Applications of Deep Learning

Automated Assessment of Vascular Tortuosity in Mouse Models of Oxygen-Induced Retinopathy

Jimmy S. Chen, MD,[1] Kyle V. Marra, MD, PhD,[2,3] Hailey K. Robles-Holmes, BS,[1] Kristine B. Ly, BS,[5] Joseph Miller, MS,[1] Guoqin Wei, PhD,[2] Edith Aguilar, MD,[2] Felicitas Bucher, MD, PhD,[4] Yoichi Ideguchi, BS,[2] Aaron S. Coyner, PhD,[6] Napoleone Ferrara, MD,[1] J. Peter Campbell, MD, MPH,[6] Martin Friedlander, MD, PhD,[2] Eric Nudleman, MD, PhD[1]

Objective: To develop a generative adversarial network (GAN) to segment major blood vessels from retinal flat-mount images from oxygen-induced retinopathy (OIR) and demonstrate the utility of these GAN-generated vessel segmentations in quantifying vascular tortuosity.

Design: Development and validation of GAN.

Subjects: Three datasets containing 1084, 50, and 20 flat-mount mice retina images with various stains used and ages at sacrifice acquired from previously published manuscripts.

Methods: Four graders manually segmented major blood vessels from flat-mount images of retinas from OIR mice. Pix2Pix, a high-resolution GAN, was trained on 984 pairs of raw flat-mount images and manual vessel segmentations and then tested on 100 and 50 image pairs from a held-out and external test set, respectively. GAN-generated and manual vessel segmentations were then used as an input into a previously published algorithm (iROP-Assist) to generate a vascular cumulative tortuosity index (CTI) for 20 image pairs containing mouse eyes treated with aflibercept versus control.

Main Outcome Measures: Mean dice coefficients were used to compare segmentation accuracy between the GAN-generated and manually annotated segmentation maps. For the image pairs treated with aflibercept versus control, mean CTIs were also calculated for both GAN-generated and manual vessel maps. Statistical significance was evaluated using Wilcoxon signed-rank tests ($P \leq 0.05$ threshold for significance).

Results: The dice coefficient for the GAN-generated versus manual vessel segmentations was 0.75 ± 0.27 and 0.77 ± 0.17 for the held-out test set and external test set, respectively. The mean CTI generated from the GAN-generated and manual vessel segmentations was 1.12 ± 0.07 versus 1.03 ± 0.02 ($P = 0.003$) and 1.06 ± 0.04 versus 1.01 ± 0.01 ($P < 0.001$), respectively, for eyes treated with aflibercept versus control, demonstrating that vascular tortuosity was rescued by aflibercept when quantified by GAN-generated and manual vessel segmentations.

Conclusions: GANs can be used to accurately generate vessel map segmentations from flat-mount images. These vessel maps may be used to evaluate novel metrics of vascular tortuosity in OIR, such as CTI, and have the potential to accelerate research in treatments for ischemic retinopathies.

Financial Disclosure(s): The author(s) have no proprietary or commercial interest in any materials discussed in this article. Ophthalmology Science 2024;4:100338 © 2023 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/).

Oxygen-induced retinopathy (OIR) is a mouse model for ischemic retinopathies, such as retinopathy of prematurity (ROP), a leading cause of childhood blindness. In the OIR model, infant mice are exposed to hyperoxic conditions at postnatal day 7 (P7), resulting in the vaso-obliteration (VO) of retinal vasculature. These pups are subsequently transferred to room air at P12, resulting in relative hypoxia that drives aberrant retinal neovascularization (NV), akin to the 2-phase pathophysiology seen in ROP.[1–3] Historically, clinical and laboratory evaluation of ROP and OIR has been highly subjective and variable.[4–6] Automated methods, such as artificial intelligence (AI) models, have been developed to identify features of ROP, such as zone,[7] stage,[8,9] and plus disease.[10,11] However, fewer algorithms exist to quantitatively evaluate disease features, such as NV and VO, in OIR.[5,12–14] This represents an important gap in knowledge where AI-assisted analysis of disease features in OIR could increase our understanding of ischemic retinopathies and assist with developing novel therapeutics.

Prior algorithms for quantitative evaluation of NV and VO have been developed using small data sets[13] or have required manual input[14] and remain prone to issues of variability and subjectivity in grading disease severity.[12,14] Recently, Xiao et al[5] developed a fully automated AI-based pipeline to automatically segment and quantitatively evaluate NV and VO using a sample of approximately 1000 flat-mount images of retinals from OIR mice.[5] Although these algorithms have greatly increased the ability to objectively evaluate disease in the OIR model, only one study has attempted to evaluate vascular tortuosity,[15] a key feature of disease severity in OIR mice and with plus disease in ROP.[16,17] In this study, Scott et al[15] validated an integral-based tortuosity index using a computer-based image analysis algorithm on manually segmented flat-mounted mice retina at P12, P14, and P17. However, these experiments were performed on a small sample size at each time interval and involved manual segmentation, which is time-consuming and not scalable. Validation and automated computation of a quantifiable metric of vascular tortuosity would add a useful tool for the evaluation of OIR disease severity that would mirror clinical evaluation of ROP.

