


Review

Artificial intelligence and machine learning in emergency medicine: a narrative review

Brianna Mueller,¹ Takahiro Kinoshita,² Alexander Peebles,³ Mark A. Graber,³ and Sangil Lee³ 

¹Department of Business Analytics, The University of Iowa Tippie College of Business, Iowa City, Iowa, USA, ²Philips Research North America, Cambridge, Massachusetts, USA, and ³Department of Emergency Medicine, The University of Iowa Carver College of Medicine, Iowa City, Iowa, USA

Aim: The emergence and evolution of artificial intelligence (AI) has generated increasing interest in machine learning applications for health care. Specifically, researchers are grasping the potential of machine learning solutions to enhance the quality of care in emergency medicine.

Methods: We undertook a narrative review of published works on machine learning applications in emergency medicine and provide a synopsis of recent developments.

Results: This review describes fundamental concepts of machine learning and presents clinical applications for triage, risk stratification specific to disease, medical imaging, and emergency department operations. Additionally, we consider how machine learning models could contribute to the improvement of causal inference in medicine, and to conclude, we discuss barriers to safe implementation of AI.

Conclusion: We intend that this review serves as an introduction to AI and machine learning in emergency medicine.

Key words: Artificial intelligence, deep learning, emergency medicine, machine learning, prediction

INTRODUCTION

IN RECENT YEARS, advancements in artificial intelligence (AI) technologies have resulted in the rapid growth of machine learning (ML) research in medicine.¹ Specifically, the development of unprecedented ML applications has shown great potential to significantly impact the field of emergency medicine. These applications address prevailing challenges in the emergency department such as triage and disposition, early detection of conditions and outcomes, emergency department operations, and therapeutic intervention. With the increasing availability of clinical data, it is exceedingly advantageous for emergency medicine clinicians to understand computational techniques like ML that are able to meaningfully process large quantities of complex data. This review aims to provide a conceptual introduction

to AI/ML and increase awareness of emerging clinical tools derived from ML methods. We present examples of ML models used in clinical research and highlight recent applications in the field of emergency medicine. Specifically, we focus not only on predictive studies, which the vast majority of ML research has targeted until the present, but also on causal inference studies because the goal of clinical research is often determining the effects of interventions on clinical outcomes. To conclude, we discuss challenges to the implementation of AI and consider reasons why only a few ML solutions have been applied in actual clinical practice despite the proliferation of applications in clinical literature. By examining barriers of clinical adoption, we also intend for this review to encourage more discussion on how to practically address these concerns and integrate machine learning into routine clinical operations.

Fundamentals of AI, ML, and deep learning

In the 1950s, Stanford Professor John McCarthy coined the term artificial intelligence as “the science and engineering of making intelligent machines”. Artificial intelligence enables machines to imitate human cognitive functions

Corresponding: Sangil Lee, MD, MS, Department of Emergency Medicine, The University of Iowa Carver College of Medicine, 200 Hawkins Dr, Iowa City, IA 52242. E-mail: sangil-lee@uiowa.edu.

Received 28 Nov, 2021; accepted 6 Feb, 2022

Funding information

No funding information provided.

such as problem-solving and learning. Machine learning is a branch of AI focused on leveraging data to develop computer systems that are able to learn and improve from experience without being explicitly programmed.² Statistical methods and algorithms are used to recognize patterns and learn relationships from data in order to build models capable of making predictions or decisions. Machine learning algorithms fall into three main categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised ML is defined by the use of labeled data to learn a mapping between input variables and an outcome variable of interest (e.g., positive diagnosis or negative diagnosis). The process of developing a supervised model involves three datasets. The algorithm first learns on the training dataset by adjusting weights to minimize a loss function that computes the distance between the predicted outcome and the true outcome for a given data point. After fitting the model, a validation set is used for optimization through tuning model parameters. The validation set can also detect overfitting, which is observed when model performance is significantly better on the training set. Finally, a test set is used to provide an estimate of how well the model can generalize to new data.

Conversely, unsupervised learning refers to methods that use unlabeled data to find naturally occurring groups or clusters. These clusters are analyzed to identify similarities and differences between data points, and understand the

distribution of data in the feature space. In reinforcement learning, a computer agent learns to achieve a goal in an interactive environment by trial and error. Unlike supervised learning where data labels serve as model feedback, reinforcement learning uses rewards or penalties as feedback based on the actions the agent performs. Over time, the agent learns action sequences that maximize the reward.

