Privacy versus artificial intelligence in medicine

Taryn J Rohringer (BMSc)[1]; Akshay Budhkar (BASc)[2,5]; Frank Rudzicz (PhD)[2,3,4,5]

[1]Faculty of Medicine, University of Toronto, Medical Sciences Building, 1 King's College Circle, Toronto, ON, Canada, M5S 1A8.
[2]Department of Computer Science, University of Toronto, 27 King's College Circle, Toronto, ON, Canada, M5S 3H7.
[3]Li Ka Shing Knowledge Institute, St Michael's Hospital, 209 Victoria St, Toronto, ON, Canada M5B 1T8.
[4]Surgical Safety Technologies, 250 Yonge St, Toronto, Ontario, Canada, M5G 1B1.
[5]Vector Institute for Artificial Intelligence, 661 University, Suite 710, Toronto, ON, Canada, M5G 1M1.

Abstract

As artificial intelligence is increasingly integrated into clinical practice, various crucial challenges will persist, especially with regards to data acquisition, reporting, and potential re-identification of patient data. This paper outlines these challenges and suggests some open questions and potential solutions. Given recent news of companies overstepping their bounds in the pursuit of patient data to train their systems, and new regulations around privacy of those data, this discussion is especially pertinent. Here, we suggest that a common good can be achieved in which data can be kept private while also useful for artificial intelligence in the practice of medicine.

Introduction

Recent advances in artificial intelligence (AI) have accelerated their use in healthcare, from remote monitoring and wearables to clinical decision support.[1] Specifically, the subset of AI known as machine learning (ML) has been driving much of this change. In ML, statistical models improve from experience without being explicitly programmed through algorithms that process data in the task of interest. However, if these technologies are to increasingly record and measure the actions of (and interaction between) doctors and patients, it is important to determine how the ethical constraints of confidentiality and privacy apply, and whether they are at odds with ML research.

In May 2018, Europe's new General Data Protection Regulation (GDPR) changed how ML fits into healthcare by requiring companies to obtain consumer consent before using their data.[2] GDPR also requires companies to explain their automated decisions meaningfully, if requested. This will encourage interpretable ML models and technologies that provide post-decision explanations of black-box models on demand. GDPR amplifies the importance of hybrid solutions that augment humans in healthcare, enabling solutions that neither the human nor the computer could achieve on their own.[3] The same month that GDPR came into effect, Canada issued new guidance for the Personal Information Protection and Electronic Documents Act (PIPEDA) detailing guidelines for obtaining meaningful consent, and against "[s]urveillance by an organization through audio or video functionality of the individual's own device."[4] More specifically, subsection 5(3) of PIPEDA states that "An organization may collect, use or disclose personal information only for purposes that a reasonable person would consider are appropriate in the circumstances." Given that consensus has not been widely achieved with regards to the details of surveillance of this type (e.g., what risks to personal information are necessary, given the technology, to achieve some perceived benefit to the person involved), it is not yet clear what a "reasonable person would consider appropriate."

As AI is increasingly integrated into clinical practice, various challenges will persist (e.g. data acquisition, reporting, and re-identification) and these emphasize a potential struggle between patient privacy and the promise of these systems.

Challenges to Data Acquisition

Personal health data is extremely valuable; for example, the $6 billion acquisition of Medco Containment Services by Merck was substantially based on how Medco's data could influence physician prescriptions.[5] The value of personal health data is only inflating as commercial companies require increasingly large data sets to train and refine their AI systems.[1] Health care institutions are also motivated to collect personal health information, as they may reap economic benefits from increased efficiency and cost-effectiveness; for example, a deep neural network has recently been developed that can detect referable diabetic retinopathy with high sensitivity and specificity, freeing up clinic time and space.[1] The physicians working in healthcare institutions may feel pressure to deliver this data to their employers, and may additionally be incentivized to collect data.

