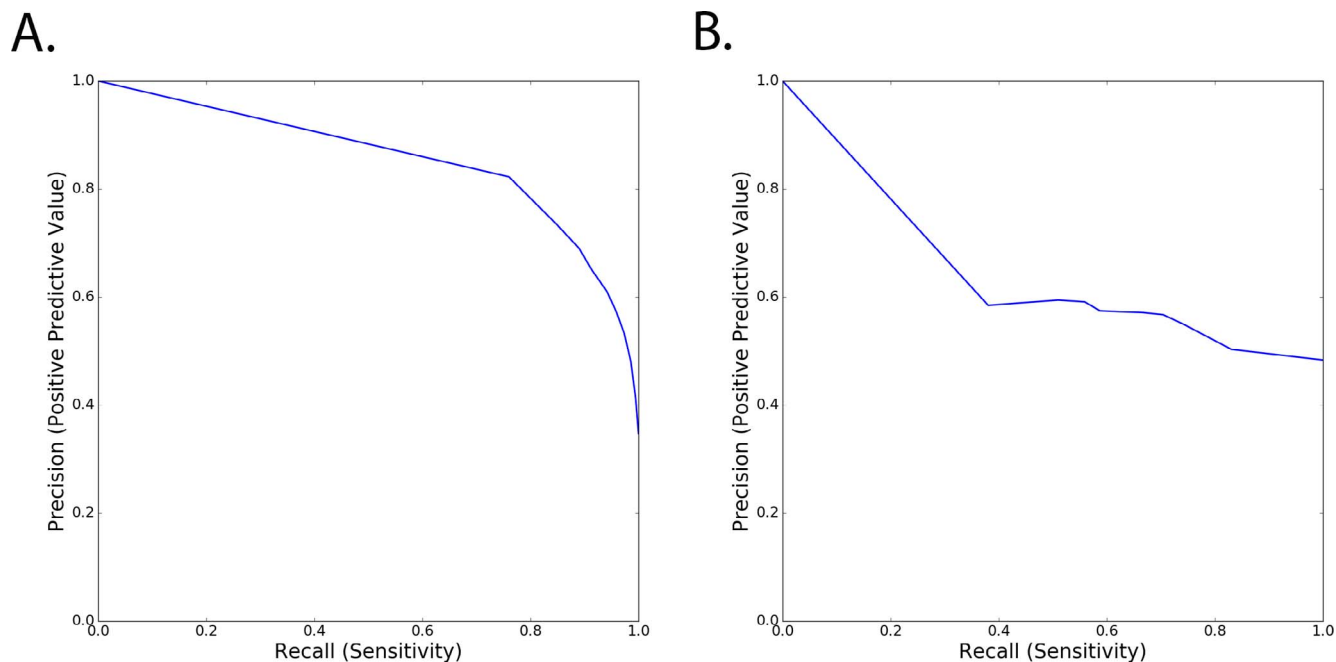
Figure 6 plots the precision and recall for the same 148 images containing microaneurysms and 47 images containing exudates for the detection task formulation of the problem. The area under the precision recall curve was 0.86 for microaneurysms and 0.64 for exudates. Overall the method performs better in terms of sensitivity than in specificity or positive predictive value.

**DISCUSSION**

CNNs are surprisingly effective at image classification, but CNNs trained on the entire retinal scan do not effectively detect the subtle pathologic changes present in early-stage retinopathy. One reason may be that the CNNs used have been optimized to recognize macroscopic lesions, such as those present in the ImageNet dataset, rather than microscopic lesions, such as microaneurysms.<sup>18</sup> This difference is highlighted by recent studies applying CNN’s architecture optimized on the ImageNet dataset to whole retinal images of the retina for DR screening.<sup>7,19</sup> Excellent performance is reported in the detection of moderate or worse DR containing macroscopically visible lesions, but detection of microscopic lesions such as microaneurysms are much more difficult. This is a problem that is not necessarily remedied with additional data. For example, Gulshan et al.<sup>7</sup> reported a 93% to 96% recall for their binary classification tasks; however, this was not improved when training with 60,000 samples versus 120,000



**FIGURE 5.** Receiver operating characteristic curves. True-positive rate and false-positive rates were calculated across all pixels in each image against ground truth binary masks produced by two ophthalmologists for 148 and 47 images containing microaneurysms and exudates, respectively, in the eOphta dataset. (A) AUC-ROC curve depicted for microaneurysms; AUC-ROC was 0.94. (B) AUC-ROC for exudates was 0.95.



**FIGURE 6.** Precision recall curves in detection. Positive predictive value and sensitivity of detection across thresholds for every lesion in 148 and 47 images containing microaneurysms and exudates, respectively, in the eOphtha dataset. (A) For microaneurysms, the area under the precision recall curve was 0.86. (B) For exudates, the area under the precision recall curve was 0.64.

samples. To build on existing techniques, groups have instead used deep learning to augment existing methods rather than deploy CNNs alone.<sup>20</sup>

Visualizations of the features learned by CNNs reveal that the signals used for classification are represented at a macroscopic scale.<sup>21</sup> Moderate and severe diabetic retinopathy contain macroscopic lesions at a scale that current CNN architectures are optimized to classify. However, the lesions that distinguish normal retinas from mild disease reside in less than 1% of the total pixel volume, a level of subtlety that is difficult for both human interpreters and CNNs to detect. Several groups have sought to locate these subtle lesions specifically. In the past, microaneurysm detection has been studied using a smaller dataset of 50 training and test images annotated by  $x$ ,  $y$  coordinates and radii to identify these tiny lesions. The performance criteria in this study evaluated the proportion of  $x$  and  $y$  coordinates falling within a set distance of the ground truth coordinates. Although evaluated by different criteria from the current study, area under the curve of the receiver operating characteristic for six methods in this study varied between 0.80 and 0.88, with a human expert scoring 0.96.<sup>22</sup>

The use of many square image patches for training and detection using a sliding window technique has been employed in pathology<sup>23</sup> with success and performs a task very similar to convolution and deconvolution in concept. Feature extraction from image patches using sparse stacked autoencoders has been trained on image patches from the DIARETDB dataset and has collected normal patches by sliding windows.<sup>24</sup> The method was internally validated on the same dataset it was trained on to obtain a patch classification AUC-ROC of 96% on a subset of patches from the DIARETDB used for testing.

