Model-to-Data Approach for Deep Learning in Optical Coherence Tomography Intraocular Fluid Segmentation

NIHAAL MEHTA, MD; CECILIA S. LEE, MD, MS; LUÍSA S. M. MENDONÇA, MD; KHADIJA RAZA, MD; PHILLIP X. BRAUN, MD; JAY S. DUKER, MD; NADIA K. WAHEED, MD, MPH; AARON Y. LEE, MD, MSCI

**OBJECTIVE** To determine whether a model-to-data deep learning approach (ie, validation of the algorithm without any data transfer) can be applied in ophthalmology.

**DESIGN, SETTING, AND PARTICIPANTS** This single-center cross-sectional study included patients with active exudative age-related macular degeneration undergoing optical coherence tomography (OCT) at the New England Eye Center from August 1, 2018, to February 28, 2019. Data were primarily analyzed from March 1 to June 20, 2019.

**MAIN OUTCOMES AND MEASURES** Training of the deep learning model, using a model-to-data approach, in recognizing intraretinal fluid (IRF) on OCT B-scans.

**RESULTS** The model was trained (learning curve Dice coefficient, >80%) using 400 OCT B-scans from 128 participants (69 female [54%] and 59 male [46%]; mean [SD] age, 77.5 [9.1] years). In comparing the model with manual human grading of IRF pockets, no statistically significant difference in Dice coefficients or intersection over union scores was found (P > .05).

**CONCLUSIONS AND RELEVANCE** A model-to-data approach to deep learning applied in ophthalmology avoided many of the traditional hurdles in large-scale deep learning, including data sharing, security, and privacy concerns. Although the clinical relevance of these results is limited at this time, this proof-of-concept study suggests that such a paradigm should be further examined in larger-scale, multicenter deep learning studies.
we retrained an existing deep learning network to segment intraretinal fluid (IRF) on OCT B-scans using a distinct data set at its source center without any data transfer, thus bringing the model to the data (Figure 1).

Methods

The study protocol was approved by the Tufts Medical Center institutional review board, the source center for the imaging data, and adhered to the tenets of the Declaration of Helsinki and the Health Insurance Portability and Accountability Act of 1996. The protocol approved the collection of previously obtained and deidentified clinical data without specific consent. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

We reviewed all OCT volumes of patients with a diagnosis of exudative age-related macular degeneration; images were obtained on the spectral-domain OCT device (Cirrus HD-OCT 5000; Carl Zeiss Meditec) using a macular cube protocol during a 6-month period from August 1, 2018, to February 28, 2019, at the New England Eye Center (NEEC). This scan protocol consisted of 512 A-scans per B-scan and 128 B-scans per volume. The spectral-domain OCT system has an 840-nm central wavelength and the following operational parameters: 68,000 A-scans per second, an A-scan depth of 2.0 mm, an axial resolution of 5 μm, and a transverse resolution of 15 μm. From all eligible OCT volumes in this patient cohort, 400 scans from 58 patients were selected for the training set. An additional 70 scans from 70 patients (ie, 1 scan per patient) that were not included in the training set and collected from a different period than the training set were selected for the test set. Partitioning of the test set was performed at the patient level, and the test data were temporarily segregated from the training set. All scans were reviewed by a trained image reader (N.M.), and individual B-scans displaying IRF were exported as portable network graphics files. Each image was then manually segmented by 3 trained readers (N.M., L.S.M.M., and P.X.B.) who traced areas with IRF, resulting in a binary segmentation map for each image (Figure 2). All tracing was completed using ImageJ, version 2.0.0 (National Institutes of Health). Of the 400 training images, 70% were randomly designated for training and 30% for validation. Data were segregated at the patient level during the retraining phase.

