International Journal of Clinical Medical Research

Artificial intelligence in healthcare

Jakub Dominik[1]

Correspondence:
Jakub Dominik
Na Homolce Hospital, Roentgenova 37, 150 00 Praha 5, Czechia
Email: Jakubdominik@finmail.com

[1]Na Homolce Hospital, Roentgenova 37, 150 00 Praha 5, Czechia

Volume number 3
Issue number 2
Pages 22-23

Abstract

In recent years, enhanced artificial intelligence algorithms and more access to training data have enabled artificial intelligence to augment or supplant certain functions of physicians. Nonetheless, the interest of diverse stakeholders in the application of artificial intelligence in medicine has not resulted in extensive acceptance. Numerous experts have indicated that a primary cause for the limited adoption is the lack of openness surrounding certain artificial intelligence algorithms, particularly black-box algorithms. Clinical medicine, particularly evidence-based practice, depends on transparency in decision-making. If there is no medically explicable artificial intelligence and the physician cannot adequately elucidate the decision-making process, the patient's trust in them will diminish. To resolve the transparency concern associated with specific artificial intelligence models, explainable artificial intelligence has arisen.

Keywords: Artificial intelligence; healthcare; algorithms; Clinical medicine

Artificial intelligence in healthcare

Marzyeh Ghassemi and colleagues have contended that existing explainable artificial intelligence applications are flawed and offer merely a partial elucidation of the mechanisms underlying artificial intelligence algorithms. [1] They have urged stakeholders to abandon the demand for explainability and to pursue alternative approaches, e.g. validation, to foster trust and confidence in black-box models. [2] Their criticism of certain explainable frameworks, e.g. post-hoc explainers, holds some weight. These explainers primarily approximate the fundamental machine learning techniques to elucidate the decision-making process. Nonetheless, given the constraints of specific explainable artificial intelligence methodologies, the assertion to limit explainable artificial intelligence in favor of alternative validation methods, e.g. randomized controlled trials, is fallacious. [3]

Models or systems with decisions that lack clear interpretability can be challenging to accept, particularly in domains e.g. medicine. [4] Dependence on the rationale of black box models contravenes medical ethics. [4] The opaque nature of medical practice obstructs practitioners from evaluating the quality of model inputs and parameters. If clinicians fail to comprehend the decision-making process, they may infringe upon patients' rights to informed consent and autonomy. [5] When clinicians are unable to understand the derivation of results, they are unable to communicate effectively with the patient, so compromising the patient's autonomy and capacity for informed consent. There have been a growing number of instances when high-performing black-box models have been identified as utilizing erroneous or confounding variables to attain their outcomes. [6] A deep learning model determined that asthma patients are at low risk of pneumonia-related mortality, based on a training dataset that included asthma patients receiving active clinical care. [6] A deep learning model designed to screen x-rays for pneumonia utilized complicating variables, e.g. the scanner's position, to identify the condition. [7] A third example included a deep learning model designed to differentiate high-risk patients from lower-risk patients using x-rays, which utilized hardware-related metadata for risk prediction. [8] These instances indicate that dependence on the precision of the models is inadequate. Supplementary trust-enhancing frameworks, e.g. explainable artificial intelligence, are necessary. [9]

Despite the increasing criticism of explainable artificial intelligence methods in recent years, there appears to be remarkably little examination of the underlying cause for the necessity of explainable artificial intelligence: deep learning models. These models lack an explicit declarative knowledge representation, complicating the formulation of an explanatory narrative.

Numerous high-performing deep learning models possess millions or even billions of parameters that are identified solely by their positions within a complicated network, lacking human-interpretable labels, resulting in a black-box scenario. [9] Furthermore, numerous deep learning models that perform effectively on training datasets often exhibit poor performance on independent datasets. Moreover, deep learning techniques necessitate substantial data for training in both interpolation and extrapolation. The challenges associated with deep learning models remain unresolved and continue to affect different applications, including medicine.

