



Explainable Artificial Intelligence for Predictive Modeling in Healthcare

Christopher C. Yang¹

Received: 1 September 2021 / Revised: 29 December 2021 / Accepted: 3 January 2022 /

Published online: 11 February 2022

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

The principle behind artificial intelligence is mimicking human intelligence in the way that it can perform tasks, recognize patterns, or predict outcomes through learning from the acquired data of various sources. Artificial intelligence and machine learning algorithms have been widely used in autonomous driving, recommender systems in electronic commerce and social media, fintech, natural language understanding, and question answering systems. Artificial intelligence is also gradually changing the landscape of healthcare research (Yu et al. in *Biomed Eng* 2:719–731, 25). The rule-based approach that relied on the curation of medical knowledge and the construction of robust decision rules had drawn significant attention in diagnosing diseases and clinical decision support since half a century ago. In recent years, machine learning algorithms such as deep learning that can account for complex interactions between features is shown to be promising in **predictive modeling in healthcare** (Deo in *Circulation* 132:1920–1930, 26). Although many of these artificial intelligence and machine learning algorithms can achieve remarkably high performance, it is often difficult to be completely adopted in practical clinical environments due to the lack of explainability in some of these algorithms. Explainable artificial intelligence (XAI) is emerging to assist in the communication of internal decisions, behavior, and actions to health care professionals. Through explaining the prediction outcomes, XAI gains the trust of the clinicians as they may learn how to apply the predictive modeling in practical situations instead of blindly following the predictions. There are still many scenarios to explore how to make XAI effective in clinical settings due to the complexity of medical knowledge.

Keywords Explainable artificial intelligence · Predictive modeling · Healthcare informatics · Human–computer interactions

✉ Christopher C. Yang
chris.yang@drexel.edu

¹ College of Computing and Informatics, Drexel University, Philadelphia, USA

1 Predictive Modeling in Healthcare

Digital transformation has speeded up predictive modeling in healthcare in areas such as patient deterioration, readmissions, mortality, documentation improvement, disease recognition, end-of-life care, patient movement, and chronic care management. The advance in cloud computing, big data, and the Internet of Things enable us to aggregate data from various sources, provide large-scale computing and storage of shared resources, connect, and exchange data through devices with sensors, software, and other technologies. The success of artificial intelligence in various applications creates many autonomous systems and accurate predictions for decision support. Predictive modeling in healthcare includes predicting disease states and trajectories of diabetes, predicting the survival of sepsis patients using the simultaneity of organ dysfunctions, predicting hospital readmission, predicting the adverse drug reactions and drug-drug reactions, and identifying the off-label uses of drugs.

Predictive modeling in healthcare can be complex, unintuitive, and often hard to explain. Because of these characteristics, predictive solutions have limited adoption and utility in the clinical environment due to their opaque nature in the healthcare setting. Predictive solutions in healthcare need an effective approach to enhance their explainability. Enhancing model explainability may lead to a better understanding of the model, increased utility of its output, and better patient care outcomes.

Artificial intelligence can assist physicians to make better clinical decisions or even replace human judgment in certain functional areas of healthcare such as radiology [1]. The advanced artificial intelligence algorithms can unlock clinically relevant information hidden in the large volume of healthcare data guided by relevant clinical questions formulated with the health care professionals. In the clinical trials of most medical studies, hypotheses or specific questions are clearly defined. The artificial intelligence algorithms are trained to predict certain outcomes given a set of features and insights can be developed through the prediction. Recent advance in deep learning has drawn significant attention. Deep learning is a neural network with multiple hidden layers so that more complex nonlinear patterns can be explored to improve prediction performance. The challenges are the interpretation of the prediction outcomes, explaining the process of the artificial intelligence algorithms in the predictions. Unfortunately, the explainability of deep learning is low, and it is difficult to trace. It is often referred to as an opaque model.

In this section, we first discuss the various data sources that support predictive modeling in healthcare as well as their advantages and challenges. We also provide some examples in healthcare predictive modeling. In the next section, we discuss the importance of explainability in artificial intelligence to gain the trust of health care professionals on predictive modeling. A good AI system needs to include an explanation model to communicate the internal decisions, behaviors, and actions to the interacting humans. We discuss the information-based explanation and instance-based clarification in explainable AI for healthcare predictive modeling.