The primary data analysis occurred from March 1 to June 30, 2019. The model-to-data approach was taken by freezing the model parameters from the prior study in which a deep learning model was trained to segment IRF on Heidelberg Spectralis OCT B-scans. The model parameters, retraining code, data preprocessing, and code for evaluation were packaged from the University of Washington and transferred using GitHub (https://github.com/uw-biomedical-ml/oct-irf-train). The researchers at the NEEC then downloaded the retraining code with the deep learning model and executed the retraining of the deep learning model. At no point did the researchers at the University of Washington have access to the computer or the OCT images at the NEEC. The model was retrained using OCT images from the spectral-domain OCT device, and segmentations were derived from the NEEC.

Before retraining, the transferred model was first evaluated directly on the test data. During training, OCT images were preprocessed in a manner similar to previously described work. Briefly, each OCT B-scan in the training set was vertically sliced into narrow (32 × 432 pixels) strip windows. In the training set, overlapping vertical windows were included as a form of offline augmentation. In the validation set, nonoverlapping windows were used for periodically evaluating the performance and determining when overfitting was occurring. The intensities were normalized using the values derived from the prior study. The normalized OCT intensities were then used directly as input into the first convolutional layer of the deep learning model. The output layer of the deep learning model provides a probability of each pixel being IRF.

The deep learning model was retrained using the following hyperparameter settings, similar to the settings used in the previous work: a batch size of 10, learning rate of $1 \times 10^{-5}$, Adam optimizer, and a smoothed Dice coefficient loss. After 30 iterations through the training set, the training was halted and the model was frozen using the highest performance on the validation set. All training was completed on a desktop workstation (Ubuntu, version 18.10; Canonical Ltd) at the NEEC equipped with a commercially available central processing unit (Core i7-5930K 3.50 Ghz; Intel Corp) and a graphics processing unit (GeForce GTX 960; Nvidia Corp).

The model with the lowest validation loss was then frozen, and the final test set of 70 images was evaluated. For each A-scan location, overlapping B-scan windows were used as input into the final model, and for each pixel, the mean of the overlapping inferences was calculated. To create the final binary segmentation masks, a predefined threshold of 0.5 was used to determine whether the mean of the overlapping inference output predicted IRF. The test set was evaluated using whole B-scan level Dice coefficients and intersection over union scores. Differences compared with manual segmentation as the criterion standard were assessed using the Wilcoxon rank sum test and 1-way Kruskal-Wallis test, given the nonnormal distribution of the data corrected for multiple comparisons. All statistics were performed using R, version 3.3.2 (R Project for Statistical Computing).

Key Points

**Question** Can a model-to-data approach be applied to deep learning studies in ophthalmology?

**Findings** This cross-sectional study applied a model-to-data approach to deep learning using 400 optical coherence tomography B-scans from patients with active exudative age-related macular degeneration. Without any data transfer between institutions, the model was trained to recognize areas of intraretinal fluid on the scans, and no difference was found in the comparison of the model with manual grading.

**Meaning** Although clinical application is at present limited, these results suggest that model-to-data approaches can obviate many of the traditional hurdles in large-scale deep learning projects and may increase application in future ophthalmology studies.
Results

The deep learning model was successfully packaged, transported, and retrained using the 400 segmented IRF images from 128 patients (69 female [54%] and 59 male [46%]; mean [SD] age, 77.5 [9.1] years) without any transfer of imaging data between institutions. Figure 1 displays a schematic of the model-to-data framework we used. Figure 2 shows an example image with IRF after manual human tracing and segmentation by the deep learning model.

The learning curve of the network during retraining as measured by Dice coefficient is displayed in Figure 3. Dice scores were not statistically different between deep learning and human graders by reference standard rotation ($P = .09$ for L.S.M.M., $P > .99$ for N.M., and $P = .12$ for P.X.B. by Bonferroni-corrected Kruskal-Wallis test). Similarly, intersection over union scores was not statistically different between deep learning and human graders ($P = .10$ for L.S.M.M., $P > .99$ for N.M, and $P = .12$ for P.X.B.). (Figure 4). The performance of the model before retraining is also shown in Figure 4. The differences in Dice coefficient and intersection over union scores for the model before vs after retraining were all statistically significant ($P < 2.2 	imes 10^{-16}$) and showed that the performance of the model after retraining were all statistically significantly better than before retraining, regardless of which human grader was chosen as the criterion standard. All training and evaluation code, model architecture, and model weights are available at GitHub (https://github.com/uw-biomedical-ml/oct-irf-train).

