![frontiers | Frontiers in Medicine]

Check for updates

The Ethical and Societal Considerations for the Rise of Artificial Intelligence and Big Data in Ophthalmology

T. Y. Alvin Liu[1]* and Jo-Hsuan Wu[2]

[1] *Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD, United States,* [2] *Shiley Eye Institute and Viterbi Family Department of Ophthalmology, University of California, San Diego, La Jolla, CA, United States*

Medical specialties with access to a large amount of imaging data, such as ophthalmology, have been at the forefront of the artificial intelligence (AI) revolution in medicine, driven by deep learning (DL) and big data. With the rise of AI and big data, there has also been increasing concern on the issues of bias and privacy, which can be partially addressed by low-shot learning, generative DL, federated learning and a "model-to-data" approach, as demonstrated by various groups of investigators in ophthalmology. However, to adequately tackle the ethical and societal challenges associated with the rise of AI in ophthalmology, a more comprehensive approach is preferable. Specifically, AI should be viewed as sociotechnical, meaning this technology shapes, and is shaped by social phenomena.

Keywords: ethics, bias, artificial intelligence, fairness, privacy

INTRODUCTION

The rise of artificial intelligence (AI) and big data has been hailed as the 4th Industrial Revolution. Recent advancement in AI, in the form of deep learning (DL) which is a subtype of machine learning (ML), and improvement in hardware such as graphic processing units (GPU), have propelled medical AI applications to the forefront of the public discourse. This is because DL has been shown to be on par with human experts in analyzing medical images across different specialties, especially in medical specialties that interact with and have access to a large number of images, such as dermatology, radiology, and ophthalmology (1–10). In addition, "super-human" feats achieved by DL, such as the robust prediction of age, gender, blood pressure and smoking status of a person from a color fundus photograph alone (11), have captured the public's imagination and sparked a debate on the role and impact of AI on medicine.

Ophthalmology, being at the forefront of this AI revolution in medicine, is well-positioned to actively participate in and be a thought-leader on the societal implications for the rise of AI and big data in medicine. In the following perspective piece, we will highlight the ethical controversies and considerations from an ophthalmological perspective. The two major concerns regarding the rise of AI in medicine and ophthalmology center on bias and privacy.

DISCUSSION

Bias and Fairness

AI has the potential to entrench, or even exacerbate, existing biases in the healthcare system *via* unfair recommendations or decision-making. Fairness can be defined as "the absence of any prejudice or favoritism toward an individual or a group based on their inherent or acquired characteristics" (12). A prominent example of a medical AI algorithm providing unfair recommendations and exacerbating biases was highlighted by a study by Obermeyer at al. (13) showing that an AI algorithm systematically biased against Black patients, by erroneously using previous health costs as a proxy for predicting health needs and illness severity.

Bias in the training data is one of the most common reasons for a ML algorithm to produce unfair downstream predictions or recommendations. Many types of bias in ML exist. A comprehensive discussion of the different kinds of bias is beyond the scope of the current paper, but is nicely summarized here (14, 15). Specifically, within the context of ophthalmology DL studies, imbalance in training images is a common, yet addressable, reason that can lead to biases against a patient subgroup, such as patients of a certain race. For example, the AREDS image dataset (16), generated from a landmark longitudinal clinical trial and used in numerous important ophthalmology DL studies, was derived primarily from Caucasian patients (about 96% of participants). While age-related macular degeneration (AMD) is more prevalent in Caucasian patients (prevalence of 5.4% vs. 4.2% in Hispanic, 2.4% in Black and 4.6% in Asian) (17–19), the difference in prevalence on a population level does not explain fully the extreme imbalance in the AREDS dataset. Additional factors, such as unequal access to or interest in participating in clinical trials, likely also played a role.

However, such imbalance in training data can be addressed in three different ways. First, patient recruitment in prospective studies can be planned to ensure equal enrollment numbers for different pre-specified patient subgroups, e.g., based on sex, age, race, ethnicity, socioeconomic status and disease severity, etc. Second, if the recruitment of a certain patient subgroup is limited by practicality or natural prevalence of the disease, e.g., Black patients with AMD, then low-shot DL can be attempted. Low-shot DL, in contrast to traditional DL which requires a large amount of data for training, can be trained with relatively few samples (20), and can outperform traditional DL approaches when the available training dataset is small (5). Third, the patient subgroup that is under-represented in the training samples can be augmented by generative DL, a DL technique that can generate synthetic data. It has been shown that retinal images, created by generative DL, can be used to train a robust DL system for AMD classification (21). Specifically, in the context of DL-based detection of referable diabetic retinopathy, generative DL has been used to increase the training image samples of an under-represented patient subgroup and has been shown to decrease the bias against that particular under-represented patient subgroup during testing (22).

