

Application of Artificial Intelligence in the Health Care Safety Context: Opportunities and Challenges

American Journal of Medical Quality
1–8

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1062860619878515

ajmq.sagepub.com



Samer Ellahham, MD^{1,2}, Nour Ellahham¹,
and Mecit Can Emre Simsekler, PhD³

Abstract

There is a growing awareness that artificial intelligence (AI) has been used in the analysis of complicated and big data to provide outputs without human input in various health care contexts, such as bioinformatics, genomics, and image analysis. Although this technology can provide opportunities in diagnosis and treatment processes, there still may be challenges and pitfalls related to various safety concerns. To shed light on such opportunities and challenges, this article reviews AI in health care along with its implication for safety. To provide safer technology through AI, this study shows that safe design, safety reserves, safe fail, and procedural safeguards are key strategies, whereas cost, risk, and uncertainty should be identified for all potential technical systems. It is also suggested that clear guidance and protocols should be identified and shared with all stakeholders to develop and adopt safer AI applications in the health care context.

Keywords

artificial intelligence, machine learning, safety, patient safety, quality

Artificial intelligence (AI) is revolutionizing health care. The primary aim of AI applications in health care is to analyze links between prevention or treatment approaches and patient outcomes. AI applications can save cost and time for the diagnosis and management of disease states, thus making health care more effective and efficient. AI enables fast and comprehensive analysis of huge data sets to effectively enable decision making with speed and accuracy. AI is largely described to be of 2 types: virtual and physical. Virtual AI includes informatics from deep learning applications, such as electronic health records (EHRs) and image processing, to assist physicians with diagnosis and management of disease states. Physical AI includes mechanical advances, such as robotics in surgery and physical rehabilitation.¹ Algorithms have been developed to train data sets for statistical applications to enable data processing with accuracy. These principles underlie machine learning (ML), which enables computers to make successful predictions using past experiences.^{2,3} Although both AI and ML can provide these advances, such technology also may raise safety concerns, which may cause serious issues for both patients and all other health care stakeholders. Data privacy and security is one such concern because most AI applications rely on a huge volume of data to make better decisions. Furthermore, ML systems

usually use data—often personal and sensitive data—to learn from and improve themselves. This makes them more at risk for serious issues such as identity theft and data breach. AI also may be associated with low prediction accuracy, which raises safety concerns. For instance, convolutional neural networks (CNNs) are trained and validated using data sets in clinical settings, which may not translate well to a larger population: for example, in one particular study, the surveillance of skin lesions for detection of skin cancer because these may be more diverse in the general population.⁴ Therefore, such an AI system may make false or inaccurate predictions. To address potential issues, the research team presents an overview of implications of AI and ML for the health care safety context. Furthermore, the team discusses the opportunities and challenges for the development and safe deployment of AI in health care.

¹Cleveland Clinic Abu Dhabi, Al Falah St, Abu Dhabi, UAE

²Cleveland Clinic, Cleveland, OH

³Khalifa University of Science and Technology, Abu Dhabi, UAE

Corresponding Author:

Samer Ellahham, Cleveland Clinic Abu Dhabi, PO Box 112412,
Al Falah St, Abu Dhabi, UAE.

Email: ellahas@clevelandclinicabudhabi.ae

Literature Review

AI in Health Care

In health care, AI is defined as the mimicking of human cognitive functions by computers.⁵ AI has been inspired by the functioning of biological neurons and includes the basics of sensing, recognition, and object recognition to enable machines to perform as good as or even better than humans. However, with the inherent lack of articulation and generation of insights, AI cannot replace physicians in health care.⁶ With no universally applicable rules in health care, AI must be supplemented with physician judgment in many instances. An extensive correlation of history and clinical findings is needed for the diagnosis or monitoring of any disease state. The physician–patient relationship is guided by associative and lateral thinking and can influence management decisions. Moreover, the influence of several factors (eg, psychosocial, emotional) on disease outcomes falls outside the scope of AI.

Machines can be more precise, reliable, and comprehensive and have relatively lower risk of bias; however, they still lack the elements of trust and empathy.⁷ There is a growing concern that AI systems learn by doing and, with repeat training, can outperform humans. AI holds a promising future in health care but only when used with diligence for the right purposes.