We previously validated a computer-based image algorithm called iROP-Assist,[18] developed to calculate vascular tortuosity in ROP, on flat-mount images from the OIR model.[19] We demonstrated that the cumulative tortuosity index (CTI), a marker of overall vascular tortuosity in an image, reflected the natural course of OIR and observed a significant decrease in vascular tortuosity in retinas treated with aflibercept (Eylea) compared with controls. Although our study validated these experiments on larger sample sizes of flat-mounted retinas, input into this algorithm also required manual segmentation. Automating the quantification of vascular tortuosity in OIR mice using AI remains a key gap in knowledge that has the potential to establish a standardized method for obtaining tortuosity measurements that may improve the reproducibility and efficiency of OIR experiments within and between laboratories.

The purpose of this proof-of-concept study was to address this gap in knowledge by developing an AI model capable of segmenting major blood vessels in retinal flat-mount images from the OIR model, evaluating this model's generalizability, and demonstrating that generated segmentation maps can be used in assessment of quantitative vascular tortuosity metrics. In summary, the current study aimed to create an automated pipeline for the standardized calculation of vascular tortuosity from a flat-mounted image. This fully automated pipeline has the novel potential to accelerate the quantification process of vascular severity in OIR in a reproducible, objective manner, increasing experimental throughput, and reducing potential biases.

Methods

Training and Testing Data Set

Animal studies were previously reviewed and approved by the Institutional Animal Care and Use Committees at both the Scripps Research Institute and the University of California, San Diego.

Age-matched C57BL/6 mice (The Jackson Laboratory [JAX]) mice were subjected to normoxic (NOX) conditions or to the OIR model as previously described.[3,20] The data set used to train the generative adversarial network (GAN) consisted of previously published flat-mounted mice retinas[20] as well as additional mice raised using the same methodology and by the same authors. Briefly, pups were exposed to an atmosphere containing 75% oxygen from P7 to P12 and then returned to room air until euthanized at prespecified time points: P12 (immediately), P17, and P25. Many of the mice sacrificed at P17 may have also been subjected to various therapeutics at various doses designed to rescue the OIR model.[20–22] Retinas were then dissected from enucleated eyes by using fine brushes to separate and clean retina from choroid and sclera. Retinas were fixed in 4% para-formaldehyde on ice for 1 hour before overnight incubation with $Ca^{2+}Mg^{2+}$ and 10 µg Griffonia simplicifolia-Isolectin B4 (GS-IB4, I21412, Thermo Fisher Scientific). After being cut into 4 leaflets, retinas were flat-mounted for imaging using a Zeiss 710 confocal laser-scanning microscope with ZEN 2010 software (Zeiss) at $10\times$ magnification and tile scanning (6 × 6 tiles).

Images from this data set were split into a training and held-out, internal test set in a 90:10 ratio. These data sets were stratified based on time points and disease (NOX vs. OIR). Additionally, the training and test set were stratified such that all images from a particular mouse were limited to the training or test set only. A second, external test set was compiled to assess the generalizability of our model and consisted of a random subset of flat-mount images from OIRSeg, a database of OIR images collected by Xiao et al.[5] Data regarding the age of the mice at euthanization were not collected for the OIRSeg data set.

Manual Segmentations

All images from the training set, held-out test set, and external test set were manually segmented for large vessels emerging from the optic disc by 4 graders (J.S.C., H.K.B., J.M., and K.L.). This definition was chosen to capture vessels that researchers are most likely to examine in evaluating the extent of vascular disease in OIR. Capillaries, neovascular tufts, and areas of VO were not segmented. To ensure that all graders segmented similarly, a small cross-validation study was previously performed with 10 flat-mount images and was evaluated both subjectively with manual review and objectively using the dice coefficient, a measure of overlapping pixels between manual segmentations by each grader.[19,23] Ultimately, this crossvalidation study demonstrated that there was no statistically significant difference between each graders segmented large vessels when assessed subjectively and objectively and that these manual segmentations were suitable as gold-standard labels for training an AI to automatically generate vessel maps.