Discussion

Without any data transfer, we trained, validated, and tested a previously reported deep learning model that quantifies the area of IRF in OCT using a new data set. The deep learning network performance did not differ statistically compared with that of human graders (ie, compared with each single human grader as a criterion standard). The ability of this model to accurately segment IRF in structural OCT images of patients with diabetic macular edema, exudative age-related macular degeneration, and retinal vein occlusions has been previously demonstrated, with the model achieving a maximal cross-validation Dice coefficient of 0.911 compared with manual segmentation. Such an approach could be extended to quantify fluid volumes in 3-dimensional cube scans, allowing for tracking of fluid volume over time and response to treatment in various retinal diseases. It is important to note that this forms only 1 of the building blocks that will allow approaches such as these to be used clinically. A recent review of deep learning studies of medical imaging highlighted several shortcomings in the existing literature, chiefly a lack of randomized clinical trials, which need to be addressed before approaches such as these can be used in the clinical setting. As an example, using our approach, robust clinical studies correlating fluid volume to treatment efficacy or to visual outcomes would be needed before this algorithm could be of clinical value.
In the present study, our approach represents, to our knowledge, the first application in the field of ophthalmology of a packaged, transportable deep learning model to the data between 2 institutions (Figure 1). In the past several years, amid an explosion of artificial intelligence, similar interest has grown in decentralized approaches to deep learning training. Libraries such as UNet and TensorFlow (https://www.tensorflow.org/) have greatly accelerated the development and accessibility of deep learning. More recently, the Open Neural Network Exchange format (https://onnx.ai/), a collaborative effort between Facebook and Microsoft Corporation, created an open-source platform to facilitate interoperability and exchange of artificial intelligence models, and Google’s announcement of their Deep Learning Containers (https://cloud.google.com/ai-platform/deep-learning-containers/) is an effort toward the same goal.

Despite the incredible promise and successes in deep learning approaches, a major challenge in medical applications of deep learning remains the large amount of accurate data needed for successful training—an “insatiable appetite for training data.”19(p118) Deep learning studies specific to medical imaging have shown improved model performance with increasingly large data sets and unacceptably poorer performance with small data sets.20,21 The precise amount of data required is largely contingent on the complexity of the

![Figure 2. Example Segmentation of Intraretinal Fluid (IRF)](image)

The original unsegmented B-scan is shown before (A) and after (B) areas of IRF were manually traced by the human grader. The deep learning (DL)-generated probability mask for areas of IRF (C) and the DL-generated segmentation map (D) are also shown.

![Figure 3. Deep Learning Network Learning Curve Generated During Training of the Model](image)
deep learning model and the structure of the data; however, it is clear that deep learning approaches are unlikely to be successful without large data sets, especially in medical applications in which a high degree of accuracy is required. Data scarcity thus represents one of the chief rate-limiting steps for applying deep learning within medical and health care contexts despite the incredible promise.22-25