1.1 Healthcare Data

Multiple sources of healthcare data have been widely used in predictive modeling. Some healthcare data are collected in clinical environments or generated by clinical trials and research, while some others are generated by health consumers or collected through sensor devices.

Electronic health records (EHRs) are the most often used data in predictive modeling for healthcare. EHRs are the electronic version of the patient medical history maintained by the health care providers. It covers all the key administrative clinical data and is formatted for easy retrieval and analytics. However, the integration of EHRs from different providers remains an issue. Although some open-source EHRs such as Medical Information Mart for Intensive Care (MIMIC), are available to the public, other EHRs are only available to the researchers of the health care providers or through collaboration with the providers. Missing data in EHRs is also a challenge in predictive modeling because the clinical data in EHRs are only recorded when the patients visit the clinics or hospitals.

Clinical trial data is a collection of data generated from clinical studies. Participants in clinical trials receive specific interventions or treatments according to the research protocol. The outcomes of the interventions are measured and recorded. Clinical trials are often used in precision medicine to identify variables that can successfully differentiate individuals who are benefitted or are harmed by a given treatment. ClinicalTrials.gov is a database of the summarized results of clinical studies provided by the US National Library of Medicine. On the other hand, clinical trials usually take a long time to collect. In addition, not everyone is eligible to participate in clinical trials. For example, children or patients with comorbidities are usually not included in clinical trials.

Scientific literature such as PubMed records from the National Library of Medicine (NLM) and MEDLINE are two popular portals that data scientists used for predictive modeling in healthcare. Natural language processing and text mining techniques are utilized to extract the medical entities and their associations. However, the results of the medical research take years to produce and publish, and therefore may not be as timely as the data available from other sources. On the other hand, the findings in the scientific literature are discovered through rigorous studies and experiments and are hence more reliable.

Health Consumer-Generated Content (HCGC) is a nontraditional data source that is drawing attention in predictive modeling for healthcare in recent years. EHRs, clinical trials, and scientific literature are collected or produced by clinicians, health professionals, or medical researchers that are believed to have high quality. However, the data is only collected when patients visit the clinics or hospitals, or the subjects are recruited in the clinical studies. Due to the popularity of online health communities and social media, health consumers are active in seeking support from their peers and sharing their firsthand experiences. The large volume of HCGC is timely, and recent research has shown that it can detect adverse drug reactions effectively [2].

Sensor data is another source of data that is collected for tracking the health conditions continuously from health consumers directly. With the advance of the

Internet of Things (IoT), many commercial wearable sensors are integrated with mobile computing. IoT wearable devices cover a wide range of different smart wearable tools such as smart watches, smart thermometers, smart helmets, smart glasses, IoT-Q-Band, etc., and many of them have been adopted in the global pandemic COVID-19 [3].

There are also other image data, genomic data, and epi data. For example, Med-Pix is a database of medical images provided by the National Library of Medicine. The National Human Genome Research Institute provides the Genome-Wide Association Studies (GWAS) to share resources for associating genetic variations with diseases. The National Health and Nutrition Examination Survey (NHANES) is an example of epi data collected for accessing the health and nutritional status of adults and children.

Other than the health consumer-generated content available in the online platforms and the open-source data, most healthcare data are only available for internal use. That means the scale of data and patient population is limited by the data available within the clinical institutions for predictive modeling. Recently, some researchers are proposing federated learning to share a global model in a central server, while the sensitive data are kept in the local institutions so that robust results can be generated across populations [4].

1.2 Examples of Predictive Modeling in Healthcare with Different Diseases and Pharmaceutical Applications

Predictive modeling with advanced artificial intelligence and machine learning algorithms is used in many different health conditions and healthcare applications. Below are some recent examples.