Training GANs for Automated Vessel Segmentation

Pix2PixHD, an AI-based GAN developed to generate large, high-resolution synthetic images from label maps, was used in this study.[24] We chose Pix2PixHD over U-Nets, which are convolutional neural networks designed for segmentation, because literature has demonstrated that both AI methods have comparable segmentation accuracy[24] and utility in generating vessel segmentations from fundus images.[25] In our study, we trained Pix2PixHD in Python[26] using a Nvidia K80 GPU.[27] The GAN was optimized using the Adam optimizer at a learning rate of 10^{-6} for 100 iterations, followed by 100 additional iterations at a learning rate linearly decaying to 0. All images in the training set were loaded into the model pairwise (the original flat-mount
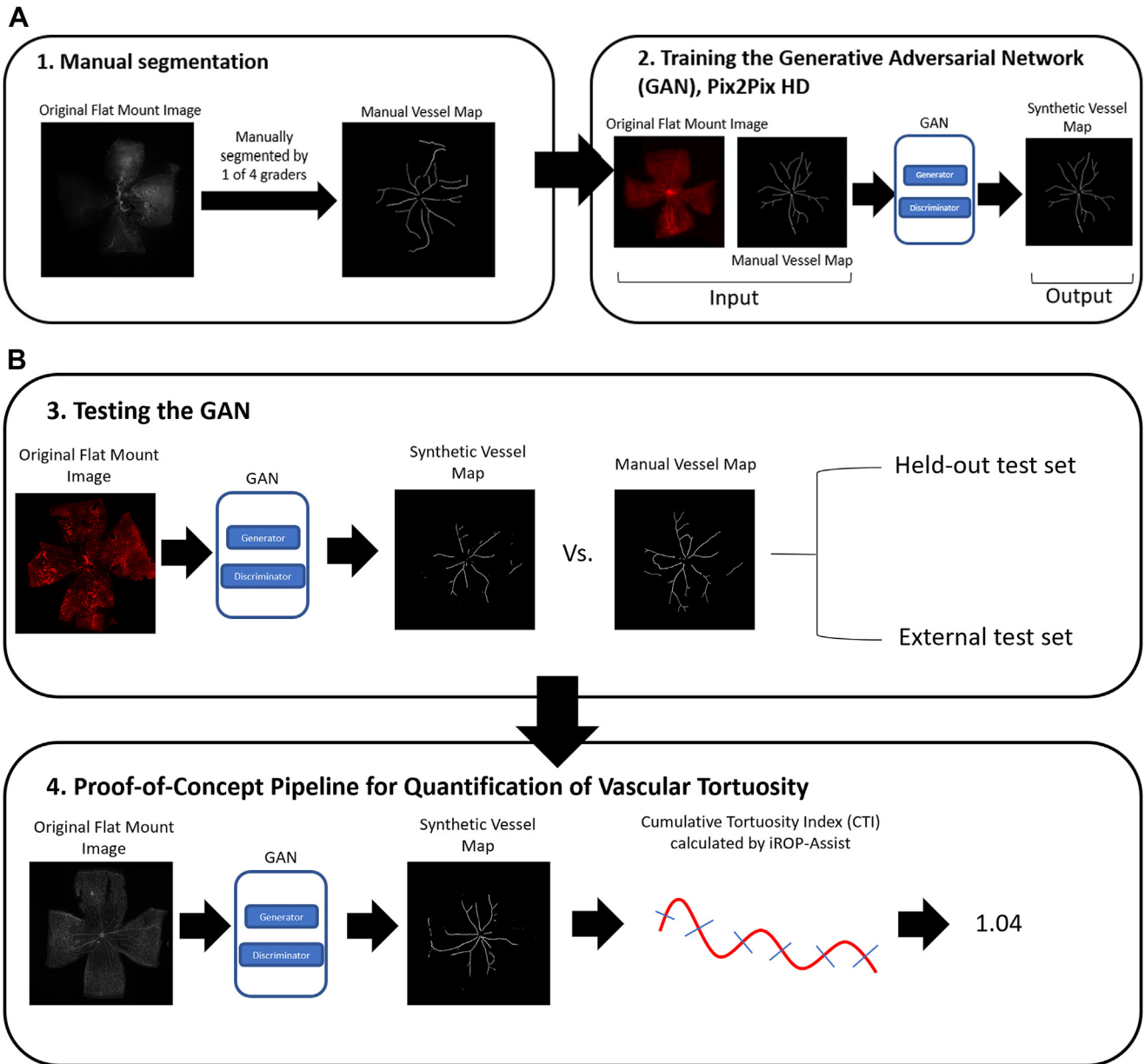
Figure 1. Study workflow. **A,** Flat-mounted images were manually segmented by 4 graders, and images underwent crossvalidation to ensure that they were similarly graded. These images were then used to train a generative adversarial network (GAN), a type of deep learning algorithm, to automatically segment large blood vessels. **B,** This GAN was validated on a held-out test set and external test set. A proof-of-concept experiment was performed to demonstrate that the vessel maps generated by the GAN could be used in quantification of vascular tortuosity.

image and manual vessel segmentation) at sizes of 512 × 512 pixels. Losses were monitored to ensure that no overfitting occurred.

Testing

After GAN training was completed, flat-mounted images and their accompanying manual segmentations from the held-out and external test sets were loaded pairwise into the GAN. These vessel maps were evaluated by an expert ophthalmologist (J.S.C.) and objectively by comparison to the manual vessel maps using the

dice coefficient. Outputted images were reviewed for error (i.e., missed vessels, inaccurate segmentation, etc.). After automated segmentation, the mean dice coefficient and associated 95% confidence intervals were reported for all images in a given data set.

Proof-of-Concept Assessment of Vascular Tortuosity

We used a third data set subsetted from a previously published study by Xin et al,[28] which contained flat-mount retinal images from OIR mice, which were injected intravitreally with an anti-

Table 1. Distribution of Images in the Training and Testing Sets

Data Set	Number of Images at P12 NOX	Number of Images at P12 OIR	Number of Images at P17 NOX	Number of Images at P17 OIR	Number of Images at P25 NOX	Number of Images at P25 OIR	Total
Training set	38	27	28	815	45	31	984
Held-out testing set	4	3	3	83	4	3	100
External testing set (OIRSeg)	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	50

GAN = generative adversarial network; NOX = normoxic; OIR = oxygen-induced retinopathy.
Flat-mount retinal images of mice euthanized at P12, P17, and P25 raised under both NOX and OIR were included in the training and held-out data set. Although the data set was imbalanced toward P17 OIR mice, several of these eyes were subject to various therapeutics and doses designed to rescue the eyes from OIR, representing a broad range of vascular tortuosity. Because the images in the OIRSeg were solely used to assess the segmentation performance of the GAN, the age of the mice was not collected for the OIRSeg images.