Deep learning is a subfield of ML that has gained massive popularity in health care the past few years due to its success on a variety of complex classification tasks.³ This can partly be attributed to increases in computational power and access to ever-growing amounts of data. Inspired by the structure and function of neurons in the cerebral cortex, a neural network is the backbone of deep learning algorithms. A neural network architecture consists of layers of interconnected nodes that are analogous to neurons (Fig. 1). In a process called forward propagation, data is fed into an input layer and flows through the system of hidden nodes connected by weights. The input to each node is a weighted linear combination of node outputs from the previous layer and a nonlinear transformation is applied to the node's output. A loss function evaluates the difference between the predicted value from the output layer and the true value. An optimization algorithm called backpropagation uses the prediction error to iteratively adjust the weights to learn the structure of the training data, gradually improving model accuracy.

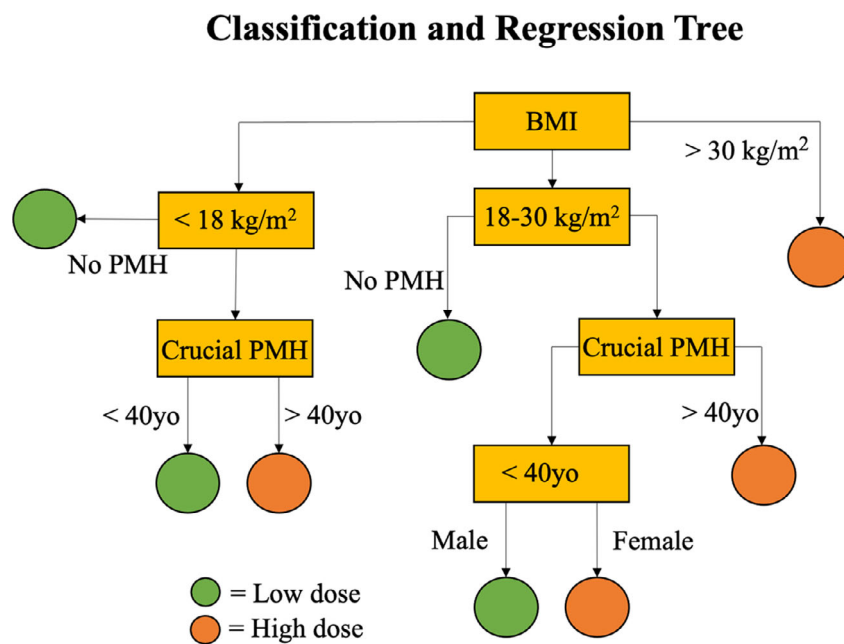


Fig. 1. Artificial neural network, the basis of deep learning algorithms.

Factors that distinguish deep learning from traditional ML

One of the primary ways ML and deep learning techniques differ is in the learning process. Deep learning algorithms have the ability to automatically learn feature hierarchies, whereas traditional ML algorithms require manual feature extraction and engineering. Second, deep learning often requires a considerable amount of data to make predictions, whereas traditional ML methods could reach a level where model performance no longer scales with the amount of data. Another major difference between the two techniques is the execution time. Due to the large number of parameters to learn, deep learning methods often take significantly longer to train. Finally, deep learning methods involve a large number of matrix multiplication operations, which can result in a heavy dependence on high-end machines.

TYPES OF MACHINE LEARNING ALGORITHMS IN CLINICAL RESEARCH

ONE OF THE most interpretable ML models used in clinical research is a decision tree. A decision tree is a supervised learning algorithm structured like a flowchart that can be used for both classification and regression tasks. The goal is to develop a model that can be used to predict the target variable of future instances based on a set of decision rules. The algorithm recursively partitions the data into subsets (decision nodes) based on the value of the feature that reduces the impurity of the resulting subsets the most. A node is considered to be “pure” if all of the data points in the node are of the same class. When a node contains equal percentages of each class, impurity is maximized. Figure 2 shows an example of a decision tree to predict medication dosage. The first split is based on body mass index (BMI), indicating BMI is the best predictor for dosage level. Each of the final subsets (leaf nodes) is assigned class membership probabilities for each data point in the node.

Ensemble learning is a ML method that aims to increase accuracy and reduce variance by combining multiple algorithms. A random forest is a commonly used ensemble model that aggregates the outputs of several decision trees to make a single prediction. In general, ensembles have higher predictive power than their constituents do individually. Other widely used algorithms include linear regression, naive Bayes, support vector machine (SVM), k-nearest neighbors, and various ensemble methods such as gradient and adaptive boosting.

