Is Artificial Intelligence Better Than Human Clinicians in Predicting Patient Outcomes?

Joon Lee1,2,3, PhD

1Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
2Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
3Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Joon Lee, PhD
Data Intelligence for Health Lab
Cumming School of Medicine
University of Calgary
3280 Hospital Dr NW
TRW SE17
Calgary, AB, T2N 4Z6
Canada
Phone: 1 403 220 2968
Email: joonwu.lee@ucalgary.ca

Abstract

In contrast with medical imaging diagnostics powered by artificial intelligence (AI), in which deep learning has led to breakthroughs in recent years, patient outcome prediction poses an inherently challenging problem because it focuses on events that have not yet occurred. Interestingly, the performance of machine learning–based patient outcome prediction models has rarely been compared with that of human clinicians in the literature. Human intuition and insight may be sources of underused predictive information that AI will not be able to identify in electronic data. Both human and AI predictions should be investigated together with the aim of achieving a human-AI symbiosis that synergistically and complementarily combines AI with the predictive abilities of clinicians.

(J Med Internet Res 2020;22(8):e19918) doi: 10.2196/19918

KEYWORDS
patient outcome prediction; artificial intelligence; machine learning; human-generated predictions; human-AI symbiosis

Absence of Human-Generated Predictions in Patient Outcome Research

In recent years, there has been a proliferation of patient outcome prediction research that applies machine learning (ML) and artificial intelligence (AI) to electronic health records (EHRs) and other clinical and administrative health data. The central premises are that 1) complex health data contains predictive information that ML can effectively extract and transform into a predictive algorithm and 2) accurate prediction of patient outcomes can facilitate early, preventative intervention and more efficient health care resource allocation through identification of high-risk patients. For example, predicting which intensive care unit patients are likely to develop sepsis can prompt early initiation of fluid resuscitation, vasopressor therapy, or antibiotics, which can reduce damage from insufficient organ perfusion [1,2]. Although AI has been enormously successful in medical imaging diagnostics, where the medical condition of interest is already present or absent in the images (eg, diagnosis of diabetic retinopathy [3] and classification of skin legions [4]), patient outcome prediction poses an inherent challenge of predicting events that have not yet occurred (eg, mortality, length of stay, and readmission) [5]. This challenge is common to both AI and human clinicians.

Interestingly, while human and AI predictions are often directly compared in medical imaging research [6-8], patient outcome prediction studies tend to focus only on ML and seldom investigate human predictions. This gap in the literature is coherent across a wide range of medical specialties and diseases, including trauma [9], cancer [11], neurosurgery [10], depression [12], acute gastrointestinal bleeding [13], sepsis [14], acute liver failure [15], ischemic stroke [16], thermal injury [17], etc. This gap is corroborated by a number of systematic reviews and meta-analyses, which target only ML methods [9-14] or empirical methods [15-19].
Humans and AI Should Work as a Team

However, there is no reason to rule out the possibility that human clinicians can outperform AI in patient outcome prediction, at least in some clinical scenarios. While AI can only access information that can be recorded in the form of electronic data, human clinicians interact face-to-face with their patients and have access to both clinical and contextual information. The qualitative information collected via clinicians’ five senses can be critical in patient outcome prediction; however, this information is mostly absent in EHRs, if it is possible to record it at all. Although some qualitative observations can be recorded in EHRs as free-text notes, such as nursing notes, these data are logged in a limited, inconsistent fashion. Human intuition and insight may well be the most underused resources in patient outcome prediction.

While the performance of ML-based patient outcome prediction models appears impressive on paper, the most accurately predicted cases tend to be “easy” cases where the likely outcomes are already obvious to human clinicians [26]. This further supports the hypothesis that human clinicians perform well in patient outcome prediction.

