MAYO CLINIC

MAYO CLINIC PROCEEDINGS: DIGITAL HEALTH

REVIEW

# Retinal Scans and Data Sharing: The Privacy and Scientific Development Equilibrium

Luis Filipe Nakayama, MD; João Carlos Ramos Gonçalves de Matos, BSc; Isabelle Ursula Stewart, BSc; William Greig Mitchell, MBBS, MPH; Nicole Martinez-Martin, JD, PhD; Caio Vinicius Saito Regatieri, MD, PhD; and Leo Anthony Celi, MD, MPH

## Abstract

In ophthalmology, extensive use of ancillary imaging has enabled the development of artificial intelligence models, for which data are crucial. A data-sharing environment promotes external validation, collaborative research, and bias assessment before implementation in the real world; however, legal and ethical concerns need to be addressed in this process. The proposed solutions for improving the security of ophthalmic data sharing are patient consent and data-sharing agreements with third parties. Federated learning enables decentralized algorithm development, however, with limited results and unknown risks. Deidentification techniques through image manipulations and synthetically generated images are possible alternatives to improve security. Still, there is no single solution available. The challenge is to determine the appropriate level of risk and ensure accountability for the use of data. Sharing data, including retinal scans, can and should be performed within a trusted research environment, where there are data use agreements and credentialing of researchers, including requirements for training in responsible conduct of data use. In this review, we discuss the challenges and consequences surrounding limited sharing of ophthalmic datasets in the development of digital innovations and explore potential solutions that will enable safer sharing of retinal scan data.

 ■ Mayo Clin Proc Digital Health 2023;1(2):67-74

In 1935, pioneers Carleton Simon and Dr Isadore Golstein radically conceived the use of a retinal scanner to identify people based on their unique retinal vascular patterns.[1] This technique has been subsequently dramatized in films (Star Trek, Batman, X men) and video games (Half-life, Splinter Cell). Nowadays, iris architecture has replaced retinal biometrics for unique authentication, offering better cost and image capture that yields utility beyond high-security environments.[2]

The sharing of retinal tissue biometric data in medical congresses, journals, presentations, and deep learning studies has historically been blighted by the inability to deidentify retinal images. However, to avoid biased predictions, it is vital that machine learning (ML) algorithms are able to make use of accurate, representative retinal data sets.

Currently, limited sharing of ophthalmic retinal data sets, which are themselves sourced from very few countries and are representative of relatively homogeneous populations, represents a drawback in the ophthalmic data ecosystem and risks that are paradoxically widening health care disparities and perpetuating health care inequality.[3]

Here, we review the challenges and consequences of limited sharing of ophthalmic data sets in the development of digital health care and provide potential solutions for deidentification methods that may enable safer sharing of retinal scans.

## Artificial Intelligence (AI) in Ophthalmology

Ophthalmology is unique, in that, extensive use of ancillary imaging in diagnosis, follow-up, and treatment provides immense potential for ML models to detect, segment, and recognize patterns in images that can augment robust clinical decision making.[4]

From the Massachusetts Institute of Technology, Institute for Medical Engineering and Science, Cambridge (L.F.N., J.C.R.G.d.M., L.A.C.), Department of Biostatistics, Harvard TH Chan School of Public Health, Boston (L.A.C.), and Department of Medicine, Beth Israel Deaconess Medical Center, Boston (L.A.C.), MA; Department of Ophthalmology, São Paulo Federal University, Sao Paulo, SP, Brazil (L.F.N., C.V.R.); University of Porto, Porto, Portugal (J.C.R.G.d.M); University of Melbourne, Melbourne (I.U.S), and Royal Victorian Eye and Ear Hospital, Melbourne (W.G.M.), Victoria, Australia; and Department of Pediatrics (N.M.-M.) and Department of Psychiatry (N.M.-M.), Center for Biomedical Ethics, Stanford School of Medicine, Stanford, CA.