In addition to addressing the data distribution, the model itself can be fine-tuned to improve fairness. For example, instead of minimizing the average error across all statistics, we could aim to minimize the maximum error of a subset of statistics as evaluated across different demographic groups of interest.

A recent scoping review on digital health solutions (23) found that AI health applications generally lacked vigorous pragmatic prospective real-world validations. Addressing training data imbalance during model development should produce more generalizable ophthalmic AI applications that perform more robustly in real-world validations.

Privacy

DL models typically require a large amount of data for training, and the rise of DL in ophthalmology coincided with the rise of big data, both in the form of images and tabular data. The training and testing of DL models often involve combining ophthalmic images from different sources, and there is increasing concern that such transfer of data represents an unacceptable risk of privacy breach, especially since fundus images are now considered protected health information.

Such concerns can be addressed in two ways: federated learning and differential privacy. The training of DL models can be facilitated by federated learning, which allows model training in a decentralized fashion, takes advantage of the data heterogeneity from disparate sources, and does not require actual transfer of data between the sources (24). This approach has been successfully implemented in the context of retinal microvasculature segmentation and referable diabetic retinopathy detection on optical coherence tomography (OCT) and OCT angiography images. The authors demonstrated that a federated learning approach achieved similar results as a traditional centralized learning approach (25). Similarly, instead of transferring data to train a DL model, the model itself can be "brought" to the data for retraining. This concept has been successfully demonstrated in the context of DL-based intraretinal fluid segmentation on OCT images, in which the parameters of a pre-trained DL model were frozen, transferred to and retrained at a different institution. The authors showed that such a "model-to-data" work flow could update a model and improve the model's performance, without the transfer of actual data (26).

Besides image databases, ophthalmology is also at the forefront of establishing massive tabular databases. The Intelligent Research in Sight (IRIS) Registry, spearheaded by the American Academy of Ophthalmology, is the largest specialty database in all of medicine in the world. The data collected to date has been invaluable, and led to numerous new insights and publications. Without a question, the IRIS Registry will be indispensable in developing the next-generation predictive ML algorithms. The data collected in IRIS is first de-identified, before being distributed to researchers. Traditional data de-identification methods include complete removal of all unique identifiers or coarsening of the original dataset. Data coarsening is achieved by providing the exact values of only a subset of the original sample and thus creating an incomplete dataset (27, 28). What remains to be seen is whether traditional data de-identification methods will be sufficient for protecting the privacy of data in the IRIS registry or similar tabular databases. Traditional de-identification methods are vulnerable to linkage

and other re-identification attacks, in which third parties correlate the supposedly anonymized data with unanticipated sources of auxiliary information to learn sensitive information about data participants. Examples of de-identification failure include the re-identification of "anonymized" hospital records released by Massachusetts' Group Insurance Commission and the re-identification of Netflix users' movie reviews from a dataset released as part of a ML challenge that Netflix hosted in 2006. A promising avenue of research is the application of differential privacy to large ophthalmic databases, such as IRIS.

Differential privacy is the only principled solution for releasing aggregate information about a statistical database, with provable guarantees that no information attributable to any individual in the dataset will be revealed. Briefly, differential privacy employs randomization to guarantee that the log odds ratio of any output of the analysis is bounded by and compared to a counterfactual world, in which any given participant has been entirely removed from the dataset, thereby formally limiting what inferences an arbitrarily well-informed observer can make about the data of any single participant (29). By definition, differential privacy prevents membership inference attacks as discussed above and provides a general umbrella of protection. However, the exact methods to create a differentially private dataset of unstructured data, e.g., ophthalmic images, are not currently available. This a major limitation of differential privacy as most recent advances in ML applications to ophthalmology have been in DL applications to ophthalmic images.

