Primary open-angle glaucoma (POAG) is one of the leading causes of blindness worldwide.1 In the United States, POAG is the most common form of glaucoma and is the leading cause of blindness among African Americans2 and Hispanics.3 Unfortunately, POAG is asymptomatic until advanced loss of peripheral vision occurs very late in the disease. However, it is possible to screen POAG at a stage where early treatment and intervention may alter its course and preserve vision that would otherwise be lost.4–6

Optic disc photography has proven to be very useful for diagnosing glaucoma, showing a classic glaucomatous appearance to expert graders. While convenient and inexpensive, the low prevalence of glaucoma and screening limitations make it challenging to conduct meaningful screenings.7 Therefore, it is important to develop an automatic model to assist clinicians in screening for and diagnosing incident POAG with high accuracy from optic disc photographs.

Developments in artificial intelligence have made automatic POAG diagnosis using optic disc photographs possible. Singh et al8 obtained the vertical cup-to-disc ratio by segmenting optic disc and optic cup and performed the
POAG classification after extracting the handcrafted feature from vertical cup-to-disc ratio. Acharya et al detected POAG using support vector machines and Naïve Bayesian classifier based on the texture features and higher-order spectral features. Dua et al also used support vector machines and Naïve Bayesian to classify POAG based on the wavelet-based energy features. Issac et al adopted an adaptive threshold-based image processing method for POAG classification. However, the segmentation accuracy of optic disc and optic cup greatly impacts these methods. In addition, these methods only consider handcrafted features, and thus are hard to learn to generalize.

Recently, deep learning methods have demonstrated promising results in biology and medicine. In the ophthalmology domain, several methods have been proposed to detect POAG at its early stage. However, these approaches all used a single image as input. In contrast, in clinical practice, glaucoma specialists often compare the follow-up image with the baseline image to trace the relevant features and assess for glaucomatous change (Fig 1). It is worth noting that the baseline image must be non-POAG.

To simulate this process, we propose a Siamese neural network model, POAGNet, in this work. The Siamese neural network uses the same weights in a twin network while working on 2 different images to compute comparable output. In our case, the output is the baseline image against which the other output is compared. To the best of our knowledge, it is the first time in the ophthalmology domain that 2 optic disc images have been utilized and compared for automated glaucoma detection via Siamese networks.

Unlike previous Siamese work, POAGNet used a convolution operation to study the feature difference between 2 outputs. Compared to the traditional absolute distance, our inner neural network approximates the difference between the 2 inputs more precisely by updating the parameter of convolution during the training. In addition, the POAGNet consists of side output to ease the vanishing gradient problems in training and force the hidden layers to favor discriminative features.

Our study also aimed to evaluate the model’s generalization capability for different data sets not employed in the training process. To this end, we assessed POAGNet on 2 large-scale, independent data sets, the Ocular Hypertension Treatment Study cohort (OHTS) and Sequential fundus Images for Glaucoma data set (SIG), with > 35 000 optic disc images in total.

The second data set is obtained from the SIG database (https://github.com/XiaofeiWang2018/DeepGF). The study protocol was approved by the institutional review board at each clinical center and Beihang University. All research adhered to the tenets of the Declaration of Helsinki. All participants provided informed consent. All risk factors were measured at baseline before the onset of the disease and collected for approximately 16 years.

The participants in this data set were selected according to eligibility and exclusion criteria. Briefly, the eligibility criteria include intraocular pressure (between 24 mm Hg and 32 mm Hg in one eye and between 21 mm Hg and 32 mm Hg in the fellow eye) and age (between 40 and 80 years). The visual field (VF) tests were interpreted by the Visual Field Reading Center, and the optic discs at clinical examination and stereoscopic photographs were interpreted by the Optic Disc Reading Center. Exclusion criteria included previous intraocular surgery, visual acuity worse than 20/40 in either eye, and diseases that may cause optic disc deterioration and VF loss (such as diabetic retinopathy). The gold standard POAG labels were graded at the Optic Disc Reading Center. In brief, 2 masked certified readers were arranged to independently detect the optic disc deterioration. If there was a disagreement between 2 readers, a senior reader reviewed it in a masked fashion. The POAG diagnosis in a quality control sample of 86 eyes (50 normal eyes and 36 with progression) showed test-retest agreement at $\kappa = 0.70$ (95% confidence interval, 0.55–0.85). More details of the reading center workflow have been described by Gorden et al.