AI and Safety

Safety in health care implies the reduction or minimization of risks and uncertainty of harmful events.^{8,9} The dimensions of safety are changing with the adoption of AI in health care. AI and ML, with a low likelihood of both expected and unexpected harms, have been applied to reinforce safety. Risk minimization is thus key to AI-based applications.

ML applications are largely classified as type A (eg, medical diagnosis) and type B applications (eg, speech transcription systems), depending on safety and risk minimization. Although safety is of paramount importance in type A applications, risk minimization is the focus in type B applications. Epistemic uncertainty, the scientific uncertainty in the model, is much less in type B applications. Errors are less common in type B applications and, hence, safety is of lesser relevance in type B applications.⁸ Besides risks and safety, the costs of unwanted outcomes also are used as a parameter of assessing outcomes as being harmful.

ML has gained importance in the prevention, diagnosis, and management of various disease conditions. Safety of these novel strategies has been described in abstract parameters according to the disease area and expected outcomes. In a study conducted by Swaminathan et al,¹⁰ an ML algorithm was developed to predict flare-ups and

provide at-home decision support for patients with chronic obstructive pulmonary disease; it successfully triaged patients with high accuracy and in favor of patient safety in a validation study (n = 101). The algorithm never undertriaged a patient who should be sent to a doctor and undertriaged patients for emergency room visits in <14% cases. In comparison, for the same decisions, physicians undertriaged patients in 22% and 30% of the cases. This model was trained using physician-labeled data sets. Model performance was validated by comparing its decisions with consensus decisions of a panel of physicians using an out-of-sample representative patient set.¹⁰

Opportunities and Challenges of AI in the Health Care Safety Context

AI plays an important role in augmenting knowledge and improving outcomes in health care. AI has widespread applications for the prediction and diagnosis of disease, handling of large quantities of data and synthesis of insights, and maximizing efficiency and outcomes in medical management of disease states.¹¹ Benefits of AI have been described for various diseases and outcomes; for example, in the prediction of sepsis in intensive care and diagnosis and classification of malignant lesions, retinal diseases, and pneumonia, among others.¹²⁻¹⁶ Principles of AI have been deployed in precision medicine to build precise, safe, and targeted therapies.¹⁷

There are enormous benefits to utilizing AI in health care. AI can be a great assist in routine clinical practice and research. Quick and easy access to information, increased outreach, and reduction of errors in diagnosis and treatment of disease are the key benefits of AI. Predictive diagnosis, precision medicine, and delivery of targeted therapies are some key areas in which AI has introduced significant improvements. Virtual follow-up and consultations provide effectiveness in terms of costs and time. For instance, AI-based telemedicine applications provide quality of care and reduce wait times and chances of infection acquired during hospital visits to patients. This ultimately results in high patient satisfaction during treatment.^{18,19}

AI has several applications in diagnosis and decision support. AI enables decision makers to access the right and up-to-date information to help make better decisions in real time.²⁰ Application of AI has brought about an evolutionary change in radiological diagnosis by improving the value and accuracy of image analysis.^{21,22} Designs based on deep learning have enabled digital image analysis for the early detection of breast pathologies with precision.²³ In another example, an ML software library has been trained to detect changes in Parkinson's disease by DaTscan image analysis. This library can be a useful adjunct to clinical diagnosis.²⁴

Table 1. Safety Issues for Artificial Intelligence (AI) in Health Care.

Safety Issue	Elements of Hazard	Key Steps to Mitigation
Distributional shift	Out-of-sample predictions	Training of AI systems with large and diverse data sets
Quality of data sets	Poor definition of outcomes Nonrepresentative data sets	Build more inclusive training algorithms using balanced data sets, correctly labeled for outcomes of interest
Oblivious impact	High rates of false-positive and false-negative outcomes	Include outliers in training data sets Enable systems to adjust for confidence levels
Confidence of prediction	Uncertainty of predictions Automation complacency	Sustained and repeated use of AI algorithms Transparent and easily accessible AI algorithms
Unexpected behaviors	Calibration drifts	Design and train systems to learn and unlearn and have more predictable behaviors
Privacy and anonymity	Identification of patient data	Define layers of security and rules for data privacy Anonymize data before sharing
Ethics and regulations	Poor ethical standards and regulatory control for development and deployment of AI	Build regulatory reforms to support integration of AI in health care