Among breakthroughs in deep learning, the convolutional neural network (CNN) has gained significant attention from researchers because of its high performance in computer

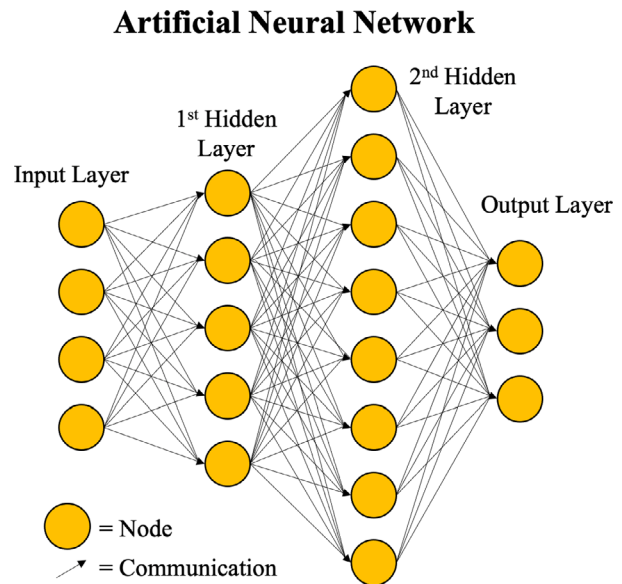


Fig. 2. Classification and regression tree to predict medication dosage. BMI, body mass index; PMH, previous medical history.

vision tasks. A CNN is a specialized type of deep neural network that adaptively learns features through backpropagation, eliminating the need for manual feature extraction. Several studies in recent years have illustrated the potential of deep learning for medical imaging tasks.⁴⁻⁶ Mzoughi *et al.*, Khan *et al.*, and Qummar *et al.*, have reported applications of a CNN for brain tumor classification, COVID diagnosis from chest X-ray images, and diabetic retinopathy detection, respectively.⁴⁻⁶ With the emergence of high performing deep learning models, neural networks have become an attractive tool for radiologists due to their ability to automatically learn feature hierarchies. For some classification tasks, CNNs have shown success in overcoming the limitations of traditional ML models.

In addition to computer vision, deep learning has made large contributions to the field of natural language processing (NLP), which is concerned with the development of machines to analyze and derive meaning from human language.⁷ Natural language processing has become a part of the clinical flow in emergency medicine as a way to harness the vast amounts of textual data in electronic health records. Certain neural network architectures are designed to effectively extract valuable information from unstructured text data in electronic health records such as clinical reports and health-care provider notes.^{8,9} Natural language processing methods have shown potential in leveraging medical records for various clinical tasks such as identifying sepsis, appendicitis, and influenza.¹⁰⁻¹³ Examples of deep learning methods that have garnered interest for NLP tasks include

recurrent neural networks, CNNs, and deep generative models. Unsupervised learning and reinforcement learning methods have also been applied for NLP.

Supervised ML algorithms have rapidly replaced traditional methods in predictive studies aiming to forecast the occurrence of outcomes using patient characteristics that are measured prior to the outcomes. As these models allow us to capture relationships between features (predictors) and outcomes flexibly, the prediction performance is expected to outweigh simple scoring systems. In addition to predictive studies, supervised ML algorithms have come into use in causal inference studies targeting the investigation of the effects of interventions on outcomes of interest. As causal inference in observational studies generally needs models to estimate treatment effects, we can expect to reduce bias using sophisticated ML models.

CURRENT RESEARCH USING AI IN EMERGENCY MEDICINE

Triage and disposition

TRIAGE REFERS TO the process by which patients are assessed upon arrival to the emergency department (ED) and prioritized based on the severity and urgency of their medical condition. Traditionally, a triage nurse will carry out the evaluation using vital signs, demographics, and ordered tests. Proficiency in triage takes time and experience. When immediate life-threatening conditions are not identified, emergency severity index level 3 is a default choice, which could leave a large number of patients waiting long hours for a provider and evaluation.¹⁴ The use of ML models in the ED can facilitate triage with more accuracy and efficiency, requiring only information routinely collected by the triage staff. In addition to predicting the urgency of medical conditions, ML techniques can be applied to develop screening tools for disease specific risk prediction.