To date, algorithms have been applied to retinal images to aid screening, decision making, and outcome prediction for ocular diseases such as diabetic retinopathy, age-related macular degeneration, glaucoma, and retinopathy of prematurity.[5]

Certain retinal imaging modalities, including optical coherence tomography, color retinal photography, and fluorescein angiography, allow for screening of vision- and life-threatening diseases and prediction of patients' sex and cardiac risk as well as provide treatment suggestions.[6–8] Publicly available ophthalmic data sets include EYEPACS from the United States, APTOS from India, and MESSIDOR from France.[9–12] However, the lack of publicly available retinal images has resulted in data sets that represent only a small number of countries, patient populations, and diseases, which lack important demographic and disease-associated information. Thus, the potential benefits of their use in ML are yet to be realized.[12]

By limiting data sharing, predictions made by ML are threatened by bias because they are drawn from an unrepresentative data set that typically favors those already benefiting from advancements in health care.[3]

## Data Sharing Ecosystem

Individual privacy is essential in a data-sharing ecosystem: insufficient attention to data protection can cause personal harm through exposure of sensitive personal data and erosion of personal dignity.[13] Data are a crucial element of AI development. For example, deep learning algorithms learn relationships from input data in multiple hidden layers with numerous levels of abstraction and produce predictive results.[4,14] The increasing number of electronic medical devices and implementation of electronic health record systems, coupled with increased cloud storage and enhanced computer speed and power, have made the collection of massive volumes of data possible, enabling more complex and robust deep learning models.[15–17]

The hazards of biased predictions by these algorithms and the need for generalizability and fairness in ML development is a significant concern in AI research.[18,19] Evaluation of fairness goes way beyond ensuring that AI models perform well in underrepresented groups. In fact, relying on a ground truth that is not fair, because of different sources of bias, may end up legitimating and perpetuating existing outcome disparities among marginalized groups.[3,20–24] Therefore, before deploying AI algorithms in clinical practice, more multidisciplinary and collaborative efforts toward understanding the pitfalls of data collection, external validation processes, and, ultimately, the need for diverse data sets are necessary to mitigate algorithmic bias.[25]

Data sharing, in particular, enables external validation and model recalibration to elucidate how and when findings might be applied to more heterogeneous, underrepresented populations and prevent perpetuating existing health inequities.[19]

Together with transparent data sharing, publicly available datasets and reproducible codes are integral components of fair and generalizable digital health care. Furthermore, scientific advancement is enabled through secondary analysis,[25,26] which is advocated by the National Institute of Health.[27]

## Retina as a Biometric Identifier

Biometric authentication is used to determine personal identity on the basis of unique

### ARTICLE HIGHLIGHTS

- In the ophthalmology specialty, ancillary imaging examinations have enabled the development of artificial intelligence models.
- A data-sharing ecosystem is crucial to address bias before model deployment in the real world; however, legal and ethical concerns need to be addressed in this process.
- The proposed solutions to data sharing are patient consent and user agreements, federated learning, image manipulation, and synthetically generated images, with no single solution available.
- Sharing of data should be performed in a trusted research environment.


Mayo Clin Proc Digital Health ■ June 2023;1(2):67-74 ■ https://doi.org/10.1016/j.mcpdig.2023.02.003
www.mcpiqojournal.org

physical or behavioral characteristics.[28] The use of biometrics is more reliable than traditional authentication systems because it requires the physical presence of the person, equating to difficulties in copying, sharing, or forging identity.[28] Biometric systems rely on pattern recognition that identifies a person according to a feature vector.[29] It needs to be secure, universal, trustworthy, consistent, and user friendly, enabling functionalities such as pattern verification, individuality identification, and identification screening.[28,29] The commonest biometric modalities that are currently employed are the hand (fingerprint, palm print, or geometry), face, iris, voice, keystroke, and signature.[28]

The retina is an ocular tissue composed of cells, neurons, and vascular layers responsible for translating optical images into complex visual outputs.[6] The vascular structures are responsible for metabolically sustaining the inner retinal tissues and their development, which starts at 6 weeks of pregnancy, and almost all processes are completed in utero.[30] The vascular pattern of every individual is unique, stable during life, and difficult to change or replicate. Individual identification premises rely on retinal venous and arterial patterns, vascular end points, bifurcations, crossings, and optic disk as minutiae information.[31–33]

There are currently no commercially available retinal scanners.[31] Outside of health records, there is currently no data set of retinal scans linked with personal information. There is confusion between retinal and iris scans, which are commonly featured in movies. Although both images are unique to an individual, the absence of a data set of retinal scans linked with personal information poses the risk of reidentification, which leads to discovery of new information about an individual, almost negligible from sharing retinal scans.