AI also finds an application in patient triage. Wearable devices have been designed to enable remote monitoring and analysis of vital signs and consciousness index. Algorithms have been trained to classify disease conditions based on severity. Models have been developed to triage patients and predict survival in the pre-hospital environment.²⁵ Electronic triage (e-triage) finds utility in emergency departments.²⁶ In a multi-center, retrospective, cross-sectional study of 172 726 emergency visits in urban and community emergency departments, e-triage was more accurate than the Emergency Severity Index (ESI) for triage. When compared to ESI, e-triage identified more than 10% (14 326 patients) of ESI level 3 patients who needed critical care or an emergency procedure and were up-triaged (6.2% e-triage vs 1.7% ESI) and needed hospitalization (45.4% e-triage vs 18.9% ESI).

However, despite such benefits, there are several limitations to the successful adoption and seamless implementation of AI. There is a paucity of evidence-based studies on the efficacy and safety of AI in health care.²⁷ For instance, clinicians often display resistance and reluctance to the adoption of AI in medical practice. Moreover, there are privacy, anonymity, ethical, and medicolegal concerns for the adoption of AI-enabled systems in medical practices and research.^{28,29} The machine performs the task according to the specific instructions given; however, it gradually learns to be flexible and to work in various situations through using newly acquired data. This triggers an increasing demand for data collection and data sharing of private and public information at the expense of the user's privacy.²⁸

Medical ethics, such as an individual's right to privacy, may be threatened by AI and big data features

because of the collection and storage of data from various sources. Furthermore, security and safety of vital information may be put at risk by misuse of medicolegal algorithms by hackers for developing autonomous techniques. To prevent such issues, AI research should comply with norms and ethics.²⁸

Errors may be inherent in AI algorithms, which may lead to unfair and adverse outcomes based on race and socioeconomic status. Furthermore, contextual interpretation and unclear ethical standards result in huge issues for the coding of AI systems.²⁸ Any direct or indirect impact of AI on patients or physicians should be minimized using preventive and precautionary safeguards. Implementation of AI and ML without thorough validation can harm patients and challenge clinicians' trust in technology.

Impact of AI on Quality Care

The impact of AI-driven applications on health care is challenged with various limitations. Key issues around the safety of AI in health care and steps to extenuate the same are listed in Table 1. These issues are likely to arise at various stages of deployment of AI.

Whereas Table 1 provides some elements of hazards and mitigation strategies, Figure 1 shows the linkage of AI-based applications and safety issues. The following sections will discuss each safety issue further.

Distributional Shift. ML can make out-of-sample predictions and raise safety concerns. This may occur because of changes in disease patterns and characteristics, trained and performance data sets, and application to varied populations.

