Machine learning and medicine-A brief introduction

Benjamin Lee BS, Christopher J Peterson MD, MS

ABSTRACT

Artificial Intelligence (AI) and Machine Learning (ML) have advanced rapidly in recent years and now have the potential to change medicine. This review provides an introduction to AI and the potential it has to affect medical practice. Specific examples of past milestones particularly in the domain of critical care are presented, including ML models that can interpret chest x-rays or predict clinical outcomes such as extubation failure or ICU mortality. Included is a brief general discussion of what AI is, how it is made, and how physicians will be involved with it. Arguments are then presented as to why AI will likely not leave physicians without a job, including expectations vs. reality, that AI still requires human supervision, that new discoveries bring new challenges, and that AI cannot design itself. Far from displacing physicians, AI, if implemented well, stands poised to automate repetitive tasks, making physicians more accurate, and allowing them to spend more time with patients.

Keywords: Machine learning; artificial intelligence; medicine; technology

INTRODUCTION

The reader may have noticed the recent uptick in words like "artificial intelligence," "machine learning," and "deep learning," in both the press and academic publications. The technologies these words represent are unique and distinct but are often referred to simply under the umbrella term "artificial intelligence" (AI). The field of AI has brought forth powerful tools that have found application in many modern devices such as Tesla's self-driving cars, Apple's Siri assistant, and Google Maps. Given the broad range of fields that have found uses for AI, many in the medical field may wonder, "What will be the impact of AI on medicine?"

To understand the future of AI, it may be helpful to understand its past. The earliest discussions about AI began in the mid-twentieth century, with Alan Turing's work on computational machines. In 1956, the term "Artificial Intelligence" was coined in a seminal meeting at Dartmouth College.[1] Shortly thereafter, Frank Rosenblatt's work laid the foundation for the neural network, which has been influential and continues to shape AI even today. More recently, the image classification record set by Hinton et al. at the ImageNet competition in 2015 generated renewed interest in the AI field by introducing what is now called deep learning.[2,3] Since then, AI has been integrated into academia, industry, and everyday life, from predictive text messaging to recommending friends on social media.

Broadly defined, AI encompasses computer systems capable of performing a wide variety of tasks that usually require human-like intelligence to perform.[4] Currently, however, the field of AI could be better described as machine learning (ML), entailing systems that can learn from data with minimal human intervention and excluding more higher order intelligent functions such as emotion or self-awareness.[5] These tasks include functions like transcribing audio to text, object identification, data categorization, and recommendation systems for ads and purchases. This is the area of focus for the majority of modern research projects, many having the potential to benefit physicians. This paper will focus primarily on

Corresponding author: **Christopher Peterson**
Contact Information: Cjpeterson1@carilionclinic.org
DOI: 10.12746/swrccc.v10i45.1043

Machine Learning (ML), which is the most relevant to medicine in the near future.

Medicine has many areas amenable to ML, specifically the high volume of high-quality medical data and advanced imaging available. Electronic medical records provide databases that physicians and data scientists can use to develop new tools to assist with patient care. Machine learning has the potential to create computational tools that can detect patterns and insights in the data that physicians cannot now discern. For example, one ML model can detect breast cancer in mammograms that would otherwise be read as benign.[6] Another model was able to outperform cardiology residents in recognizing electrocardiogram abnormalities.[7] In his book *Deep Medicine*, cardiologist Eric Topol paints a vivid picture of the types of machine learning tools that physicians would most appreciate, such as Alexa-like speech recognition apps that can document encounters or place orders based on verbal commands.[8] While the potential for these tools is great, physician insight and input will be needed to both develop and implement these tools in clinical practice. This review will provide both an introduction to ML technology and a perspective on its future in medicine, particularly pulmonary critical care, as well as encourage physicians to be involved in its integration into medicine.

How does it work?