VEGF medication, aflibercept (Eylea, 2.5−25 μg), in 1 eye versus immunoglobulin G control in the fellow eye.[28] All images were manually segmented, input into the GAN, and a segmented vessel map was generated. These vessel maps were then input into a computer-based image analysis algorithm, i-ROP Assist,[18,29] for calculation of the CTI. Statistical analysis on these outputs was performed using R 4.0.5.[30] A Wilcoxon signed-rank test thresholded at P value of ≤ 0.05 was performed to assess statistically significant differences between the CTIs from the images treated with aflibercept versus IgG control. The overall workflow of this study is shown in Figure 1.

Results

Data Set

Overall, 1084 image pairs of flat-mount retinal images and manual segmentations from the primary data set were included in this study. This data set was further stratified into the training set and held-out test set, consisting of 984 and 100 image pairs, respectively. Additionally, 50 images from the OIRSeg data set were also included as an external
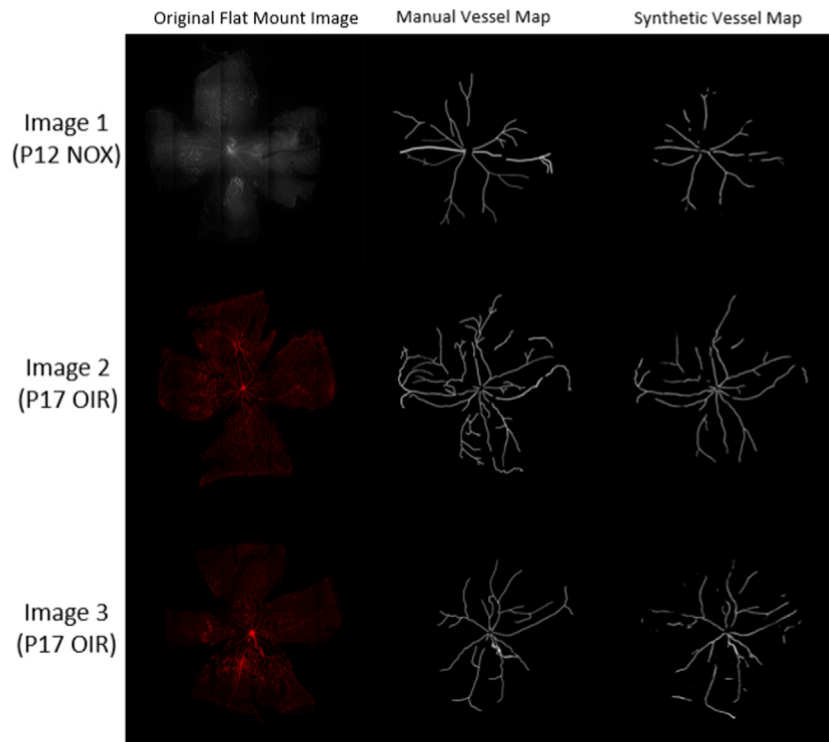

Figure 2. Examples of vessel segmentation maps produced by a generative adversarial network (GAN). After the GAN was trained on the training set, its performance was tested on a held-out test set. The dice coefficient was used to evaluate similarity in segmentation performance between the GAN-generated and manual vessel segmentations. For the held-out test set, the dice coefficient was 0.75 ± 0.27, demonstrating that the GAN had learned to accurately segment large vessels from the flat-mount images.
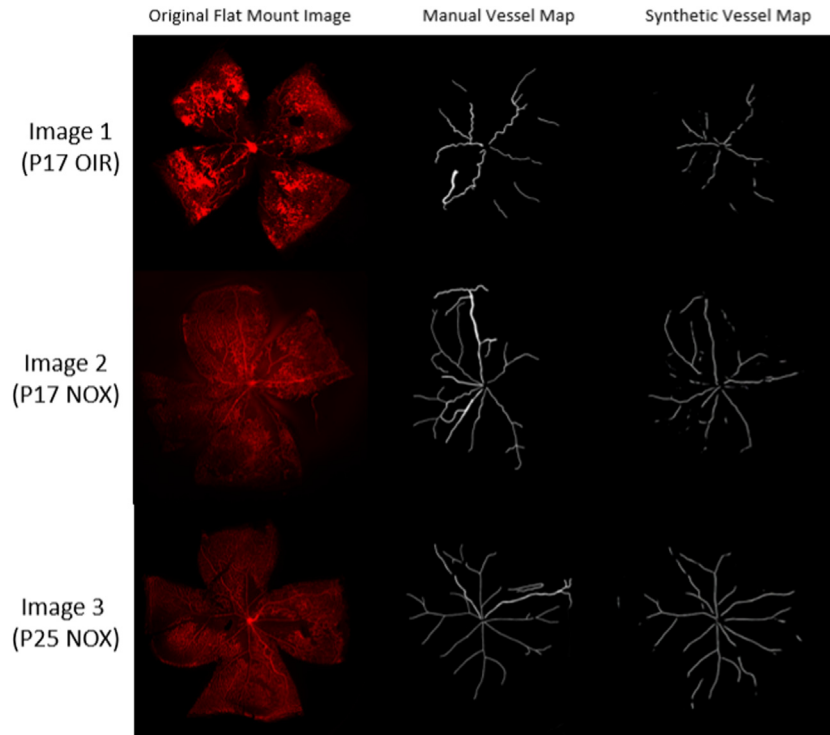
Figure 3. Examples of segmented vessel maps produced by a generative adversarial network (GAN). The trained GAN was additionally tested on an external test set, containing flat-mount images from the OIRSeg database that were not included in the training or held-out test set. The dice coefficient was 0.77 ± 0.17, demonstrating that the GAN's ability to segment large vessels from flat-mount images may have good generalizability to data sets from different institutions.

test set. The distribution of images in both of these data sets is shown in Table 1.