Medical imaging

The ED provider does not always have timely access to radiology interpretation. An accurate identification of a fracture in X-ray images or a stroke in magnetic resonance imaging scans conventionally requires timely access to avoid misdiagnosis and a delay in treatment. This is especially critical when working in a smaller ED with limited access to specialists. Deep learning models for medical imaging with high sensitivities could help clinicians quickly identify life-threatening pathologies. Recent reports suggest that the quality of AI interpretation is not inferior to an expert radiologist.^{15–18}

Emergency department operations and management

Stochasticity in ED operations, such as patient arrivals, types of medical treatments and diagnostic tests required, and the duration of treatments and tests imposes unique challenges to predict future service demands. Artificial intelligence has the potential to transform ED operations and hospital leadership at multiple steps in the patient care process from arrival to discharge. The integration of ML could improve ED operations by better matching resources to patient needs, ultimately reducing costs and improving patient outcomes.¹⁹ Emergency department overcrowding is an increasing issue in health care that can have negative implications for the quality of patient care.²⁰ Predictive models for ED volume could help plan staffing models and prepare for surge and disaster situations. Second, many complicating factors make it difficult to estimate ED wait times. Machine learning algorithms with the ability to identify patterns in complex feature sets have the potential to produce more accurate ED wait times. On a larger scale, if multiple nearby hospitals report accurate wait times, low acuity patients could have a choice for ED service based on wait time and travel distance.

Not all AI applications will survive and win trust from clinicians and patients. It seems that triage and radiology models are likely to be adopted faster than other applications in ED operations.¹⁹ Table 1 outlines the selected works covered in this narrative to provide an overview of the most recent ML techniques in emergency medicine that have shown promise to improve patient outcomes.

CAUSAL INFERENCE, RISK, PREDICTION, AND METRICS

HERNÁN *et al.* has categorized medical and epidemiological research data science approaches in three ways: description, prediction, and causal inference (counterfactual prediction).³⁴ Although descriptive and predictive studies are essential to understanding the frequency, determinants, and prognosis of diseases or conditions, clinicians usually cannot achieve the ultimate goal by solely using these types of research methods when we aim to improve patient outcomes through interventions. Thus, causal inference, a type of study that compares hypothetical potential outcomes using two or more different treatments in a targeted population, attracts great interest in medical research. For example, we aspire to know whether resuscitative endovascular balloon occlusion of the aorta (REBOA) improves mortality in patients with life-threatening trauma.

Machine learning models can contribute to the improvement of causal inference in several ways.

Table 1. Machine learning models reported in emergency medicine publications

Study	Year	Outcome	Methods
Triage			
Raita <i>et al.</i> ²¹	2019	Develop machine learning models to predict critical care and hospitalization outcomes; compare model performance with the ESI	Lasso regression, random forest, gradient boosting machine, and a deep neural network
Ivanov <i>et al.</i> ²²	2021	Use variables collected at triage and free text from patient records to produce predictive models for acuity; compare model performance with clinical gestalt	Gradient boosting model
Chen <i>et al.</i> ²³	2020	Employ deep learning methods to predict disposition using variables collected at triage and clinical notes; compare model performance to rapid emergency medicine score (REMS)	Deep neural network
Disease-specific risk prediction			
Obeid <i>et al.</i> ²⁴	2019	Identify altered mental status during the assessment of patients by applying natural language processing techniques to ED provider notes	Naïve Bayes, lasso regression, decision tree, random forest, SVM, and convolutional neural networks
Patel <i>et al.</i> ²⁵	2018	Detect pediatric asthma during triage using clinical data combined with information about weather, neighborhood characteristics, community viral load, and socioeconomic status	Gradient boosting machine, decision tree, random forest, and lasso regression
Klang <i>et al.</i> ²⁶	2021	Predict admission to the neurosciences intensive care unit within 30 min of ED arrival using clinical, demographic, and unstructured text data from nurse and physician notes	Gradient boosting machine
Kim <i>et al.</i> ²⁷	2020	Explore the use of a machine learning model as a triage screening tool for septic shock; compare model performance to quick sepsis-related organ failure assessment (qSOFA) and modified early warning score (MEWS)	SVM, gradient boosting machine, random forest, ridge regression, lasso regression, multivariate adaptive regression splines, ensembles
Taylor <i>et al.</i> ²⁸	2018	Address the high diagnostic error rates for UTI in the emergency department with machine learning predictive models	Random forest, SVM, gradient boosting machine, adaptive boosting, elastic net, neural network, and logistic regression
Imaging			
Lindsey <i>et al.</i> ²⁹	2018	Develop deep learning models to predict fractures in wrist radiographs and compare model performance to clinicians' ability to detect fractures	Convolutional neural networks
Feng <i>et al.</i> ³⁰	2018	Echocardiogram and predicting mortality	Multivariate regression and gradient boosted model to draw causal inference
Chilamkurthy <i>et al.</i> ¹⁵	2018	Critical head CT finding	Deep learning
Ginat <i>et al.</i> ¹⁶	2019	Identification of intracranial hemorrhage	Deep learning
Rao <i>et al.</i> ¹⁷	2021	AI to serve as a peer review tool to reduce the false-negative rate of radiologists for intracranial hemorrhage detection	Convolutional neural network
ED operations			
Jilani <i>et al.</i> ²⁰	2019	Address overcrowding by developing a heuristic-based time series model to obtain a prediction for ED attendance	Time series forecasting and neural networks