SIG Database. The second data set is obtained from the SIG data set (https://github.com/XiaofeiWang2018/DeepGF). The study protocol was approved by the institutional review board at each clinical center and Beihang University. All research adhered to the tenets of the Declaration of Helsinki. All participants provided informed consent. The SIG contains 3837 optic disc images, of which 153 (3.99%) have POAG. In the SIG data set, all optic disc images are annotated with binary labels of glaucoma, that is, positive or negative glaucoma. The samples are labeled glaucomatous when they satisfy any of the 3 criteria, that is, retinal nerve fiber layer defect, rim loss, and optic disc hemorrhage.

Model Development

Overall Architecture. The POAGNet comprises 2 convolutional blocks that share the weight and are followed by 7 layers (Fig 2). The prediction block in Figure 2 shows the detail of the 7 layers. In the beginning, 2 optic disc images, $x_1$ and $x_2$, are passed through...
the convolutional neural network, DenseNet-201, respectively. We used the output of last ($F_{ds}$ and $F_{dsel}$) and second last Dense Blocks ($F_{dbs}$ and $F_{dbsel}$). For each output, we concatenated two outputs, followed by using $1 \times 1$ convolution, a batch normalization, and rectified linear units. The advantage of a convolution is that, instead of vectors, we can compute the similarity between 2 feature maps at all kernels. Therefore, the output is not a single value but another feature map with spatial support. In the end, a global average pooling and a fully connected layer with sigmoid activation are attached.

The POAGNet connects the side-output layer to the last 2 prediction blocks. The side-output layer generates the inherent scales of discriminative features and relieves the vanishing gradient problems in training. Finally, all side-output layers are averaged to generate a POAG prediction.

**Loss Function.** In this study, we use binary cross-entropy as the loss function in the POAGNet. In addition, to overcome the severe class imbalance for the POAG classification, we apply the weighted cross-entropy, a commonly used loss function in classification. The adopted weighted cross-entropy was as follows:

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^{N} \left[ \beta y_n \log \left( \hat{y}_n(x_n, \theta_s) \right) + \left(1 - \beta \right) \left(1 - y_n \right) \log \left(1 - \hat{y}_n(x_n, \theta_s) \right) \right]$$

where $N$ is the number of training examples and $\beta$ is the balancing factor between positive and negative samples. Here, we used the inversely proportional to POAG frequency in the training data.

The overall loss function is the average of the losses associated with a prediction from the last 2 blocks:

$$\mathcal{L}_s = \alpha \mathcal{L}^{(1)} + \left(1 - \alpha\right) \mathcal{L}^{(2)}$$

**Image Augmentation.** In this study, the following augmentation techniques were applied on the fly during training: (1) random rotation between 0° and 10°, (2) random translation: an image was randomly translated along the x- and y-axes by distances ranging from 0% to 10% of width or height of the image, and (3) random flipping. These augmentation operations increase the diversity of the data set.

**Evaluation Metrics**

Our experiments report accuracy, sensitivity (recall), and specificity. In addition, we report the area under the curve (AUC) receiver operating characteristics curve. A receiver operating characteristics curve plots true positive rate (also called sensitivity) versus false-positive rate at different classification thresholds.

**Experimental Settings**

We first used a single image as input without manual cropping in the POAG detection task to fine-tune a DenseNet-201, which has been pretrained on ImageNet. Then, we used this DenseNet-201 to initialize the subnets (DenseNet-201) in the POAGNet and fine-tuned the entire network in an end-to-end manner. Therefore, the loss is propagated back to the individual neural networks, creating better feature representations for each training iteration.

All images are resized to $224 \times 224 \times 3$ as input of the proposed model. The models were implemented by Keras with a backend of TensorFlow. The proposed network was optimized using the Adam optimizer method. The learning rate is $5 \times 10^{-5}$, $\alpha$ is 0.8. The experiments were performed on Intel Core i9-9960 X 16 cores processor and NVIDIA Quadro RTX 6000 GPU.

For the OHTS data set, we split the entire data set randomly at the patient level. We took 1 group (20% of total subjects) as the hold-out test set and the remaining as the training set. For the SIG data set, we used the official training, development, and testing split in this study. In addition, we performed cross-data set bias and evaluation: (1) model trained on OHTS and tested on SIG; (2) model trained on SIG and tested on OHTS; (3) model trained on SIG and OHTS jointly and tested on SIG, OHTS, and their combination, respectively.