Validation Results on the Held-out and External Tests

Overall, the average dice coefficient between all manually segmented and GAN-generated vessel maps was 0.75 ± 0.27 and 0.77 ± 0.17 for the held-out and external test sets, respectively. Examples of GAN-produced vessel maps are shown in Figure 2 and Figure 3. Additionally, manual review of the test set images revealed reasons for disagreement between the manual versus GAN-generated vessel maps. The reasons included were identification of vessels by the AI algorithm not segmented by the grader, and vice versa, as well as discrepancies in tortuosity among vessels segmented manually and automatically. Examples of these errors are also seen in Figure 2 and Figure 3.

Vascular Tortuosity Comparison of Mice treated with Aflibercept vs. Control

A total of 36 images pairs of flat-mount retina images and their associated manually segmented vessel maps, consisting of 18 flat-mount retina images from eyes treated with aflibercept and 18 images from eyes treated with immunoglobulin G control, were input into the GAN. Cumulative tortuosity indices for each GAN-generated vessel map were then calculated using iROP-Assist. The mean CTI calculated using the manual vessel segmentations when comparing eyes treated with aflibercept versus control was 1.12 ± 0.07 versus 1.03 ± 0.02 ($P < 0.001$). The mean CTI calculated using the GAN-generated vessel segmentations when comparing eyes treated with aflibercept versus control was 1.06 ± 0.04 versus 1.01 ± 0.01 ($P = 0.003$; Fig 4). There was a statistically significant difference between CTIs calculated from the GAN-generated and manually segmented vessel maps for the aflibercept group ($P < 0.001$) and control group ($P = 0.01$). On error analysis, the GAN was noted to segment fewer vessels than manual segmentations and tended to exclude dimly stained vessels. Variations in number of vessels segmented between the representative control and aflibercept images may be because of grader variation and variations in mice vasculature.

Discussion

In this study, we developed a GAN to automatically segment large blood vessels from flat-mounted retina images from the OIR model and used these images to calculate the CTI for eyes treated with aflibercept and IgG control. This study has following 3 key findings: (1) GANs can accurately segment large blood vessels from flat-mount images; (2) our GAN performed well on a
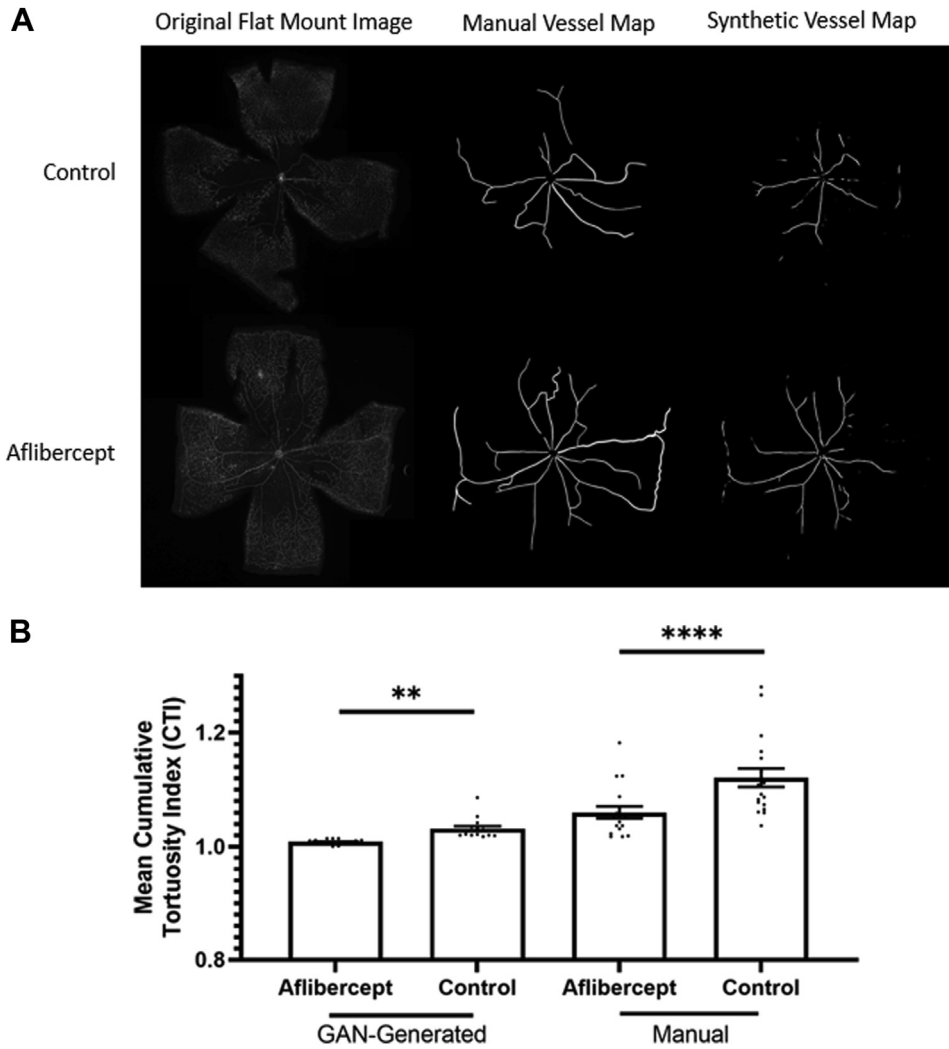
Figure 4. Comparison of cumulative tortuosity indices (CTIs) for flat-mount images of mice treated with and without aflibercept. **A,** Representative images of synthetic vessel maps and their corresponding flat-mount images and manual vessel maps. Variations in number of vessels segmented between the representative control and aflibercept images may be because of grader variation and variations in mice vasculature. **B,** Despite a statistically significant difference in generated CTIs between manual and synthetic vessel maps, synthetic vessel maps generated both manually and by the GAN both demonstrated statistically significant differences in vascular tortuosity in mice who were treated with aflibercept versus that of controls. These differences between GAN-generated and manually segmented vessels maps were likely in part due to the GAN consistently segmenting fewer vessels, likely learning that some of the vessels segmented by humans were not necessarily major vessels. Error bars represent standard error of the mean. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

