International Journal of Research in Health and Allied Sciences
Journal home page: www.ijrhas.com

Official Publication of "Society for Scientific Research and Studies" [Regd.]
ISSN: 2455-7803

Original Research

Artificial intelligence and health records: challenges for protecting health data privacy in current times

[1]Vinit Shashikant Patil, [2]Jeslin V James, [3]Junaid Bin Ahmed, [4]Rafeeque P A, [5]Siraj M M, [6]Azhar Mubarak K, [7]Jawad Ebn Mohammed Abdulla PP

[1]Consultant Oral Pathologist, Kozhikode, Kerala, India;
[2,4]Consultant Physician and Intensivist, Kozhikode, Kerala, India;
[3,5,6,7]Consultant Anesthesiologist and Intensivist, Kozhikode, Kerala, India

ABSTRACT:
Artificial intelligence (AI) aims to mimic human cognitive functions. Recent advances in healthcare artificial intelligence (AI) are occurring rapidly and there is a growing discussion about managing its development. Many AI technologies end up owned and controlled by private entities. The nature of the implementation of AI could mean such corporations, clinics and public bodies will have a greater than typical role in obtaining, utilizing and protecting patient health information. This raises privacy issues relating to implementation and data security. This will require innovation, and there will also be a regulatory component to ensuring that private custodians of data are using cutting edge and safe methods of protecting patient privacy.

Keywords: Artificial intelligence, health care, health data, privacy protection

Received: 21 September, 2021 Accepted: 28 October, 2021

Corresponding author: Vinit Shashikant Patil, Consultant Oral Pathologist, Kozhikode, Kerala, India

This article may be cited as: Patil VS, James JV, Ahmed JB, A Rafeeque P, MM Siraj, K Azhar M, PP Jawad EMA. Artificial intelligence and health records: challenges for protecting health data privacy in current times Int J Res Health Allied Sci 2021; 7(6):16-18.

INTRODUCTION

In the modern era, maintaining the privacy of your personal information has become more challenging than ever. Cyberattacks and social media have resulted in the average person sharing more information than ever before, in ways that they may not be aware of. One area of data privacy that isn't discussed often, however, is health data. The potential of artificial intelligence (AI) to promote better health care has taken the centre stage in modern debates on public health and health policy. Although AI is considered a contemporary innovation, it has been in development for more than a half century. AI research began in the 1950s, when Alan Turing raised the idea that machines could one day think as humans.[1] Then came, in 1959, the first instance of 'machine learning' (ML), where computer scientists created a program capable of solving puzzles on its own.[2] Now, AI promises to lead the next major technological revolution, similar in stature to electricity and the internet. In the field of health care, AI has already led to improvements, particularly in areas such as precision medicine, diagnosis tools, psychological support, and help for the elderly.[3,4] AI technologies generally require large amounts of both personal and non-personal data to function. In health care specifically, AI technologies rely on personal information, including health-related data extracted from medical files or research participants' results.[5] Promoting AI and capturing its benefits for the health care system yet depend, in large part, on procuring a convenient access to this sensitive data.[6] Ensuring that privacy protections are in place appears essential, especially with individuals showing substantial concerns about sharing their data in the medical and clinical context.[7]

AI DEVICES IN HEALTH CARE

AI devices mainly fall into two major categories. The first category includes machine learning (ML) techniques that analyse structured data such as imaging, genetic and EP data. In the medical applications, the ML procedures attempt to cluster patients' traits, or infer the probability of the disease outcomes. The second category includes natural language processing (NLP) methods that extract information from unstructured data such as clinical notes/ medical journals to supplement and enrich structured medical data. The NLP procedures target at turning texts to machine-readable structured data, which can then be analysed by ML techniques. [8-12] As powerful as AI techniques can be, they have to be motivated by clinical problems and be applied to assist clinical practice in the end.