Table 1. Characteristics of the OHTS and SIG Data Sets

| Data set | OHTS | SIG |
|----------|------|-----|
|          | Train | Dev | Test | Train | Dev | Test |
| Eyes Pairs | 2440 | 180 | 652 | 300 | 35 | 70 |
| POAG Normal | 1481 | 326 | 478 | 110 | 15 | 28 |
| OHTS = Ocular Hypertension Treatment Study; POAG = primary open-angle glaucoma; SIG = Sequential fundus Images for Glaucoma.
Results

We compared our method with 6 models on POAG diagnosis on the OHTS data set, including the DenseNet-201 with a single image as input, EfficientNetB0 with a single image as input, MobileV2 with a single image as input that was used in Thakur et al., ResNet-50 with a single image as input that was used in Fan et al., the traditional Siamese network with absolute distance, and POAGNet using the last DenseNet Block (POAGNet w/o side output). DenseNet-201, EfficientNetB0, MobileNetV2, and ResNet-50 were pretrained on ImageNet and we modified the fully connected layers so that they could meet our binary classification requirement. We fine-tuned the entire network in an end-to-end manner.

POAG Diagnosis on the OHTS Data Set

We first trained and validated the models on the OHTS data set. Table 2 shows the performance comparison. Our model achieved the best results, with an accuracy of 0.9283, a sensitivity of 0.7469, a specificity of 0.9421, and an AUC of 0.9587. The performance of POAGNet was then compared with that of the models with a single image as input. The performance of POAGNet was superior to that of the best single model, DenseNet-201, with 9.63% higher accuracy, 1.78% higher sensitivity, 1.28% higher specificity, and 8.37% higher AUC.

In addition, the performance of POAG was compared with 2 variations, 1 without side output and 1 with absolute distance. The POAGNet obtained an AUC of 0.9587, resulting in an improvement of 5.1% over the POAGNet with the absolute distance (row 4) and 3.5% over the POAGNet w/o side output (row 5). To check how many side outputs are optimal, we also performed the experiments where the POAGNet used the side output of the last 3 Basic Blocks as well as all Basic Blocks. It turns out that the proposed structure, which used the side output of the last 2 Basic Blocks, achieved the best results (Table S1). We also tried the contrast loss instead of binary cross-entropy loss, and the results decreased (Table S2).

Table 2. The Results of Models Trained and Validated on the OHTS Data Set

| Method                        | Accuracy | Sensitivity | Specificity | AUC    |
|-------------------------------|----------|-------------|-------------|--------|
| DenseNet-201                  | 0.8320   | 0.7291      | 0.9393      | 0.8750 |
| EfficientNetB0                | 0.8899   | 0.5071      | 0.9199      | 0.8370 |
| MobileV2                     | 0.7426   | 0.7531      | 0.7418      | 0.8124 |
| ResNet-50                    | 0.8653   | 0.6823      | 0.8795      | 0.8650 |
| POAGNet (absolute distance)   | 0.9070   | 0.6841      | 0.9240      | 0.9075 |
| POAGNet (w/o side output)     | 0.9059   | 0.7803      | 0.9154      | 0.9236 |
| POAGNet                      | 0.9283   | 0.7469      | 0.9421      | 0.9587 |

AUC = area under the curve; OHTS = Ocular Hypertension Treatment Study; POAG = primary open-angle glaucoma.

POAG Diagnosis on the SIG Data Set

We then trained and validated POAGNet on the SIG data set. Table 3 compares the performance of POAGNet on the SIG data set. The POAGNet obtained the best results, with an accuracy of 0.9176, a sensitivity of 0.1786, a specificity
of 0.9519, and an AUC of 0.7518. Same as on the OHTS data set, the performance of POAGNet was superior to DenseNet-201, the best model with a single image input (10.8% in AUC), POAGNet with absolute distance (7.4% in AUC), and POAGNet w/o side output (5.5% in AUC).

**Cross-Data Set Bias and Evaluation**

In separate experiments, to assess the generalizability and transferability of POAGNet, we compared the performance of models trained on OHTS, SIG, and their combination (OHTS + SIG) (Fig 3).