Critics of explainable artificial intelligence contend that validity measures should take precedence over explainability frameworks.6. [10] The justification is that, at present, several pharmaceuticals and medical devices utilize validation techniques (e.g. randomized controlled trials) to demonstrate efficacy; therefore, artificial intelligence-enabled medical equipment or software should adhere to the same standards. Nonetheless, we contend that this argument is unsuitable. The efficacy of artificial intelligence systems is typically evaluated based on predictive accuracy metrics. [11] Despite optimal attempts, artificial intelligence systems are improbable to attain flawless accuracy owing to several sources of inaccuracy. [12] Theoretical 100% accuracy does not ensure that the artificial intelligence system is devoid of biases, particularly when trained on heterogeneous and complex data, as is common in medicine.

Disregarding or limiting explainable artificial intelligence hampers the integration of artificial intelligence in medicine, as few alternatives adequately address accountability, trust, and regulatory issues while fostering confidence and transparency in the technology. Employing explainable frameworks may facilitate the alignment of model performance with the aims of therapeutic recommendations. [12] Consequently, facilitating enhanced integration of artificial intelligence models in clinical practice. Transparent algorithms or explanatory methodologies can mitigate the risks associated with the implementation of artificial intelligence systems for clinical practitioners. [12] An increasing number of instances demonstrate how explainable frameworks in several medical specialties promote transparency and insight. [13] These case studies can inform the integration of explainable artificial intelligence with medical artificial intelligence systems. This integration facilitates a secondary level of explainability and numerous advantages, e.g. enhanced interpretability, increased understanding for clinicians promoting evidence-based practice, and superior therapeutic results.

References

1. Yoon, C.H., Torrance, R., and Scheinerman, N., *Machine learning in medicine: should the pursuit of enhanced interpretability be abandoned?* Journal of Medical Ethics, 2022. **48**(9): p. 581-585.
2. Sadeghi, Z., Alizadehsani, R., Cifci, M.A., et al., *A review of Explainable Artificial Intelligence in healthcare.* Computers and Electrical Engineering, 2024. **118**: p. 109370.
3. Ghassemi, M., Oakden-Rayner, L., and Beam, A.L., *The false hope of current approaches to explainable artificial intelligence in health care.* The Lancet Digital Health, 2021. **3**(11): p. e745-e750.
4. Amann, J., Blasimme, A., Vayena, E., et al., *Explainability for artificial intelligence in healthcare: a multidisciplinary perspective.* BMC Medical Informatics and Decision Making, 2020. **20**(1): p. 310.
5. Kundu, S., *AI in medicine must be explainable.* Nature Medicine, 2021. **27**(8): p. 1328-1328.
6. Reddy, S., Fox, J., and Purohit, M.P., *Artificial intelligence-enabled healthcare delivery.* Journal of the Royal Society of Medicine, 2019. **112**(1): p. 22-28.
7. Hassija, V., Chamola, V., Mahapatra, A., et al., *Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence.* Cognitive Computation, 2024. **16**(1): p. 45-74.
8. Vellido, A., *Societal Issues Concerning the Application of Artificial Intelligence in Medicine.* Kidney Diseases, 2018. **5**(1): p. 11-17.
9. Kelly, C.J., Karthikesalingam, A., Suleyman, M., et al., *Key challenges for delivering clinical impact with artificial intelligence.* BMC Medicine, 2019. **17**(1): p. 195.
10. Grzybowski, A., Jin, K., and Wu, H., *Challenges of artificial intelligence in medicine and dermatology.* Clinics in Dermatology, 2024. **42**(3): p. 210-215.
11. Ong, J.C.L., Seng, B.J.J., Law, J.Z.F., et al., *Artificial intelligence, ChatGPT, and other large language models for social determinants of health: Current state and future directions.* Cell Reports Medicine, 2024. **5**(1).
12. Cutillo, C.M., Sharma, K.R., Foschini, L., et al., *Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency.* npj Digital Medicine, 2020. **3**(1): p. 47.
13. Desai, A.N., *Artificial Intelligence: Promise, Pitfalls, and Perspective.* JAMA, 2020. **323**(24): p. 2448-2449.