The four dimensions of contestable AI diagnostics - A patient-centric approach to explainable AI

Ploug, Thomas; Holm, Søren

Published in:
Artificial Intelligence in Medicine

DOI (link to publication from Publisher):
10.1016/j.artmed.2020.101901

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Artificial Intelligence In Medicine 107 (2020) 101901

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

The four dimensions of contestable AI diagnostics - A patient-centric approach to explainable AI

Thomas Ploug[a,*], Søren Holm[b,c]

[a] Aalborg University, Centre for Applied Ethics and Philosophy of Science, Department of Communication and Psychology, A. C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark
[b] University of Manchester, Centre for Social Ethics and Policy, School of Law, Manchester M13 9 PL, United Kingdom
[c] Center for Medical Ethics, Faculty of Medicine, University of Oslo, Norway

ARTICLE INFO

Keywords:
AI diagnostics
Contestability
Explainability
Health data
Bias
Performance
Organisation of diagnostic labour

ABSTRACT

The problem of the explainability of AI decision-making has attracted considerable attention in recent years. In considering AI diagnostics we suggest that explainability should be explicated as 'effective contestability'. Taking a patient-centric approach we argue that patients should be able to contest the diagnoses of AI diagnostic systems, and that effective contestation of patient-relevant aspect of AI diagnoses requires the availability of different types of information about 1) the AI system's use of data, 2) the system's potential biases, 3) the system performance, and 4) the division of labour between the system and health care professionals. We justify and define thirteen specific informational requirements that follows from 'contestability'. We further show not only that contestability is a weaker requirement than some of the proposed criteria of explainability, but also that it does not introduce poorly grounded double standards for AI and health care professionals' diagnostics, and does not come at the cost of AI system performance. Finally, we briefly discuss whether the contestability requirements introduced here are domain-specific.

1. Introduction

The development of machine learning/deep learning models holds great potential for medical diagnosis and treatment planning. There is great hype surrounding these developments, but also some substance to the claims. Several AI diagnostic algorithms have already been granted regulatory approval by the FDA [1], and the research is progressing fast in other areas. The balance between hype and substance is illustrated by a recent meta-analysis of AI diagnostic systems in the context of medical imaging and histopathology which identified 20,530 unique papers published between 2012–2019, but only 25 that directly compared the performance of AI and health care professionals (HCPs) and contained sufficient data to perform a quantitative meta-synthesis. The authors found "the diagnostic performance of deep learning models to be equivalent to that of health-care professionals" [2].

But machine learning/deep learning models are 'black-boxes' [3–9]. The decision-procedure is notoriously hard to interpret and explain in detail. This poses an ethical and practical problem. Without the ability to interpret or explain AI diagnostics, it becomes hard to determine if differences in diagnoses reflect diagnostically relevant differences between patients or if they are instances of bias or diagnostic errors and over/under-diagnosis. A recent study found that a prediction algorithm widely used in health care exhibited a racial bias that if remedied would increase the percentage of black patients receiving extra care from 17.7–46.5 % [10].

In response to these problems, public committees and expert groups, research institutions and private companies have in recent years issued reports on and guidelines for responsible use of AI. A systematic review of the corpus of such guidelines found 84 documents containing ethical principles and guidelines for the use of AI [11]. The review identifies a significant global convergence on the importance of, among others, the transparency or explainability of AI decision-making. The review also finds, however, a significant variance in the posited features of transparency. Thus, for instance, transparency is posited to 1) serve very different purposes, e.g. minimise harm, foster trust or democracy, and 2) concern many different properties of AI decision-making, e.g. data use, level of automated decision, and access to source-code. Transparency may 3) be defined relative to different groups of stakeholders, e.g. developers, users, oversight committees, and 4) it may vary from one context to another. Transparency across all possible features may be desirable but not ethically required. In any case, there is need of determining a set of criteria for picking out the essential features of a

* Corresponding author.
E-mail addresses: ploug@hum.aau.dk (T. Ploug), soren.holm@manchester.ac.uk (S. Holm).

https://doi.org/10.1016/j.artmed.2020.101901
Received 30 January 2020; Received in revised form 30 April 2020; Accepted 3 June 2020
Available online 09 June 2020

minimal conception of transparency or explainability.