Diabetes is a chronic disease and requires long-term treatment management. Liu et al. [5] proposed reinforcement learning to learn and recommend sequential treatments including oral antidiabetic drugs and insulins to optimize the long-term patient outcomes of type 1 diabetes. Krieg et al. [6] proposed to use high-order networks to model complex relationship between diseases. The proposed method is used to predict the disease states and reproduce the disease trajectories for type 2 diabetes. The results show that it produces better performance than the first-order network. Zhu et al. [7] proposed a dilated recurrent neural network (DRNN) to forecast the future glucose levels of type 1 diabetes. Glushan et al. [8] developed deep learning algorithms to detect diabetic retinopathy from retinal fundus images and found that it achieved high sensitivity and specificity. The images were graded by a panel of ophthalmologists with an annotation tool.

Sepsis is a life-threatening organ dysfunction caused by the body's response to infections. The immune system releases the chemical to the bloodstream and causes inflammation throughout the entire body while fighting the infections. Jia et al. [9] adopted the deep reinforcement learning method to learn the optimal treatment strategy and investigated the safety issues associated with sepsis treatment. Yu et al. [10] proposed deep inverse reinforcement learning with the Mini-Tree (DIRL-MT) model to infer the best reward functions from a set of presumably optimal treatment

trajectories. Jazayeri et al. [11] used the network-based model to identify the simultaneity of organ dysfunction, which is useful in predicting sepsis with or without septic shocks and their survival.

The hospital readmission rate is often used as a measure of the health service outcome. The goal of the Hospital Readmissions Reduction Program (HRRP) is improving communication and care coordination to reduce avoidable readmissions. Hu et al. [12] proposed the deep learning model combining wavelet transform and deep forest to predict hospital readmission of diabetic patients. Xue et al. [13] proposed a subgraph mining-based method to analyze temporal patterns and to extract multivariate temporal trends for the prediction of ICU readmission.

Detecting adverse drug reactions through postmarketing surveillance is crucial due to the limitations of the premarketing clinical trials. It is particularly important for the patients who are excluded in the clinical trials including children and patients with severe conditions. Kim et al. [14] extracted social media data and adopted entity extractions, natural language processing, and proportional report ratio to detect the adverse drug reactions for ADHD patients who were at the preschool age. Kavuluru et al. [15] proposed the character-level recurrent neural network (Char-RNN) architectures to detect and classify DDIs from unstructured text. Yang and Yang [16] used triad prediction in heterogeneous network mining to predict drug-drug interactions. Yang and Yang [17] used association mining and temporal analysis to detect adverse drug reactions from the health consumer-generated content and was able to detect adverse drug reactions early than the FDA's alerts.

Off-label drug uses are common in medical practice. At the same time, the pharmaceutical industry is actively developing strategies to conduct drug repositioning effectively to accelerate drug development. Zhao and Yang [18] proposed the tensor decomposition and the meta-path method in heterogeneous networking for off-label drug use detecting with health consumer-generated data. Yang and Zhao [19] used the phenotypic information extracted from pharmaceutical databases and social media data and heterogeneous networking mining to identify potential drugs for repositioning.

These successful predictive modeling examples have proven that advanced artificial intelligence and machine learning algorithms are promising in unlocking clinically relevant information hidden in the large volume of healthcare data, making accurate predictions, and discovering valuable insights for various healthcare applications.

2 Explainable Artificial Intelligence

We often measure the performance of an artificial intelligence system by metrics such as sensitivity and specificity to determine if it achieves an acceptable performance in predictive modeling for healthcare. The performance is usually expected to be higher than those applied in other domains such as electronic commerce. However, from the clinical perspective, not only the predictive modeling needs to achieve high performance, but it also needs to provide explanations to gain the trust of the health care professionals, so that they can relate how the results can be adopted in

the clinical practices. Questions such as the population and the features that are being used in the healthcare models are often asked because they make a significant impact on the performance and the outcomes. There may be bias in the data collected for the prediction, and the algorithms may also perform differently with certain data properties. Elul et al. [20] identified several unmet needs of health care practitioners. These unmet needs include lack of explanations in clinically meaningful terms, coping with the unknown medical conditions, and transparency of the system's limitations. All these unmet needs stress the importance of the explainability, transparency, and interactions between the practitioners and the AI systems.

