Symposium on Artificial Intelligence - Review Article

Data Privacy in Healthcare: In the Era of Artificial Intelligence

Abstract

Data Privacy has increasingly become a matter of concern in the era of large public digital respositories of data. This is particularly true in healthcare where data can be misused if traced back to patients, and brings with itself a myriad of possibilities. Bring custodians of data, as well as being at the helm of disigning studies and products that can potentially benefit products, healthcare professionals often find themselves unsure about ethical and legal constraints that undelie data sharing. In this review we touch upon the concerns, leal frameworks as well as some common practices in these respects.

Keywords: *Artificial intelligence, data sharing, medical ethics*

Neel Yadav, Saumya Pandey, Amit Gupta, Pankhuri Dudani[1], Somesh Gupta[1], Krithika Rangarajan

Departments of Radiodiagnosis and Interventional Radiology, [1]Dermatology, All India Institute of Medical Sciences, New Delhi, Delhi, India

Background

Digital data in healthcare is a double-edged sword. While on the one hand, digitalization has allowed for a wide variety of advancements, including teleconsultations, easy retrieval and duplication of data for records, and development of applications such as machine learning, it has also allowed for the possibility that the personal medical records of a patient can be accessed by a number of individuals. Data-hungry processes such as machine learning have necessitated the maintenance and sharing of large data repositories and these can have significant consequences on individual patients if this sensitive health information can be linked to the patient and shared with others without the permission of the concerned patient. As compared to data involved in telemedicine, artificial intelligence (AI) applications necessitate the use of much larger volumes of data, which makes its security even more crucial. Also, the data used for AI applications usually has to be uploaded to one or more cloud servers or Graphics Processing Units (GPUs), which adds another level in data processing where potential data compromise can occur.

The growing advancements in AI in healthcare have ramifications in nearly every field, with several of these technologies already undergoing field trials for mass deployment.[1,2] The scope of AI has widened to assist diagnostics and clinical decision-making in many other fields including dermatology, pathology, and genetics.[3-12] Not restricted to diagnostics, recent advances have made AI a game-changer in surgical branches like ophthalmology, robotic surgery, and transplant surgery with a potentially significant impact on the detection, clinical decision-making, and training.[13-15]

AI techniques, however, inherently require a large amount of data.[3] Thus, protecting patient information in a fool-proof way is an essential prerequisite to proceeding with any research related to AI. There is currently no centralized protocol for data encryption and sharing for AI-based research. However, such a protocol is decided on an individual project basis after approval from an ethical angle by the concerned institutional ethics committee. For example, for a specific AI-based study using anonymized patient treatment data retrospectively, informed patient consent may be waived off if deemed appropriate by the ethics committee. There are multiple open-source large medical data repositories (like Kaggle andThe Cancer Imaging Archive (TCIA)) available for public access, which can be used for AI-based studies for the development of standardized protocols and reproducible results. For example, Digital Database for

How to cite this article: Yadav N, Pandey S, Gupta A, Dudani P, Gupta S, Rangarajan K. Data privacy in healthcare: In the era of artificial intelligence. Indian Dermatol Online J 2023;14:788-92.

Received: 15-Jul-2023. **Revised:** 01-Sep-2023. **Accepted:** 08-Sep-2023. **Published:** 27-Oct-2023.

Address for correspondence:
Dr. Krithika Rangarajan,
Room No 160D, Department of Radiology (Institute of Rotary Cancer Hospital), All India Institute of Medical Sciences,
New Delhi - 110 608, Delhi, India.
E-mail: krithikarangarajan86@gmail.com

Access this article online

Website: https://journals.lww.com/idoj

DOI: 10.4103/idoj.idoj_543_23

Quick Response Code:

Screening Mammography (DDSM) and Optimam are freely accessible large databases containing meticulously curated and annotated mammographic images that can be used for training and testing of deep learning algorithms.[16,17] The recent COVID-19 pandemic also underlined the importance of such open-source datasets which helped in the rapid development of algorithms with various applications for helping COVID-19 patients and treating doctors.

Therefore, it is necessary to make hospitals capable of handling large amounts of data, and allow for their protection, while at the same time allowing their utilization in a safe way for the purpose of design and development of such advanced diagnostic techniques.