For accuracy and specificity (Fig 3A, C), our models achieved comparable results in different training and testing scenarios, indicating that our model can generalize and transfer across different data sets.

For sensitivity and AUC (Fig 3B, D), POAGNet achieved comparable or better performance when trained on OHTS or OHTS + SIG but weak performance when trained on SIG only. When POAGNet is trained on SIG only, the training size is relatively small, which makes it difficult to fully train the model. We will explain observations in more detail in the "Discussion" section.

We also sampled the OHTS testing data set, making the number of instances in the OHTS testing data set the same as those in the SIG testing data set. The AUC is 0.8797 (Table S3).

**Discussion**

This study proposed a new end-to-end deep learning network that simulates the clinical process of automatic POAG detection from optic disc photographs. Two data sets were used to evaluate the proposed model. The results demonstrated that the proposed network was superior to the state-of-the-art model. The proposed model has potential for eye services in the future after extensive validation across multiple and diverse image data sets. Unlike the previous networks that use a single image as input, our proposed model simulates the clinical process by comparing the differences between 2 input images (baseline and follow-up).

The POAGNet achieved superior performance to the state-of-the-art methods for 3 reasons. First, the proposed network used the Siamese neural network and 2 images (baseline and follow-up) as input. This architecture simulates the clinical process of identifying POAG from optic disc photographs by comparing the follow-up image with the baseline image. Tables 2 and 3 show that the Siamese neural networks were superior to DenseNet-201, suggesting that the Siamese network can leverage the differences between 2 images for a more accurate POAG diagnosis.

Second, POAGNet used a convolution operation to measure the difference between 2 outputs. The advantage of the convolution operation is that it computes the similarity on all kernels, and the parameters in the convolution can be updated during the training. The output of this operation is not a single score but rather a feature map with spatial support. As shown in Tables 2 and 3, POAGNet w/o side output achieved a better result than the POAGNet with absolute distance, indicating that utilizing convolution to measure the similarity between 2 outputs could boost the performance. Third, the POAGNet used side output to relieve the vanishing gradient problems in training and encourage the hidden layers to favor discriminative features. By comparing rows 5 and 6 in Tables 2 and 3, POAGNet can leverage multiscale information to boost the model's performance.

We also analyzed the causes of misclassification. We found that the major reason for misclassification is due to the optic disc unobvious discrepancy in the early stage. Fig S1 shows the positive predictive value, sensitivity, specificity, accuracy, and AUC in each time interval. Specificity, accuracy, and AUC are almost identical during these time intervals. Additionally, there are only 1 and 4 POAG cases for the first 2 time intervals, which makes the sensitivity relatively high. The sensitivity is also relatively stable for the rest time intervals. On the other hand, the positive predictive value has a continuing upward trend, indicating that nonPOAG-POAG pairs are easier to detect at the late stage than at the early stage. Second, we counted the number of misclassified patients in each relative year after the POAG onset year. Table 4 shows that more patients are misclassified when their POAG onset year is closer to the baseline visit. We also studied if the proposed model can focus on the important region to detect POAG. Because class activation maps are usually applied for the model with a single input, we used another method to reflect the regions that were more important for POAG diagnosis. Specifically, we used a 5 × 5 window size with a step size of 1 to mask the image and obtain the prediction probability for each window. We obtained the probability of each pixel by averaging all the probabilities of the windows it belongs to. Therefore, we can get the saliency map by drawing the probability of each pixel, which can reflect the important region for POAG detection. Four examples in Figure 4 demonstrate that the proposed model focused on the region between the optic disc and cup to detect POAG.

We conducted further experiments to evaluate the POAGNet performance on external validation data sets. Comparing the models trained on OHTS, SIG, and their combination, AUCs are similar on the SIG test set. This observation demonstrates the relative generalizability of POAGNet. On the other hand, we observed that the AUC on
OHTS drops from 0.9587 (trained on OHTS) to 0.7410 (trained on SIG). One potential reason is that the ophthalmologists carefully selected the images in SIG, and all glaucomatous images are due to the structural glaucomatous optic nerve abnormalities. In contrast, some optic disc photographs in OHTS were graded as POAG due to the change of VF without glaucomatous disc changes over time. Therefore, detecting POAG due to VF defects remains challenging.