A requirement of transparency or explainability can also be found in legislation. Article 13 and 14 of the EU General Data Protection Regulation (GDPR) stipulate that if data subjects are profiled, they have a right to "meaningful information about the logic involved" [12]. This applies to the medical context as explained in the official EU guideline [13]. While it is clear that this is a requirement of some form of explainability of AI decision-making, it is a rather vague requirement [3,14,15].

We believe that the GDPR requirement of explainability should be understood as a requirement of contestability. That is, AI decision-making must be explainable to a degree that makes it possible for an individual to contest the decision of the system. Interpreting AI explainability as substantiating a requirement of contestability, it becomes of cardinal importance to understand this requirement in greater detail.

While the notion of AI contestability has attracted attention in recent writings, the focus has first and foremost been 'contestability by design', i.e. providing design principles for algorithmic systems that will enable professionals/expert users to challenge the reasoning of these systems in an ongoing process, and thus to enable a fruitful reasoning together of man and machine [16–19]. Among the suggested design principles are 1) that the accuracy of system predictions are improved through phased deployment on the basis of incentivised feedback from expert users, 2) that the legibility of the systems is heightened through mechanisms that allow expert users to "unpack aggregate measures, tracing system decision all the way down", and 3) that the potential misuse and bias of AI systems is countered by providing mechanisms for recording expert users "disagreements with system behaviour" [16]. The 'contestability by design' approach is an *ex ante* approach focusing on system development and optimisation, i.e. it concerns the process before an individual, e.g. a patient, is subjected to an AI decision [15,18]. Consequently, it does not specify the contestability right of an individual subjected to AI decision-making, including the information requirements of such a right.

In this article, we explicate the practical implications of the idea that the legal and ethical notion of explainability should be understood, at least partly in terms of contestability. In the literature explainability is often understood as the ability to answer 'why' questions [20]. Contesting a decision is asking a type of why question. The kinds of 'why' questions that different stakeholders are likely to pose may differ. The 'why' questions raised by systemdevelopers, expert users and patients are, for instance, likely to be significantly different. We develop a patient-centric, value-based framework for evaluating the contestability of AI diagnostics and for identifying the explanatory elements necessary for effective contestation *ex post*. In short, we approach the explainability of AI diagnostics from the perspective of the patient. This approach allows us to define a minimal set of criteria of effective contestability. These criteria provide practical guidance in the health care context and explicate the legal notion of transparency or explainability.

2. Explainability and effective contestability

Providing patients with an explanation of their AI generated diagnosis may be guided by different principles and interests. A maximal interpretation of the requirement of explainability would require that an explanation should spell out why the diagnosis was the scientifically best possible explanation of the set of signs, symptoms and indicators. A minimal interpretation would require a statement to the effect that the diagnosis was arrived at by a machine on the basis of health data. But which of these interpretations should guide explanation of AI driven diagnostics?

The GDPR provides some – yet again rather vague – guidance on this issue. Article 22 states that in those cases, where a data subject may legitimately be subjected to automated decision-making including profiling, the data controller should safeguard the data subject's right "to express his or her point of view and to contest the decision" [12]. (Although the GDPR is specific to the EU, other legal systems also contain mechanisms by which patients can contest health care decisions). Interpreting the requirement of explainability along these lines, it would imply that an explanation of a diagnosis must allow an individual to contest the diagnosis. The bare right to contest is, however, empty. Just having the right to say 'I disagree with this decision and contest it' does not help the data subject. Only a right to effective contestation is worth having, i.e. a right to contest a decision through a demand for an adequate explanation. What is needed here is a more substantial notion of contestability in relation to AI driven diagnostics. A notion that it is embedded in a wider ethical framework of individual rights and interests in relation to diagnosis.

This approach to contestability is distinct from the 'Causability' approach recently developed by Holzinger et al. [21,22]. Holzinger et al. define causability as a relation between an expert user and an AI system where causability is "… the extent to which an explanation of a statement to a user achieves a specified level of causal understanding with effectiveness, efficiency and satisfaction in a specified context of use." (Holzinger et al. 2020). This explicitly requires a causal model of the diagnostic reasoning, or in the case of machine learning a mapping from the machine model to a causal model understandable by the human expert. Our approach differs in two ways. We focus on contestability by the patient and on contestability without assuming or defining any specific requirements for explainability in advance.