diverse set of images and have good generalizability; and (3) GAN-produced vessel segmentation maps can be used to quantify vascular tortuosity in experiments using the OIR model to investigate the effects of therapeutics on this parameter.

The first key finding is that GANs can accurately segment major blood vessels from flat-mount images. Our model achieved a dice coefficient of 0.75 ± 0.27 and 0.77 ± 0.17 on our held-out and external data sets, representing excellent segmentation performance compared with manual segmentation. These results suggest that GANs can learn complex patterns of vasculature from flat-mount retinal images accurately. Although AI models have previously been developed to segment vessels from human fundus photographs,[31,32] ultra-widefield fundus photographs,[33] and fluorescein angiography,[34,35] there is a dearth of AI models trained to segment vasculature in flat-mount retinal images. Prior work has largely focused on vessel segmentation based on computer-based image analysis algorithms,[13,15] but widespread use of these algorithms has been limited because of sample size, disease diversity, and need for manual input. The model presented here was trained with a robust data set. A majority of images in the data set were collected at P17 at the peak of NV in the OIR model, which is the most common time point when quantifying the effects of therapeutics. Many of these images ranged in disease and vascular severity based on the therapeutic intervention. Additionally, we included age-matched controls at P12, P17, and P25 for both NOX and OIR to diversify our

training data set. This work presents numerous benefits, including obviating the need for manual segmentation, which is time-consuming, difficult, and labor intensive. Additionally, synthetic vessel maps generated from this GAN may be used in the future to analyze vascular tortuosity and dilation, 2 key features of plus disease in ROP and OIR. Future work may evaluate how these vessel maps can be used to augment our understanding of the pathophysiology of OIR and its response to therapeutics.

The second key finding is that our GAN performed well on a diverse set of images and may have good generalizability. High segmentation performance was achieved on 2 different testing data sets by 2 different laboratories. Therefore, our model may be useful in a variety of laboratory settings, although further evaluation of this model on more diverse data sets is needed to validate this claim. This is incredibly important because previous generalization studies on high-performing AI models for the diagnosis of diabetic retinopathy have demonstrated significantly decreased performance in diverse human populations.[36] The generalization performance of our model confers additional benefits to OIR research, including the ability to potentially compare vascularity between images from different laboratories, which may promote data sharing among researchers. Of note, there were discrepancies between the GAN-generated and manually segmented vessel maps. For example, in both Figure 2 and Figure 3, the GAN consistently segmented fewer vessels than the human graders. Additionally, the GAN was observed to have slightly decreased performance with dimly stained vessels (Fig 2). These discrepancies may be in part because human graders included branches of major vessels that the GAN did not view as important, which may also affect calculations of CTI between the GAN-generated and manual vessel segmentations. Additionally, although all vessels were manually segmented in a binary manner (i.e., pixels represented vessels or not), the GAN outputted segmentations in a probabilistic manner (i.e., likelihood a pixel represented a vessel). Although we did not modify any hyperparameters regarding thresholding in our GAN, our images likely seem more binary because of our segmentation methodology, which may limit the performance of the GAN during training (i.e., learning to replicate the binary input). Training on probabilistically labeled segmentations has recently shown to improve segmentation performance and represents an avenue for future work.[37]