Finally, we compared our work with 2 previous works based on the OHTS data set. We found 2 experimental differences when comparing our work with the prior work of Thakur et al. First, Thakur et al. discarded 24% of the optic disc photographs with poor image quality, while we used the whole OHTS data set. Second, Thakur et al. manually cropped the images in the data preprocessing stage, but we did not. It is possible that the use of these techniques might have improved the accuracy of the model. However, we deliberately avoided this extensive preprocessing to minimize the labor required and make our model as generalizable as possible. To make a fair comparison, we applied the same method in Thakur et al. to our data set (MobileV2 in Table 2). The AUC was 0.8124, lower than POAGNet under the same setting. In addition, our proposed model achieved an AUC of

---

**Table 4. The Misclassified in Each Relative Year After the Year When the Participants Truly Converted to POAG**

| POAG Onset Year | Number | %    |
|-----------------|--------|------|
| 0               | 37     | 30.58|
| 1               | 15     | 12.40|
| 2               | 19     | 15.70|
| 3               | 13     | 10.74|
| 4               | 12     | 9.92 |
| 5               | 6      | 4.96 |
| 6               | 6      | 4.96 |
| 7               | 8      | 6.61 |
| 8               | 4      | 3.31 |
| 9               | 1      | 0.83 |
| 10              | 0      | 0.00 |
| 11              | 0      | 0.00 |

POAG = primary open-angle glaucoma.
0.9578, higher than the AUC obtained by Thakur et al without cropping and discarding parts of the poor-quality image. Both results suggest that our approach may be more suitable to deploy to health care centers with a quick turnaround time.

We have performed the same method as provided in the study by Fan et al (ResNet-50) for our data for 3 models. The 3 models are Optic disc changes attributable to POAG by Endpoint Committee (Model 1), VF changes attributable to POAG by Endpoint Committee (Model 2), and Optic disc or VF changes attributable to POAG by Endpoint Committee (Model 3). The AUCs are 0.8483, 0.8115, and 0.8650 for these 3 models, respectively, and the result of Model 3 is similar to the result obtained by Fan et al (0.88). Because we did not use the same training and testing data split and preprocessing methods, the results are not strictly comparable. We also used these 3 models to test our proposed model, and the AUCs are 0.9378, 0.9286, and 0.9587, respectively.

**Limitations and Future Work**

One limitation of our proposed model comes from the data imbalance. Only 6.2% and 3.99% of all images had POAG on OHTS and SIG, respectively. The low portion of POAG images may result in relatively lower sensitivity. We plan to incorporate data sets from different countries and populations to improve the model in the future.

Another potential limitation is that, as previously discussed, it remains challenging to only detect POAG due to VF defects. This is mainly because theVF defects lack the obvious structural sign of glaucomatous optic neuropathy. It would be interesting to study the difference between the results derived from VF defects and glaucomatous optic nerve separately and incorporate VF into the deep learning models for a multimodal study.

In this work, we used baseline images and “follow-up images” as input to obtain substantial differences between images for POAG detection. The longitudinal images between the baseline image and the “follow-up image” may also play an important role in tracing the temporal patterns of POAG progression. Thus, we recommend using sequential models such as long short-term memory or transformer for POAG detection in the future.

In conclusion, this study proposed a new end-to-end deep learning network that simulates the process for automatic POAG detection from optic disc photographs. Two data sets were used to evaluate the proposed model. The results demonstrated that the proposed network performs well on POAG diagnosis. The cross-data set validation further proves the generalizability and robustness of our method. To maximize this study’s transparency and reproducibility and provide a benchmark for further refinement and development of the algorithm, we make the deep learning model, training codes, and data partition publicly available (https://github.com/bionlplab/poagnet).

**Footnotes and Disclosures**

Originally received: March 29, 2022.
Final revision: August 1, 2022.
Accepted: August 5, 2022.
Available online: August 13, 2022. Manuscript no. XOPS-D-22-00058.

1 Department of Population Health Sciences, Weill Cornell Medicine, New York, New York.
2 Institute for Public Health, Washington University School of Medicine, St. Louis, Missouri.
3 Department of Ophthalmology and Visual Sciences, Washington University School of Medicine, St. Louis, Missouri.
4 Department of Ophthalmology, Weill Cornell Medicine, New York, New York.
*Co-corresponding authors.
Disclosure:
All authors have completed and submitted the ICMJE disclosures form.
The authors have made the following disclosures:

No conflicting relationship exists for any author.