3. The four dimensions of effective contestability of AI diagnostics

There are at least four aspects of AI involvement in medical diagnosis that an individual might want to contest and which would be covered by a right to contest. Below we 1) show the relevance of a right to contest these four aspects of AI diagnosis, 2) suggest the ethical reasons in terms of protecting important interests for granting individuals a right to contest each of these four aspects of AI diagnosis, and 3) explicate the information that individuals must be provided in order to exercise this right.

3.1. The use of personal health data in AI diagnostics

First, an individual should be able to contest the use of personal health data in AI driven diagnostics.

An AI diagnostic model is applied to personal health data derived from various different sources. Such personal health data may not only be sensitive, it may also be outdated, one-sided, erroneous and incomplete [23]. Consequently, AI diagnostics may not only lead to invasions of privacy in the form of undesirable use of sensitive data, but also to harms following the use of inaccurate personal health data. In the future AI systems may also use non-health data, e.g. data about social conditions, use of social media etc. [24]. This raises additional concerns about data sources and quality and reliability.

Individuals have a right to privacy, and they have a right to protect themselves against harm, and this implies a right to contest the use of health and other personal data. Exercising this right to contest the use of personal health and other data requires that individuals have access to 1) information about the types of personal data used in AI diagnostics, e.g. clinical tests, tissue, scans etc. However, since both the sensitivity and quality of data may be dependent on the source of the data, individuals must also be provided with 2) information about these sources e.g. Electronic Patient Record etc.

3.2. The potential bias of AI diagnostics

Second, an individual should be able to contest potential bias in AI diagnostics.

AI diagnostics may be biased due to bias in the training data or in the prior human categorisation of the training data [25–28]. Bias may

Box 1
The four dimensions of contestable AI and associated informational requirements

Contestability – Object	Contestability – Variables	Contestability – Questions	Contestability – Required Explanation
Data	Types of data	What types of personal data are used?	1) The decision D was made on the basis of data of type X, Y, Z about you
	Data sources	Where do your data come from?	2) The decision D was based on data from sources X, Y, Z
Bias	Training data	On which data was the AI trained?	3) The decision D was made by a system trained on existing data of type X, Y, Z
	'Tagging groups'	Who tagged training data?	4) The decision D made by a system trained on data tagged by X, Y, Z
	Tested for bias	Was training data or AI system tested for bias?	5) The decision D was made by a system tested for bias of type X, Y, Z
Diagnostic Performance	Performance	What is the performance of the AI system?	6) The decision D was made by a system with a performance of X, Y, Z
	Performance testing	How was the performance determined?	7) The decision D was made by a system with a performance determined by tests X, Y, Z
	Essential indicators	What are key variables of AI decision-making?	8) The key input data resulting in decision D was X, Y, Z
	Alternatives	Are alternatives considered?	9) The alternatives to decision D are X, Y, Z with a probability of x, y, z (< d)
	Longevity	When is decision reconsidered?	10) The decision D will be reconsidered if conditions X, Y, Z obtain
Decision	AI involvement	To what degree is AI making the decision?	11) The decision D involved an AI system with respect to X, Y, Z
	Human involvement	To what degree are humans making the decision?	12) The decision D was wholly/partly made by health professionals X, Y, Z
	Responsibility	Who is responsible for the decision?	13) The objective/legal responsibility for decision D is held by X, Y, Z

lead to discrimination defined as unfair differential treatment if it causes unwarranted differences in diagnostic patterns for particular individuals or groups [29–31]. Prior to the clinical deployment of an AI system potential bias in AI decision-making may be tested for by applying the AI model to relevantly different datasets. Prior testing in this way requires a set list of known 'triggers' of discrimination, e.g. gender, age, ethnicity etc. It may also be possible to test for bias in an individual case, e.g. by using a suitably modified set of input data to investigate whether the system provides the same result. Finally, indications of biased system diagnostics may also come from 'counterfactually' testing a system against previous or alternative systems or simply against HCP diagnostics [32]. In both cases the testing cannot be exhaustive, but it will clearly be a way of minimising the risk of harmful and discriminatory bias.

Individuals have a right to protect themselves against discrimination, and therefore should be granted a right to contest bias in AI diagnostics. Exercising the right to contest bias requires that individuals have access to information about 1) the character of the dataset on which the model is built, 2) how the data were categorised by humans, and 3) the character and level of testing the AI model has undergone. In some cases where an initial general claim of potentially relevant bias can be made out following disclosure of these three elements, an individual would also have a right to have bias investigated at the individual level.