### Legislation on Personal Data Protection

The Health Insurance Portability and Accountability Act (HIPAA) is the US national standard to safeguard sensitive patient information and prevent unwanted disclosure.[34] It was signed as a law in 1996 with the purpose of improving health insurance coverage. The HIPAA guidelines have defined 18 personal health information (PHI) identifiers; however, medical imaging examinations and retinal images are not included in the definition.[35] In 2017, the Office of Human Research Protections of the Department of Health and Human Services reconsidered the definition of what is considered identifiable information (common rule).[36] The decision after deliberating on responses from the public was that biospecimens and data derived from biospecimens, including whole-genome sequences and medical images, are not considered identifiers that can lead to discovery of new information about an individual unless they can be linked with another data set where the biospecimens and data derived from biospecimens are mapped to named individuals.[36]

In the European Union, personal data of patients are regulated by General Data Protection Regulation (GDPR). The GDPR, established in 2018, is often regarded as one of the strictest regulations of its kind. Because of its "extra-territorial effect," it protects any European Union citizen anywhere in the world.[37] Because any information that can lead to either direct or indirect identification of an individual is considered personal data under the GDPR, retinal scans also fall into this category.

Although prohibiting, in general, the use of biometric data processing for the purpose of uniquely identifying persons, this is authorized if there is explicit consent or it is necessary for public health reasons (eg, ensuring the safety of medical device standards).

Even though retinal scans are unique to each person, they cannot, per se, be linked to the identity of a person. The GDPR requires data sets to be pseudonymized, ie, the data can no longer be attributed to a specific patient without the use of additional information. Therefore, there must be layers of security to ensure that data containing PHI are not embedded in these images.

When it comes to sharing identifiable data, it is allowed as long as the intention for its use is communicated at the time the data are collected. To allow third parties to process data on one's behalf, data processing agreements must be signed by all parties.

## POSSIBLE SOLUTIONS

Data transmission needs to be secure, reliable, and available exclusively to intended

recipients.[38] Privacy safeguards are invariably detrimental to the overall utility of data, and the best solutions need to balance individual privacy with the ability to facilitate scientific development through broad, collaborative use of data.

### Patient Consent

Patient consent allows researchers to access patient data for medical, scientific, and educational purposes. In biomedical AI research, consent for data sharing is required to include privacy information, the intended use of data, and possible risks. However, reliance on patient consent for data usage is predisposed to data that are only representative of a small (biased) subset of the population. Because informed consent documents are often long and riddled with complex medical vernacular, only 1% of the patient population reads the consent form for more than 5 seconds, and some (ie, older patients, nonnative speaking patients, or those with comorbidities) are less likely to read it.[39,40]

Moreover, in the development of research databanks, it is not feasible to recontact all original participants for subsequent reconsent for the use of secondary data, with a trend toward the use of broad consent describing the scope and aim of the biobank and its governance.[41,42] Because of this, reliance on consent forms and personal preferences can produce data sets that are poorly representative of those who most stand to benefit from AI in health care.