The third key finding is that GAN-produced vessel segmentation maps can be used to quantify vascular tortuosity in experiments using the OIR model to investigate the effects of therapeutics on this parameter. In our proof-of-concept experiment, we demonstrated that GAN-generated vessel maps can be used to calculate a quantitative metric of vascular tortuosity, CTI. Cumulative tortuosity index calculations using both GAN-generated and manual vessel segmentation maps both demonstrated that aflibercept rescued vascular tortuosity compared with control. Altogether, our work represents a first step into introducing a novel vascular tortuosity metric, CTI from the iROP-Assist model, as an adjunct to NV and VO in evaluating the efficacy of therapeutics in OIR. Vascular tortuosity is an important pathophysiologic hallmark in ischemic retinopathies that current models of OIR do not assess. In previous work, we demonstrated that vascular tortuosity was correlated with NV in OIR mice at pathophysiologically relevant time points (P12, P17, and P25).[19,23] The standardized quantitative assessment of vascular tortuosity using CTI developed here may provide an important additional metric while assessing the efficacy of various therapeutics across diverse laboratory settings. This work also addresses the gap of subjective disease assessment that has existed in several clinical ophthalmic diseases, including glaucoma[38,39] and ROP.[40,41] In infants, AI-assisted assessment of plus disease has resulted in a vascular tortuosity metric that may decrease subjectivity in clinical disease assessment internationally.[42] We expect that these data will reduce subjectivity in this commonly used experimental model. Nevertheless, future work is needed to assess the fidelity of quantitative vascular tortuosities based on synthetic vessel maps from other laboratories.

This study has additional limitations that future work may address. Our data set was imbalanced toward P17 images of OIR mice. Although this is in part due to the retrospective use of images collected to evaluate therapeutics at the peak of disease in OIR, future work is needed to evaluate the performance of this model across all spectrums of disease and age. Second, because of the vascular complexity within flat-mount images, we chose to segment larger vessels only. More research is needed to assess whether complete segmentation of the vasculature, including capillaries, is feasible and adds to assessment to vascular tortuosity.

Overall, GANs can be trained to accurately segment large blood vessels from flat-mount images that can be integrated into a pipeline for calculating vascular tortuosity from these synthetic images. This work has the potential to accelerate our ability to assess responses to novel therapeutics developed in the laboratory using vascular tortuosity as a biomarker, which in turn may improve our ability to treat ischemic retinopathies. Further validation of this tool and the vascular tortuosity metric will be needed in diverse settings before widespread implementation.

Footnotes and Disclosures

[1] Shiley Eye Institute, Viterbi Family Department of Ophthalmology, University of California San Diego, San Diego, California.

[2] Molecular Medicine, the Scripps Research Institute, San Diego, California.

[3] School of Medicine, University of California San Diego, San Diego, California.

References

1. Kim CB, D'Amore PA, Connor KM. Revisiting the mouse model of oxygen-induced retinopathy. *Eye Brain*. 2016;8:67−79.
2. Scott A, Fruttiger M. Oxygen-induced retinopathy: a model for vascular pathology in the retina. *Eye (Lond)*. 2010;24:416−421.
3. Smith LE, Wesolowski E, McLellan A, et al. Oxygen-induced retinopathy in the mouse. *Invest Ophthalmol Vis Sci*. 1994;35:101−111.
4. Campbell JP, Ataer-Cansizoglu E, Bolon-Canedo V, et al. Expert diagnosis of plus disease in retinopathy of prematurity from computer-based image analysis. *JAMA Ophthalmol*. 2016;134:651−657.
5. Xiao S, Bucher F, Wu Y, et al. Fully automated, deep learning segmentation of oxygen-induced retinopathy images. *JCI Insight*. 2017;2:e97585.
6. Choi RY, Brown JM, Kalpathy-Cramer J, et al. Variability in plus disease identified using a deep learning-based retinopathy of prematurity severity scale. *Ophthalmol Retina*. 2020;4:1016−1021.
7. Zhao J, Lei B, Wu Z, et al. A deep learning framework for identifying zone I in RetCam images. *IEEE Access*. 2019;7:103530−103537.
8. Chen JS, Coyner AS, Ostmo S, et al. Deep learning for the diagnosis of stage in retinopathy of prematurity: accuracy and generalizability across populations and cameras. *Ophthalmol Retina*. 2021;5:1027−1035.
9. Mulay S, Ram K, Sivaprakasam M, Vinekar A. Early detection of retinopathy of prematurity stage using deep learning approach. In: *Medical Imaging 2019*. Computer-Aided Diagnosis; 2019:10950.
10. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136:803−810.
11. Tan Z, Simkin S, Lai C, Dai S. Deep learning algorithm for automated diagnosis of retinopathy of prematurity plus disease. *Transl Vis Sci Technol*. 2019;8:23.
12. Connor KM, Krah NM, Dennison RJ, et al. Quantification of oxygen-induced retinopathy in the mouse: a model of vessel loss, vessel regrowth and pathological angiogenesis. *Nat Protoc*. 2009;4:1565−1573.
13. Mazzaferri J, Larrivée B, Cakir B, et al. A machine learning approach for automated assessment of retinal vasculature in the oxygen induced retinopathy model. *Sci Rep*. 2018;8:3916.
14. Simmons MA, Cheng AV, Becker S, et al. Automatic analysis of the retinal avascular area in the rat oxygen-induced retinopathy model. *Mol Vis*. 2018;24:767−777.
15. Scott A, Powner MB, Fruttiger M. Quantification of vascular tortuosity as an early outcome measure in oxygen induced retinopathy (OIR). *Exp Eye Res*. 2014;120:55−60.
16. Good WV. Early Treatment for Retinopathy of Prematurity Cooperative Group. Final results of the Early Treatment for Retinopathy of Prematurity (ETROP) randomized trial. *Trans Am Ophthalmol Soc*. 2004;102:233−250.
17. Reynolds JD, Dobson V, Quinn GE, et al. Evidence-based screening criteria for retinopathy of prematurity: natural history data from the CRYO-ROP and LIGHT-ROP studies. *Arch Ophthalmol*. 2002;120:1470−1476.
18. Yildiz VM, Tian P, Yildiz I, et al. Plus disease in retinopathy of prematurity: convolutional neural network performance using a combined neural network and feature extraction approach. *Transl Vis Sci Technol*. 2020;9:10.
19. Marra KV, Chen JS, Robles-Holmes H, et al. Application of deep learning to quantify vascular tortuosity in mouse models