Machine learning models are comprised of trained algorithms that attempt to identify underlying patterns in data sets and make predictions accordingly. An algorithm, in the machine learning context, may be defined as the mathematical formula and architecture used to perform iterative calculations on data to train a model. Algorithms are not magic; under the hood, they resemble a web of interconnected statistical and mathematical principles used to perform calculations that can learn from patterns in the data. Algorithms vary in complexity, with one of the simplest algorithms being logistic regression, an algorithm that iteratively learns the likelihood of a binary (1 or 0) outcome, e.g., using time spent in the sun and the altitude to predict whether an individual will be sunburnt or not. More complex algorithms like Neural Networks

can identify higher-order relationships, such as combinations of pixel colors and intensities that would predict whether a photo was taken on a sunny day or a cloudy day. The model, or weights, can be thought of as a redux of what an algorithm has learned after having been trained for a specific task. In the case of the earlier logistic regression, the model would be the final weighted values one would multiply the values for time in the sun and altitude by to best predict whether an individual will be sunburnt. The accuracy of these models is then determined by how accurately they make predictions on unfamiliar data sets; most models still make numerous mistakes, but typically the model that makes the least mistakes is then used.

Physicians already use their own types of predictive models when creating diagnostic differentials. For example, a patient presenting with cough, malaise, myalgia, and fever during the winter months would be suspected to have a respiratory virus. Likewise, an ML model could also assess these and other data points to diagnose influenza. Both experienced physicians and ML models identify which combination of factors would produce the most accurate prediction and weigh them accordingly. The constellation of symptoms and patterns that are consistently predictive of an illness are often developed into a model. In clinical medicine, this takes the form of diagnostic guidelines. These guidelines can be transferred among physicians and adapted as new research further refines understanding of the underlying clinical patterns.

There are also patterns beyond established guidelines that physicians can rely on, often referred to as "intuition" or "clinical gestalt." Physicians develop this intuition based on years of experience, and these experiences are often assembled, sometimes subconsciously, into patterns that can direct clinical practice. While these cognitive "models" can be effective, they are limited by both the experience of the physician and the inability to be transferred among persons. Machine learning algorithms offer a way to analyze these types of patterns with much greater computational precision and then encapsulate these patterns with computer models, allowing others to add to and access the insights learned.

The ability of ML to deal with this complexity comes from both the large amounts of data that are fed into these algorithms as well as the flexibility of the algorithms themselves. The selection of each algorithm is crucial, as each has strengths and weaknesses, and thus may perform differently on certain types of datasets. For example, Convolutional Neural Nets (CNNs) are particularly effective at analyzing images, while Random Forest algorithms have the unique strength of generating easily interpretable flowcharts and decision trees. Several other algorithms most widely used in publications include Support Vector Machines (SVM), Artificial Neural Nets (ANN), and Bayesian Networks. In some cases, researchers have found that simpler models perform just as well at a task as bigger more computationally expensive deep learning models, allowing for analyses to be made much faster, even in real-time, rather than taking minutes to complete.[9] To interpret the results of machine learning advances in medical literature, it is important to recognize these model-specific behaviors and limitations.

Selecting an algorithm is but the first step in many. Every model is trained, tested, and reworked by human users to ensure that each matches the specific parameters of the task at hand. If the necessary conditions, assumptions, and desired outcomes are not fully understood, the results of the model performance may suffer or even fail altogether. Another limitation is overfitting, instances in which the model is too specific to features found only in the dataset it was trained on. For example, a model that erroneously predicts that all 45-year-old males named John have influenza is focusing on details that will clearly not apply to other patients. The opposite is also true; overgeneralizing that every patient with a runny nose has the "flu" is also inaccurate as there are many other causes of nasal discharge. The optimal model for predicting the "flu" will likely lie between these two extremes, and the skill of applying machine learning is being able to locate the point at which the model is specific enough to be helpful but generalizes well enough to be used on other patients.