3.3. The performance of AI diagnostics

Third, an individual should be able to contest the performance of AI diagnostics.

A 'locked' AI model for diagnostics will have a set performance when applied to a patient population which is qualitatively identical to the original learning set. Although there are several examples of AI models outperforming humans, [33–35] they are still not completely accurate [2]. They will inevitably produce diagnostic errors and over/under-diagnosis for particular individuals, which may ultimately be harmful. The performance of an AI model – e.g. the number of true and false positives, and the number of true and false negatives [36] – may be determined through tests on sets of data of various size and composition, and this will influence the reliability of the estimated accuracy of the model when it is applied in the routine clinical context. The performance of a self-modifying AI system may change over time, as may the risk of bias. However, the patient can be provided evidence about performance and bias when the system was implemented for routine clinical use and when diagnostic accuracy was last tested. Even if self-modifying AI systems were to be introduced in routine clinical use, this would therefore not change the basic contestability requirements.

The right to contest the performance is ultimately rooted in individuals' right to protect themselves against harm. Exercising this right to contest the performance of AI diagnostics requires 1) information about the performance of the AI model, and 2) information about the tests used to determine the performance. However, contesting a diagnosis simply on grounds of possible poor performance makes little sense since it does very little towards identifying what could and should reasonably have been considered. It does not enable individuals to defend themselves against harm in an informed way. Informed contestation of a diagnosis minimally also requires 3) information about the key indicators of the diagnosis, 4) alternatives to the suggested diagnosis, and 5) information about the changes that will lead to a reconsideration of the diagnosis.

3.4. The organisation and division of diagnostic labour

Fourth, an individual should be able to contest the organisation and division of the diagnostic labour between HCPs and AI.

The level of AI input into a diagnostic process may vary according to the division of diagnostic labour between the system and the HCP. AI may be used for initial screening prior to diagnosis, input directly into the diagnostic process, make a diagnosis that only need to be ratified by a HCP, be used as a 'second opinion', or in some other way. The organisation and division of diagnostic labour between AI and professionals may promote or impede the quality of the diagnostic processes, and promote or impede the HCPs' ability to make their own diagnostic decisions. Evidence suggests that clinical decision support systems may improve practitioner performance, [37,38] but also may lead to over-reliance on the performance of these systems (automation bias) and deskilling [39–44].

The right to contest the division and organisation of diagnostic labour is also a matter of shielding individuals against harm. If the organisation and division of diagnostic labour can affect the quality of diagnostics, individuals should have a right to contest this. Exercising this right requires 1) information about the role of AI in the diagnostic process, 2) information about the role of HCPs in the diagnostic process. Challenging the organisation and division of diagnostic labour also requires 3) information about the objective/legal responsibility for diagnostic procedures.

The four aspects of contestability in relation to AI diagnostics and the associated information/explanation requirements are summarised

in Box 1 above.

4. Implications of contestability

4.1. Contestability and the medical encounter

The right to contest and the correlative duty to provide the information needed for effective contestation does not imply that every patient should be provided with all of the information needed for contestability in all of the four dimensions. Most patients are probably unlikely to want to contest the advice provided by the AI system to the HCP, and will be satisfied with the explanation of their diagnosis provided by the HCP. The right to contest does, however, generate one duty that is relevant whenever an AI system has provided advice, that is the duty to inform that patient that AI advice has been provided and used by the HCP.

Contestability does, however, create duties for developers of AI systems and for organisations purchasing and using such systems. Developers and user organisations have to be able to provide all the elements of information outlined above if a patient contests the AI advice. And user organisations have a further duty to train their employees to provide this information to patients and help them understand whether there is a justifiable basis for their contestation of the AI advice.

4.2. Contestability requirements apply to both AI and HCPs

Contestability requirements apply to AI and HCP diagnostics alike. There seems to be no relevant difference between AI and HCP diagnostics that would justify double standards [45]. HCPs can also be biased, make mistakes, or not work optimally with colleagues or AI systems. And, HCPs are arguably also 'black boxes' [9]. The exact reasoning of HCPs – every aspect of it – cannot be fully replicated, scrutinized and simulated. Only key factors behind their diagnostics may be reconstructed. A set of contestability requirements for HCP diagnostics is therefore also needed.

