Department of Public Health and Caring Sciences

Ethical Informatics:
Towards an ethical framework for Clinical Decision Support Systems

Douglas Spangler

Abstract

Clinical Decision Support Systems (CDSS) can be used to predict patient outcomes, identify likely diagnoses and allocate healthcare resources. The literature investigating the ethical aspects of rule-based CDSS has identified issues relating to standards of care, the appropriate use and users of such systems, and their impact on patient-provider relationships. This paper extends this discussion to machine-learning based systems with particular attention given to their effects on distributive justice. Frameworks for examining these issues based on consequentialist, procedural and deontologoical normative theories of justice are proposed and discussed.

Introduction

Clinical Decision Support Systems (CDSS) are increasingly common in the practice of modern medicine. Such medical informatics systems are used to diagnose patients, detect medical errors, and appropriately allocate scarce resources, and CDSS have been found to improve practitioner performance in a wide range of applications (1). As such systems continue to impact human decision making, their effect on not only individual patient care, but also on the principles which the healthcare system strives uphold must be considered. The ethical issues arising from the use of CDSS have been investigated previously. The literature on these issues hes been characterized by Goodman as relating broadly to standards of care, the appropriate use and users of such systems, and their impact on patient-provider relationships. (2)

A key feature of most CDSS today is the use of formal empirical evidence in creating the rules used to predict outcomes or recommend interventions. The rise of deep learning algorithms to classify and interpret data holds both great promise and significant risks. IBMs Watson system for instance is already in use to identify likely diagnoses based on electronic health records (3). The reasoning used by this class of algorithms is difficult or impossible to interpret, and they are not immune to bias. A similar system used to evaluate the recidivism risk of parolees in Florida was, for instance, found to misclassify black defendants as “high-risk” at nearly twice the rate of their white counterparts (4). While most foreseeable CDSS applications of this technology involve the use of algorithms which are able to be externally understood and interpreted, this may not continue to be the case.

The widespread use of such sophisticated CDSS is in the mid- to long-term is indeed a likely scenario, and it would be desirable to begin developing methods for externally evaluating the ethical aspects of these systems. It is not a given that these systems will be free of bias, and important public health factors must be taken into account to ensure the continued development of a just healthcare system. The ethical development of increasingly advanced CDSS may require the explicit enumeration of the values which we desire to uphold, in a manner which can be used to constrain and direct the application of machine learning algorithms. There exist several normative theories of justice which may be applied to suggest methods for analyzing this difficult problem. Beauchamp and Childress (5) describe three important fields of thought which will be considered here, including the consequentialist, procedural and deontological approaches to characterizing justice. Each of these frameworks will be applied to propose methods for analyzing CDSS and ensuring their use results in an ethical and just distribution of healthcare.

Purpose

This paper will review the ethical issues relating to CDSS presented by Goodman and investigate the implications of advances in the field of medical informatics. These ethical questions have been examined in the literature with regards to current-generation CDSS, but little formal literature exists in regards to the ethical problems arising from the use of deep learning algorithms whose decision making processes are difficult to interpret. Normative theories of justice as described by Beauchamp and Childress (2001) will be applied in search of models which might adequately guide the resolution of these issues.

Discussion

Three main areas of ethical concern have arisen in regards to CDSS (2). The first of these is the effect of the system on the standard of care. That is to say, whether the system improves patient care and helps providers fulfill the responsibilities placed upon them by society. It may also be the case that such systems add some set of additional responsibilities to providers. In healthcare, providing a predetermined standard of care is closely linked to the problem of error avoidance. Such errors are defined in relation to an existing empirical construct, and these are themselves subject to revision over time as new evidence comes to light. In terms of the standard of care then, a CDSS may be considered to be ethical if it improves providers' ability to maintain adherence to a standard. Such adherence can be measured as the system's performance. For instance, a CDSS may be designed to use a validated pediatric dehydration score to recommend rehydration for specific acute gastroenteritis patients (6). The use of an evidence-based standard allows for the valid use of deviation from the system recommendation as a measurement of performance.

The implementation of systems using algorithms whose decision making processes cannot be interpreted must involve a redefinition of the standard of care. There is an inherent trade-off in model complexity and interpretability, and beyond a point, recommendations made by the system can only be characterized in terms of the outcome produced. As in the study referenced above, real-world outcomes in studies of CDSS are relatively difficult to identify (7). The risk of harm caused by lack of model interpretability is difficult to characterize as well. Goodman recommends the principle of progressive caution to encourage the development of advances in the field, while maintaining a robust base of evidence for the beneficence thereof. Using this principle, it may be that such uninterpretable underlying models must meet a higher standard of evidence based on outcomes under a wide range of circumstances before they can be broadly implemented with acceptable levels of risk.