In recent years, XAI has drawn significant attention to improving the presentation of the predictive modeling results as well as providing better communication between human and AI systems. The AI systems need to include an explanation model to communicate the internal decisions, behaviors, and actions to the interacting humans. The successful explanation involves both cognitive and social processes [21]. Healthcare professionals need to understand how the AI algorithms reach a decision and how the AI algorithms behave in different scenarios through communicating with the AI systems by question answering, case analysis, illustrations using examples, and visualization of data properties and associations. Given the in-depth knowledge of the health care professionals, the communications may dig into specific medical entities and situations through continuous dialogues between human and AI systems.

The process of explanation is socio-cognitive [22]. It involves the cognitive process and the social process. The cognitive process determines an explanation for a given event. It is called the explanandum, in which the causes for the event are identified, and a subset of these causes is selected as the explanation (explanans). The social process is transferring knowledge between the explainer and the explainee through interaction. The goal is to provide the explainee with sufficient information to understand the causes of the event.

Explanations are contextual, and they cannot be achieved simply by the presentation of associations and causes. Only a small subset, which is relevant to the context, is needed to be provided to the explainee by the explainer. Without the selection process, the information can be overloaded. It may cause more confusion to the explainee rather than explaining what the explainee needs to understand the AI systems. The critical process is capturing the context when the explainee is asking questions to the AI systems. It is why the social process is so important in explainable AI. The interactions between the health care professionals and the AI systems provide the context to the AI systems to support the selection of the causes or information to offer explanans to the health care professionals. The knowledge is being transferred as the interactions continue to drill into the specific concerns of the health care professionals. The interaction is continuous, and it may involve some arguments about the explanation that may need further clarifications to ensure the right piece of information is provided. The health care professionals may express contrasting perspectives about the received information to request further explanations or provide counterexamples to understand how the AI systems may behave differently. The iterative explanations offer richer information and better satisfaction to the health care professionals as opposed to the

recommended decision, the performance report, and the overloaded information of the AI algorithms. Through the knowledge transfers, the health care professionals will be able to understand how to adopt the decisions recommended by the AI systems in clinical applications.

Deep learning has drawn significant attention in predictive modeling in recent years. With the hidden layers and the hierarchical structure in deep learning, it can achieve remarkably high performance compared with other AI algorithms if a large volume of high-quality data is available for training the neural network. However, many consider deep learning as an opaque model because it lacks an explicit declarative knowledge representation and difficult to generate the underlying explanation structure. Some researchers have attempted to provide explanations of deep learning by producing a model to approximate so that the decision is easier to understand, discovering key features to learn an interpretable model, explaining the role of single neurons or groups of neurons in encoding certain concepts, using heatmaps to highlight the most sensitive parts of inputs [23], explaining the model behavior, or using representative examples for illustrations. However, it is still difficult to explain how deep learning produces the prediction.

Liao et al. [24] developed an XAI question bank through interviewing UX and design practitioners working on AI systems to bridge the gap between the user needs in interpreting the results of predictive models and the technical capabilities of the AI systems. Ten types of questions are identified in the study, including (1) input/data, (2) output, (3) performance, (4) how, (5) why, (6) why not, (7) what if, (8) how to be that (9) how to still be this, and (10) others.

The first four types of questions are frequent questions that users try to understand the AI systems at the initial stage. The input are questions related to the dataset used in the predictive modeling. Specific questions include the data source, sample size, variables, and ground truth. The output questions are related to what the AI systems are predicting and how they can be useful. The performance questions are related to the reliability of the prediction that can be measured by metrics and how do the AI systems make mistakes. The how questions are addressing the process/logic in the prediction.

The other six types of questions are trying to drill down to the details when arguments arise or the users are trying to understand the AI systems through specific situations. The why questions are trying to address why the prediction is made when a specific instance is given. The why not questions are the opposite of the Why questions that address why the prediction is not made given a specific instance. The users may also try to understand why two instances resulted in different predictions through the why and why not questions. The what if questions are trying to understand what the AI systems are predicting if an instance is changed to what the users specify. It is another way to understand how the AI systems work when the instance is not available in the dataset, but it is important to know what the AI systems will predict. The how to be that questions are trying to understand what the minimum changes for the instance are required to obtain a different prediction. In contrast, the how to still be these questions are trying to understand what the maximum changes are allowed to obtain the same prediction. The other questions are

follow-up questions regarding how the AI systems can be improved or what the AI terminology means.

