AMERICAN ACADEMY OF OPHTHALMOLOGY®

Editorial

Big Data and Artificial Intelligence in Ophthalmology: Where Are We Now?

Cecilia S. Lee, MD, MS - *Seattle, Washington*
James D. Brandt, MD - *Sacramento, California*
Aaron Y. Lee, MD, MSCI - *Seattle, Washington*

The successful marriage of big data and artificial intelligence (AI) has been one of the greatest steps forward in medical research this century. The term *big data*, defined to encompass both large quantities and diverse types of data, was coined in the 1990s,[1] and the term *AI* emerged even earlier: the first academic conference on AI occurred in 1956.[2] However, these concepts did not become revolutionary drivers of medical research until recently, when all types of health care settings began passively collecting massive amounts of medical data. Electronic health records (EHRs) data and medical imaging data are major examples of these types of routinely collected big data. At the same time, multiomics have given rise to complex datasets with multidimensional data for each patient. These dramatic changes in data collection have enabled AI to progress rapidly toward solving challenging medical research questions.

Ophthalmology has been at the forefront of AI research, in particular machine learning and deep learning approaches, because of the ubiquitous availability of noninvasive, rapid, and relatively inexpensive ophthalmic imaging. These advances in ophthalmic imaging have provided the big data required for successful applications of computer vision advances.[3,4] The Intelligent Research in Sight (IRIS®) Registry, the largest clinical database in the world led by the American Academy of Ophthalmology, is an additional example of why ophthalmology is uniquely positioned for big data questions.[5,6] Both big data and AI research are data driven, rather than based on the traditional observation-to-hypothesis approach, allowing discoveries outside the boundaries of traditional research methodology. Examples include novel biomarkers in the midst of multiomics datasets,[7] subtle associations that would not be detectable with smaller scale datasets,[8,9] and identification of imaging features that were undetected with human eyes, such as the hyporeflective outer retinal band associated with delayed rod-mediated dark adaptation, an established marker of early age-related macular degeneration.[10] This process drives the generation of new hypotheses and better insights into disease mechanisms.

Some of the many exciting study results include novel associations between progression of glaucoma and pulmonary function. This was an unexpected finding from multimodal machine learning models that combined retinal imaging, demographics, and systemic variables from the UK Biobank database, which achieved an area under the receiver operating characteristic curve of 0.97 in predicting progression of glaucoma.[11] Additional examples include deep learning algorithms to detect drusen, neovascular age-related macular degeneration, and diabetic macular edema after training with 101 418 OCT images of the retina from 5761 patients,[12] creating OCT angiography images from structural OCT images alone,[13] forecasting future visual field deficits from a single baseline visual field,[14] and fully automated, open-sourced quantification of oculoplastics measurements using routine facial photographs.[15]

Despite these numerous successes, several inherent and data-specific limitations must be kept in mind before interpreting study results. First, statistical significance becomes less meaningful when handling millions of data points because even minor differences will show statistical significance. To address this, clinical significance and biological plausibility should be emphasized when designing the study and analyzing their results. Second, a clear understanding of the data used to train AI algorithms is critical because the resulting algorithms may not be generalizable to the intended populations.[16] Third, validation of both big data and AI study results in independent study cohorts are key.[17]

As this field of research continues to advance, other critical issues remain. Although incorporating data from digital imaging systems is one thing, incorporating data from clinical notes and manually entered simple data like visual acuity and IOP is another, especially because today's clinical chart is filled with entry errors and imported text.[18] How can data cleaning methods be applied effectively and accurately to large EHR datasets such as the IRIS® Registry? How do we assess the validity of the EHR data? What are scalable statistical methods for a dataset of similar size and extent as the IRIS® Registry? How do we define clinically significant differences, rather than simply identifying statistical significance, in the setting of large datasets? Despite the data-rich ophthalmic environment, we remain hampered by limited interoperability between diagnostic systems and EHRs.[19] How do we incentivize manufacturers of new diagnostic methods and EHRs to share data reliably and seamlessly from the outset? While doing so, how do we protect the privacy of participants when we continue to add complex linkages between many dimensional datasets?[20] What are the ethical implications of big data and AI research? How do we standardize the datasets and promote data sharing?[21,22] Who is responsible for data breaches or failures of clinically deployed algorithms?

https://doi.org/10.1016/j.xops.2021.100036
ISSN 2666-9145/21

Ophthalmology Science is pleased to announce the Big Data & AI Special Issue with the hope of providing a home for studies in this growing field. In this special issue, we are particularly interested in research that highlights novel methodologies, such as nontraditional statistical approaches that can be applied to big data and machine learning or deep learning studies. Another goal is to initiate discussions of current challenges and potential strategies to overcome them. Some examples include standardization of the datasets, data sharing processes, data privacy, and obstacles to transferring research findings into clinical care. Understanding both the potential and the limitations of big data and AI approaches through an array of diverse studies and commentaries is the primary goal of this *Ophthalmology Science* special issue.

As vision science researchers and clinicians, we are uniquely positioned to take advantage of big data and AI research. As editors, we are grateful for this opportunity and look forward to engaging actively with the scientific community through this remarkable special issue.