On the other hand, AI easily outperforms humans in processing, analyzing, and finding patterns in complex, high-dimensional data [27]. As demonstrated by IBM Watson [28] and AlphaGo [29], the memory, attention, and information processing abilities of AI vastly exceed the capabilities of human cognition [30]. This AI advantage is crucial for extracting and using data-driven insights from big data [31]; it is also key to the recent successful breakthroughs in ML, particularly in deep learning [32], in a number of problem domains, including medical imaging [33]. In addition, AI does not suffer from fatigue [34] or cognitive biases (eg, recall bias) [35] as humans do. However, even if AI outperforms human clinicians in patient outcome prediction, human performance represents a more meaningful benchmark that puts AI performance in better perspective. Understanding the superiority of AI in comparison with humans can facilitate adoption of AI technology in real patient care.

The bottom line is that both AI and humans can make unique contributions to patient outcome prediction, and they should help each other to maximize predictive performance. Patient outcome prediction research should aim for human-AI symbiosis, where the respective predictive abilities of AI and human clinicians are combined in a synergistic and complementary way [36]. Given the challenging nature of patient outcome prediction, creating an AI to act alone without human help will simply lead to suboptimal predictive performance because even state-of-the-art ML technology cannot leverage information that is not present in the data [26].

Another way for AI and humans to work together is via the human-in-the-loop model, where humans directly inform machines on how to learn from the data at hand by providing guidance based on human intuition and knowledge. The term “interactive machine learning” [37] was coined to describe this paradigm; it encompasses more well-known branches of ML, such as active learning, where humans select which data points should be labelled. This human-in-the-loop approach can greatly reduce the computational complexity of some ML problems; for example, it has shown promising results in protein folding [38]. Moreover, in the field of human-computer interaction, the human-in-the-loop concept has been studied in the context of vehicle control [39], security [40,41], and decision-making [40,42]. Knowledge from these application areas can potentially inform the design of human-AI symbiosis in patient outcome prediction.

AI and human prediction performance may vary across different types of patients. Complex patterns in data can be more predictive than human intuition in certain patient subgroups, and the opposite may be true in other subpopulations. An investigation of how AI and human predictions can be optimally combined for different types of patients could directly contribute to advancing precision medicine. A better understanding of the respective predictive powers of AI and humans in various clinical scenarios can also help increase human trust in AI (eg, “For this type of patient, I need to trust AI more because most predictive information is buried in the complex data”). This can facilitate evidence-based adoption of AI technology.

For human clinicians to completely trust AI, it is necessary to understand why an algorithm arrives at a given conclusion; this requires transparency, traceability, and causality. The active field of explainable AI has been producing useful methods, such as SHapley Additive exPlanations (SHAP) [43], that can help explain how ML models work at an algorithmic level (this explanation is almost always based on correlation rather than causation); however, human clinicians ultimately want to elevate this algorithmic explainability to a model that is understandable by humans with sufficient causal understanding, also known as causability [44]. Therefore, mapping explainability to causability will be key in achieving true human-AI symbiosis.

One major roadblock to the proposed human-AI symbiosis is the need to collect a large number of human predictions in a variety of clinical scenarios, which is labor-intensive and adds...
to clinicians’ workloads. Seamlessly integrated electronic prediction collection platforms (eg, embedded in a multi-center EHR system) can minimize this burden and enable large-scale prediction collection.

**From Patient Outcome Prediction to Real Impact**

Once predictive performance is optimized via human-AI symbiosis, the next important step is to formulate clinical guidelines so that the predictive information is actionable. This is a crucial step, as accurate predictions alone will not lead to any real impact; rather, the combination of accurate predictions and appropriate interventions by clinicians will have a greater effect [5,26].

The ultimate goal of patient outcome prediction is to improve patient outcomes and decrease health care costs through early intervention and efficient use of health care resources. To prove that this goal has been met, we will need to perform randomized clinical trials of AI-driven patient care [45], such as that conducted by Wijnberge and colleagues [46]. In addition to simply comparing AI with human work alone, these randomized clinical trials should investigate a promising third species: human-AI symbiosis.

**Acknowledgments**

The author would like to thank the University of Calgary for institutional support.

**Conflicts of Interest**

None declared.

**References**


Abbreviations
- **AI**: artificial intelligence
- **APACHE**: Acute Physiology And Chronic Health Evaluation
- **EHR**: electronic health record
- **ML**: machine learning
- **SHAP**: SHapley Additive exPlanations

©Joon Lee. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 26.08.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.