Table 1. (Continued)

Study	Year	Outcome	Methods
Pak <i>et al.</i> ³¹	2021	Use queuing and service flow variables to build models to predict wait time for low acuity patients assigned to the waiting room after triage	Lasso regression, ridge regression, random forest
Lee <i>et al.</i> ³²	2020	Develop an optimal scheduling policy to minimize patient wait times where the wait time value differs between high and low acuity patients	Reinforcement learning
Xu <i>et al.</i> ³³	2014	Group ED patients with common features to improve resource management	Unsupervised learning

Abbreviations: AI, artificial intelligence; CT, computed tomography; ED, emergency department; ESI, emergency severity index; SVM, support vector machine; UTI, urinary tract infection.

First, unsupervised learning models can identify groups of patients that share specific characteristics. A recent article identified four clinically meaningful phenotypes of sepsis using data from several observational studies and randomized control trials.³⁵ In this study, k-means-based consensus clustering was used for the grouping. Although this study did not explicitly conduct causal inference, the authors suggested that the effects of early goal-directed therapy differed across these identified phenotypes.

Second, predictions from supervised learning models can find high-risk patients more accurately than the previous approaches. As an example, a study referenced earlier (Table 1) predicted critical care (intensive care unit admission or in-hospital death) and hospitalization of patients presented to the ED using baseline demographics, vital signs, chief complaints, and patient comorbidities.²¹ The authors found that the discrimination accuracies of the four ML models (lasso regression, random forest, gradient boosting, and deep neural network) were higher than that of the prediction model using logistic regression. Even though the purpose of this study was not causal inference, earlier detection of high-risk patients might lead to the identification of a subgroup that benefits from immediate aggressive interventions.

Finally, ML models can be directly used in causal inference to improve the model fit of either a treatment model to construct a propensity score or an outcome model or both. Although logistic regression is almost always used for propensity score estimation, it is plausible to use more sophisticated methods for this purpose. A representative example was a study that evaluated the effect of transthoracic echocardiography on 28-day mortality in intensive care unit patients with sepsis.³⁰ The authors used gradient boosting, rather than logistic regression, for the treatment model to estimate propensity scores to receive transthoracic echocardiography. Similarly, we can also use ML models for

the outcome model to estimate treatment effect to draw causal inference.

A challenge in using ML models for causal inference is that there are not enough reliable ways to verify that these methods are better than the traditional parametric models using linear or logistic regression. In the predictive studies, we can compare the performance of ML models with that of parametric models in preserved test datasets by some metrics, including the area under the receiver operating curve (AUROC) and the area under the precision-recall curve (AUPRC). Additionally, there is no guarantee that ML algorithms better eliminate confounding, even though the prediction accuracy for the treatment model and outcome model is improved. As scientists do not know the ground truth of the causal effect, there is always a risk of overfitting when we use complicated models. Hernán *et al.* advocated for using a sophisticated epidemiological method named doubly robust estimators, which combines a model to predict the outcome using multiple covariates with a model for the exposure (i.e., the propensity score model), to estimate the causal effect of an exposure on an outcome. They also suggest that sample splitting with cross-fitting could overcome the risk of overfitting of ML models. Machine learning models are not “magic wands” that automatically answer causal questions. However, it can help researchers estimate the effects of interventions accurately.

BARRIERS TO SAFE IMPLEMENTATION OF AI

Artificial intelligence is not a panacea for diagnostic and therapeutic dilemmas. Many prediction models using AI are presented in academic articles. However, the number of algorithms that have been used for the improvement of patient care is still limited. Yin *et al.*³⁶ found that only 51 relevant studies reported implementing and

evaluating AI applications in clinical practice. One explanation for why numerous studies are not translated into real practice is that many researchers design ML applications without the goal of implementation in mind. That is, some work is “research for research’s sake”. Beyond this reasoning, there are also unique ethical and practical considerations for the actual implementation of AI technologies in health care.