In general, obtaining adequately sized data sets requires compiling data from multiple clinical sites and transferring to a single site where the computational model is housed. Although this is technically feasible, compiling data from multiple sources into a single large data set has several challenges, including the time and resources needed to transfer data. The computational resources and time needed to transfer the large data needed for deep learning applications, such as imaging or video data, can be considerable, in addition to the often massive computational needs for deep learning training itself. Of note, the largest-scale deep learning study in ophthalmology thus far used data from 32 clinical sites within the Moorfields Eye Hospital NHS (National Health Service) Foundation Trust.13 This study also used Google DeepMind, a historically powerful system.26 Achieving data sets of this scale (particularly without the benefits of a nationalized health system, as is the case in the United States) will likely pose major challenges, and computational power on the scale of DeepMind is out of the reach of all but the largest organizations and institutions. Minimizing the resources needed for deep learning studies could thus significantly increase its accessibility. Additional hurdles include security and privacy risks in transferring potentially identifiable data, the potential need for multiple data sharing agreements, and obtaining appropriate permissions; both the International Committee of Medical Journal Editors27 and the Institute of Medicine28 have recently emphasized the importance of ethical data sharing. The European Union’s General Data Protection Regulation29 and the United States Health Insurance Portability and Accountability Act30 are examples of the international legal hurdles to sharing of clinical data. Various methods of improving the efficiency of data sharing and of augmenting existing data training features have been proposed and implemented.31 These approaches, however, still rely on an essentially data-to-model paradigm. Other techniques include federated learning, in which training data remain distributed and updates to a centralized model are shared, and differential privacy, whereby a model is indirectly trained on multiple data sets that are disjointed from the underlying sensitive data.32,33

A more advantageous solution would be a model-to-data approach, as applied in the present study (Figure 1). The model-to-data paradigm, in which “data remain stationary with models moving to the data,” has been previously described by Guinney and Saez-Rodriguez34(p392) and has been successfully applied in a number of large-scale data challenges, including the Digital Mammography DREAM Challenge. The model-to-data framework can obviate many of the hurdles intrinsic to the traditional data sharing approach and thereby not only improve the efficiency with which deep learning projects could be completed but also potentially allow data that were not initially collected with permission for sharing to be used. This in turn would increase data availability, improve the diversity of data sets, and augment the abilities of resulting deep learning models. Particularly in a clinical context, a model-to-data approach mitigates many of the major challenges for clinical use of deep learning, including data transfer. We trained a deep learning model on a central system at one institution and then retrained it on a distinct computer system at a separate institution; no clinical data were transferred between institutions, and the study was completed without researchers exposed to the other institution’s clinical data. In the future, model-to-data could allow for central construction of a
A robust deep learning model, followed by distribution to multiple clinical sites where distributed training can take place.

A similar distributed deep learning approach was described and simulated, using a single data set and without actual model distribution to a distinct institution, by Chang et al. In that study, a model was trained to classify images—including color fundus photographs of patients with diabetic retinopathy—using model distribution among 4 simulated institutions. Our study extends the approach proposed by Guinney and Saez-Rodriguez and simulated by Chang et al by demonstrating the feasibility of a model-to-data approach by retraining deep learning networks across 2 distinct institutions, with training data from separate patient populations and different machines as the input source. The present study is thus primarily a proof of concept for the model-to-data approach in ophthalmology: that of packaging an easily shared deep learning model and bringing it to the data. Our findings show that collaborative machine learning can be completed in ophthalmology without any data leaving the home institution. If extended, this strategy could allow for open-source, collaborative training networks among multiple institutions, each of which would generate models with distinct training weights. These weights could then be compiled to create a single, multicenter, validated deep learning model.

Emerging ethical issues are related to data privacy for deep learning applications. Our approach satisfies and has distinct advantages under the conventional standard for data sharing and patient confidentiality, namely that data, even if unidentifiable, should not be shared outside an institution without specific permissions. Bringing the model to the data is from a privacy, confidentiality, and security perspective—much easier in this regard. However, as deep learning models become more accurately reflective of the data from which they were trained, the question has arisen of whether training data could be reconstructed from the trained deep learning model. This possibility, termed model inversion, was first introduced in the context of medical applications for deep learning and is part of the broader set of emerging privacy and security concerns. Model inversion has recently been explored further, with several studies suggesting that deep learning models are vulnerable to such “inversion attacks” unless specifically designed to mitigate their susceptibilities. However, the extent of the risk posed by model inversion in allowing for reconstruction of training data is not entirely clear; other information about the original data may be needed to reconstruct it to any meaningful degree. Further studies are needed to better understand model inversion threats. For example, the threat posed by using model inversion to partially reconstruct highly identifiable information, such as a photograph of a face, may be less consequential for imaging data such as an OCT B-scan. However, to the extent that deep learning models are susceptible, our model-to-data approach does not obviate vulnerability to inversion. Unless other approaches to security and privacy, such as differential privacy, are incorporated, a deep learning model produced by a model-to-data approach could still theoretically be inverted. Consequently, a multifaceted approach will likely be necessary in the construction of adequately secure future deep learning models.