The contestability requirements cannot, however, be the same. It simply does not make sense to require information about the 'training data' for a HCP. Contestability requirements must reflect how a diagnostic system – whether it be a HCP or an AI system – is trained and process data in the diagnostic context. In short, contestability requirements must concern types of information that it makes sense to require in relation to a specific diagnostic setup. Developing contestability requirements for HCPs is beyond the scope of this paper.

4.3. Contestability requirements do not impede performance

A key concern in the literature on explainable AI is the potential trade-off between diagnostic performance and explainability [3,9]. It has been suggested that AI decision-making should be understood as simulatability, i.e. it should be possible for a human "to take the input data together with the parameters of the model and in reasonable time step through every calculation required to produce a prediction" [3]. A requirement of simulatability would imply that all sufficiently complex AI models, including some of the best performing types of machine learning such as deep learning models, are inexplainable. Hence, requiring explainability as simulatability would be at the cost of performance.

Contestability is a weaker requirement than simulatability. Contestability as introduced here does not come at the cost of system performance for any type of AI because it makes very weak demands concerning the transparency of the actual AI decision-making procedure – or in the GDPR parlance – the logics of the automated processing. It only requires the availability of information about the development of the AI system, and about key indicators or variables that lead to a certain diagnosis, cf. requirement 8 in Table 1.

Satisfying the requirements of contestability do come at a cost. There are and will be costs associated with collecting and making available the types of information enabling effective contestation of AI diagnostics, e.g. costs related to the training of personnel to collect and provide this information (see above). However, these costs do not affect system performance. They are costs that will be incurred by any attempt of providing transparency in AI decision-making and diagnostics.

5. Contestability is domain-specific

The requirements of contestable AI suggested in this article has been made with specific reference to a patients' rights and interests in relation to biomedical diagnostics. The question is, however, if these requirements are domain-specific. Do they only apply to the health care context or do they also apply to AI decision-making in other contexts? Do the same requirements apply, for instance, to AI supported decision-making concerning liability or guilt in legal matters or to the use of AI for making personal credit assessments in the financial domain?

There are important similarities between all contexts. The harm caused by biased and/or erroneous AI decision-making may potentially be significant on both individual and societal level. In health care it may cause bodily harm, and in the legal and financial contexts it may reduce personal liberty and economic freedom. In all contexts it may undermine public trust in 'system decisions' and lead to growing inequality.

There are, however, also important differences between these contexts. An important difference is the level of conflicting interests. In health care decision-making is and must be in the best interest of patients. In legal and financial matters decision-making must weigh individual interests against the interests of society and companies. In contexts defined by conflicting interests the requirements of contestability (explainability and transparency) could reasonably be argued to be stronger than in other contexts. In consequence, the contestability requirements introduced in this article could be argued to be a necessary, but insufficient part of the set of contestability requirements for AI decision-making in these contexts. Further discussion is warranted.

For present purposes we simply conclude that the requirements introduced in this article provide a minimal, domain-specific and patient-centric set of conditions of contestability appropriate for the health care context.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.artmed.2020.101901.

References

[1] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56.
[2] Liu X, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019;1:e271–97.
[3] Lipton ZC. The mythos of model interpretability. ArXiv160603490 Cs stat. 2016.
[4] Burrell J. How the machine 'thinks': understanding opacity in machine learning algorithms. Big Data Soc 2016;3. 2053951715622512.
[5] Doshi-Velez F, Kim B, et al. Considerations for evaluation and generalization in interpretable machine learning. In: Escalante HJ, editor. Explainable and interpretable models in computer vision and machine learning Springer International Publishing; 2018. p. 3–17. https://doi.org/10.1007/978-3-319-98131-4_1.
[6] Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 2018;6:52138–60.
[7] Pasquale F. The black box society. Harvard University Press; 2015.
[8] Caruana R, et al. Intelligible models for HealthCare: predicting pneumonia risk and Hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD' 15 1721–1730 2015. https://doi.org/10.1145/2783258.2788613.
[9] London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep 2019;49:15–21.
[10] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366:447–53.