of oxygen-induced retinopathy. *Invest Ophthalmol Vis Sci.* 2022;63:2070-F0059.

20. Marra KV, Aguilar E, Guoqin W, et al. Bioactive extracellular vesicles from a subset of endothelial progenitor cells rescue retinal ischemia and neurodegeneration. *JCI Insight.* 2022;7: e155928.

21. Marra KV, Sakimoto S, Murinello S, et al. Extracellular vesicles from endothelial colony forming cells as paracrine mediators of neurovasculotrophic repair of the retina. *Invest Ophthalmol Vis Sci.* 2018;59:5489−5489.

22. Bucher F, Aguilar E, Marra KV, et al. CNTF prevents development of outer retinal neovascularization through upregulation of CXCL10. *Invest Ophthalmol Vis Sci.* 2020;61:20.

23. Marra KV, Chen JS, Robles-Holmes H, et al. Vascular tortuosity quantification as an outcome metric of the oxygen-induced retinopathy model of ischemic retinopathy. *Ophthalmol Sci.* doi: 10.1101/2022.10.02.510568

24. Wang T, Liu M, Zhu J, et al. High-resolution image synthesis and semantic manipulation with conditional GANs. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; vol 2018. IEEE Publications; 2018:8798−8807.

25. Chen JS, Coyner AS, Chan RVP, et al. Deepfakes in opthalmology: applications and realism of synthetic retinal images from generative adversarial networks. *Ophthalmol Sci.* 2021;1: 100079.

26. Van Rossum G, Drake Jr FL. *Python Reference Manual.* [Centrum voor Wiskunde en Informatica]; 1995.

27. Triton Shared Computing Cluster. University of California, San Diego. Service. https://doi.org/10.57873/T34W2R

28. Xin H, Biswas N, Li P, et al. Heparin-binding VEGFR1 variants as long-acting VEGF inhibitors for treatment of intra-ocular neovascular disorders. *Proc Natl Acad Sci U S A.* 2021;118:e1921252118.

29. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP, et al. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: performance of the "i-ROP" system and image features associated with expert diagnosis. *Transl Vis Sci Technol.* 2015;4:5.

30. R Core Team. Austria: R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing. Vienna; 2014.

31. Yue K, Zou B, Chen Z, Liu Q. Retinal vessel segmentation using dense U-net with multiscale inputs. *J Med Imaging (Bellingham).* 2019;6:034004.

32. Gegundez-Arias ME, Marin-Santos D, Perez-Borrero I, Vasallo-Vazquez MJ. A new deep learning method for blood vessel segmentation in retinal images based on convolutional kernels and modified U-Net model. *Comput Methods Programs Biomed.* 2021;205:106081.

33. Ding L, Kuriyan AE, Ramchandran RS, et al. Weakly-supervised vessel detection in ultra-widefield fundus photography via iterative multi-modal registration and learning. *IEEE Trans Med Imaging.* 2021;40:2748−2758.

34. Ding L, Bawany MH, Kuriyan AE, et al. A novel deep learning pipeline for retinal vessel detection in fluorescein angiography. *IEEE Trans Image Proc.* 2020;29(1).

35. Sun G, Liu X, Yu X. Multi-path cascaded U-net for vessel segmentation from fundus fluorescein angiography sequential images. *Comput Methods Programs Biomed.* 2021;211: 106422.

36. Lee AY, Yanagihara RT, Lee CS, et al. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. *Diabetes Care.* 2021;44:1168−1175.

37. Agrawal T, Choudhary P. Segmentation and classification on chest radiography: a systematic survey. *Vis Comput.* 2023;39: 874−913.

38. Jampel HD, Friedman D, Quigley H, et al. Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. *Am J Ophthalmol.* 2009;147:39−44.e1.

39. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA.* 2014;311:1901−1911.

40. Kalpathy-Cramer J, Campbell JP, Erdogmus D, et al. Plus disease in retinopathy of prematurity: improving diagnosis by ranking disease severity and using quantitative image analysis. *Ophthalmology.* 2016;123:2345−2351.

41. Trese MT. Subjectivity in retinopathy of prematurity screening. *Am J Ophthalmol.* 2015;160:406−407.

42. Campbell JP, Chiang MF, Chen JS, et al. Artificial intelligence for retinopathy of prematurity: validation of a vascular severity scale against international expert diagnosis. *Ophthalmology.* 2022;129:e69−e76.