As with statistics, results can be misleading and require scrutiny before and even after being deployed. Changing disease rates in populations may render a previously adequate model out of date. Furthermore, the complexity of medicine makes it unlikely that any model will be able to be 100% accurate, necessitating ongoing reevaluations and fine-tuning.

Machine learning and critical care medicine

As a particularly data-heavy specialty with numerous signals to monitor simultaneously, critical care may benefit highly from employing ML. Many critical care physicians have experienced signal fatigue from monitoring numerous vital signs and laboratory results from seriously ill patients. This, combined with heavy workloads, is cited as a major component of the high burnout rates in critical care physicians.[10] Machine learning applications offer a way for computers to assist in monitoring complex signal patterns and then alerting physicians where to direct their limited attention. Reliable models that can instantaneously interpret patterns will help provide a stop-gap measure to help prevent adverse outcomes that a physician might not otherwise discern. Increased accuracy in the ability to predict who will require more intensive care could aid physicians in making decisions earlier in a hospital stay to reduce the need for respiratory interventions in these individuals. Many such ML tools are already under development, with no fewer than 85 ML models specific to ICU care registered on the National Institutes of Health website (clinicaltrials.gov) as of March 2021.

Chest X-ray Interpretation

Machine Learning models have demonstrated particular strength in performing imaging-based tasks. A recent landmark model called ChexNet was trained on a database of over 100,000 frontal chest X-rays. This massive dataset had 14 different diseases labeled by 4 radiologists and was the first of its kind ever assembled. ChexNet was able to perform better than the radiologists at the task of identifying pneumonia on x-ray alone.[11] Since those results in 2017 much more work has been done in this area[12] and last year saw the FDA approve an ML model to be integrated into x-ray machines that can detect pneumothorax.[13]

Several ML models for detecting pneumonia have demonstrated high levels of accuracy (up to 99%) and AUC (99%).[14,15]

SEPSIS

The sepsis early risk assessment (SERA) model has an 87% sensitivity and specificity for predicting sepsis 12 hours before its onset, a 32% improvement in accuracy compared to physician predictions.[16] Many other models have been developed to predict sepsis[17] and sepsis-induced coagulopathy.[18] However, one sepsis alarm integrated into the Epic EMR (Epsis Sepsis Model), though shown to be effective,[19] may have nonetheless contributed to alarm fatigue during the COVID-19 pandemic due to poor implementation.[20] This example was an early attempt to integrate ML into hospital systems and provides several important lessons about model deployment that led to this failure to scale. The model was later found to have a lower AUC upon external validation than initially reported.[21] Due to the cutoffs used, it was eventually only found to catch 7% of sepsis patients missed by a physician and missed sepsis in 67% of sepsis patients despite generating alerts on 18% of all hospitalized patients.[21] This further underscores the importance of rigorously testing and scrutinizing a model before deployment, as well as having the input of physicians to make sure that the tool is clinically meaningful.

INFECTION

A model predicting methicillin resistant *Staphylococcus aureus* in mechanically ventilated patients with 98% sensitivity, 47% specificity, and a positive predictive value of 0.65 has been developed. In this study, admission from the ED was found to be the most predictive factor/feature.[22] Other models have been developed to diagnose bloodstream infections early[23] as well as predict their outcomes.[24] In a study in Israel, an ML algorithm analyzed patient data and predicted patterns of bacterial resistance, resulting in researchers' being able to reduce the number of mismatched antibiotic prescriptions for urinary tract infections by 30–40%.[25] It is possible that future models may assist providers with the detection of pathogens and the selection of appropriate antibiotics.