Footnotes and Disclosures

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s): J.D.B.: Consultant – Adverum, Thea; Financial support – Santen; Lecturer – Genentech; Equity owner – Glaukos

A.Y.L.: Consultant – Genentech/Roche, Topcon, Verana Health, Johnson & Johnson, Gyroscope; Financial support – United States Food and Drug Administration, Lowry Medical Research Institute, Santen, Carl Zeiss Meditec, Microsoft, Novartis; Lecturer – Topcon

Supported by the National Eye Institute, National Institutes of Health, Bethesda, Maryland (grant nos.: K23EY029246 and R01AG060942; Latham Vision Research Innovation Award, Seattle, Washington (A.Y.L.); and Research to Prevent Blindness, Inc, New York, New York (unrestricted grant). The sponsors or funding organizations had no role in the design or conduct of this research. This article does not reflect the opinions of the Food and Drug Administration.

Correspondence:
Cecilia S. Lee, MD, MS, University of Washington, Box 359607, 325 Ninth Avenue, Seattle, WA 98104. E-mail: leecs2@uw.edu.

References

1. Lohr S. The origins of "big data": an etymological detective story. The New York Times. Available at: https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/. 01.02.13. Accessed 27.05.21.
2. Buchanan BG. A (very) brief history of artificial intelligence. *AIMag*. 2005;26(4), 53–53.
3. Yanagihara RT, Lee CS, Ting DSW, Lee AY. Methodological challenges of deep learning in optical coherence tomography for retinal diseases: a review. *Transl Vis Sci Technol*. 2020;9(2):11.
4. Wang SY, Pershing S, Lee AY, AAO Taskforce on AI and AAO Medical Information Technology Committee. Big data requirements for artificial intelligence. *Curr Opin Ophthalmol*. 2020;31(5):318–323.
5. Chiang MF, Sommer A, Rich WL, et al. The 2016 American Academy of Ophthalmology IRIS® Registry (Intelligent Research in Sight) database: characteristics and methods. *Ophthalmology*. 2018;125(8):1143–1148.
6. Parke II DW, Lum F, Rich WL. The IRIS® Registry: Purpose and perspectives. *Ophthalmologe*. 2017;114(Suppl 1):1–6.
7. Zhang G, Xue Z, Yan C, et al. A novel biomarker identification approach for gastric cancer using gene expression and DNA methylation dataset. *Front Genet*. 2021;12:644378.
8. Lee CS, Owen JP, Yanagihara RT, et al. Smoking is associated with higher intraocular pressure regardless of glaucoma: a retrospective study of 12.5 million patients using the Intelligent Research in Sight (IRIS®) Registry. *Ophthalmol Glaucoma*. 2020;3(4):253–261.
9. Owen JP, Blazes M, Lacy M, et al. Refractive outcomes after immediate sequential vs delayed sequential bilateral cataract surgery: a cohort analysis. *JAMA Ophthalmol*. In press.
10. Lee AY, Lee CS, Blazes MS, et al. Exploring a structural basis for delayed rod-mediated dark adaptation in age-related macular degeneration via deep learning. *Transl Vis Sci Technol*. 2020;9(2):62.
11. Mehta P, Petersen CA, Wen JC, et al. Automated detection of glaucoma with interpretable machine learning using clinical data and multi-modal retinal. *Am J Ophthalmol*. 2021 May 2;S0002-9394(21)00220-8. https://doi.org/10.1016/j.ajo.2021.04.021. In press.
12. Faes L, Wagner SK, Fu DJ, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *Lancet Digit Health*. 2019;1(5):e232–e242.
13. Lee CS, Tyring AJ, Wu Y, et al. Generating retinal flow maps from structural optical coherence tomography with artificial intelligence. *Sci Rep*. 2019;9(1):5694.
14. Wen JC, Lee CS, Keane PA, et al. Forecasting future Humphrey visual fields using deep learning. *PLoS One*. 2019;14(4):e0214875.
15. Van Brummen A, Owen JP, Spaide T, et al. PeriorbitAI: artificial intelligence automation of eyelid and periorbital measurements. *Am J Ophthalmol*. 2021 May 16;S0002-9394(21)00281-6. https://doi.org/10.1016/j.ajo.2021.05.007. In press.
16. Cahan EM, Hernandez-Boussard T, Thadaney-Israni S, Rubin DL. Putting the data before the algorithm in big data addressing personalized healthcare. *NPJ Digit Med*. 2019;2:78.
17. 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(5):1168–1175.
18. Henriksen BS, Goldstein IH, Rule A, et al. Electronic health records in ophthalmology: source and method of documentation. *Am J Ophthalmol*. 2020;211:191–199.
19. McKay KM, Apostolopoulos N, Dahrouj M, et al. Assessing the uniformity of uveitis clinical concepts and associated ICD-10 codes across health care systems

sharing the same electronic health records system. *JAMA Ophthalmol*. In press.
20. Tom E, Keane P, Blazes M, et al. Protecting data privacy in the age of AI-enabled ophthalmology. *Trans Vis Sci Tech*. 2020;9(2):36.
21. Lee AY, Campbell JP, Hwang TS, et al. Recommendations for standardization of images in ophthalmology. *Ophthalmology*. 2021;128(7):969−970.
22. Lee CS, Lee AY, Holland GN, et al. Big data and uveitis. *Ophthalmology*. 2016;123(11):2273−2275.