In the GDPR, patient consent is required to share identifiable data, and the intention for its use should be communicated at the time of data collection.[37]

### Data Sharing Agreement With Third Party

Data use agreements with third parties are required by the HIPAA and GDPR and have been applied in the exchange of patient data among researchers.[43] An agreement between parts that includes clauses designed to protect data and responsibilities is required before data sharing but is not a substitute for data deidentification or expert determination. Agreements for data sharing consist of details of the project, responsibilities of investigators, and information about individual participant data, complying with the HIPAA guidelines.[43] In AI, agreements on research transfer with third parties are essential for data sharing. However, they still need to address additional points, such as deidentification level and a statement indicating that users are not permitted to reidentify patients or share their data with persons other than researchers working on the research project.[44]

### Expert Determination

According to the HIPAA guidelines, there are 2 potential paths for sharing patient health information: safe harbor and expert determination.[45] The safe harbor method consists of explicitly removing all HIPAA identifiers from the patient's record before sharing data; however, given that ophthalmic images do not have the same distribution of PHI as other medical data (ie, clinical texts), this method provides less relative "anonymity" and is, thus, less applicable to sharing of ophthalmic images. Expert determination requires a specialist with "appropriate knowledge of and experience applying generally accepted statistical and scientific principles and methods for rendering information not individually identifiable" to determine that the risk of reidentification is sufficiently low[45]; however, there is no standard method or readily available open-source solution for this method, making it similarly difficult to create a standardized benchmark for sharing ophthalmic image data.[46]

### Federated Learning

The term federated learning was coined by Google in 2016 with the aim of training a shared global model with a central server while keeping all data decentralized in its storage place of origin.[47,48] This has already been applied in areas such as finance, keyboard prediction, vehicle-to-vehicle communication, and Apple's Siri and Face ID.[49,50] In health care, federated learning represents a possibly ideal alternative form of AI because it allows large amounts of data to be analyzed without dependence on large cloud or data storage capacity.

In ophthalmology, there are simulated frameworks that apply federated learning in diabetic retinopathy[51] and retinopathy of prematurity,[52] integrating multimodal


Mayo Clin Proc Digital Health ■ June 2023;1(2):67-74 ■ https://doi.org/10.1016/j.mcpdig.2023.02.003
www.mcpiqojournal.org

examinations such as retinal fundus photography and optic coherence tomography.

However, the deployment of federated learning in health care has challenges, limitations, and vulnerabilities, including the need for appropriately harmonized and adequate-quality data sets to optimize model performance; statistical challenges in normalizing data sources to compatible formats; controlling for model bias across distributed data; the risk of identifying underlying training data through inverting applied data weights; the need for local, professional, and computational resources to run AI models before sharing them with the centralized system; and the communication transfer cost.[47,53]

### Deidentification

The concept of removing PHI from health care data for secondary use is not uniform throughout legislations across countries. The term deidentification comes from the HIPAA law regarding identification and removal of PHI from data. In the GDPR, the concept of pseudonymization accounts for reducing the risk of data reidentification.[54–56]

### Image Manipulation

Traditional image protection techniques, including inserting digital watermarks and cryptography solutions, and image preprocessing techniques, such as random cropping and image augmentation, are alternatives to deidentification that may help ensure individual anonymity. Validation studies are needed to assess how efficacious these techniques are in ensuring the security of retinal scan data without compromising the features of the original retinal disease and model performance for downstream tasks.[57]

### Synthetic Data

The most common synthetic data approach uses generative adversarial networks (GANs), a framework that applies adversarial networks to generate high-quality synthetic images through a generator and discriminator, which compete against each other in the form of a zero-sum game to minimize the probability of the discriminative model predicting the sample origin.[58]

Generative adversarial networks have been used to address the shortage of large and diverse data sets as well as improve the quality of images and the accuracy of classification methods.[59–62]

Synthetically generated images have also been applied in image deidentification.[63] However, although GAN-generated images may ensure better patient privacy, there are still unresolved uncertainties regarding maintenance of the capacity to address downstream tasks.[64]

In synthetic health care data, there are no clear evaluative metrics for the assessment of data fidelity, with GANs limited by the quality and size of the original data set. The balance among synthetic image fidelity to real data, representation of the variability of real data, and the absence of identity leakage must be weighed carefully.[62,65]

## CONCLUSION

The proliferation of AI in health care brings unprecedented opportunities for efficient, accurate, and cost-effective diagnosis, prediction, and clinical decision making that performs on-par with or better than human clinicians.[3] In ophthalmology, AI models have, to date, been applied to diseases such as diabetic retinopathy, retinopathy of prematurity, and age-related macular degeneration for screening, detection, and prognosis. However, a deficit in the availability of public data sets has hindered collaborative research and external validation studies that would allow benefits to be felt in wider, heterogeneous patient populations.[5,12]