EXTUBATION FAILURE

A model predicted extubation failure using 89 clinical and laboratory variables as inputs performed with an AUC of .83 when trained on the MIMIC-IV dataset and areas under the receiver operating curve (AUC) of 0.8 in prospective validation on an external dataset.[26] Examining what the algorithm had learned, researchers found that mechanical ventilation (MV) duration and pressure support ventilation (PSV) levels were the most influential factors the model used in making these predictions. Another model was developed to predict prolonged mechanical ventilation with an AUC of .85 and tracheostomy placement with an AUC of .83 using six different severity of illness scores calculated on the first day of ICU admission.[27]

ICU ADMISSION AND MORTALITY

Predictors for mortality and disease severity are already employed in clinical medicine, such as the Acute Physiologic and Chronic Health Evaluation (APACHE-IV) and Simplified Acute Physiologic Score (SAPS). An interesting study used ML to not only predict overall ICU mortality with high accuracy but also to sort through the model parameters to explain which factors were most predictive of mortality.[23] These investigators were able to determine hospital mortality with an AUC of 87–91% compared to the AUC of 0.88 using APACHE-IV achieves. Machine learning identified factors like a significant increase in creatinine level was likely to precede mortality, even though this had been considered an unimportant factor. The model was also able to detect time-dependent patterns, such as an increase in glucose level commonly seen 16 hours prior to death. These are signals that are too complex for humans to monitor, but these models can and do excel at continuously monitoring and alerting physicians that the signal is more than a simple upward or downward trend.[28] One study suggested that models may become more predictive with greater length of stay,[29] possibly due to more data or selection for patients that have more severe illnesses.

In contrast, another study found ideal mortality prediction at approximately 2 days into ICU admission,[30] although this may be due to the frequency of data sampling provided to the model or fewer patients with longer ICU stays.[29,30] Predictors of mortality from specific pathologies, such as sepsis,[31] heart failure,[32] and acute kidney injury[33] have also been studied.

ACUTE RESPIRATORY DISTRESS SYNDROME (ARDS)

A CNN model predicted ARDS using chest x-rays with an 83% sensitivity and 88% specificity on a group of 413 images that were reviewed by 6 physicians.[34] One study found that accuracy in predicting ARDS severity was improved when using the patient condition on ICU day two rather than day one (as described in the Berlin criteria) and that a predictive model using $PaO_2/FiO_2 x PEEP$ index was superior to those using PaO_2/FiO_2.[35] Another model identified more cases of ARDS than clinicians (73.5% v 33.2%), but in cases in which clinicians did diagnose ARDS, the diagnosis was made earlier.[36]

COVID-19

Researchers were able to use a decision trees model to determine which factors were most predictive of ARDS due to COVID-19 in a sample of 600 patients. They discovered that age and BMI over 25 were the most predictive, but also discovered creatine kinase and neutrophil to lymphocyte ratio to be important novel predictors. The model performed with an overall AUC of .99 when tested on an external dataset.[37]

RISK STRATIFICATION

Lehman and colleagues used algorithms called Hierarchical Dirichlet Processes (HDP) that were designed to identify clusters of topics in unstructured progress notes to stratify ICU patients from the MIMIC-II database according to the risk of mortality. They found that including topics identified in nursing documentation of the first 24-hours of the ICU stay greatly improved the accuracy of the SAPS-1 model from an AUC of .72 to .82.[38]

LENGTH OF STAY AND READMISSION

Length of stay has also been studied to predict the length of stay in the context of traumatic brain injury (TBI),[39] predictions were also made using patient vital signs[40] as well as a gradient boosted decision trees algorithm trained using data available in the eICU and MIMIC III datasets.[41] A model was able to predict readmission from information in the MIMIC III dataset significantly better than both the Stability and Workload Index for Transfer score (AUC = .65) and the Modified Early Warning Score (AUC = .58), with an AUC of .76.[42] While more work is clearly needed, these studies nonetheless provide a proof of concept for how future versions of these models could benefit intensivists.