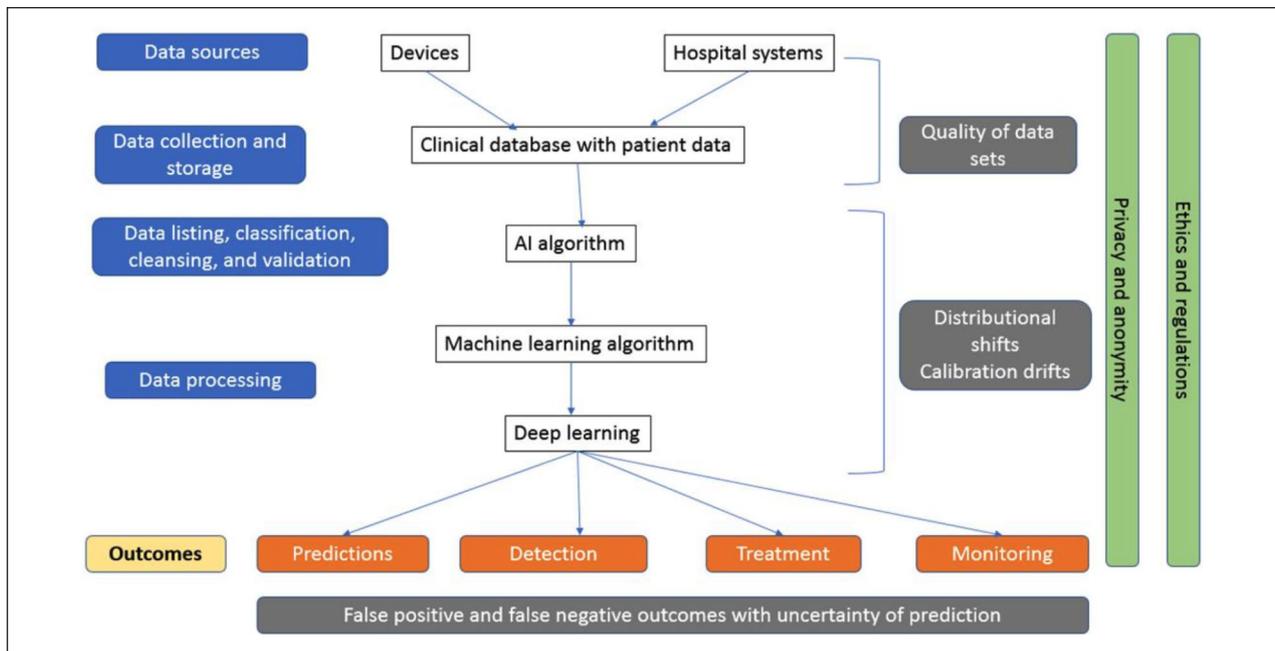


Figure 1. Safety concerns at various stages of deployment of artificial intelligence (AI).

Typically, CNNs are trained and validated using data sets in clinical settings.⁴ These applications may not perform with equal precision when applied to larger population-based samples. For example, in the case of application of ML for surveillance of skin lesions for detection of skin cancer, the appearance of skin lesions and patient characteristics may be more diverse in the general population than in the training data set. Confidence in an AI system may be low with regard to prediction accuracy when the system has insufficient information to make a decision. Such unsafe AI decision-support systems may predict a false low or high risk of disease and the prediction may not be trusted.⁸

Quality of Data Sets. High-quality data sets are key to training AI systems in health care. Data should be labeled correctly for the outcomes of interest so that the systems find a “ground truth” to learn associations. Failing this, AI applications will have poor reproducibility. Lacunae are reported for training data sets, even in high-performing AI applications. For example, CheXNet, a 121-layered CNN, outperformed 4 radiologists in the detection of pneumonia.¹⁴ This application was trained on only frontal radiographic images. Although lateral images and history of fever are key to establishing the diagnosis of pneumonia in clinical settings, these components were not factored into training the algorithm.^{30,31}

Training and operational data sets are never the same for the application of ML in medicine. There may be deficiencies in training data sets, and outliers and surprises in

operational data sets. This is referred to as the “frame problem” of AI in medicine, which implies deficiencies in updating inputs to describe the environment for autonomous agents.³²

ML algorithms are trained on balanced data sets for cases and controls. However, this may not necessarily apply to trial and real-world settings, in which the numbers may vary for cases and controls. Imbalanced data sets may erroneously overdiagnose or underdiagnose a disease condition.³³ Data sampling-based boosting frameworks may be applied to imbalanced data sets to enable accurate outputs for diagnosis.^{34,35}

Oblivious Impact. AI systems may be insensitive to impact. These systems may fail to account for the false-positive and false-negative predictions in relation to the clinical context. Li et al³⁶ applied a deep learning algorithm to detect referable diabetic retinopathy in a data set of 71 043 retinal images. In this study, the algorithms had a high sensitivity and specificity of 92.5% and 98.5%, respectively. Misclassification of mild or moderate diabetic retinopathy accounted for 85.6% of false-positive cases, whereas undetected intraretinal microvascular abnormalities accounted for 77.3% of all false-negative cases. To be able to achieve more accurate and error-free systems, the systems should be able to identify the outliers and accordingly adjust their confidence of diagnosis.

Efforts are being focused on developing effective algorithms to enable accurate diagnosis of medical conditions. A prospective study is ongoing for the application

of AI to detect diabetic retinopathy in primary care.³⁷ AI computer-aided detection systems, leveraging CNN and automatic hierarchical learning capabilities, are being adopted to decrease false-positive rates in breast cancer screening.³⁸

Confidence of Prediction. Explainability and interpretability of AI and ML algorithms is challenging. Only interpretable models can be described, and explanation is important for adoption of any system for medical decision making. AI has raised concerns about “black box” medical decisions. This implies that the predictions of the system are not known until the final outcome has occurred.³⁹ Although training data sets may be insufficient, there is no way to confirm the impact of the data set on prediction of the disease state. This is a feature that will become evident only after sustained and repeated use in practice. There will always remain inability to assess the knowledge of the machine; transparency in methods of development and sharing of algorithms can help build confidence in adopting these applications.^{40,41}