Patients have the right to "informational privacy," defined as controlled access to their personal information.[5] The ethics and legality of collecting these data hinge on the patient's informed, voluntary consent. If the patient is presented with a choice between AI-enabled systems and their less effective alternatives (if they even exist), they may treat consent with the same ambivalence as they do end-user license agreements in traditional software.[6] For

Corresponding Author:
Frank Rudzicz
frank@cs.toronto.edu

example, image-processing AI has been shown to detect malignant skin lesions with greater sensitivity and specificity than 20 of 21 human dermatologists;[7] consider extension of this technology via consumer hardware that allows for daily tracking at home. In situations where the alternatives are so disparate, the question of whether consent in such a case is truly informed or voluntary is open for debate. Consent processes to data acquisition that follow the letter of the law but not its purpose may significantly jeopardize patient privacy and autonomy over their own data, and further threaten patient trust in the healthcare system.

Challenges to Reporting and Disclosure

Though confidentiality is a core constituent of the physician-patient relationship, there are clear situations in which a physician can and must break this confidentiality. The College of Physicians and Surgeons of Ontario (CPSO), for example, requires that physicians report any evidence of child abuse/neglect, impaired driving, sexual abuse, and misconduct in long-term care homes.[8] If AI systems, including those in smart homes or wearable devices, are able to detect these occurrences, must they record them and do they have a similar responsibility to report them? Such reporting can be in the interest of public safety, but potential loss of patient trust must be carefully weighed. These clinical judgements have not yet been attempted by AI systems. The duty-to-warn statutes vary across jurisdictions, while AI systems may be accessed virtually from anywhere. How should decision-making abilities integrate duty-to-warn, if the parameters differ by region but the systems do not?

Perhaps an even more ambiguous area for AI to navigate is 'permissive reporting' – contexts in which there is no legal imperative to break confidentiality, but in which there is a legal allowance to do so. The decision of whether action is pursued or not has traditionally come down to physician judgement. This allowance is made in instances where there is an imminent risk of serious harm or death to an individual or groups of individuals.[8] Beyond the programming challenge that is associated with AI detecting these heterogeneous and ambiguous situations, there is the additional question of whether to notify a third party of their occurrence. Additionally, these situations will no longer be limited to voluntary revelations within the doctor's office, but potentially under the assumption of privacy while at home or work. How will AI be engineered to judge whether to report a perceived risk of imminent serious harm? Will there be any requirement to report at all?

To answer these questions, a re-evaluation of physician behaviour in these circumstances may be required. Consider a scenario where the physician suspects their patient has experienced domestic abuse. The CPSO guidelines for permissive reporting do not apply when the risk of harm is neither serious nor imminent, as in suspected domestic abuse. Though confidentiality cannot be violated in this example, physicians can initiate a discussion with their patient about how they themselves can take action and navigate the situation, through providing them additional resources. Should the patient not choose to take action, the physician is still able to monitor the patient. If AI can detect but not report said scenarios, what benefit can it offer? Should bounds exist on their positive predictive values?

Challenges to Re-identification and Differential Privacy

Despite efforts to de-identify sensitive data, unique patterns that remain can be used to fully re-identify patients, especially in the presence of complementary data, which risks identity theft.[9] "Differential privacy" offers a robust algorithmic solution by injecting noise into the data, or at various points in the training process, without significantly degrading model performance.[10] If appropriately used, the resulting models are theoretically secure and provably unable to be reverse-engineered.[11] Though promising as research, this has not yet been implemented in healthcare settings.

While GDPR applies to any data that can be re-identified, it does not apply to synthetic data that has statistical properties identical to such data, such as simulated data points sampled randomly from a statistical distribution. Methods that synthetically mimic the general distribution of data may allow researchers to globally access models without breaching privacy.

Conclusion

Recently, the partnership between the Royal Free NHS Foundation Trust and Google's subsidiary DeepMind was ruled illegal by the UK's data protection regulator and terminated, due to insufficient informed consent with regards to secondary use of private data. This example demonstrates an apparent struggle between regulators and industry with regards to the drilling for the private medical data that fuels modern algorithms.