The continuous interaction between the AI systems and the health care professionals is crucial in [predictive modeling in healthcare](#). Given the initial explanations of the prediction, it allows the health care professionals to clear doubts by further interrogation with user-driven questions. The contrasting views of the health care professionals allow them to drill down specific instances for an argumentation-based interaction. We further classify the ten types of questions identified by Laio et al. (2020) into two categories, (1) information-based explanation and (2) instance-based clarification. The information-based explanation can be extracted from the documentation of the implemented predictive model. However, the instance-based clarification will need to be generated by examining the instances through executing the predictive model. In some cases, the instances are not available in the training data or testing data. By executing the predictive model with the suggested instances, it may identify the gap between the AI system and the healthcare professionals. The new instances provided by the health care professionals may further enhance the predictive modeling to achieve better performance.

Table 1 Information-based explanation questions

Input	What database is being utilized for the source of the data? What are the limitations of the utilized databases? What are the inclusion/exclusion criteria for the patient population? What are the characteristics of the patient population? What is the sample size? Is the data documentation consistent? What is the data latency? Is the data validated by clinical stakeholders? Is there any missing data? What is the ratio of training and testing data? How is the ground truth labeled?
Output	What type of output does the predictive model produce? What is the definition and logic of the output in the specific health problem? How do you interpret the presentation of the model output by the end users? What are the features that contributed to the model output? Which part of the medical workflow is the appropriate place to present the output of the model? How is the risk analysis affected by a model output?
Performance	What is the accuracy, sensitivity, specificity, and reliability of the model? How often do errors occur? What are the common errors in the model? Can the errors be identified quickly and how? How does the system monitor the model performance? How does the model perform with different clinical scenarios or patient populations? How can we improve the performance of the model?
How	How are the variables selected in the predictive model? How are the parameters in the model configured? How does the model deal with missing or inconsistent data? How can the end users access the features that contributed to a specific model output? What kind of algorithm is used? How does the model impact the current medical workflow?

Table 2 Instance-based clarification questions

Why	Why does a particular patient have this prediction? Why do patient A and patient B have the same prediction?
Why not	Why is a particular patient not predicted in this class? Why is a particular patient predicted in this class but not another class? Why do patient A and patient B have different predictions?
What if	What does the prediction model predict if there are changes in the patient (e.g., treatment)? What does the prediction model predict if there is a patient with certain features (e.g., demographic features)? What if a parameter is adjusted in a certain variable? What if a different combination of variables is utilized in the algorithm?
How to be that	What kind of patient gets a different prediction? How should the features change for this patient to get a different prediction?
How to still be this	What is the change permitted for this patient to still get the same prediction? What is the limit of the feature(s) a patient can have to still get the same prediction? What is the necessary feature(s) present or absent to guarantee the same prediction? What kind of patient gets this prediction?

We provide some examples of the questions in the information-based explanation and instance-based clarification for predictive modeling in healthcare in Table 1 and Table 2 and illustrate the dialogue between the XAI system and health care professionals in Fig. 1. Table 3 illustrates the other questions.