This project was supported by the National Library of Medicine under award number 4R00LM013001. This work was also supported by awards from the National Eye Institute, the National Center on Minority Health and Health Disparities, National Institutes of Health (grants EY09341, EY09307), Horncrest Foundation, awards to the Department of Ophthalmology and Visual Sciences at Washington University, the NIH Vision Core Grant P30 EY 02687, Merck Research Laboratories, Pfizer, Inc, White House Station, New Jersey, and unrestricted grants from Research to Prevent Blindness, Inc, New York, NY.

HUMAN SUBJECTS: Human Subjects were used in this study. The study protocol was approved by Institutional Review Board at each clinical center and Weill Cornell Medicine. All research adhered to the tenets of the Declaration of Helsinki. All participants provided informed consent.

No animal subjects were used in this study.

Author Contributions:

Research design: Lin, Wang, Van Tassel, Peng

Data acquisition and/or research execution: Liu, Gordon, Kass

Data analysis and/or interpretation: Lin, Van Tassel

Obtained funding: N/A

Manuscript preparation: Lin, Liu, Gordon, Kass, Wang, Van Tassel, Peng

Abbreviations:

AUC = area under the curve; OHTS = Ocular Hypertension Treatment Study; POAG = primary open-angle glaucoma; SIG = sequential fundus images for Glaucoma; VF = visual field.

Keywords:

Deep learning, Fundus photographs, Primary open-angle glaucoma (POAG), Siamese network.

Correspondence:

Yifan Peng, PhD, Weill Cornell Medicine, 425 E 61st DIV 305, New York, NY 10065. E-mail: yip4002@med.cornell.edu

Reference

1. Bourne RR, Stevens GA, White RA, et al. Causes of vision loss worldwide, 1990–2010: a systematic analysis. Lancet Glob Health. 2013;1(6):e339–e349.

2. Sommer A, Tielsch JM, Katz J, et al. Racial differences in the cause-specific prevalence of blindness in east Baltimore. N Engl J Med. 1991;325(20):1412–1417.

3. Jiang X, Torres M, Varma R, Los Angeles Latino Eye Study Group. Variation in intraocular pressure and the risk of developing open-angle glaucoma: the Los Angeles Latino Eye Study. Am J Ophthalmol. 2018;188:51–59.

4. Doshi V, Ying-Lai M, Azen SP, et al. Sociodemographic, family history, and lifestyle risk factors for open-angle glaucoma and ocular hypertension: the Los Angeles Latino Eye Study. Ophthalmology. 2008;115(4):639–647.e2.

5. Quigley HA, Katz J, Derick RJ, et al. An evaluation of optic disc and nerve fiber layer examinations in monitoring progression of early glaucoma damage. Ophthalmology. 1992;99(1):19–28.

6. Fleming C, Whitlock EP, Beil T, et al. Screening for primary open-angle glaucoma in the primary care setting: an update for the US Preventive Services Task Force. Ann Fam Med. 2005;3(2):167–170.

7. Kolomeyer NN, Katz LJ, Hark LA, et al. Lessons learned from two large community-based glaucoma screening studies. J Glaucoma. 2021;30:875–877.

8. Singh A, Dutta MK, Partha Sarathi M, et al. Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image. Computer Methods Programs Biomed. 2016;124:108–120.

9. Acharya UR, Dua S, Du X, Chua CK. Automated diagnosis of glaucoma using texture and higher order spectra features. IEEE Trans Inf Technol Biomed. 2011;15(3):449–455.

10. Dua S, Acharya UR, Chowriappa P, Sree SV. Wavelet-based energy features for glaucomatous image classification. IEEE Trans Inf Technol Biomed. 2011;16(1):80–87.

11. Issac A, Sarathi MP, Dutta MK. An adaptive threshold based image processing technique for improved glaucoma detection and classification. Computer Methods Programs Biomed. 2015;122(2):229–244.

12. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interf. 2018;15(141):20170387.

13. Lin M, Momin S, Lei Y, et al. Fully automated segmentation of brain tumor from multiparametric MRI using 3D context deep supervised U-net. Med Phys. 2021;48:4365–4374.