Modern uses of AI may serve to shift the responsibility for care and monitoring from healthcare professionals to patients themselves. For example, advanced AI is currently being used on consumer-grade 'smart watches' to intensively monitor patients with chronic obstructive pulmonary disease in a way not possible in situ but which also requires that patients care for the equipment and follow protocol themselves.[12] What kinds of patients are favored in this new dynamic, and might patients not well-equipped to manage and maintain their own data receive substandard care? How do we establish the degree to which consent is informed and non-coerced? Answers to these questions, and others, will require legal guidelines that supersede the piecemeal guidelines that currently fragment different jurisdictions. Increased emphasis on international technical standards may be a crucial step towards patient safety, including standards for data management.[13]

Unfortunately, there may always be a tradeoff between achieving the maximal accuracy in ML models and preserving complete patient privacy. One solution may be to pursue methods that can filter out private demographic information while only compromising model performance minimally. For example, dementia can be detected by ML classifiers trained on linguistic features of speech.[14] While advanced age may strongly condition the likelihood of developing dementia, one's precise age may not always be permissibly recorded. In response, modern neural networks have been developed to isolate and filter out confounding information of this type, while only modestly reducing accuracy. [14] In this way, we may reframe an apparent tug-of-war over private patient data so that all parties are pulling in the same direction.

Acknowledgements

This research is partially funded by the Vector Institute for Artificial Intelligence and the Canadian Institute for Advanced Research.

References

1. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402–10. doi:10.1001/jama.2016.17216
2. European Commission. 2018 reform of EU data protection rules. European Commision. https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en. Published 24 January 2018. Accessed 12 July 2018.
3. Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Informatics. 2016;3(2):119-131. doi:10.1007/s40708-016-0042-6.
4. Office of the Privacy Commissioner of Canada. Privacy Commissioner issues new guidance to help address consent challenges in the digital age. https://www.priv.gc.ca/en/opc-news/news-and-announcements/2018/nr-c_180524/ Published 24 May 2018. Accessed 20 December 2018.
5. Committee on Regional Health Data Networks. Confidentiality and Privacy of Personal Data. In: Donaldson MS, Lohr KN, eds. Health Data in the Information Age. Washington DC: National Academies Press; 1994. doi:10.17226/2312.
6. Bakos Y, Marotta-Wurgler F, Trossen DR. Does Anyone Read the Fine Print? Consumer Attention to Standard Form Contracts. Ssrn Published Online First: 2009. doi:10.2139/ssrn.1443256
7. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115. doi:10.1038/nature21056
8. The College of Physicians and Surgeons of Ontario. Mandatory and Permissive Reporting. The College of Physicians and Surgeons of Ontario. https://www.cpso.on.ca/Policies-Publications/Policy/Mandatory-and-Permissive-Reporting. Published June 21, 2018. Accessed July 12, 2018.
9. El Emam K, Jonker E, Arbuckle L, Malin B. A Systematic Review of Re-Identification Attacks on Health Data. PLoS One. 2011;6(12):1-12. doi:10.1371/journal.pone.0028071.
10. Abadi M, Chu A, Goodfellow I, et al. Deep Learning with Differential Privacy. Paper presented at the ACM SIGSAC Conference on Computer and Communications Security; October 2016; Vienna, Austria/ Pages 308-318, Vienna, Austria. doi:10.1145/2976749.2978318.
11. Beaulieu-Jones BK, Wu ZS, Williams C, et al. Privacy-preserving generative deep neural networks support clinical data sharing. bioRxiv Prepr Published Online First: 2017. doi:http://dx.doi.org/10.1101/159756
12. Wu R, Liaqat D, De Lara E, Son T, Rudzicz F, Alshaer H, Abed P, Gershon A. Feasibility of using a smartwatch to intensively monitor patients with COPD. JMIR Mhealth Uhealth: 2018; 6(6). doi: 10.2196/10046
13. Rudzicz F, Paprica PA, Janczarski M. Towards international standards for evaluating machine learning. In Proceedings of SafeAI at AAAI19.
14. Zhu Z, Novikova J, Rudzicz F. Isolating effects of age with fair representation learning when assessing dementia. arXiv:1807.07217v3