On the other hand, safety of AI in health care may be threatened by automation complacency.⁴² With repeated use, physicians may construe AI to be infallible and entrust it with blind faith. Automation complacency, seen in both naïve and expert physicians, occurs when the automated task competes with manual tasks for the physician’s attention. Physicians may subconsciously learn to avoid available alternatives when an AI system repeatedly demonstrates agreement with their diagnoses.^{43,44}

Quality of AI systems can be defined in terms of interpretable predictions with an estimate of confidence. The knowledge about the certainty of prediction can help clinicians minimize automation bias.

Unexpected Behaviors and Unscalable Oversight. The behavior of AI is difficult to predict and control. Calibration drifts are common in regression and ML models for personalized risk estimates. Examples of these drifts include models for acute kidney injury and hospital mortality.^{45,46}

AI is a continuously learning autonomous system that may make unpredictable efforts to implement its learning. This may happen because of the phenomenon of “wireheading.” ML algorithms can replicate past decisions. This may lead to several challenges. A common experience with unexpected behavior is the use of automation for heparin dosing. With continuous learning, the system may deliver a higher dose of heparin leading to a possible increase in adverse events. This calls for training systems to guard against dangerous overdosing. The systems, being mechanical, may be focused on drug delivery alone with no concern for long-term outcomes.⁴⁷ Signal processing algorithms in mechanical ventilators may

achieve optimal oxygenation irrespective of the long-term possibility of lung damage.⁴⁸ The learning phase of the systems may be difficult to measure and predict. This may lead to unscalable oversight requiring inconvenient and expensive measurements to monitor bleeding. Another example can be an autonomous insulin pump that may need exhaustive information on food intake before the system can determine the correct insulin regimen for optimal control of blood glucose. These are essentially challenges of “reinforcement learning” in automated systems, which have an inherent ability to maximize a defined reward.⁴⁹ The continuous supply of data for sustained training of AI systems also may become limiting.⁵⁰

Bias, Ethics, and Anonymity. AI, an effort to mimic biological intelligence, lends itself to bias. AI reciprocates training, and any bias in training data sets can train the AI application to adopt bias in the operational data set. Human-like biases are a known and recognized component of automated systems.⁵¹ These unintentional biases can compromise the safety of AI systems. Erroneous judgments can be made by AI applications: for example, in the interpretation of radiological imaging.⁵²

Ethics are a key concern in health care, and AI is no exception. Adoption of AI, its use for research, its impact on outcomes, and susceptibility to bias are growing concerns as AI is finding extended applications in medicine and health care. AI and ML are not yet fully mature to adopt and qualify for basic biomedical principles of autonomy, beneficence, justice, and non-maleficence.⁴² Clinicians and medical researchers, though not the lead developers of AI-based applications, should ensure that these principles are not transgressed. This onus lies solely on humans and not on AI systems.

Privacy and anonymity of AI systems is a common challenge. Sensitive patient data are now used in digital format and fed into networked systems. The levels of security of these systems often are not clearly defined. A review of EHRs’ security and privacy (n = 49 articles) reported role-based access control as the most preferred access control model. Digital signatures, and logins and passwords were the other authentication approaches. The authors reported relative lack of training in security and privacy for health care workers and system users.⁵³ Security and privacy of data are one of the major concerns while using AI in a medical setup. Various research articles of systematic literature revealed the importance of maintaining security and privacy of data. Some studies also indicated who should provide access to EHR data.⁵³ A recent study described ranges of features fundamental to the security and privacy of EHRs, such as access control, and compliance with security requirements, among others.⁵⁴

Another potential issue is loss, leak, and manipulation of information that may imply risks associated with AI-based mobile health applications. Consumer data that are aggregated and shared across apps pose risks to individual privacy and security. The centrally connected app families, including multiple industries outside of health care, may potentially raise data privacy concerns.⁵⁵ It can be said that AI-based mobile medical application developers should secure users' data confidentiality.