Garbage in, garbage out

One of the greatest barriers to the safe implementation of AI is the accuracy of input. Datasets used for ML training will be chosen and scored by “expert” clinicians. These datasets could contain cases that are incorrectly diagnosed, leading to fundamental flaws in decision making. Additionally, the composition of these datasets might be subject to bias. Cases that stand out in the programmer/clinician’s mind may be over-represented (e.g., “availability bias”). There can also be “spectrum bias” in ML. For example, a computed tomography (CT) dataset of biopsy-proven lung cancer could be visually different than that diagnosed incidentally in the ED, leading to degraded AI performance. Additionally, there may be “base rate neglect”. A weighting of “cancer” versus “not-cancer” should be based not only on lesion characteristics but on the pretest probability in the population to which the AI is being applied. This pretest probability will differ when, for example, AI is applied to patients in an academic referral center where a cancer

CT ML/AI dataset is developed than it will in a community ED.

Errors can be amplified in ML iterations (Table 2). While the inclusion of a misdiagnosis in the first dataset can lead to less-than-ideal diagnostic accuracy, further ML based on this initial dataset can reinforce this bias. One can think of it as a form of “confirmation bias”; the AI is looking for patterns it already knows even if they are erroneous. For example, several mislabeled echocardiograms or radiographs could lead to the incorporation of similar, erroneously interpreted studies into the ML process. This is not just theoretical; several algorithms have become “self-fulfilling prophecies”. In one case, questionable race-based adjustments for glomerular filtration rate biased the process of referral for kidney transplants against Black patients.³⁷ There are cautionary tales of algorithms used to identify “drug seekers” that include spurious information.³⁸ Concern has also been raised about the accuracy of AI in those with a disability.³⁹

The accuracy of AI can also be hampered at the bedside based on the subjective nature of required data. Even with something so fundamental to the diagnostic process as patient history, there is often a lack of interobserver agreement. This can lead to variable scoring of predictive models. For example, interobserver agreement of the patient history is poor even with something as straightforward as the HEART score, designed to predict the 6-week risk of major cardiac events.⁴⁰ As history will necessarily make up part of a predictive model, the prediction for any individual patient will be dependent on the accuracy of this data.

Table 2. Errors related to artificial intelligence and human heuristic equivalent

Source of error	Human heuristic equivalent	Error
Inclusion of notable cases in ML database	Availability bias	Cases that are exceptional “come to mind” and may be over-represented in the ML database leading to missing “usual” cases
Application of ML to a population in which it was not derived	Base rate neglect	Baseline rate of disease will differ in community settings versus academic settings in which ML datasets are derived. Not taking this into account can lead to diagnostic errors
ML will reinforce what it is taught over multiple iterations and <i>not</i> recognize patterns it is not taught	Confirmation bias	Turning a “blind eye” to information that is not consistent with what one believes or is taught
Types of information in the ML dataset are limited	Unpacking principle/bias	The more specific the information we have, the higher we judge the likelihood of an event. Lack of information could hinder our diagnostic accuracy (e.g., the lack of specific aspects of a history)

Abbreviation: ML, machine learning.

Most AI is proprietary

Another barrier to the safe implementation of AI is the proprietary nature of most systems. External validation must be assured as part of the quality control process. For example, the sepsis decision support tool in the EPIC electronic medical record was found to be neither sensitive nor specific when applied to an external validation set.⁴¹ Data sharing is another issue when we undertake ongoing training, validation, and improvement of AI algorithms. The model can be quickly outdated due to dynamically evolving clinical practices. Thus, users need to continuously provide data to fine-tune the model to fit the current situation. Massive efforts should also be made for anonymization and de-identification of the data to protect patients' privacy.

Probabilities

We cannot expect diagnostic perfection; whether made by an AI or human, diagnosis and treatment decisions are probabilistic. Errors will be made. Ideally, we would accept the same rate of "misses" by an AI as by a human provider. This remains a fertile area for research.