The second area of concern involves the appropriate use and users of the CDSS. As with other tools in healthcare, CDSS is only effective if used in appropriate circumstances, by appropriately trained personnel for the purpose it was designed for. Goodman suggests that a control on qualifications for the use of such systems may be desirable to ensure appropriate use. Setting minimum qualifications for the use various forms of CDSS may serve not only to mitigate risks, but also to allay concerns regarding practitioner autonomy. Decisions made by rule-based CDSS are relatively simple and readily interpreted by the clinician, and can be analyzed to detect possible errors. More sophisticated models may be determined to, for instance, predict pregnancy outcomes for lupus patients with 90% accuracy (8), but if a clinician disagrees with this assessment in a clinical setting, it may be difficult to determine whether the error is hers or the algorithm's.

It is in the interaction between humans and computers where machine learning algorithms are likely to have the most short-term impact. Machine learning tools have, for instance, been effectively implemented to code medical documents (9), and perform other classification tasks peripheral to the clinical decision making process (10). Machine-learning based systems to relieve cognitive burden and improve productivity involve less risk, and are likely to precede widespread adoption of CDSS as strictly defined. While improvements in the usability of informatics systems is likely beneficial on the whole, they also involve increasing the risk of misuse of these systems by staff or patients. If not carefully designed, CDSS can cause users to overestimate the reliability and accuracy of algorithmically generated decisions.

This "Computational fallacy" contributes to Goodman's final ethical issue, that of the effect on professional relationships. Such a system must not, in Goodman's view, be allowed to unduly influence hesitant patents or undermine trust in their provider, going so far as weighing the benefits of concealing the use of these systems from patients. The risk of patients using such tools as surrogates for formal medical evaluation is considered, and is weighed against the benefits of including patients in the decision making process. In implementing deep learning algorithms designed to interface with humans, the problems of the computational fallacy are amplified. The uncertainty associated any prediction must be made clear to the user, and final decisions should be left to humans where appropriate.

In addition to the ethical issues presented by Goodman, problems relating to liability and responsibility for the appropriate use of CDSS are discussed. Goodman argues that CDSS could be conceptualized either as a service or a product, resulting in different applicable standards of liability. While services such as diagnosis by a medical doctor are held to a standard of negligence, the establishment of causality is sufficient to meet the standard of strict liability used to determine the liability of manufacturers of medical products.(11) There exists no consensus regarding the legal standing of these systems. Proponents of the "learned intermediary" doctrine hold that an adequately informed clinician may be held accountable for undetected malfunctions in a system. Another approach is to consider computer systems as agents to whom responsibility can be attributed on par with human agents. There is a need to develop sound regulation to clarify the position of these systems, while balancing the ability to ensure appropriate levels of accountability against the desire to encourage the continued design and implementation of beneficent CDSS.

As CDSS algorithms increase in complexity and become more similar to humans in decision making capacity, is seems increasingly reasonable to apply similar standards to human- and computer-generated decisions. While human clinicians may be queried in regards to the motivations behind their decisions, extracting this information from deep learning systems is non-trivial and typically limited to estimations of variable importance.(12) While obvious deviations from an established standard of care are detectable in post hoc review, it may be that in unclear cases, mistakes and appropriate assessments are indistinguishable. Given this difficulty of establishing negligence in individual cases, there is a special onus on these systems to establish their effectiveness and compliance with ethical standards at the aggregate level.

Normative theories

Of the many potential effects of implementing machine learning algorithms of the type discussed in CDSS, the undermining of the traditional method for establishing adherence to a standard of care by demonstrating compliance with evidence-based guidelines is perhaps the most fundamental. If we cannot demonstrate that a system bases decisions on some discrete set of rules which have been shown through research to be free of bias, a framework for ensuring that these systems lead to just outcomes at the population level is necessary. Beauchamp and Childress describe a number of theories relating to distributive justice, of which three were chosen based on their applicability to the topic. A consequentialist approach to evaluating the ethics of CDSS involves a focus on maximizing outcomes. This theory of justice is the most natural to apply to allocative systems, as the process of maximizing or minimizing some specified output value is the goal of most predictive algorithms. Addressing concerns of distributive justice then involves ensuring that population-level outcomes are considered when making individual-level decisions. This would involve identifying appropriate measures of distributive equity and ensuring that properties relevant to the equity measure are available in the decision-making process. For instance, with all else being equal between two potential recipients of home nursing services, we may want a CDSS to assign the resource to the patient living in the poor, underserved neighborhood for reasons of equity. Such a scheme could however be criticized for sacrificing optimal patient care in the name of a sort of affirmative action. A reliance on outcome measures may also be challenging to implement owing to the difficulty of obtaining high-quality outcome data as discussed above.