Current scenario and concerns

As outlined previously, evolving and deploying AI-based health innovations involves dealing with big data sets of information. Big data involves large volumes of data accessed and analyzed at high speed with substantial heterogeneity across individuals and data types.[18,19] Even though such data is necessarily de-identified before sharing with a third-party data aggregator, the risk that new ways of data linkage may be developed, which may end up recognizing the sources, remains real. This has been demonstrated and highlighted by several studies of how newer algorithms are capable of identifying people from public and private data repositories.[1,18,20] For example, a study conducted in 2018 that analyzed data sets from National Health and Nutrition Examination Survey found that an algorithm could be used to re-identify 85.6% of adults and 69.8% of children in a physical activity cohort study, despite the supposed removal of identifiers of protected health information.[21] The problem of identification of patient may be particularly consequential in fields such as dermatology, where even though meta-data of the patient can be delinked, it would be impossible to de-identify photos of the patient, particularly if the lesion is one that is seen on the face or some other easily identifiable part of the body.

Many applications of AI in healthcare involve the consumption of protected health information as well as unprotected data generated by the users themselves, such as health trackers on smart devices, Internet search history and inferences from shopping patterns, or by entities not covered by protective laws such as Health Insurance Portability and Accountability Act (HIPAA).[18,22,23] Removal of the required identifiers in compliance with such laws can be rendered redundant if such data can be re-identified through triangulation with these other identifiable data sets. This is especially true in cases of AI backed by information technology behemoths like Google, Apple, and Meta.[18]

The impact of such a breach in privacy can be consequentialist, deontological, or both. Consequentialist concerns are adverse effects that are measurable and tangible. A few commonplace examples would include facing workplace discrimination if one's medical history is made public, or facing inflated health insurance premiums as a result of additional information accessible due to a breach of privacy.[18] This may prove to be a bigger problem in AI applications involving predictions based on certain inferences drawn from behavioural and lifestyle patterns, as the probability of a certain health event that may occur in the future may have many clinical, social, and occupational ramifications.[24-26] Deontological effects are subjective, unmeasurable, and manifest even if this personal information is not actually misused, or if the person is never made aware of such breach. These concerns stem from the feeling of loss of control that comes with such an invasion of privacy. The mental trauma that such knowledge may cause may be unquantifiable, but nevertheless, undeniable.[18]

There is also the concern arising from data sharing across jurisdictions, such as cross-continent data sharing. A different set of laws may govern personal health information in the area where it is generated vis-à-vis where it is analyzed and used for deep learning algorithms. For example, in Europe, health-related information is protected by the same set of rules that govern all data sharing, European Union's General Data Protection Regulation,[27] whereas there are more health data-specific laws in the United States, like HIPAA.[18] This may leave loopholes in terms of data sharing regulations which may be exploited.[28,29] More importantly, deep learning is inherently sensitive to biases in the underlying data. This means that obtaining training data from multiple continents, across geographies, with a multicultural background is absolutely necessary, in order to ensure these networks are usable across the globe. This necessitates training of the network with such diverse data, thus bringing to the fore the need to find solutions to such legal challenges.

The distribution of data used for training such algorithms can by themselves be a source of bias, raising some ethical concerns. Since most input is reliant on electronic health records, AI applications based on such data are disproportionately perceptive and sensitive to findings seen in the socio-economic class that can afford formal healthcare and health insurance. Socio-economic minorities and marginalized populations that are missing from non-health data such as credit card use or Internet history may also be absent from big health data due to lack of affordability or insurance.[18,30] When trained on such data, for instance, the AI may recommend suboptimal treatment to a community, that is traditionally (as seen in the EMR) marginalized, simply because the network learns that such patients have undergone suboptimal treatment, without understanding the socio-economic basis of such treatment. This is demonstrated amply by the experience of the company Amazon, which discovered that their recruitment system was biased against women, simply because of the gender bias in the data it was trained on.

Indian context

India's rapidly growing economy has also been undergoing a digital transformation. This has brought to the fore the need to develop a safe and reliable infrastructure for the storage and transfer of data.

The importance of such infrastructure is amply demonstrated by the attack on the country's apex and most equipped medical institute in late 2022. The hospital was crippled by a cyber-attack that targeted its services, ranging from patient registration, online appointments, diagnostic report generation, billing, and administrative systems, such as salary disbursal and drug procurement. For over 2 weeks, these services had to be managed manually, leading to long queues and adding to patient waiting times. Even while online services have now resumed, with data restored from a backup server, the personal data of more than 30 million patients and healthcare workers may have been compromised.[31]

Frameworks for Protecting Data Privacy

Acts and Policies

Global initiatives:

1) **General Data Protection Regulation:** It came into effect in May 2018. Though only applicable in the European Union, the norms can be used as a guide by other nations. The GDPR promotes the creation of digital systems that respect users' privacy.[32]
2) **Global Initiative on Ethics of Autonomous and Intelligent Systems:** It is aimed at formulation of a set of standards and principles for Autonomous and Intelligent Systems, to make them secure, ethical, and advantageous to society at large. It also aims to stimulate public participation in the creation of ethical frameworks to increase public understanding of the ethical concerns surrounding this technology.[33]
3) **HIPAA:** It came into effect in 1996**,** it was enacted as a federal law for formulating national standards for handling sensitive patients' health information and prohibiting its disclosure without the patient's consent or knowledge.[34]

National Policies:

1) **Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011:** This is a component of India's Information Technology Act, aimed at preventing unauthorized access to sensitive personal information, including health information**.** businesses and organizations are required to implement suitable security measures in accordance with these requirements.
2) **Ayushman Bharat Digital Mission:** This was launched by the Prime Minister in September 2021. It has multiple components and Ayushman Bharat Health Account (ABHA) number is one of them; it is a unique 14-digit number provided to each individual giving access to the cardholder for threading their health records (only with the informed consent of the patient) across multiple systems and stakeholders. Other components include Healthcare Professionals Registry (HPR), Health Facility Registry (HFR), and Unified Health Interface (UHI).[35]
3) **Digital Personal Data Protection Bill, 2023:** This bill was introduced in 2019, subsequently referred to a joint parliamentary committee, underwent public consultation, and then finally passed in August 2023. The bill governs all digital personal data and says that such data may only be used after informed consent of the person the data belongs to, however, provides several exemptions depending on the purpose and authority using the data. The bill also obligates data fiduciaries to maintain the accuracy of data, keep data secure, and delete data once its purpose has been met. The bill does allow the transfer of personal data outside India, except to certain countries as notified by the central government. The bill introduced the constitution of the Data Protection Board of India, the members of which are appointed for 2 years, and allowed for the imposition of a fine of up to Rs. 250 crores for failure to take adequate security measures to prevent breaches.[36]

AI models-based Privacy Protection

1. **Federated learning:** Since data transfer can result in leaks of data and is particularly problematic in case of transfer across the border, attempts have been made at transferring networks, rather than data. Federated learning is a sort of distributed learning in which several clients work together to jointly develop a model, while maintaining the confidentiality of their input. Here the learning happens separately, each time with a separate set of data, and the model trained ultimately can draw from knowledge across all datasets.[37,38]
2. **Differential privacy:** A mathematical approach known as differential privacy attempts to add randomness or noise to sensitive data to conceal the contributions of each participant.[39]
3. **Cryptographic techniques:** Cryptographic techniques allow for encryption of data prior to training and testing. These cryptographic techniques can be broadly categorized as Secure Multi-Party Computation (SMPC) or Homomorphic Encryption (HE).
4. **Hybrid Privacy-Preserving Techniques:** Combine all the above methods to ensure data security in the biomedical domain.[40]