Limitations

There were several limitations to our study. Training of the existing model was completed at a single site using data from 1 device. To explore the model’s potential, this approach needs to be extended in multicenter and multisite studies beyond this initial proof-of-concept study. Moreover, although our deep learning model showed no statistically significant differences in performance vs human grading, its performance may have been improved with an even larger data set. Future studies should further explore the potential of a model-to-data approach, including with larger data sets, different types of imaging data, and multicenter data. Studies completed in a more clinical setting will be valuable in assessing whether this approach has application in clinical practice.

Conclusions

A model-to-data approach to deep learning was demonstrated for the first time, to our knowledge, in ophthalmology. Using this approach, the performance of the deep learning model was trained and showed no statistically significant difference in quantifying the intraretinal fluid pockets in OCT compared with human manual grading. Such a paradigm has the potential to more easily facilitate large-scale and multicenter deep learning studies.
Gyroscope Therapeutics. Dr. A. Y. Lee reported receiving grants from the National Eye Institute and Research to Prevent Blindness outside the submitted work, consulting for Genentech and Verana Health, Inc; receiving grants from Carl Zeiss Meditec, Inc, and Microsoft Corporation and honorarium from Topcon Medical Systems, Inc; and being an employee of the US Food and Drug Administration. No additional disclosures were reported.

Funding/Support: This study was supported by the Macula Vision Research Foundation; the Massachusetts Lions Clubs; grants 5-R01-EY012893-31, K23-EY029246, and R01-AGO60942-01A1 from the National Institutes of Health; grant FA9550-15-1-0473 from the Air Force Office of Scientific Research; the Champalimaud Vision Award; the Beckman-Arygros Award in Vision Research; the Yale School of Medicine Medical Student Fellowship; a scholarship provided by the CAPES Foundation—Ministry of Education Brazil, as part of the CAPES-Print program (process B8887.369769/2019-00); and an unrestricted grant from Research to Prevent Blindness.

Role of the Funder/Sponsor: The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Information: This article does not reflect the opinions of the US government or the US Food and Drug Administration.