HOW PHYSICIANS CAN GET INVOLVED

The physician, though typically not trained in computer science, has a crucial role as both producer and interpreter of clinical data. Clinical datasets containing imaging and documentation on an individual patient level, are in high demand for ML. Machine learning projects typically require thousands of images or subjects, with algorithms typically performing better when given more subjects and when given more granular details about those subjects. For example, the ChexNet model was trained on a dataset of over 100,000 chest x-rays furnished by Stanford University physicians.[6] Publicly available databases, like the one Stanford produced, are key ways in which physicians can provide researchers with de-identified, high-quality labeled data. The Medical Information Mart for Intensive Care (MIMIC-IV),[42] operated by the Massachusetts Institute of Technology (MIT), is a database of de-identified data from over 40,000 patients admitted to the Beth Israel Deaconess medical center in Boston, MA. This and previous MIMIC databases have already proven an invaluable resource for studying care in the ICU.[43] Future databases like these will need to be collected under the guidance of physicians, who are uniquely positioned to ensure both their accuracy and clinical relevance.

In addition to data collection efforts, physicians are also needed to guide the ML projects themselves.

Both physicians and engineers are experts in their own fields, but typically lack sufficient understanding to effectively develop ML for a field like medicine. Physicians are needed to both identify relevant problems and effectively integrate ML tools into healthcare. Only by working together will both groups be able to improve technology in medicine, and nowhere is this more apparent than with ML. Only with good, labeled data and guidance from physicians will research teams and companies be able to build the tools that medicine most needs. As the primary end-users of many of these models, physicians will also be needed to monitor the performance and the accuracy of these models with prospective studies,[44] and give necessary feedback.

For those interested in a more thorough introduction, we recommend starting with *deep Medicine* by Eric Topol,[8] *Intelligence-Based Medicine* by Anthony Chang,[45] and reviews by Sanchez-Pinto et al.[46] For more formal training in AI (or ML), the American Board of Artificial Intelligence in Medicine (ABAIM) offers introductory courses in medicine-based machine learning as well as a network of other physicians interested in ML.

WILL IT TAKE MY JOB?

One of the most often expressed concerns about AI, ML, and technology in general is how it will affect (or replace) various occupations. History shows that technology drastically altered agricultural and manufacturing jobs in the US. Indeed, many cite the oncoming shift to AI as a fourth industrial revolution[47] and some fear a similar outcome in medicine with AI. However, physicians' being fully replaced by AI is unlikely for several reasons.

EXPECTATIONS VS. REALITY

Medicine has experienced technological shifts before and has often leaned toward sensationalism when predicting the end effect. When the MRI was first developed, some predicted that it would make radiologists obsolete; clearly, this has not been the case.[48] Other predictions, from the cure for the common cold[49]

to a rapidly developed HIV vaccine,[50] have been postulated and proven inaccurate. As far back as the 1960s, AI researchers were making wild predictions that within their lifetimes all scientific problems would be handed over to AI and that human labor would be obsolete.[51] While the future is not always clear, it is often less spectacular than predicted, especially in a complex field like medicine in which fundamental interactions among humans will likely prove themselves irreplaceable in the healing process.

AI REQUIRES HUMAN SUPERVISION

Despite its complexity, AI is ultimately a tool and, like other tools, requires human input. While AI can identify complex patterns in data, it requires humans to determine whether or not these patterns are useful or clinically relevant. Furthermore, AI is still error-prone and will likely always require supervision from a trained user to interpret the data within the broader landscape of clinical knowledge. For example, by changing only a small number of pixels, MIT students were able to trick a Google image search into labeling a picture of a cat as "guacamole."[52] Machine learning algorithms are also highly susceptible to biases, and constant evaluation is necessary to eliminate these from the dataset to prevent training these into the model. For example, a model used to detect skin cancer was more likely to interpret the image as malignant if a ruler was present in the image.[53] Rulers proved to be present more often in cases in which the dermatologist had concern for cancer, and thus the physicians' interpretations were perpetuated through a bias in the dataset.[53] Similar confounders have been observed in other models for melanoma,[54] pneumonia,[55] and hip fractures.[56]

Care must also be taken especially before deploying models to make sure that they will generalize well to represent the population at large. Physicians ultimately will be the only ones able to identify some of these biases. External validation and scrutiny of models become increasingly important to ensure that confounders and biases are mitigated, especially considering that much medical data is dependent on physicians who may or may not follow best practices.[44] Patients will look to physicians to understand what to

do with the information the models put out. Artificial intelligence is not aptly placed to address ethical issues, build rapport, and empathize with patients. This is and likely will always be reserved for people and not machines. Though models may aid in decision-making, the final decision and responsibility will, as always, rest with the physician.