Malpractice issues are likely with advances in telemedicine devices, necessitating the acquisition of applicable licenses and adequate trainings and certifications. This may add to the cost of adopting telemedicine and may lead to mistrust of these systems.⁵⁶

Approaches to Achieving Safety in AI

Accuracy of prediction, causality of predictive models, human effort for labeling out-of-sample cases, and reinforcement and learning of systems contribute to making applications safe for use in health care. Four key strategies of safety engineering apply to safety of AI in health care: (1) safe design, (2) safety reserves, (3) safe fail, and (4) procedural safeguards.⁸ Inherently safe design implies that potential hazards will be excluded and not merely controlled in systems. Systems for use in health care can be made safe by eliminating the chance of training data sets not being sampled from the test data sets. Although this can boost system accuracy, shifts in data domains continue to be a challenge. Safety reserves should be built into AI applications to enable detection of uncertainty in the training and test systems and handle the average and maximum test errors. The systems should be designed to fail safely (ie, the systems should continue to remain safe even when they fail in the intended operation). The model should be trained to confidently reject when it cannot make an intended prediction. In the event of such rejections, human interventions can be supplemented to make predictions. Procedural safeguards include user experience design to guide the users for setting up and running the application, which increases safety. Open data and open source software also are said to increase safety.

Irrational extrapolation of ML algorithms should be avoided. Algorithms trained on easy-to-obtain patient samples will not be in the best interests of the safety of patients when applied to wider and more diverse populations. Algorithms should be designed and trained to reason for disease severity and trajectory.⁵⁷

Defining ethical standards will enable wider application of AI and also will help surmount the fears of AI overpowering human capabilities. Societal benefits of AI will emerge if AI tools are made open source, user friendly, and of proven clinical benefits and economic value.⁵⁸

Deployment of AI in health care is not well regulated. Assessment of efficacy and safety of AI applications should be standardized. Both the US Food and Drug Administration (FDA) and the European Medicines Agency have set up working groups to develop and validate technical and digital applications in health care.⁵⁹ In addition, the European patent office has recently added guidance for patent applications for AI-based devices and ML.⁶⁰ The first deep learning clinical platform, Arterys' medical imaging platform, was approved by the FDA in 2017. The FDA supports the use of real-world evidence and adaptive design in clinical trials to assess the performance and operation of AI in health care.⁵⁰ Regulatory reforms are needed to enable more efficient and safe exchange and sharing of data.

Conclusions

Strategies for safety of AI and ML in health care are evolving and are not yet fully developed. So far, safety of AI in health care is focused on predictions and outcomes based on predictions. Systems and applications with substantial and nominal safety should be handled with required protocol. Models for personalized risk estimates should be well calibrated and efficient, and effective updating protocols should be implemented. Cost, risk, and uncertainty should be defined for all possible applications. Automated systems and algorithms should be able to adjust for and respond to uncertainty and unpredictability. Efforts should be targeted toward decreasing epistemic uncertainty.

Finite numbers of test samples before deployment is a common challenge in AI-based learning systems. Training samples are not always representative of test samples. The "frame problem" can be addressed by introducing a human component to AI applications, including continuous calibration of the systems depending on human feedback, clinician review of atypical data sets, and inclusion of diverse populations in training sets.³²

Training is needed not only for AI-based systems but also for clinicians, who can be groomed as information specialists to further train and develop accurate and dependable AI solutions.⁵ AI-augmented clinicians should be more efficient and confident and not faced with the uncertainty of risks associated with technical advances in medicine. Physicians should understand, develop, adopt, and leverage AI to improve patient care.

Efforts should be made to maximize the benefits of AI in health care. Experts recommend 4 critical aspects in this regard: quantifying benefits to enable measurement, building trust for adoption of AI, building and enhancing technical skills, and organizing a system of governance.⁵⁹ Data protection legislation should be formulated and strengthened for the collection and processing of data in clinical research. With objective and demonstrable safety,

AI can enable value-based and patient-centric health care. Quality standards for AI applications in medicine should be clearly defined to add value, accuracy, efficiency, and satisfaction to AI in health care.