[11] Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019;1:389–99.
[12] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General data protection regulation) (Text with EEA relevance). OJ L Vol. 119 (2016).
[13] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and profiling for the purposes of regulation 2016/679. 2018.
[14] Goodman B, Flaxman S. European union regulations on algorithmic decision making and a 'Right to explanation' AI Mag Can 2017;38:50–7.
[15] Edwards L, Veale M. Enslaving the algorithm: from a 'Right to an explanation' to a 'Right to better decisions'? IEEE Secur Priv 2018;16:46–54.
[16] Hirsch T, Merced K, Narayanan S, Imel ZE, Atkins DC. designing contestability: interaction design, machine learning, and mental health. Proceedings of the 2017 Conference on Designing Interactive Systems - DIS' 17 95–99 2017. https://doi.org/10.1145/3064663.3064703.
[17] Mulligan DK, Kluttz DN, Kohli N. Shaping our tools: contestability as a means to promote responsible algorithmic decision making in the professions. 2020. p. 16.
[18] Almada M. Human intervention in automated decision-making: toward the construction of contestable systems. Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law - ICAIL' 19 2–11 2019. https://doi.org/10.1145/3322640.3326699.
[19] Vaccaro K, Karahalios K, Mulligan DK, Kluttz D, Hirsch T. Contestability in algorithmic systems. Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing - CSCW' 19 523–527 2019. https://doi.org/10.1145/3311957.3359435.
[20] Miller T. Explanation in artificial intelligence: insights from the social sciences. Artif Intell 2019;267:1–38.
[21] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. WIREs Data Min Knowl Discov 2019;9:e1312.
[22] Holzinger A, Carrington A, Müller H. Measuring the quality of explanations: the system causability scale (SCS). Comparing Hum Mach Explanations 2019. ArXiv191209024 Cs.
[23] Scott IA. Hope, hype and harms of Big Data. Intern Med J 2019;49:126–9.
[24] Ghani NA, Hamid S, Targio Hashem IA, Ahmed E. Social media big data analytics: a survey. Comput Hum Behav 2018. https://doi.org/10.1016/j.chb.2018.08.039.
[25] Cabitza F, Ciucci D, Rasoini R. A giant with feet of clay: on the validity of the data that feed machine learning in medicine. ArXiv170606838 Cs Stat. 2018.
[26] Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. JAMA Intern Med 2018;178:1544–7.
[27] Chen JH, Asch SM. Machine learning and prediction in medicine — beyond the peak of inflated expectations. N Engl J Med 2017;376:2507–9.
[28] Char DS, Shah NH, Magnus D. Implementing machine learning in health care — addressing ethical challenges. N Engl J Med 2018;378:981–3.
[29] Bobrowski D, Joshi H. Unmasking A.I.'S bias in healthcare: the need for diverse data. Univ. Tor. Med. J. 2019;96.
[30] Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018;169:866.
[31] Goodman SN, Goel S, Cullen MR. Machine learning, health disparities, and causal reasoning. Ann Intern Med 2018;169:883.
[32] Cowgill B, Tucker C. Algorithmic Bias: a counterfactual perspective. Work. Pap. NSF trust. Algorithms. 2020. p. 3.
[33] Bedi G, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 2015;1:15030.
[34] Rajpurkar P, et al. CheXNet: radiologist-level pneumonia detection on chest X-Rays with deep learning. ArXiv171105225 Cs stat. 2017.
[35] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 2017;12:e0174944.
[36] Lever J, Krzywinski M, Altman N. Classification evaluation. Nat Methods 2016. https://doi.org/10.1038/nmeth.3945https://www.nature.com/articles/nmeth.3945.
[37] Garg AX, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005;293:1223–38.
[38] Sullivan F, Wyatt JC. How decision support tools help define clinical problems. BMJ 2005;331:831–3.
[39] Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA 2017;318:517–8.
[40] Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. J Am Med Inform Assoc 2003;10:478–83.
[41] Povyakalo AA, Alberdi E, Strigini L, Ayton P. How to discriminate between computer-aided and computer-hindered decisions: a case study in mammography. Med Decis Making 2013;33:98–107.
[42] Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19:121–7.
[43] Goddard K, Roudsari A, Wyatt JC. Automation bias: empirical results assessing influencing factors. Int J Media Inf 2014;83:368–75.
[44] Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. J Am Med Inform Assoc 2017;24:423–31.
[45] Zerilli J, Knott A, Maclaurin J, Gavaghan C. Transparency in algorithmic and human decision-making: is there a double standard? Philos Technol 2019;32:661–83.