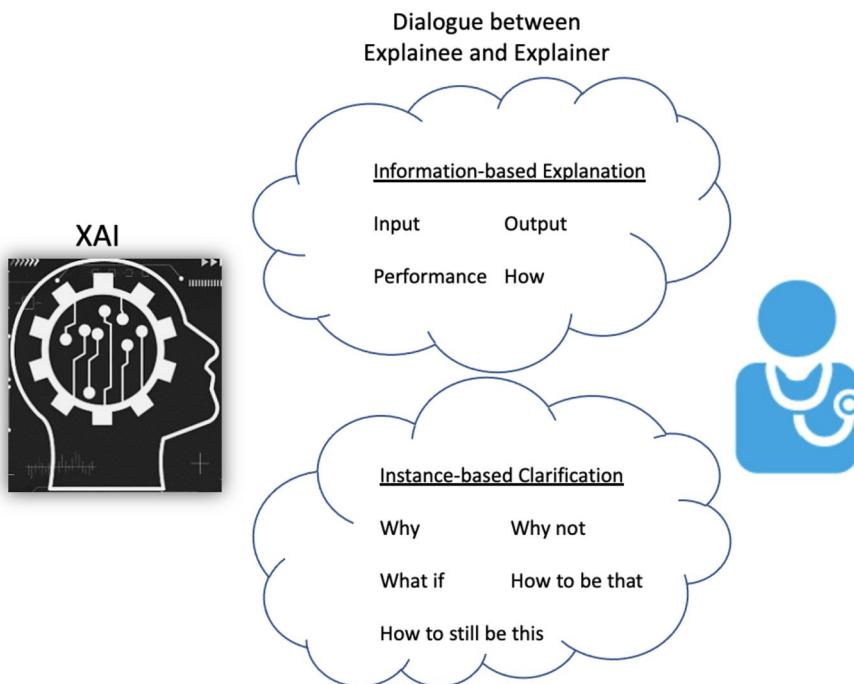


Fig. 1 Dialogue between the XAI system (explainer) and health care professionals (explainees): information-based explanation and instance-based clarification

Table 3 Other questions

Others	What is the problem the model is trying to solve? What measures can be utilized to demonstrate the impact of the model? What is the education and communication plan for the impacted users? What is the reporting strategy for the model? Are the metrics and measures of the model application defined? Is the solution scalable?
--------	--

3 Conclusion

Explainability is essential for a successful predictive modeling system in healthcare. Without transparency, it is difficult to gain the trust of health care professionals and adopt the predictive models into their daily operations. XAI has drawn significant attention in recent years. Information-based explanation questions should be addressed by the XAI system to answer questions related to the input, output, performance, and process (how) of the predictive models. Instance-based clarification questions should be integrated into the user interfaces to address questions regarding why, why not, what if, how to be that, and how to still be this when the users provide instances for clarifying the predictions. Users should be able to create instances and understand how the predictive model obtains the results. Healthcare institutions are actively developing predictive modeling systems to support their operations. XAI can be integrated to enhance the transparency of healthcare predictive modeling. The interactions between the health care professionals and the AI system are important to transfer the knowledge and adopt the models in the healthcare operations.

The pace of adopting artificial intelligence, machine learning, and data science in healthcare is accelerating. A robust link between AI and meaningful clinical and operational capabilities is imperative [27]. Lindsell et al. [27] emphasized the importance of engaging end users (including clinicians, patients, and operational leaders) at the outset of data interrogation, the iterative interaction of change-informed AI and AI-Informed change, and transforming the AI pipeline. The success of AI in healthcare not only relies on the advance in AI algorithms but also on the human in the loop involving all stakeholders.