The use of retinal scanners as a biometric authentication method relies on the unique and stable architectural characteristic of retina vascular tissues. Unfortunately, real-world AI applications on retinal scans are still limited by high costs of hardware and software, distrustful public perception of the technology, lack of publicly available data sets representative of populations that most stand to gain from clinical AI, and challenges in image capture, thus making hand and facial structures as well as iris characteristics the more commonly used identifiers.[1,66]

The use of patient consent is a valid solution to data collection and secondary use for medical, scientific, and educational purposes. Nevertheless, as discussed, consent itself can potentially create biased data sets and hamper

the inclusion of retrospective data in algorithms.

Federated learning is an alternative that would enable collaborative research without explicitly sharing patient data; however, there are challenges in multicentric data harmonization and bias control, limiting data representation to a few capable centers.

Health Insurance Portability and Accountability Act expert determination is another possible approach for sharing medical imaging; however, the statistical and scientific method is not straightforward and lacks standards and open-access available solutions.

In deidentification, image manipulation and GAN-generated images are alternatives, with a balance between private security and the capability to preserve downstream tasks and with the ability to identify clinical findings.

There is presently no viable single solution for safe retinal scan deidentification that still allows for retinal images to be effectively utilized in public databases for clinical models. More studies are needed to assess risks and improve the security of sharing retinal data.

Finally, there is always a risk of reidentification with any high-dimensional health data, whether it is an image, genomic data, or some combination of clinical and social demographic data. The challenge is in determining the appropriate level of risk and ensuring accountability for the use of health data. Sharing data, including retinal scans, can and should be performed within a trusted research environment, where there are data use agreements and credentialing of researchers, including requirements of responsible data use training.

## POTENTIAL COMPETING INTERESTS

The authors report no competing interests.

**Abbreviations and Acronyms:** **AI,** artificial intelligence; **GAN,** generative adversarial network; **GDPR,** General Data Protection Regulation; **HIPPA,** Health Insurance Portability and Accountability Act; **ML,** machine learning; **PHI,** personal health information

**Correspondence:** Address to Luis Filipe Nakayama, MD, Massachusetts Institute of Technology, Institute for Medical Engineering and Science, 77 Massachusetts Ave, Cambridge, MA02139 (luisnaka@mit.edu).

**ORCID**

Luis Filipe Nakayama: https://orcid.org/0000-0002-6847-6748

## REFERENCES

1. Simon C. A new scientific method of identification. *N Y State J Med*. 1935;35(18):901-906.
2. Jain A, Bolle R, Pankanti S. *Biometrics: Personal Identification in Networked Society*. Springer; 1999.
3. Celi LA, Cellini J, Charpignon ML, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—a global review. *PLoS Digit Health*. 2022;1(3):e0000022. https://doi.org/10.1371/journal.pdig.0000022.
4. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. https://doi.org/10.1038/nature14539.
5. Ting DS, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167-175. https://doi.org/10.1136/bjophthalmol-2018-313173.
6. Hoon M, Okawa H, Della Santina L, Wong RO. Functional architecture of the retina: development and disease. *Prog Retin Eye Res*. 2014;42:44-84. https://doi.org/10.1016/j.preteyeres.2014.06.003.
7. Korot E, Pontikos N, Liu X, et al. Predicting sex from retinal fundus photographs using automated deep learning. *Sci Rep*. 2021;11(1):10286. https://doi.org/10.1038/s41598-021-89743-x.
8. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2(3):158-164. https://doi.org/10.1038/s41551-018-0195-0.
9. Decencière E, Zhang X, Cazuguel G, et al. Feedback on a publicly distributed image database: the Messidor database. *Image Anal Stereol*. 2014;33(3):231-234. https://doi.org/10.5566/ias.1155.
10. Diabetic retinopathy detection. California Healthcare Foundation. https://www.kaggle.com/c/diabetic-retinopathy-detection. Accessed January 17, 2023.
11. APTOS 2019 blindness detection. Asia Pacific Tele-Ophthalmology Society. https://www.kaggle.com/c/aptos2019-blindness-detection/data. Accessed January 17, 2023.
12. Khan SM, Liu X, Nath S, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *Lancet Digit Health*. 2021;3(1):e51-e66. https://doi.org/10.1016/S2589-7500(20)30240-5.
13. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf11. Accessed January 17, 2023.
14. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018;19(6):1236-1246. https://doi.org/10.1093/bib/bbx044.
15. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24-29. https://doi.org/10.1038/s41591-018-0316-z.
16. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56. https://doi.org/10.1038/s41591-018-0300-7.
17. Beaulieu-Jones B, Finlayson SG, Chivers C, et al. Trends and focus of machine learning applications for health research. *JAMA Netw Open*. 2019;2(10):e1914051. https://doi.org/10.1001/jamanetworkopen.2019.14051.
18. Gichoya JW, Banerjee I, Bhimireddy AR, et al. AI recognition of patient race in medical imaging: a modelling study. *Lancet Digit Health*. 2022;4(6):e406-e414. https://doi.org/10.1016/S2589-7500(22)00063-2.
19. Mitchell WG, Dee EC, Celi LA. Generalisability through local validation: overcoming barriers due to data disparity in