This method in its current state has several limitations. Due to a training set in which roughly half of patches are abnormal in the dataset of patches, the CNN trained on this set assumes a higher than normal prior probability for lesion detection, which may cause high false-positive rates. With a few more training

examples of cotton wool spots, the separation of cotton wool spots and exudates into separate categories likely would improve the false-positive rate as well. Another way to decrease false-positive rates, normal patches can be extracted in an automated fashion, and future studies will seek to determine if increased data can increase the specificity while preserving the high sensitivity. Bias is introduced into the deep learning model by factors such as the clinician-identified regions of interest and the predefined patch size. Furthermore, the sliding window technique was very sensitive to changes in the window size and stride. Although this sliding window CNN showed promising results, its clinical applicability is limited by the computational time required to run several forward passes per image. Although this study uses a single-size window, it is likely that lesions are detected at scales optimal to their size unless those lesions can be detected based on their texture alone.

Future steps for this method are to develop screening algorithms for other causes of retinal disease, including such common entities as macular degeneration and glaucoma, to uncommon but time-sensitive emergent diseases such as retinal tears, retinitis, and optic nerve swelling, and rare pathologies such as white dot syndrome, dystrophies, and ocular tumors.

In summary, this study demonstrates that a sliding window approach using neural networks trained on clinician-selected regions of interest is able to detect subtle pathologic lesions with significantly fewer examples than traditional CNNs. Outside its applications in DR, this proposed method is a promising step toward development of screening algorithms for other, less-common retinal diseases.

### Acknowledgments

The authors thank Darwin Yi for access to the FirstAid Deep Learning Code repository, assistance in curation of the manuscript, and experimentation and evaluation design.

Supported in part by grants from the National Cancer Institute, National Institutes of Health (U01CA142555, 1U01CA190214, 1U01CA187947).

Disclosure: **C. Lam**, None; **C. Yu**, None; **L. Huang**, None; **D. Rubin**, None

## References

1. Abramoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. *IEEE Rev Biomed Eng.* 2010;3:169-208.
2. Zhang X, Saaddine JB, Chou CF, et al. Prevalence of diabetic retinopathy in the United States, 2005-2008. *JAMA.* 2010;304:649-656.
3. Vijan S, Hofer TP, Hayward RA. Cost-utility analysis of screening intervals for diabetic retinopathy in patients with type 2 diabetes mellitus. *JAMA.* 2000;283:889-996.
4. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst.* 2012:1097-1105.
5. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway: IEEE; 2016:2818-2826.
6. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: *Computer Vision and Pattern Recognition.* IEEE Computer Society Conference on June 20, 2009 in Miami, Florida. Piscataway; IEEE; 2009:248-255.
7. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316:2402-2410.
8. Niemeijer M, van Ginneken B, Russell SR, Suttorp-Schulten MS, Abramoff MD. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Invest Ophthalmol Vis Sci.* 2007;48:2260-2267.
9. Antal B, Hajdu A. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Trans Biomed Eng.* 2012;59:1720-1726.
10. Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging.* 2016;35:2369-2380.
11. Fang L, Cunefare D, Wang C, Guymer RH, Li S, Farsiu S. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomed Opt Express.* 2017;8:2732-2744.
12. Cunefare D, Fang L, Cooper RF, Dubra A, Carroll J, Farsiu S. Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks. *Sci Rep.* 2017;7:6620.
13. van Grinsven MJ, van Ginneken B, Hoyng CB, Theelen T, Sánchez CI. Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images. *IEEE Trans Med Imaging.* 2016;35:1273-1284.
14. Kaggle Diabetic Retinopathy Detection Competition Report, 2015. Kaggle website. Available at: <https://www.kaggle.com/c/diabetic-retinopathy-detection>. Accessed July 27, 2015.
15. Decencière E, Cazuguel G, Zhang X, et al. TeleOphta: machine learning and image processing methods for teleophthalmology. *IRBM.* 2013;34:196-203.
16. First Aid for Medical Image Deep Learning. Available at: <https://github.com/yidarvin/firstaid>. Accessed March 2, 2017.
17. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. *Int J Comput Vis.* 2010;88:303-338.
18. Xu Y, Xiao T, Zhang J, Yang K, Zhang Z. Scale-invariant convolutional neural networks. Available at: <https://arxiv.org/abs/1411.6369>. Accessed November 24, 2014.
19. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology.* 2017;124:962-969.
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. Yosinski J, Clune J, Nguyen A, Fuchs T, Lipson H. Understanding neural networks through deep visualization. Available at: <https://arxiv.org/abs/1506.06579>. Accessed June 22, 2015.
22. Niemeijer M, Van Ginneken B, Cree MJ, et al. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE Trans Med Imaging.* 2010;29:185-195.
23. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep learning for identifying metastatic breast cancer. Available at: <https://arxiv.org/abs/1606.05718>. Accessed June 18, 2015.
24. Shan J, Li L. A deep learning method for microaneurysm detection in fundus images. In: *Connected Health: Applications, Systems and Engineering Technologies (CHASE).* New York: IEEE; 2016:357-358.