Acknowledgements This work was supported in part by the National Science Foundation under the Grant NSF-1741306 and IIS-1650531. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y (2017) Artificial intelligence in healthcare: past, present, and future. *Stroke Vasc. Neurol* 2:e000101
2. Yang CC, Yang H, Jiang L (2014) Postmarketing drug safety surveillance using publicly available health consumer contributed content in social media. *ACM Trans Manag Inf Syst* 5:1
3. Nasajpour M, Pouriyeh S, Parizi RM, Dorodchi M, Valero M, Arabnia HR (2020) Internet of things for current COVID-19 and future pandemics: an exploratory study. *J. Healthc. Inform Res* 4:325–364. <https://doi.org/10.1007/s41666-020-00080-6>
4. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F (2021) Federated learning for healthcare informatics. *J. Healthc. Inform Res* 5:1–19
5. Liu Z et al (2020) A deep reinforcement learning approach for type 2 diabetes mellitus treatment. *IEEE Int. Conf. Healthc* 2020:1–9. <https://doi.org/10.1109/ICHI48887.2020.9374313>
6. Krieg SJ, Robertson DH, Pradhan MP, Chawla NV (2020) Higher-order networks of diabetes comorbidities: disease trajectories that matter. *IEEE Int. Conf. Healthc* 2020:1–11. <https://doi.org/10.1109/ICHI48887.2020.9374390>
7. Zhu T, Li K, Chen J et al (2020) Dilated recurrent neural networks for glucose forecasting in type 1 diabetes. *J. Healthc. Inform. Re* 4:308–324. <https://doi.org/10.1007/s41666-020-00068-2>
8. Gulshan V, Peng L, Coram M et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316(22):2402–2410. <https://doi.org/10.1001/jama.2016.17216>
9. Jia Y, Burden J, Lawton T, Habli I (2020) Safe reinforcement learning for sepsis treatment. *IEEE Int. Conf. Healthc* 2020:1–7. <https://doi.org/10.1109/ICHI48887.2020.9374367>
10. Yu C, Ren G, Liu J (2019) Deep inverse reinforcement learning for sepsis treatment. *IEEE Int. Conf. Healthc* 2019:1–3. <https://doi.org/10.1109/ICHI.2019.8904645>
11. A. Jazayeri, M. Capan, C. C. Yang, F. Khoshnevisan, M. Chi, and R. Arnold, 2019 “Network-based modeling of sepsis: quantification and evaluation of simultaneity of organ dysfunctions,” *Proceedings of ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, Niagara Falls, September 9 – 10
12. Hu P, Li S, Huang Y, Hu L (2019) Predicting hospital readmission of diabetics using deep forest. *IEEE Int. Conf. Healthc* 2019:1–2. <https://doi.org/10.1109/ICHI.2019.8904556>
13. Xue Y, Klabjan D, Luo Y (2019) Predicting ICU readmission using grouped physiological and medication trends. *Artif Intell Med* 95:27–37. <https://doi.org/10.1016/j.artmed.2018.08.004>
14. Kim H, Liang OS, Yang CC (2020) Detecting potential adverse drug reactions of preschool ADHD treatment using health consumer-generated content. *IEEE Int. Conf. Healthc* 2020:1–6. <https://doi.org/10.1109/ICHI48887.2020.9374395>
15. Kavuluru R, Rios A, Tran T (2017) Extracting drug-drug interactions with word and character-level recurrent neural networks. *IEEE Int. Conf. Healthc* 2017:5–12. <https://doi.org/10.1109/ICHI.2017.15>
16. H. Yang and C. C. Yang, 2016 “Discovering drug-drug interactions and associated adverse drug reactions with triad prediction in heterogeneous healthcare networks,” *Proceedings of IEEE International Conference on Healthcare Informatics*, Chicago, IL, October 4 – 7
17. Yang H, Yang CC (2015) “Using health consumer contributed data to detect adverse drug reactions by association mining with temporal analysis,” *ACM Trans. Intell. Syst. Technol* 6:4
18. Zhao M, Yang CC (2018) Exploiting OHC data with tensor decomposition for off-label drug use detection. *IEEE Int. Conf. Healthc* 2018:22–28. <https://doi.org/10.1109/ICHI.2018.00010>
19. Yang CC, Zhao M (May 2019) Mining heterogeneous networks for drug repositioning using phenotypic information extracted from social media and pharmaceutical databases. *Artif Intell Med* 96:80–92
20. Yonatan E, Rosenberg AA, Schuster A, Bronstein AM, Yaniv Y (2021) “Meeting the unmet needs of clinicians from AI systems showcased for cardiology with deep-learning-based ECG analysis.” *Proc. Natl. Acad. Sci* 118(24):e2020620118. <https://doi.org/10.1073/pnas.2020620118>
21. Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. 2019. A grounded interaction protocol for explainable artificial intelligence. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '19)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1033–1041.

22. Miller T (2019) Explanation in artificial intelligence: insights from the social sciences. *Artif Intell* 267(2019):1–38
23. Viton F, Elbattah M, Guérin J-L, Dequen G (2020) Heatmaps for visual explainability of CNN-based predictions for multivariate time series with application to healthcare. *IEEE Int. Conf. Healthc* 2020:1–8. <https://doi.org/10.1109/ICHI48887.2020.9374393>
24. Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. DOI:<https://doi.org/10.1145/3313831.3376590>
25. Yu K, Bean AL, Kohane IS (October 2018) “Artificial intelligence in healthcare”, *Nature*. Biomed Eng 2:719–731
26. Deo RC (2015) Machine learning in medicine. *Circulation* 132:1920–1930
27. Lindsell CJ, Stead WW, Johnson KB (2020) Action-informed artificial intelligence—matching the algorithm to the problem. *JAMA* 323(21):2141–2142. <https://doi.org/10.1001/jama.2020.5035>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.