Regulation

Some regulation is already in place. The European Union ranks medical AI applications as "high risk," making them subject to stricter oversight than, for example, the AI that fills out your music playlist or recommends your next binge-watch. Artificial intelligence/ML is also subject to regulation by the Food and Drug Administration as a "medical device" in the United States. We have argued in a prior paper that, given the role of AI/ML in patient care, programmers/designers of diagnostic software should be considered medical providers and should be subject to traditional principles of medical ethics such as beneficence and nonmaleficence.⁴²

LIMITATIONS

MANY RESEARCH STUDIES at the forefront of innovation could be found in articles published in nonmedical journals or preprints that have yet to be peer-reviewed. Therefore, we are not able to capture all state-of-the-art ML technologies in emergency medicine.

CONCLUSION

THIS REVIEW SUMMARIZED the current status of ML research in emergency medicine. Although many applications have demonstrated efficacy in academic

literature, few have been implemented in practice due to barriers such as potential bias in datasets, the proprietorship of systems, and regulation. Quality measures and ethical controls need to be developed, including appropriate external oversight. Although this might make AI development more cumbersome, ensuring accuracy when potentially life-changing decisions are being made is critical. Future research should work towards overcoming these challenges to bridge the gap between academic research and clinical integration.

DISCLOSURE

Approval of the research protocol: N/A.

Informed consent: N/A.

Registry and registration no. of the study/trial: N/A.

Animal studies: N/A.

Conflict of interest: None.

REFERENCES

- 1 Lee S, Mohr NM, Street WN, Nadkarni P. Machine learning in relation to emergency medicine clinical and operational scenarios: an overview. *West J. Emerg. Med.* 2019; 20: 219–27.
- 2 Helm JM, Swiergosz AM, Haeberle HS *et al.* Machine learning and artificial intelligence: definitions, applications, and future directions. *Curr. Rev. Musculoskelet Med.* 2020;13: 69–76.
- 3 Esteva A, Robicquet A, Ramsundar B *et al.* A guide to deep learning in healthcare. *Nat. Med.* 2019;25:24–9.
- 4 Mzoughi H, Njeh I, Wali A *et al.* Deep Multi-Scale 3D Convolutional Neural Network (CNN) for MRI gliomas brain tumor classification. *J. Digit. Imaging* 2020;33:903–15.
- 5 Khan AI, Shah JL, Bhat MM. CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput. Methods Programs Biomed.* 2020; 196: 105581.
- 6 Qummar S, Khan FG, Shah S *et al.* A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access.* 2019;7:150530–9.
- 7 Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing [review article]. *IEEE Comput. Intell. Mag.* 2018; 13: 55–75.
- 8 Wunnava S, Qin X, Kakar T, Sen C, Rundensteiner EA, Kong X. Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding. *Drug Saf.* 2019; 42: 113–22.
- 9 Wu S, Roberts K, Datta S *et al.* Deep learning in clinical natural language processing: a methodical review. *J. Am. Med. Inform. Assoc.* 2020;27:457–70.