The cost and distribution of outcomes in AI-based systems are not precisely known. Large feasibility studies and cost-effectiveness assessments can help improve adoption of AI in health care. Privacy, sharing, and disclosure of safety data relating to AI applications should be strengthened. High standards should be defined for validation of AI and ML applications in health care. Methods, guidelines, and protocols should be formulated to enable the safe and effective development and adoption of AI and ML in health care. Trust and training will allow the full functional integration of AI into research and practice in health care.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Samer Ellahham  <https://orcid.org/0000-0003-1779-2414>

Mecit Can Emre Simsekler  <https://orcid.org/0000-0002-1555-5012>

References

- Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism*. 2017;69S:S36-S40.
- Baştanlar Y, Ozuysal M. Introduction to machine learning. *Methods Mol Biol*. 2014;1107:105-128.
- Deo RC. Machine learning in medicine. *Circulation*. 2015;132:1920-1930.
- Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhhl T, Blum A. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*. 2018;29:1836-1842.
- Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. *JAMA*. 2016;316:2353-2354.
- Shah NR. Health care in 2030: will artificial intelligence replace physicians? *Ann Intern Med*. 2019;170:407-408.
- Goldhahn J, Rampton V, Spinaz GA. Could artificial intelligence make doctors obsolete? *BMJ*. 2018;363:k4563.
- Varshney KR. Engineering safety in machine learning. <https://ieeexplore.ieee.org/document/7888195>. Accessed September 4, 2019.
- Varshney KR, Alemzadeh H. On the safety of machine learning: cyber-physical systems, decision sciences, and data products. *Big Data*. 2017;5:246-255.
- Swaminathan S, Qirko K, Smith T, et al. A machine learning approach to triaging patients with chronic obstructive pulmonary disease. *PLoS One*. 2017;12(11):e0188532.
- Duggal R, Brindle I, Bagenal J. Digital healthcare: regulating the revolution. *BMJ*. 2018;360:k6.
- McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. *BMJ Open Qual*. 2017;6(2):e000158.
- Bae S-H, Yoon K-J. Polyp detection via imbalanced learning and discriminative feature learning. *IEEE Trans Med Imaging*. 2015;34:2379-2393.
- Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T. CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning. <https://arxiv.org/pdf/1711.05225.pdf>. Accessed September 4, 2019.
- De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24:1342-1350.
- Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-118.
- Seyhan AA, Carini C. Are innovation and new technologies in precision medicine paving a new era in patients centric care? *J Transl Med*. 2019;17(1):114.
- Yellowlees PM, Chorba K, Burke Parish M, Wynn-Jones H, Nafiz N. Telemedicine can make healthcare greener. *Telemed J E Health*. 2010;16:229-232.
- Young K, Gupta A, Palacios R. Impact of telemedicine in pediatric postoperative care [published online December 5, 2018]. *Telemed J E Health*. doi:10.1089/tmj.2018.0246
- Phillips-Wren G, Jain L. Artificial intelligence for decision making. In: Gabrys B, Howlett RJ, Jain LC eds. *Knowledge-Based Intelligent Information and Engineering Systems*. Berlin, Heidelberg: Springer; 2006. *Lecture Notes in Computer Science*; vol 4252.
- Recht M, Bryan RN. Artificial intelligence: threat or boon to radiologists? *J Am Coll Radiol*. 2017;14:1476-1480.
- Mayo RC, Leung J. Artificial intelligence and deep learning—radiology's next frontier? *Clin Imaging*. 2018;49:87-88.
- Robertson S, Azizpour H, Smith K, Hartman J. Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. *Transl Res*. 2018;194:19-35.
- Zhang YC, Kagen AC. Machine learning interface for medical image analysis. *J Digit Imaging*. 2017;30:615-621.
- Kim D, You S, So S, et al. A data-driven artificial intelligence model for remote triage in the prehospital environment. *PLoS One*. 2018;13(10):e0206006.
- Levin S, Toerper M, Hamrock E, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the Emergency Severity Index. *Ann Emerg Med*. 2018;71:565-574.
- Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. *PM R*. 2017;9(5S):S106-S115.