healthcare. *BMC Ophthalmol*. 2021;21(1):1-3. https://doi.org/10.1186/s12886-021-01992-6.

20. Wong AI, Charpignon M, Kim H, et al. Analysis of discrepancies between pulse oximetry and arterial oxygen saturation measurements by race and ethnicity and association with organ dysfunction and mortality. *JAMA Netw Open*. 2021;4(11):e2131674. https://doi.org/10.1001/jamanetworkopen.2021.31674.
21. Bhavani SV, Wiley Z, Verhoef PA, Coopersmith CM, Ofotokun I. Racial differences in detection of fever using temporal vs oral temperature measurements in hospitalized patients. *JAMA*. 2022;328(9):885-886. https://doi.org/10.1001/jama.2022.12290.
22. Yüksel S, Altun-Uğraş G, Altınok N, Demir N. The effect of cuff size on blood pressure measurement in obese surgical patients: a prospective crossover clinical trial. *Florence Nightingale J Nurs*. 2020;28(2):205-212. https://doi.org/10.5152/FNJN.2020.19119.
23. Etienne A, Laroia T, Weigle H, et al. Novel electrodes for reliable EEG recordings on coarse and curly hair. Paper presented at: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada; July 20, 2020. pp. 6151-6154, doi: 10.1109/EMBC44109.2020.9176067.
24. Gershengorn HB, Garland A, Kramer A, Scales DC, Rubenfeld G, Wunsch H. Variation of arterial and central venous catheter use in United States intensive care units. *Anesthesiology*. 2014;120(3):650-664. https://doi.org/10.1097/ALN.0000000000000008.
25. Seastedt KP, Schwab P, O'Brien Z, et al. Global healthcare fairness: we should be sharing more, not less, data. *PLoS Digit Health*. 2022;1(10):e0000102. https://doi.org/10.1371/journal.pdig.0000102.
26. Ficek J, Wang W, Chen H, Dagne G, Daley E. Differential privacy in health research: a scoping review. *J Am Med Inform Assoc*. 2021;28(10):2269-2276. https://doi.org/10.1093/jamia/ocab135.
27. Data management and sharing policy. National Institutes of Health. https://sharing.nih.gov/data-management-and-sharing-policy. Accessed October 11, 2022.
28. Jain AK, Ross A, Pankanti S. Biometrics: a tool for information security. *IEEE Trans Inf Forensics Secur*. 2006;1(2):125-143. https://doi.org/10.1109/TIFS.2006.873653.
29. Prabhakar S, Pankanti S, Jain AK. Biometric recognition: security and privacy concerns. *IEEE Secur Privacy*. 2003;1(2):33-42. https://doi.org/10.1109/MSECP.2003.1193209.
30. Selvam S, Kumar T, Fruttiger M. Retinal vasculature development in health and disease. *Prog Retin Eye Res*. 2018;63:1-19. https://doi.org/10.1016/j.preteyeres.2017.11.001.
31. Farzin H, Abrishami-Moghaddam H, Moin MS. A novel retinal identification system. *EURASIP J Adv Signal Process*. 2008;2008(1):1-10. https://doi.org/10.1155/2008/280635.
32. Fatima J, Syed AM, Usman Akram M. A secure personal identification system based on human retina. Paper presented at: 2013 IEEE Symposium on Industrial Electronics & Applications; Kuching, Malaysia; September 22, 2013. pp. 90-95, doi: 10.1109/ISIEA.2013.6738974.
33. Aleem S, Sheng B, Li P, Yang P, Feng DD. Fast and accurate retinal identification system: using retinal blood vasculature landmarks. *IEEE Trans Ind Inform, Kuching, Malaysia*. 2019;15(7):4099-4110. https://doi.org/10.1109/TII.2018.2881343.
34. Health Insurance Portability and Accountability Act of United States. *US Government Printing Office*. 1996.
35. Disclosures for emergency preparedness — a decision tool: limited data set (LDS). Office for Civil Rights (OCR). https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/limited-data-set/index.html. Accessed January 18, 2023.