14. Lin M, Jiang M, Zhao M, et al. Cascaded triplanar autoencoder M-Net for fully automatic segmentation of left ventricle myocardial scar from three-dimensional late gadolinium-enhanced MR images. IEEE J Biomed Health Inform. 2022;26:2582–2593.

15. Zhang Y, Li X, Lin M, et al. Deep-recursive residual network model for automated classification of clinically significant prostate cancer in mp-MRI images based on an end-to-end deep neural network. IEEE Trans Med Imaging. 2018;37(5):1217–1139.

16. Wang Z, Liu C, Cheng D, et al. Automated detection of clinically significant prostate cancer in mp-MRI images based on an end-to-end deep neural network. IEEE Trans Med Imaging. 2020;39(16):12935–12947.

17. Peng Y, Keenan TD, Chen Q, et al. Predicting risk of late age-related macular degeneration using deep learning. NPJ Digit Med. 2020;3(1):1–10.

18. Peng Y, Dharssi S, Chen Q, et al. DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. Ophthalmology. 2019;126(4):565–575.

19. Lin M, Wyme YE, Zhou B, et al. Artificial intelligence in tumor subregion analysis based on medical imaging: a review. J Appl Clin Med Phys. 2021;22(7):10–26.

20. Tan H, Shi H, Lin M, et al. Vessel wall segmentation of common carotid artery via multi-branch light network. In: Proc SPIE 11313, Medical Imaging 2020: Image Processing, 11313, Accessed March 10, 2020. https://doi.org/10.1117/12.2549599.

21. Hou B, Zhang H, Ladhizinsky G, et al. Clinical evidence engine: proof-of-concept for a clinical-domain-agnostic decision support infrastructure. arXiv. 2021. https://doi.org/10.48550/arXiv.2111.00621.

22. Hou B-J, Zhou Z-H. Learning with interpretable structure from gated RNN. IEEE Trans Neural Netw Learn Syst. 2020;31(7):2267–2279.

23. Chen X, Xu Y, Wong DWK, et al. Glaucoma detection based on deep convolutional neural network. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). New York: IEEE; 2015:715–718.
24. Li L, Xu M, Liu H, et al. A large-scale database and a CNN model for attention-based glaucoma detection. *IEEE Trans Med Imaging*. 2019;39(2):413–424.

25. 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(8):1199–1206.

26. Thakur A, Goldbaum M, Youssefi S. Predicting glaucoma before onset using deep learning. *Ophthalmol Glaucoma*. 2020;3(4):262–268.

27. Christopher M, Belghith A, Bowd C, et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci Rep*. 2018;8(1):1–13.

28. Fu H, Cheng J, Xu Y, et al. Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Trans Med Imaging*. 2018;37(11):2493–2501.

29. Li A, Cheng J, Wong DWK, Liu J. Integrating holistic and local deep features for glaucoma classification. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. New York: IEEE; 2016:1328–1331.

30. Chicco D. Siamese neural networks: an overview. In: *Artificial Neural Networks*. Berlin/Heidelberg, Germany: Springer; 2021:73–94.

31. Li L, Wang X, Xu M, et al. DeepGF: glaucoma forecast using the sequential fundus images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin/Heidelberg, Germany: Springer; 2020:626–635.

32. Gordon MO, Kass MA. The Ocular Hypertension Treatment Study: design and baseline description of the participants. *Arch Ophthalmol*. 1999;117(5):573–583.

33. Kass M, Heuer D, Higginbotham E, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002;120(6):701–713.

34. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE; 2017:4700–4708.

35. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*. New York: PMLR; 2015:448–456.

36. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*: JMLR Workshop and Conference Proceedings. 323rd. New York: JMLR; 2011:315.

37. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86(11):2278–2324.

38. Ho Y, Wookey S. The real-world-weight cross-entropy loss function: modeling the costs of mislabeling. *IEEE Access*. 2019;8:4806–4813.

39. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv*. 2014. https://doi.org/10.48550/arXiv.1412.6980.

40. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning*. New York: PMLR; 2019:6105–6114.

41. Sandler M, Howard A, Zhu M, et al. MobiNetv2: inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE; 2018:4510–4520.

42. Fan R, Bowd C, Christopher M, et al. Detecting glaucoma in the ocular hypertension study using deep learning. *JAMA Ophthalmol*. 2022;140(4):383–391.