New Discoveries Bring New Challenges

Just as the invention of MRI did not solve all of radiology's challenges, AI discoveries may simultaneously solve some problems and lead to other much deeper, more complicated questions. Artificial intelligence may liberate physicians from the burden of monotonous tasks like reading chest X-rays and EKGs, but the fact remains that the compendium of medical knowledge, though prodigious, is incomplete. Discoveries driven by AI will likely reveal a veritable hydra of medical conundrums that still await proper attention, e.g., one AI study identified three new potential multiple sclerosis subtypes.[57] Addressing these discoveries may enhance the way even field experts practice medicine. For example, Lee Sedol, the world's premier player of the traditional board game Go, was recently beaten by Google's milestone AlphaGo program that employed ML algorithms. Although he lost 4 of 5 rounds to the computer, Sedol was able to win one round with a novel strategy catalyzed by his struggle with his AI opponent, and this changed the way the game is played by professionals today.[58] Similarly, human chess players assisted by AI, a so-called "Centaur" player, proved more effective than either humans or AI programs alone.[59] The same has and will likely continue to be true for medicine as physicians learn to use new tools.

Machine learning cannot design itself

Machine learning algorithms are still very much dependent upon a human operator to carefully design the algorithm architecture. While it is true that ML involves an algorithm improving its performance on a specific task with minimal human input, human input is still crucial. An algorithm can only improve at the task that it has been set to, and, while programs have been able to outperform humans at certain tasks,

the program must be told what to do by an operator. Indeed, as in statistics, it would be unexpected for a statistics program to select the appropriate statistical test, ask the right research question, or even collect the research data automatically. A data scientist is required to select the right algorithm to investigate the specific question, and then through a process called data cleaning ensure that the data will run through the algorithm correctly.[57] As entropy would suggest, even once established, there is a continual need to fine-tune and update models, much as how Microsoft or Apple products receive updates. These adjustments do not happen automatically and require work from developers and new data from those using these models. Finally, for those concerned over a potential fully autonomous and self-aware AI, given the extremely limited understanding of human cognition, it is unreasonable to expect humans to design a machine with capabilities that exceed our current understanding. As physicist and philosopher Ragnar Fjelland observed, "The overestimation of technology is closely connected with the underestimation of humans."[60]

Conclusions

Machine learning has potential to augment clinical medicine. Though complex, it is based on principles familiar to and practiced by physicians. Furthermore, its effectiveness depends on the supervision and contribution of physicians who can bridge the gap between data engineers and the patient. Like any tool, ML will require training and adaptation but can confer a significant advantage to those who use it. Its effective integration into medicine will depend not only on technological advancements but on physicians who will develop and use it in daily practice. Though unfamiliar, the capacity for ML to improve medical practice and patient care is increasing and should encourage physicians in integrating this capability into medicine.