REFERENCES
1. Schwaab K. The Fourth Industrial Revolution: what it means and how to respond. World Economic Forum. Published January 14, 2016. Accessed May 6, 2019. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-
what-it-means-and-how-to-responder/
2. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med. 2016;375(13):1216-1219. doi:10.1056/NEJMp1606181
3. Lee A, Taylor P, Kalpathy-Cramer J, Tufail A. Machine learning has arrived: Ophthalmology. 2017;124(12):1276-1278. doi:10.1016/j.ophtha.2017.08.046
4. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-444. doi:10.1038/nature14539
5. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88. doi:10.1016/j.media.2017.07.005
6. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019;103(2):167-175. doi:10.1136/bjophthalmol-2018-313173
7. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. Prog Retin Eye Res. 2019;72:100759. doi:10.1016/j.preteyeres.2019.04.003
8. Sandhu HS, Elmqig M, El-Adawy N, et al. Automated diagnosis of diabetic retinopathy using clinical biomarkers, optical coherence tomography (OCT), and OCT angiography. Am J Ophthalmol. Published online January 23, 2020. doi:10.1016/j.ajo.2020.01.016
9. Lee CS, Baugham DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. Ophthalmol Retina. 2017;1(4):322-327. doi:10.1016/j.orx.2016.12.009
10. Wu Y, Lutrill S, Feng S, et al. Development and validation of a machine learning, smartphone-based tonometer. Br J Ophthalmol. Published online December 23, 2019. doi:10.1136/bjophthalmol-2019-315446
11. Shah M, Roomans Lede A, Rittscher J. Automated classification of normal and Stargardt disease optical coherence tomography images using deep learning. Acta Ophthalmol. Published online January 24, 2020. doi:10.1111/aos.14433
12. Cui T, Wang Y, Jis, et al. Applying machine learning techniques in nomogram prediction and analysis for SMILE treatment. Am J Ophthalmol. 2020;210:71-77. doi:10.1016/j.ajo.2019.07.015
13. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342-1350. doi:10.1038/s41591-018-0067-6
14. Lee CS, Teying AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning-based, automated segmentation of macular edema in optical coherence tomography. Biomed Opt Express. 2017;8(7):3440-3448. doi:10.1117/BOE.8.034440
15. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053
16. Zheng Y, Sahni J, Campa C, Stangos AN, Raj A, Harding SP. Computerized assessment of intraretinal and subretinal fluid regions in spectral-domain optical coherence tomography images of the retina. Am J Ophthalmol. 2013;155(2):277-286.e1. doi:10.1016/j.ajo.2012.07.030
17. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ. 2020;368:m689. doi:bmj/m689
18. Verbraeken J, Wolting M, Katzy J, Klippenburg J, Verbeelen T, Refferele JS. A survey on distributed machine learning. arXiv. Preprint posted online December 20, 2019. Accessed February 13, 2020. https://arxiv.org/abs/1912.09789
19. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys. 2019;29(2):102-127. doi:10.1007/s00139-018-01103-1
20. Zhu X, Vondrick C, Fowlkes C, Ramanan D. Do we need more training data? Cornell University Library. March 5, 2015. Accessed February 13, 2020. https://dl.acm.org/doi/10.5555/3608529
21. Papenott N, Abad M, Erlingsson Ü, Goodfellow I, Talwar K. Semi-supervised knowledge transfer for deep learning from private training data. Cornell University Library. October 18, 2016. Accessed February 13, 2020. https://arxiv.org/abs/1605.07575
22. Guinney J, Saez-Rodriguez J. Alternative models for sharing confidential biomedical data. Nat Biotechnol. 2018;36(5):391-392. doi:10.1038/nbt.4128
23. Chang K, Ballachand N, Lam C, et al. Distributed deep learning networks among institutions for medical imaging. J Am Med Inform Assoc. 2018;25(8):945-954. doi:10.1093/jamia/ocy017
24. Kallman S, Mostert M, Gerlinger C, van Delden JMJ, van Thiel GJMW. Responsible data sharing in international health research: a systematic review of principles and norms. BMC Med Ethics. 2019;20(1):21. doi:10.1186/s12910-019-0355-9
25. Fedrickson M, Lanz E, Jha S, Lin S, Page D, Ristenpart T. Privacy in pharmacogenetics. In: Proceedings of the 23rd USENIX Conference on Security Symposium; August 2014. Accessed February 10, 2020. https://dl.acm.org/doi/10.5555/2671225.2671227

jamaophthalmology.com

JAMA Ophthalmology October 2020 Volume 138, Number 10
Deep Learning Algorithms and the Protection of Data Privacy

Kurt K. Benke, MSc, PhD; Janan Arslan, MSc, MBiostat

The application of artificial intelligence to pattern recognition in medicine has been described in recent studies in machine learning.¹ In particular, the use of machine learning in the form of the convolutional neural network has underpinned the current popularity of so-called deep learning algorithms, which are essentially convolutional neural network models distinguished by having a large number of hidden layers with many adjustable parameters requiring a large quantity of data for training. A deep learning algorithm can be applied to raw data without the need for data normalization and can be used to extract the inherent nonobvious features from training images that are necessary for accurate discrimination.¹,²