A procedural method for ensuring the ethical use of CDSS may involve a reliance on the openness and transparency of the algorithms and data used to generate decisions. While much of the research being done in the field of artificial intelligence is performed using open-source tools and datasets, this is not consistently the case with implementations of the technology in practice. The COMPAS recidivism model has for instance been validated in terms of predictive power (13), but the core algorithms are proprietary and not available for review. It may be that the risk of bias is sufficiently great that only fully transparent, reviewable systems may be tolerated. While such an approach would not directly ensure the ethical application of CDSS, it may ensure that researchers have the tools necessary to investigate ethical questions surrounding the use of these systems. Another interpretation of the procedural approach may involve designing systems to require some form of final approval by a human in a wide range of decision-making applications.

Finally, a deontological approach may involve setting certain a priori criteria for the system based on given maxims. For instance, we may desire that a CDSS make no distinction as to race in clinical decision-making. In practice, this may be accomplished by excluding variables relating to ethnicity of the patient. While satisfying, this approach involves some difficulties. By eliminating such variables from consideration, we may be negatively impacting the standard of care provided. Ethnicity is an important predictor for disease (sickle cell anemia for instance), and bias may still occur due to correlations between the "protected" variable and other non-excluded factors. The feasibility of implementing Kantian ethical models in machine learning systems at a deeper level remains an open question (14).

Conclusion

The use of CDSS based on machine learning instead of static, evidence-based rules involves an exacerbation of the issues identified by Goodman. Standards of care must increasingly come to be defined by outcomes rather than processes. As machine learning improves the usability of CDSS for a wider audience, the problem of appropriate use and users is magnified and the computational fallacy must be negated though good system design. Goodman's framework is however insufficient to evaluate population-level issues of allocative justice. Three approaches to mitigating the risk of systemic bias are described. Ensuring that indicators of bias are included as outcomes when training machine learning algorithms may be seen to constitute a consequentialist approach. Ensuring that the data and algorithms used may be independently verified to be free of bias may satisfy the proceduralist. A deontological view of ethics involves the specification of categorical imperatives, the implementation of which may lead to unintended consequences. Each of these approaches have both merit and drawbacks, and no single approach appears to be dominant. Further consideration of the ethical aspects of machine learning based CDSS is necessary, and guidelines for evaluating the particularly difficult problem of their effect on distributive justice should be developed before the technology is presented to policymakers as a fait accompli.

References

1. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of Clinical Decision-Support Systems: A Systematic Review. Ann Intern Med. 2012 Jul 3;157(1):29.

2. Goodman KW. Ethical and Legal Issues in Decision Support. In: Berner ES, editor. Clinical decision support systems [Internet]. Springer; 2007 [cited 2016 Dec 3]. Available from: http://link.springer.com/content/pdf/10.1007/978-0-387-38319-4.pdf

3. Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinforma. 2015;5:4.

4. Larson J, Mattu S, Kirchner L, Angwin J. How We Analyzed the COMPAS Recidivism Algorithm. ProPublica [Internet]. 2016 May 23 [cited 2016 Dec 3]; Available from: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

5. Beauchamp TL, Childress JF. Principles of biomedical ethics. Oxford University Press, USA; 2001.

6. Geurts D, de Vos-Kerkhof E, Polinder S, Steyerberg E, van der Lei J, Moll H, et al. Implementation of clinical decision support in young children with acute gastroenteritis: a randomized controlled trial at the emergency department. Eur J Pediatr. 2016 Dec 8;

7. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review. JAMA. 2005 Mar 9;293(10):1223–38.

8. Paydar K, Niakan Kalhori SR, Akbarian M, Sheikhtaheri A. A clinical decision support system for prediction of pregnancy outcome in pregnant women with systemic lupus erythematosus. Int J Med Inf. 2017 Jan;97:239–46.

9. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated Encoding of Clinical Documents Based on Natural Language Processing. J Am Med Inform Assoc. 2004 Sep 1;11(5):392–402.

10. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. Biomed Eng OnLine. 2014;13:94.

11. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical Decision Support Systems Could Be Modified To Reduce “Alert Fatigue” While Still Minimizing The Risk Of Litigation. Health Aff (Millwood). 2011 Dec 1;30(12):2310–7.

12. Olden JD, Joy MK, Death RG. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol Model. 2004 Nov 1;178(3–4):389–97.

13. Brennan T, Dieterich W, Ehret B. Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System. Crim Justice Behav [Internet]. 2008 Oct 20 [cited 2016 Dec 18]; Available from: http://journals.sagepub.com/doi/abs/10.1177/0093854808326545

14. Powers TM. Prospects for a Kantian Machine. IEEE Intell Syst. 2006 Jul;21(4):46–51.