Finally, next-generation data infrastructure, specifically geared toward big data, ML and data privacy, is being developed, and a cutting-edge example is swarm learning. Swarm learning (30) is a decentralized data infrastructure that uses blockchain technology to ensure peer-to-peer data security. In contrast to federated learning which still requires a central parameter server, swarm learning is completely decentralized and, in addition, could inherit and be compatible with aforementioned differential privacy algorithms.

CONCLUSION

We are in the midst of the 4th Industrial Revolution, and ophthalmology has been at the forefront of the rise in AI/ML/DL and big data in medicine, and encountered various ethical and societal implications of this trend. While the concerns surrounding bias, fairness and privacy can be partially addressed by the strategies outlined above, a more comprehensive approach is preferable. This shift in mentality is best demonstrated by a recently announced special funding opportunity that was offered by the National Institute of Health as part of the Bridge2AI Common Fund[1]. The funding opportunity aims to produce Data Generation Projects that prospectively curate AI/ML ready data based on ethical principles. Multi-disciplinary teams, comprised of physicians, computer scientists and ethicists, are expected to promote a culture of ethical inquiry and consider ethical issues throughout the entire lifecycle of the project. Such an approach is grounded in the emerging view that AI is a sociotechnical issue: that is, AI shapes, and is shaped by social phenomena. The acknowledgment that the successful application of AI to medicine hinges on the holistic tackling of the associated ethical and societal implications is indeed a huge step forward, and we predict ophthalmologists in particular will play an important role in this regard in the years to come.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

---

[1]https://commonfund.nih.gov/bridge2ai

REFERENCES

1. Bridge J, Harding S, Zheng Y. Development and validation of a novel prognostic model for predicting AMD progression using longitudinal fundus images. *BMJ Open Ophthalmol.* (2020) 5:e000569. doi: 10.1136/bmjophth-2020-000569

2. Peng Y, Keenan TD, Chen Q, et al. Predicting risk of late age-related macular degeneration using deep learning. *NPJ Digit Med.* (2020) 3:111. doi: 10.1038/s41746-020-00317-z

3. Bhuiyan A, Wong TY, Ting DSW, Govindaiah A, Souied EH, Smith RT. Artificial intelligence to stratify severity of Age-Related Macular Degeneration (AMD) and predict risk of progression to late AMD. *Transl Vis Sci Technol.* (2020) 9:25. doi: 10.1167/tvst.9.2.25

4. Ludwig CA, Perera C, Myung D, Greven MA, Smith SJ, Chang RT, et al. Automatic identification of referral-warranted diabetic retinopathy using deep learning on mobile phone images. *Transl Vis Sci Technol.* (2020) 9:60. doi: 10.1167/tvst.9.2.60

5. Burlina P, Paul W, Mathew P, Joshi N, Pacheco KD, Bressler NM. Low-shot deep learning of diabetic retinopathy with potential applications to address artificial intelligence bias in retinal diagnostics

and rare ophthalmic diseases. *JAMA Ophthalmol.* (2020) 138:1070–7. doi: 10.1001/jamaophthalmol.2020.3269

6. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA.* (2017) 318:2211–23. doi: 10.1001/jama.2017.18152

7. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol.* (2018) 136:803–10. doi: 10.1001/jamaophthalmol.2018.1934

8. Campbell JP, Kim SJ, Brown JM, Ostmo S, Chan RVP, Kalpathy-Cramer J, et al. Evaluation of a deep learning-derived quantitative retinopathy of prematurity severity scale. *Ophthalmology.* (2020) 128:1070–6. doi: 10.1016/j.ophtha.2020.10.025

9. Liu TYA, Wei J, Zhu H, Subramanian PS, Myung D, Yi PH, et al. Detection of optic disc abnormalities in color fundus photographs using deep learning. *J Neuroophthalmol.* (2021) 41:368–74. doi: 10.1097/WNO.0000000000001358

10. Liu TYA, Zhu H, Chen H, Arevalo JF, Hui FK, Yi PH, et al. Gene expression profile prediction in uveal melanoma using deep learning: a pilot study for

the development of an alternative survival prediction tool. *Ophthalmol Retina.* (2020) 4:1213–5. doi: 10.1016/j.oret.2020.06.023

11. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng.* (2018) 2:158–64. doi: 10.1038/s41551-018-0195-0

12. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *arXiv:1908.09635v3.* (2019). doi: 10.48550/arXiv.1908.09635

13. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* (2019) 366:447–53. doi: 10.1126/science.aax2342

14. Olteanu A, Castillo C, Diaz F, Kiciman E. Social data: Biases, methodological pitfalls, and ethical boundaries. (2016) 2:13. doi: 10.2139/ssrn.2886526

15. Suresh H, Guttag J. A framework for understanding sources of harm throughout the machine learning life cycle. In: *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21).* New York, NY: Association for Computing Machinery (2021). p. 1–9. doi: 10.1145/3465416.3483305

16. Age-Related Eye Disease Study Research Group. The Age-Related Eye Disease Study (AREDS): design implications. AREDS report no 1. *Control Clin Trials.* (1999) 20:573–600. doi: 10.1016/S0197-2456(99)00031-8

17. Klein R, Klein BE, Knudtson MD, Wong TY, Cotch MF, Liu K, et al. Prevalence of age-related macular degeneration in 4 racial/ethnic groups in the multi-ethnic study of atherosclerosis. *Ophthalmology.* (2006) 113:373–80. doi: 10.1016/j.ophtha.2005.12.013

18. Friedman DS, Katz J, Bressler NM, Rahmani B, Tielsch JM. Racial differences in the prevalence of age-related macular degeneration: the Baltimore Eye Survey. *Ophthalmology.* (1999) 106:1049–55. doi: 10.1016/S0161-6420(99)90267-1

19. Zhou M, Duan P-C, Liang J-H, Zhang X-F, Pan C-W. Geographic distributions of age-related macular degeneration incidence: a systematic review and meta-analysis. *Br J Ophthalmol.* (2021) 105:1427–34. doi: 10.1136/bjophthalmol-2020-316820

20. Wang Y, Yao Q, Kwok J, Ni LM. Generalizing from a few examples: a survey on few-shot learning. *arXiv:1904.05046v3* (2020).

21. Burlina PM, Joshi N, Pacheco KD, Liu TYA, Bressler NM. Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. *JAMA Ophthalmol.* (2019) 137:258–64. doi: 10.1001/jamaophthalmol.2018.6156

22. Burlina P, Joshi N, Paul W, Pacheco KD, Bressler NM. Addressing artificial intelligence bias in retinal diagnostics. *Transl Vis Sci Technol.* (2021) 10:13. doi: 10.1167/tvst.10.2.13

23. Gunasekeran DV, Tseng RMWW, Tham Y-C, Wong TY. Applications of digital health for public health responses to COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies. *NPJ Digital Med.* (2021) 4:40. doi: 10.1038/s41746-021-00412-9

24. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep.* (2020) 10:12598. doi: 10.1038/s41598-020-69250-1

25. Lo J, Timothy TY, Ma D, Zang P, Owen JP, Zhang Q, et al. Federated learning for microvasculature segmentation and diabetic retinopathy classification of OCT data. *Ophthalm Sci.* (2021) 1:100069. doi: 10.1016/j.xops.2021.100069

26. Mehta N, Lee CS, Mendonça LSM, Raza K, Braun PX, Duker JS, et al. Model-to-data approach for deep learning in optical coherence tomography intraretinal fluid segmentation. *JAMA Ophthalmol.* (2020) 138:1017–24. doi: 10.1001/jamaophthalmol.2020.2769

27. Heitjan DF. Ignorability and coarse data: some biomedical examples. *Biometrics.* (1993) 49:1099–109. doi: 10.2307/2532251

28. Shardell M, El-Kamary SS. Sensitivity analysis of informatively coarsened data using pattern mixture models. *J Biopharm Stat.* (2009) 19:1018–38. doi: 10.1080/10543400903242779

29. Dwork C, Roth A. The algorithmic foundations of differential privacy. *Found Trends Theor Comput Sci.* (2013) 9:211–407. doi: 10.1561/9781601988195

30. Warnat-Herresthal S, Schultze H, Shastry KL, Manamohan S, Mukherjee S, Garg V, et al. Swarm learning for decentralized and confidential clinical machine learning. *Nature.* (2021) 594, 265–270. doi: 10.1038/s41586-021-03583-3