Conclusion

With the increasing usage of AI in medical subspecialties concerns regarding data sharing, triangulation, and ethical issues are being encountered due to a lack of heterogeneity in data representation. The impact of a data breach can be consequentialist, deontological, or both so there is a need for enforcement of federal laws focusing on health data sharing and usage. AI models like Federated learning, Differential Privacy, and Cryptographic techniques can be used to protect the privacy of patients and safety concerns can be tackled off responsibly.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Murdoch B. Privacy and artificial intelligence: Challenges for protecting health information in a new era. BMC Med Ethics 2021;22:122.
2. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, *et al*. Artificial intelligence in healthcare: Past, present and future. Stroke Vasc Neurol 2017;2:230-43.
3. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018;18:500-10.
4. Orringer DA, Pandian B, Niknafs YS, Hollon TC, Boyle J, Lewis S, *et al*. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. Nat Biomed Eng 2017;1:0027.
5. Island J. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2017); 2017.
6. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, Hermann G, *et al*. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 2015;114:345-50.
7. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, *et al*. Exploratory study to identify radiomics classifiers for Lung Cancer Histology. Front Oncol 2016;6:71.
8. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115-8.
9. Albarqouni S, Baur C, Achilles F, Belagiannis V, Demirci S, Navab N. AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Trans Med Imaging 2016;35:1313-21.
10. Djuric U, Zadeh G, Aldape K, Diamandis P. Precision histology: How deep learning is poised to revitalize histomorphology for personalized cancer care. NPJ Precis Oncol 2017;1:22.
11. Yuan Y, Shi Y, Li C, Kim J, Cai W, Han Z, *et al*. DeepGene: An advanced cancer type classifier based on deep learning and somatic point mutations. BMC Bioinformatics 2016;17(Suppl 17):476.
12. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol 2015;33:831-8.
13. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: Promises and Perils. Ann Surg 2018;268:70-6.
14. Li S, Zhao R, Zou H. Artificial intelligence for diabetic retinopathy. Chin Med J (Engl) 2021;135:253-60.
15. O'Sullivan S, Nevejans N, Allen C, Blyth A, Leonard S, Pagallo U, *et al*. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. Int J Med Robot 2019;15:e1968.
16. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data 2017;4:170177.
17. Halling-Brown MD, Warren LM, Ward D, Lewis E, Mackenzie A, Wallis MG, *et al*. OPTIMAM Mammography Image Database: A Large-Scale Resource of Mammography Images and Clinical Data. Radiol Artif Intell 2021;3:e200103.
18. Price WN 2nd, Cohen IG. Privacy in the age of medical big data. Nat Med 2019;25:37-43.
19. President USEOot, Podesta J. Big data: Seizing opportunities, preserving values: White House, Executive Office of the President; 2014.
20. Hayden EC. Privacy loophole found in genetic databases. Nature News 2013;17.
21. Na L, Yang C, Lo CC, Zhao F, Fukuoka Y, Aswani A. Feasibility of reidentifying individuals in large national physical activity data sets from which protected health information has been removed with use of machine learning. JAMA Netw Open 2018;1:e186040.
22. Digital Personal Data Protection act 2023: Ministry of Electronics and Information Technology, Government of India (no date) Digital Personal Data Protection Act 2023 | Ministry of Electronics and Information Technology, Government of India. Available at: https://www.meity.gov.in/content/digital-personal-data-protection-act-2023.
23. Philibert RA, Terry N, Erwin C, Philibert WJ, Beach SR, Brody GH. Methylation array data can simultaneously identify individuals and convey protected health information: An unrecognized ethical concern. Clin Epigenetics 2014;6:28.
24. Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff (Millwood) 2014;33:1139-47.
25. Liu NT, Holcomb JB, Wade CE, Batchinsky AI, Cancio LC, Darrah MI, *et al*. Development and validation of a machine learning algorithm and hybrid system to predict the need for life-saving interventions in trauma patients. Med Biol Eng Comput 2014;52:193-203.
26. Cohen IG, Lynch HF, Vayena E, Gasser U. Big data, health law, and bioethics: Cambridge University Press; 2018.
27. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019;2:77.
28. Iacobucci G. Patient data were shared with Google on an "inappropriate legal basis," says NHS data guardian. BMJ 2017;357:j2439.
29. Vincent J. Privacy advocates sound the alarm after Google grabs DeepMind UK health app. The Verge. 2018;14.
30. Malanga S, Loe J, Robertson CT, Ramos K. Big data neglects populations most in need of medical and public Health Research and interventions. Big Data, Health Law, and Bioethics (HF Lynch, IG Cohen, and U Gasser eds,), Forthcoming, Arizona Legal Studies Discussion Paper 2016 (16-26).
31. Malhotra S. Cyberattacks hold up India's push for digitisation of health. BMJ 2023;380:263.

32. Regulation GDP. General data protection regulation (GDPR)–official legal text. Gen Data Prot Regul (2016).
33. Chatila R, Havens JC. The IEEE global initiative on ethics of autonomous and intelligent systems. Robotics and Well-Being 2019:11-16.
34. United States. Health Insurance Portability and Accountability Act of 1996. Public Law 104-191. US Statut Large 1996;110:1936-2103.
35. Sharma RS, Rohatgi A, Jain S, Singh D. The Ayushman Bharat Digital Mission (ABDM): making of India's Digital Health Story. CSIT 2023;11:3-9.
36. Digital Personal Data Protection act 2023: Ministry of Electronics and Information Technology, Government of India (no date) Digital Personal Data Protection Act 2023 | Ministry of Electronics and Information Technology, Government of India. Available at: https://www.meity.gov.in/content/digital-personal-data-protection-act-2023.
37. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST). 2019;10:1-19.
38. Moshawrab M, Adda M, Bouzouane A, Ibrahim H, Raad A. Reviewing federated machine learning and its use in diseases prediction. Sensors (Basel) 2023;23:2112.
39. Jain P, Gyanchandani M, Khare N. Differential privacy: its technological prescriptive using big data. Journal of Big Data. 2018;5:1-24.
40. Torkzadehmahani R, Nasirigerdeh R, Blumenthal DB, Kacprowski T, List M, Matschinske J, *et al*. Privacy-Preserving artificial intelligence techniques in Biomedicine. Methods Inf Med 2022;61:e12-27.