28. Keskinbora KH. Medical ethics considerations on artificial intelligence. *J Clin Neurosci*. 2019;64:277-282.
29. Vellido A. Societal issues concerning the application of artificial intelligence in medicine. *Kidney Dis (Basel)*. 2019;5(1):11-17.
30. Raouf S, Feigin D, Sung A, Raouf S, Irugupati L, Rosenow EC III. Interpretation of plain chest roentgenogram. *Chest*. 2012;141:545-558.
31. Potchen EJ, Gard JW, Lazar P, Lahaie P, Andary M. Effect of clinical history data on chest film interpretation-direction or distraction. *Invest Radiol*. 1979;14:404.
32. Yu K-H, Kohane IS. Framing the challenges of artificial intelligence in medicine. *BMJ Qual Saf*. 2019;28:238-241.
33. Storkey AJ. When training and test sets are different: characterising learning transfer. In: Lawrence CSS, ed. *Dataset Shift in Machine Learning*. Cambridge, MA: MIT Press; 2013:3-28.
34. Bae S-H, Yoon K-J. Polyp detection via imbalanced learning and discriminative feature learning. *IEEE Trans Med Imaging*. 2015;34:2379-2393.
35. Zhao Y, Wong ZS, Tsui KL. A framework of rebalancing imbalanced healthcare data for rare events' classification: a case of look-alike sound-alike mix-up incident detection. *J Healthc Eng*. 2018;2018:6275435.
36. Li Z, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. *Diabetes Care*. 2018;41:2509-2516.
37. Vidal-Alaball J, RoyoFibla D, Zapata MA, Marin-Gomez FX, Solans Fernandez O. Artificial intelligence for the detection of diabetic retinopathy in primary care: protocol for algorithm development. *JMIR Res Protoc*. 2019;8(2):e12539.
38. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. *Clin Radiol*. 2019;74:357-366.
39. Hoffman RR, Klein G, Mueller ST. Explaining explanation for "explainable Ai." *Proc Hum Factors Ergon Soc Annu Meet*. 2018;62(1):197201.
40. Handelman GS, Kok HK, Chandra RV, et al. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *AJR Am J Roentgenol*. 2019;212(1):38-43.
41. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Cent Rep*. 2019;49:15-21.
42. Challen R, Denny J, Pitt M, et al. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 2019;28:231-237.
43. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. *Hum Factors*. 2010;52:381-410.
44. Wickens CD, Clegg BA, Vieane AZ, Sebok AL. Complacency and automation bias in the use of imperfect automation. *Hum Factors*. 2015;57:728-739.
45. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. *J Am Med Inform Assoc*. 2017;24:1052-1061.
46. Davis SE, Lasko TA, Chen G, Matheny ME. Calibration drift among regression and machine learning models for hospital mortality. *AMIA Annu Symp Proc*. 2017;2017:625-634.
47. Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D. Concrete problems in AI safety. <https://arxiv.org/pdf/1606.06565.pdf>. Accessed September 4, 2019.
48. Dellaca' RL, Veneroni C, Farre' R. Trends in mechanical ventilation: are we ventilating our patients in the best possible way? *Breathe (Sheff)*. 2017;13:84-98.
49. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health*. 2018;8(2):020303.
50. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230-243. doi:10.1136/svn-2017-000101
51. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science*. 2017;356:183-186.
52. Stone P, Brooks R, Brynjolfsson E, et al. One hundred year study on artificial intelligence (AI100). <http://ai100.stanford.edu/2016-report>. Accessed May 3, 2019.
53. Fernández-Alemán JL, Señor IC, Lozoya PÁO, Toval A. Security and privacy in electronic health records: a systematic literature review. *J Biomed Inform*. 2013;46:541-562.
54. Rezaeibagha F, Win KT, Susilo W. A systematic literature review on security and privacy of electronic health record systems: technical perspectives. *Health Inf Manag*. 2015;44(3):23-38.
55. Grundy Q, Held FP, Bero LA. Tracing the potential flow of consumer data: a network analysis of prominent health and fitness apps. *J Med Internet Res*. 2017;19:e233.
56. Pacis DMM, Subido EDC Jr, Bugtai NT. Trends in telemedicine utilizing artificial intelligence. *AIP Conf Proc*. <https://doi.org/10.1063/1.5023979>. Published February 13, 2018. Accessed September 16, 2019.
57. Saria S, Butte A, Sheikh A. Better medicine through machine learning: what's real, and what's artificial? *PLoS Med*. 2018;15:e1002721.
58. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism*. 2017;69S:S36-S40.
59. Diebolt V, Azancot I, Boissel FH. "Artificial intelligence": which services, which applications, which results and which development today in clinical research? Which impact on the quality of care? Which recommendations? *Therapie*. 2019;74:155-164.
60. European Patent Office. Guidelines for examination in the European patent office. <https://www.epo.org/law-practice/legal-texts/guidelines.html>. Accessed April 25, 2019.