- 10 Kaufman DR, Sheehan B, Stetson P *et al.* Natural language processing-enabled and conventional data capture methods for input to electronic health records: a comparative usability study. *JMIR Med. Inform.* 2016;4:e35.
- 11 Rink B, Roberts K, Harabagiu S *et al.* Extracting actionable findings of appendicitis from radiology reports using natural language processing. *AMIA Jt. Summits Transl. Sci. Proc.* 2013;2013:221.
- 12 Doan S, Maehara CK, Chaparro JD *et al.* Building a natural language processing tool to identify patients with high clinical suspicion for Kawasaki disease from emergency department notes. *Acad. Emerg. Med.* 2016;23:628-36.
- 13 Ferraro JP, Ye Y, Gesteland PH, *et al.* The effects of natural language processing on cross-institutional portability of influenza case detection for disease surveillance. *Appl. Clin. Inform.* 2017;8:560-80.
- 14 Arya R, Wei G, McCoy JV, Crane J, Ohman-Strickland P, Eisenstein RM. Decreasing length of stay in the emergency department with a split emergency severity index 3 patient flow model. *Acad. Emerg. Med.* 2013; 20: 1171-9.
- 15 Chilamkurthy S, Ghosh R, Tanamala S *et al.* Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet.* 2018;392:2388-96.
- 16 Ginat DT. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. *Neuroradiology* 2020; 62(3): 335-40. <https://doi.org/10.1007/s00234-019-02330-w>.
- 17 Rao B, Zohrabian V, Cedeno P, Saha A, Pahade J, Davis MA. Utility of artificial intelligence tool as a prospective radiology peer reviewer - detection of unreported intracranial hemorrhage. *Acad. Radiol.* 2021; 28: 85-93.
- 18 Schaffter T, Buist DSM, Lee CI *et al.* Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw. Open.* 2020;3:e200265.
- 19 Berlyand Y, Raja AS, Dorner SC *et al.* How artificial intelligence could transform emergency department operations. *Am. J. Emerg. Med.* 2018;36:1515-7.
- 20 Jilani T, Housley G, Figueredo G, Tang P-S, Hatton J, Shaw D. Short and Long term predictions of Hospital emergency department attendances. *Int. J. Med. Inform.* 2019; 129: 167-74.
- 21 Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit. Care.* 2019; 23: 64.
- 22 Ivanov O, Wolf L, Brecher D *et al.* Improving ED emergency severity index acuity assignment using machine learning and clinical natural language processing. *J. Emerg. Nurs.* 2021;47:265-78. e7.
- 23 Chen C-H, Hsieh J-G, Cheng S-L, Lin Y-L, Lin P-H, Jeng J-H. Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data. *Int. J. Med. Inform.* 2020; 139: 104146.
- 24 Obeid JS, Weeda ER, Matuskowitz AJ *et al.* Automated detection of altered mental status in emergency department clinical notes: a deep learning approach. *BMC Med. Inform. Decis. Mak.* 2019;19:164.
- 25 Patel SJ, Chamberlain DB, Chamberlain JM. A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage. *Acad. Emerg. Med.* 2018; 25: 1463-70.
- 26 Klang E, Kummer BR, Dangayach NS *et al.* Predicting adult neuroscience intensive care unit admission from emergency department triage using a retrospective, tabular-free text machine learning approach. *Sci. Rep.* 2021;11:1381.
- 27 Kim J, Chang H, Kim D, Jang D-H, Park I, Kim K. Machine learning for prediction of septic shock at initial triage in emergency department. *J. Crit. Care.* 2020; 55: 163-70.
- 28 Taylor RA, Moore CL, Cheung K-H, Brandt C. Predicting urinary tract infections in the emergency department with machine learning. *PLoS One* 2018; 13: e0194085.
- 29 Lindsey R, Daluiski A, Chopra S *et al.* Deep neural network improves fracture detection by clinicians. *Proc. Natl. Acad. Sci. U. S. A.* 2018;115:11591-6.
- 30 Feng M, McSparron JI, Kien DT *et al.* Transthoracic echocardiography and mortality in sepsis: analysis of the MIMIC-III database. *Intens. Care Med.* 2018; 44: 884-92.
- 31 Pak A, Gannon B, Staib A. Predicting waiting time to treatment for emergency department patients. *Int. J. Med. Inform.* 2021; 145: 104303.
- 32 Lee S, Lee YH. Improving emergency department efficiency by patient scheduling using deep reinforcement learning. *Healthcare (Basel)* 2020; 8: 77.
- 33 Xu M, Wong TC, Chin KS. A medical procedure-based patient grouping method for an emergency department. *Appl. Soft. Comput.* 2014; 14: 31-7.
- 34 Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: A classification of data science tasks. *CHANCE* 2019; 32: 42-9.
- 35 Seymour CW, Kennedy JN, Wang S *et al.* Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* 2019;321:2003-17.
- 36 Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. *J. Med. Internet Res.* 2021; 23: e25759.
- 37 Ahmed S, Nutt CT, Eneanya ND *et al.* Examining the potential impact of race multiplier utilization in estimated glomerular filtration rate calculation on African-American care outcomes. *J. Gen. Intern. Med.* 2021;36:464-71.
- 38 A drug addiction risk algorithm and its grim toll on chronic pain sufferers | WIRED. Accessed 23 Nov 2021. <https://www.wired.com/story/opioid-drug-addiction-algorithm-chronic-pain/>
- 39 Guo A, Kamar E, Vaughan JW, Wallach H, Morris MR. Toward fairness in AI for people with disabilities SBG@a

- research roadmap. SIGACCESS Access Comput. 2020; 125: 1–1.
- 40 Soares WE, Knee A, Gemme SR *et al.* A prospective evaluation of clinical HEART score agreement, accuracy, and adherence in emergency department chest pain patients. *Ann. Emerg. Med.* 2021;78:231-41.
- 41 Wong A, Otles E, Donnelly JP *et al.* External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern. Med.* 2021;181:1065-70.
- 42 Graber MA, Bailey O. The wizard behind the curtain: programmers as providers. *Philos. Ethics Humanit. Med.* 2016; 11: 4.