CONCERNS WITH HEALTH DATA ACCESS, USE AND CONTROL

AI have several unique characteristics compared with traditional health technologies. Notably, they can be prone to certain types of errors and biases , and some-times cannot easily or even feasibly be supervised by human medical professionals. The latter is because of the "black box" problem, whereby learning algorithms' methods and "reasoning" used for reaching their conclusions can be partially or entirely opaque to human observers. This opacity may also apply to how health and personal information is used and manipulated if appropriate safeguards are not in place. Notably, in response to this problem, many researchers have been developing interpretable forms of AI that will be easier to integrate into medical care. Because of the unique features of AI, the regulatory systems used for approval and ongoing oversight will also need to be unique. A significant portion of existing technology relating to machine learning and neural networks rests in the hands of large tech corporations. Google, Microsoft, IBM, Apple and other companies are all "preparing, in their own ways, bids on the future of health and on various aspects of the global healthcare industry." Given that there already seen have been examples of corporate abuse of patient health information, it is unsurprising that issues of public trust can arise. For example, a 2018 survey of four thousand American adults found that only 11% were willing to share health data with tech companies, versus 72% with physicians. [8-10] Moreover, only 31% were "somewhat confident" or "confident" in tech companies' data security. [11-13] In some jurisdictions like the United States, this has not stopped hospitals from sharing patient data that is not fully anonymized with companies like Microsoft and IBM. [14,15] A public lack of trust might heighten public scrutiny of or even litigation against commercial implementations of healthcare AI.

HEALTHCARE DATA AND JURISDICTIONS

Healthcare data breaches haven risen in many jurisdictions around the world, including the United States [3,4] Canada, [5,7] and Europe. [3,8] And while they may not be widely used by criminal hackers at this time, AI and other algorithms are contributing to a growing inability to protect health information. A number of recent studies have highlighted how emerging computational strategies can be used to identify individuals in health data repositories managed by public or private institutions. [11] And this is true even if the information has been anonymized and scrubbed of all identifiers. A study by Na et al., for example, 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 data aggregation and removal of protected health information."[13] A 2018 study concluded that data collected by ancestry companies could be used to identify approximately 60% of Americans of European ancestry and that, in the near future, the percentage is likely to increase substantially.[14] Furthermore, a 2019 study successfully used a "linkage attack framework"—that is, an algorithm aimed at re-identifying anonymous health information—that can link online health data to real world people, demonstrating "the vulnerability of existing online health data."[15] And these are just a few examples of the developing approaches that have raised questions about the security of health information framed as being confidential. Indeed, it has been suggested that today's "techniques of re-identification effectively nullify scrubbing and compromise privacy."[16] This reality potentially increases the privacy risks of allowing private AI companies to control patient health information, even in circumstances where "anonymization" occurs. It also raises questions of liability, insurability and other practical issues that differ from instances where state institutions directly control patient data. Considering the variable and complex nature of the legal risk private AI developers and maintainers could take on when dealing with high quantities of patient data, carefully constructed contracts will need to be made delineating the rights and obligations of the parties involved, and liability for the various potential negative outcomes. One way that developers of AI systems can potentially obviate continuing privacy concerns is through the use of generative data. Generative models develop the ability to generate realistic but synthetic patient data with no connection to real individuals. This can enable machine learning without the long term use of real patient data, though it may initially be needed to create the generative model. [17-20]

CONCLUSION

It is an exciting period in the development and implementation of healthcare AI, and patients whose data are used by these AI should benefit significantly, if not greatly, from the health improvements these technologies generate. Nonetheless, the implementation of commercial healthcare AI faces serious privacy challenges. Given personal medical information is among the most private and legally protected forms of data, there are significant concerns about how access, control and use by for-profit parties might change over time with a self-improving AI. An emphasis on patient agency and consent in the development of regulation in this space would reflect the key legal and ethical values of liberal democracies. Also, the right to withdraw data could be clearly communicated and especially made easy to exercise; where feasible, generative data could be used to fill the data gaps created by these agency-driven withdrawals and to avoid de-operationalizing AI systems. Regarding the reidentification issue, there will be a need for new and improved forms of data protection and anonymization. This will require innovation, and there will also be a regulatory component to ensuring that private custodians of data are using cutting edge and safe methods of protecting patient privacy.