Despite success in various medical applications, there have been criticisms voiced about the limitations of deep learning and the need for moderation in published claims about the potential for use in clinical practice.³ One important issue is the security of private data because a large sample size is required for training a deep learning algorithm on patient data stored in a central computer or remote location. Extensive sets of data are needed when the images of phenotypes have very subtle differences or if there is a significant level of data variability.⁴ Protection of individual tagged data and identifiable patient information, including electronic health records, is regarded as very important when data sets are shared for training the deep learning algorithm.

Another issue is the limited sample size often encountered in ophthalmology, especially for rare diseases, which is a problem that can substantially affect the statistical confidence and performance of a deep learning algorithm. The inherent issue with a small sample size when training a model is that it may lead to poor generalization on new data. A possible solution to the problem of small sample size is to combine data from several institutions to increase the size of the training set. However, sharing patient data from different institutions and storing in a single central computer have many associated problems, such as technical compatibility, legal, and ethical issues, to which one may add protection of data privacy. Combining the data from different sources increases the risk of exposing patient identities if the central computer is compromised. Using local or batch-mode training⁵ on different sites and model averaging may introduce complications associated with error propagation, logistics, and epistemic uncertainties relating to model verification and validation.

One approach to increasing sample size while protecting data privacy at the same time is to distribute the deep learning model itself to multiple institutions for local retraining but without sharing private data that can be used for patient identification.⁴ This so-called model-to-data approach is the reverse of the traditional approach of data aggregation on a central location for training. It has been the subject of increasing research in recent years and is now emerging as an area of interest in machine learning applications in ophthalmology. Distributing an initially trained model to multiple centers for retraining and updating has the potential advantage of low storage requirements and no requirement for transfer of patient data.

In this issue of JAMA Ophthalmology, Mehta et al⁶ describe the development, testing, and validation of a model-to-data approach using a deep learning model. The stated objective was to determine whether the model-to-data deep learning approach, without any data transfer, can be applied successfully on real experimental data rather than simulated data. The design was a cross-sectional study involving patients with active exudative age-related macular degeneration imaged using optical coherence tomography (OCT). The application in ophthalmology involved grading the intraretinal fluid abnormalities imaged on OCT B-scans. They retrained an existing deep learning model to segment intraretinal fluid images and were able to successfully package, transport, and retrain the model using the segmented images without transferring any data between institutions.

Mehta et al⁶ applied their methodology for the model-to-data approach by initially using the model parameters obtained from a previous study in which the deep learning algorithm was trained to segment intraretinal fluid.⁷ The model parameters, retraining code, and evaluation code were transferred using GitHub from the University of Washington in Seattle to the New England Eye Center in Boston, where retraining was undertaken using new OCT images captured by different instrumentation. They reported that the model was trained with 400 OCT B-scans with a training Dice coefficient greater than 80% and validation Dice coefficient of approximately 80%.

The Dice coefficient (0%-100%) for image segmentation performance is a measure of similarity expressed as twice the area of overlap (number of pixels) between input and output images divided by the total area (number of pixels in both images), ie, it is associated with the proportion of pixels assigned correctly to regions in the output image that correspond to target regions in the input image. It is proportional

---

38. Bae H, Jang J, Jung D, Jang H, Ha H, Yoon S. Security and privacy issues in deep learning. Cornell University Library. July 31, 2018. Accessed February 13, 2020. https://arxiv.org/abs/1807.11655
39. Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. In: 2017 IEEE Symposium on Security and Privacy, May 22-26, 2017. Accessed February 13, 2020. https://ieeexplore.ieee.org/document/7958568
40. Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security; October 2015. Accessed February 10, 2020. https://dl.acm.org/doi/10.1145/2810103.2813677