Acknowledgments

REFERENCES

1. Haenlein M, Kaplan A. A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. California Management Review 2019;61:5–14.
2. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 2012;25.
3. Tappert CC. Who is the father of deep learning? 2019 International Conference on Computational Science and Computational Intelligence (CSCI); 2019: IEEE. p. 343–8.
4. Dennis MA. Marvin Minsky. Encyclopedia Britannica 2022.
5. Machine learning. Oxford Learner's Dictionaries: Oxford University Press; 2022.
6. Yala A, Mikhael PG, Lehman C, et al. Optimizing risk-based breast cancer screening policies with reinforcement learning. Nature Medicine 2022;28:136–43.
7. Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications 2020;11:1760.
8. Topol E. Deep medicine: how artificial intelligence can make healthcare human again: Basic Books, Inc.; 2019.
9. Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. Information Fusion 2022;81:84–90.
10. Moss M, Good VS, Gozal D, et al. A critical care societies collaborative statement: burnout syndrome in critical care health-care professionals. A call for action. American Journal of Respiratory and Critical Care Medicine 2016;194: 106–13.
11. CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning. 2017. 2022,
12. Çallı E, Sogancioglu E, van Ginneken B, et al. Deep learning for chest x-ray analysis: a survey. Medical Image Analysis 2021;72:102125.
13. Bai N. Artificial intelligence that reads chest x-rays is approved by FDA. University of California San Francisco https://www.ucsf.edu/news/2019/09/415406/artificial-intelligence-reads-chest-x-rays-approved-fda. Accessed 10/4/2022.
14. Elshennawy NM, Ibrahim DM. Deep-pneumonia framework using deep learning models based on chest x-ray images. Diagnostics (Basel) 2020;10:649.
15. Kundu R, Das R, Geem ZW, et al. Pneumonia detection in chest x-ray images using an ensemble of deep learning models. PLoS One 2021;16:e0256630.
16. Goh KH, Wang L, Yeow AYK, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. Nature Communications 2021;12:711.
17. Moor M, Rieck B, Horn M, et al. Early prediction of sepsis in the ICU using machine learning: a systematic review. Frontiers in Medicine 2021;8.
18. Zhao QY, Liu LP, Luo JC, et al. A machine-learning approach for dynamic prediction of sepsis-induced coagulopathy in critically ill patients with sepsis. Front Med (Lausanne) 2020;7:637434.
19. Bennett T, Russell S, King J, et al. Accuracy of the epic sepsis prediction model in a regional health system. arXiv preprint arXiv:190207276 2019.
20. Wong A, Cao J, Lyons PG, et al. Quantification of sepsis model alerts in 24 US hospitals before and during the COVID-19 pandemic. JAMA Network Open 2021;4:e2135286-e.
21. Habib AR, Lin AL, Grant RW. The epic sepsis model falls short—the importance of external validation. JAMA Internal Medicine 2021;181:1040–1.
22. Hirano Y, Shinmoto K, Okada Y, et al. Machine learning approach to predict positive screening of methicillin-resistant *Staphylococcus aureus* during mechanical ventilation using synthetic dataset from MIMIC-IV database. Frontiers in medicine 2021;8:694520-.
23. Roimi M, Neuberger A, Shrot A, et al. Early diagnosis of bloodstream infections in the intensive care unit using machine-learning algorithms. Intensive Care Med 2020;46: 454–62.
24. Zoabi Y, Kehat O, Lahav D, et al. Predicting bloodstream infection outcome using machine learning. Scientific Reports 2021;11:20101.
25. Yelin I, Snitser O, Novich G, et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. Nature Medicine 2019;25:1143–52.
26. Zhao QY, Wang H, Luo JC, et al. Development and validation of a machine-learning model for prediction of extubation failure in intensive care units. Front Med (Lausanne) 2021;8:676343.
27. Parreco J, Hidalgo A, Parks JJ, et al. Using artificial intelligence to predict prolonged mechanical ventilation and

tracheostomy placement. Journal of Surgical Research 2018;228:179–87.

28. Na Pattalung T, Ingviya T, Chaichulee S. Feature explanations in recurrent neural networks for predicting risk of mortality in intensive care patients. J Pers Med 2021;11.

29. Thorsen-Meyer H-C, Nielsen AB, Nielsen AP, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. The Lancet Digital Health 2020;2:e179–e91.