36. Revised common rule. Office for Human Research Protections (OHRP). https://www.hhs.gov/ohrp/regulations-and-policy/regulations/finalized-revisions-common-rule/index.html. Accessed November 29, 2022.
37. General Data Protection Regulation (GDPR) compliance guidelines. General Data Protection Regulation. https://gdpr.eu. Accessed January 19, 2023.
38. Batlle JC, Dreyer K, Allen B, et al. Data sharing of imaging in an evolving health care world: report of the ACR data sharing workgroup, part 1: data ethics of privacy, consent, and anonymization. *J Am Coll Radiol*. 2021;18(12):1646-1654. https://doi.org/10.1016/j.jacr.2021.07.014.
39. Kotsenas AL, Balthazar P, Andrews D, Geis JR, Cook TS. Rethinking patient consent in the era of artificial intelligence and big data. *J Am Coll Radiol*. 2021;18(1):180-184. https://doi.org/10.1016/j.jacr.2020.09.022.
40. Ozhan MÖ, Süzer MA, Comak I, et al. Do the patients read the informed consent? *Balk Med J*. 2014;31(2):132-136. https://doi.org/10.5152/balkanmedj.2014.13212.
41. Mikkelsen RB, Gjerris M, Waldemar G, Sandøe P. Broad consent for biobanks is best — provided it is also deep. *BMC Med Ethics*. 2019;20(1):1-12. https://doi.org/10.1186/s12910-019-0414-6.
42. Sheehan M. Can broad consent be informed consent? *Public Health Ethics*. 2011;4(3):226-235. https://doi.org/10.1093/phe/phr020.
43. Polanin JR, Terzian M. A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. *J Clin Epidemiol*. 2019;106:60-69. https://doi.org/10.1016/j.jclinepi.2018.10.006.
44. Kalkman S, Mostert M, Udo-Beauvisage N, Van Delden JJ, Van Thiel GJ. Responsible data sharing in a big data-driven translational research platform: lessons learned. *BMC Med Inform Decis Mak*. 2019;19(1):1-7. https://doi.org/10.1186/s12911-019-1001-y.
45. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) privacy rule. Office for Civil Rights (OCR). https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html. Accessed September 19, 2022.
46. Benitez K, Loukides G, Malin B. Beyond safe harbor: automatic discovery of health information de-identification policy alternatives. *IHI*. 2010;2010:163-172. https://doi.org/10.1145/1882992.1883017.
47. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. *J Healthc Inform Res*. 2021;5(1):1-19. https://doi.org/10.1007/s41666-020-00082-4.
48. McMahan HB, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. *Preprint. Posted online February*. 2016;17. *arXiv [csLG]*, http://arxiv.org/abs/1602.05629.
49. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: concept and applications. *ACM Trans Intell Syst Technol*. 2019;10(2):1-19. https://doi.org/10.1145/3298981.
50. Samarakoon S., Bennis M., Saad W., Debbah M. Federated learning for ultra-reliable low-latency V2V communications. IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates. December. 2018;9. Paper presented at: 2018. pp. 1-7, doi: 10.1109/GLOCOM.2018.8647927.
51. Yu TTL, Lo J, Ma D, et al. Collaborative diabetic retinopathy severity classification of optical coherence tomography data through federated learning. *Invest Ophthalmol Vis Sci*. 2021;62(8):1029-1029.
52. Lu C, Hanif A, Singh P, et al. Federated learning for multicenter collaboration in ophthalmology: improving classification performance in retinopathy of prematurity. *Ophthalmol Retina*. 2022;6(8):657-663. https://doi.org/10.1016/j.oret.2022.02.015.
53. Nguyen TX, Ran AR, Hu X, et al. Federated learning in ocular imaging: current progress and future direction. *Diagnostics*. 2022;12(11):2835. https://doi.org/10.3390/diagnostics12112835.