REFERENCES

1. Jiang F, Jiang Y,Zhi H, Dong Y,Li H,Ma S,Wang Y,Dong Q,Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230–43
2. Evans M. Hospitals give tech giants access to detailed medical records. Wall Street J. 2020. https://www.wsj.com/articles/hospitals-give-tech-giants-access-to-detailed-medical-records-11579516200. Accessed 15 Mar 2021.
3. HIPAA Journal. Healthcare data breach statistics. https://www.hipaajourn al.com/healthcare-data-breach-statistics/. Accessed 19 Jul 2021.
4. Verizon Enterprise. 2020 Data breach investigations report. 2020. https:// enterprise.verizon.com/resources/reports/2020-data-breach-investigat ions-report.pdf. Accessed 19 Jul 2021.
5. CBC News. LifeLabs pays ransom after cyberattack exposes information of 15 million customers in B.C. and Ontario. 2019. https://www.cbc.ca/ news/canada/british-columbia/lifelabs-cyberattack-15-million-1.5399577. Accessed 15 Mar 2021.
6. Hunter J. Privacy breach in B.C. health ministry led to freeze on medical research data. The Globe and Mail. 2016. https://www.theglobeandmail. com/news/british-columbia/privacy-breach-in-bc-health-ministry-led- to-freeze-on-medical-research-data/article29767108/. Accessed 15 Mar 2021.
7. Solomon H. Cost of Canadian data breaches continues to rise, says study. IT World Canada. 2018. https://www.itworldcanada.com/article/cost-of-canadian-data-breaches-continues-to-rise-says-study/406976. Accessed 15 Mar 2021.
8. European Union Agency for Cybersecurity. From January 2019 to April 2020 Dta breach ENISA Threat Landscape. 2020. https://www.enisa. europa.eu/publications/enisa-threat-landscape-2020-data-breach/at_ download/fullReport. Accessed 19 Jul 2021.
9. University of California–Berkeley. Artificial intelligence advances threaten privacy of health data. EurekAlert! 2019. https://www.eurekalert.org/pub_ releases/2019-01/uoc--aia010319.php. Accessed 15 Mar 2021.
10. Kolata G. Your data were ‘anonymized’? These scientists can still identify you. New York Times. 2019. https://www.nytimes.com/2019/07/23/ health/data-privacy-protection.html. Accessed 15 Mar 2021.
11. Hayden EC. Privacy loophole found in genetic databases. Nature News. 2013. https://www.nature.com/news/privacy-loophole-found-in-genet ic-databases-1.12237. Accessed 15 Mar 2021.
12. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science. 2013;339(6117):321–4.
13. Na L, Yang C, Lo CC, Zhao F, Fukuoka Y, Aswani A. Feasibility of reidentify- ing 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(8):e186040.
14. Erlich Y, Shor T, Pe’er I, Carmi S. Identity inference of genomic data using long-range familial searches. Science. 2018;362(6415):690–4.
15. Ji S, Gu Q, Weng H, Liu Q, Zhou P, He Q, Beyah R, Wang T. De-health: all your online health information are belong to us. arXiv preprint. 2019. https://arxiv.org/abs/1902.00717.
16. Lubarsky B. Re-identification of “anonymized data”. UCLA L. REV. 1701;1754(2010). https://georgetownlawtechreview.org/wp-content/ uploads/2017/04/Lubarsky-1-GEO.-L.-TECH.-REV.-202.pdf.
17. Yoon J, Drumright LN, Van Der Schaar M. Anonymization through data synthesis using generative adversarial networks (ads-gan). IEEE J Biomed Health Inform. 2020;24(8):2378–88.
18. Baowaly MK, Lin CC, Liu CL, Chen KT. Synthesizing electronic health records using improved generative adversarial networks. J Am Med Inform Assoc. 2019;26(3):228–41.
19. Dietterich T. Overfitting and undercomputing in machine learning. ACM Comput Surv. 1995;27(3):326–7.
20. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol. 2017;7(4):351–67.