30. Meiring C, Dixit A, Harris S, et al. Optimal intensive care outcome prediction over time using machine learning. PLoS One 2018;13:e0206862.

31. Kong G, Lin K, Hu Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. BMC Medical Informatics and Decision Making 2020;20: 251.

32. Li F, Xin H, Zhang J, et al. Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database. BMJ Open 2021;11:e044779.

33. Liu J, Wu J, Liu S, et al. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. PLoS One 2021;16:e0246306.

34. Sjoding MW, Taylor D, Motyka J, et al. Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. The Lancet Digital Health 2021;3:e340–e8.

35. Sayed M, Riaño D, Villar J. Novel criteria to classify ARDS severity using a machine learning approach. Critical Care 2021;25:150.

36. You JY, Gong M, Chen J-T. Machine learning algorithm increases recognition of ARDS. CHEST 2020;158:A583.

37. Xu W, Sun N-N, Gao H-N, et al. Risk factors analysis of COVID-19 patients with ARDS and prediction based on machine learning. Scientific Reports 2021;11:2933.

38. Lehman L-w, Saeed M, Long W, et al. Risk stratification of ICU patients using topic models inferred from unstructured progress notes. AMIA Annu Symp Proc 2012;2012:505–11.

39. Abujaber A, Fadlalla A, Nashwan A, et al. Predicting prolonged length of stay in patients with traumatic brain injury: a machine learning approach. Intelligence-Based Medicine 2022;6:100052.

40. Alghatani K, Ammar N, Rezgui A, et al. Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation. JMIR Med Inform 2021;9:e21347.

41. Wu J, Lin Y, Li P, et al. Predicting prolonged length of ICU stay through machine learning. Diagnostics (Basel) 2021; 11:2242.

42. Johnson A, Bulgarelli L, Pollard T, et al. MIMIC-IV (version 1.0). *PhysioNet*.

43. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Scientific Data 2016;3: 160035.

44. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. New England Journal of Medicine 2019;380:1347–58.

45. Chang AC. Intelligence-based medicine: artificial intelligence and human cognition in clinical medicine and healthcare: Academic Press; 2020.

46. Sanchez-Pinto LN, Luo Y, Churpek MM. Big data and data science in critical care. Chest 2018;154:1239–48.

47. Ross P, Maynard K. Towards a 4th industrial revolution. Intelligent Buildings International 2021;13:159–61.

48. Langlotz CP. Will artificial intelligence replace radiologists? Radiol Artif Intell 2019;1:e190058-e.

49. Samuel L. A cure for the common cold and 4 other medical predictions that haven't (yet) come true. StatNews. https://www.statnews.com/2017/02/27/medical-predictions/ Accessed 10/4/2022.

50. Watkins DI. The vaccine search goes on. Scientific American 2008;299:69–77.

51. Muehlhauser L. What should we learn from past AI forecasts. Open Philanthropy Project 2016.

52. Athalye A. Fooling neural networks in the physical world with 3D adversarial objects. LabSix, October 2017;31.

53. Narla A, Kuprel B, Sarin K, et al. Automated classification of skin lesions: from pixels to practice. Journal of Investigative Dermatology 2018;138:2108–10.

54. Winkler JK, Fink C, Toberer F, et al. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. JAMA Dermatology 2019;155:1135–41.

55. Badgeley MA, Zech JR, Oakden-Rayner L, et al. Deep learning predicts hip fracture using confounding patient and healthcare variables. npj Digital Medicine 2019;2:31.

56. Zech JR, Badgeley MA, Liu M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLOS Medicine 2018;15:e1002683.

57. Eshaghi A, Young AL, Wijeratne PA, et al. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. Nature Communications 2021;12:2078.

58. Kohs G. AlphaGo. Moxie Pictures; 2017.

59. Case N. How to become a centaur. Journal of Design and Science 2018.

60. Fjelland R. Why general artificial intelligence will not be realized. Humanities and Social Sciences Communications 2020;7:10.