54. Legislation. Official. Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=EN. Accessed January 19, 2023.
55. Health Insurance Portability and Accountability Act of 1996 (HIPAA). Centers for Disease Control and Prevention. https://www.cdc.gov/phlp/publications/topic/hipaa.html. Accessed August 21, 2022.
56. Chevrier R, Foufi V, Gaudet-Blavignac C, Robert A, Lovis C. Use and understanding of anonymization and de-identification in the biomedical literature: scoping review. *J Med Internet Res*. 2019;21(5):e13484. https://doi.org/10.2196/13484.
57. Naheed T, Usman I, Khan TM, Dar AH, Shafique MF. Intelligent reversible watermarking technique in medical images using GA and PSO. *Optik*. 2014;125(11):2515-2525. https://doi.org/10.1016/j.ijleo.2013.10.124.
58. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Commun. ACM*. 2020;63(11):139-144. https://doi.org/10.48550/ARXIV.1406.2661.
59. Costa P, Galdran A, Meyer MI, et al. End-to-end adversarial retinal image synthesis. *IEEE Trans Med Imaging*. 2018;37(3):781-791. https://doi.org/10.1109/TMI.2017.2759102.
60. Chen Y, Long J, Guo J. RF-GANs: a method to synthesize retinal fundus images based on generative adversarial network. *Comput Intell Neurosci*. 2021;2021:1-17. https://doi.org/10.1155/2021/3812865.
61. Yoo TK, Choi JY, Kim HK. CycleGAN-based deep learning technique for artifact reduction in fundus photography. *Graefes Arch Clin Exp Ophthalmol*. 2020;258(8):1631-1637. https://doi.org/10.1007/s00417-020-04709-5.
62. Chen RJ, Lu MY, Chen TY, Williamson DF, Mahmood F. Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng*. 2021;5(6):493-497. https://doi.org/10.1038/s41551-021-00751-8.
63. Meden B, Emeršič Ž, Štruc V, Peer P. k-Same-Net: k-Anonymity with generative deep neural networks for face deidentification. *Entropy*. 2018;20(1):60. https://doi.org/10.3390/e20010060.
64. Cheng V, Suriyakumar VM, Dullerud N, Joshi S, Ghassemi M. Can you fake it until you make it? Impacts of differentially private synthetic data on downstream classification fairness. Paper presented at: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada; March 3, 2021; pp. 149-160. https://doi.org/10.1145/3442188.3445879.
65. Alaa AM, Van Breugel B, Saveliev ES, van der Schaar M. How faithful is your synthetic data? Sample-level metrics for evaluating and auditing generative models. Paper presented at: International Conference on Machine Learning, Baltimore, Maryland, USA; June 28, 2022. doi: 10.48550/arXiv.2102.08921.
66. Matin A, Mahmud F, Zuhori ST, Sen B. Human iris as a biometric for identity verification. Paper presented at: 2016 2nd International Conference on Electrical, Computer and Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh; December 8, 2016; doi: 10.1109/ICECTE.2016.7879610.