Artificial Intelligence in the Intensive Care Unit

Massimiliano Greco, MD[1,2] Pier F. Caruso, MD[1] Maurizio Cecconi, MD[1,2]

[1] Department of Anesthesiology and Intensive Care, Humanitas Clinical and Research Center—IRCCS, Rozzano, Milan, Italy
[2] Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy

Address for correspondence Maurizio Cecconi, MD, Department of Anesthesiology and Intensive Care, Humanitas Clinical and Research Center—IRCCS, Via Manzoni 56, Rozzano, Milan 20089, Italy (e-mail: maurizio.cecconi@hunimed.eu).

Semin Respir Crit Care Med 2021;42:2–9.

Abstract

The diffusion of electronic health records collecting large amount of clinical, monitoring, and laboratory data produced by intensive care units (ICUs) is the natural terrain for the application of artificial intelligence (AI). AI has a broad definition, encompassing computer vision, natural language processing, and machine learning, with the latter being more commonly employed in the ICUs. Machine learning may be divided in supervised learning models (i.e., support vector machine [SVM] and random forest), unsupervised models (i.e., neural networks [NN]), and reinforcement learning. Supervised models require labeled data that is data mapped by human judgment against predefined categories. Unsupervised models, on the contrary, can be used to obtain reliable predictions even without labeled data. Machine learning models have been used in ICU to predict pathologies such as acute kidney injury, detect symptoms, including delirium, and propose therapeutic actions (vasopressors and fluids in sepsis). In the future, AI will be increasingly used in ICU, due to the increasing quality and quantity of available data. Accordingly, the ICU team will benefit from models with high accuracy that will be used for both research purposes and clinical practice. These models will be also the foundation of future decision support system (DSS), which will help the ICU team to visualize and analyze huge amounts of information. We plea for the creation of a standardization of a core group of data between different electronic health record systems, using a common dictionary for data labeling, which could greatly simplify sharing and merging of data from different centers.

Keywords
- ICU
- critical care
- machine learning
- artificial intelligence
- supervised learning
- unsupervised learning
- reinforcement learning

Evidence-based medicine is the foundation of modern medicine. Since naval surgeon James Lind first published its observations on scurvy treatment in sailors, evidence-based medicine was progressively recognized as the landmark of modern medicine.[1]

Randomized controlled trials reside on the highest step on the podium of evidence, and their design takes often advantage of multicentricity to increase external validity.[2] However, while large multicentric study remain the gold standard, they are difficult and expensive to conduct in terms of both time and resources. Accordingly, only between 10 and 20% of recommendations in modern medicine are evidence based.[3] This aspect is particularly relevant in intensive care medicine, where in recent years a large number of randomized controlled trials focusing on mortality and major clinical outcomes yielded negative results.[4] The above-mentioned limitation is related with the characteristics of the population of critically ill patients, which have wide variability in comorbidities, age, and baseline mortality according to different conditions at intensive care unit (ICU) admission.

The high frequency of negative trial in critical care may thus be related to heterogeneity and confounding effects. Accordingly, ICU patients will present great variability in the benefits from a specific therapy, called heterogeneity of treatment effect, which may even result in apparent paradoxes, such as negative trials of therapies, that results beneficial in high-risk subgroups.[5]

published online November 5, 2020

Issue Theme Challenges in Critical Care; Guest Editors: Jean-Louis Vincent, MD, PhD, and Antonio Artigas, MD, PhD

DOI https://doi.org/10.1055/s-0040-1719037.
ISSN 1069-3424.

As reported in the previous chapter, the ICU is the most suitable ward, among all the hospital wards to begin the transition to big data. This is due to the high number of monitoring systems, collecting continuously and with a high granularity (i.e., collecting data every minute or even every second) respiratory, hemodynamic, neurological, and clinical data.

Big data may help to overcome some of the limitations of the evidence-based medicine applied to intensive care medicine. Randomized controlled trials are generally employed to control the risk of bias and confounder, randomly distributing them among cases and controls. While this strategy may work well for bias introduced by observer (selection bias, reporting bias, and observer bias), consequently reducing the risk of type-1 error (false-positive error), this same approach may actually increase the risk of type-2 error (false-negative error), hampering the identification of a positive treatment effect in a specific subpopulation. Risk of type-2 error is further increased by the unfeasibility and costs which would be needed to conduct randomized trials in every possible subpopulation.

Conversely, with large amount of data effortlessly collected by electronic health record (EHR), which contains the same resolution used by caring physicians to take decisions, it becomes extremely easier to conduct analysis of therapeutic interventions in smaller populations.

On the drawback, the level of resolution of these data are often so high that classic statistical explorations and analyses may be too difficult and time consuming to be performed, at least not in real time or within a short-time window. In this respect, artificial intelligence (AI) may be more than helpful, with the diffusion of powerful machine learning algorithms designed to automate and simplify data analysis.

"Can a Machine Think?": History and Definition of Artificial Intelligence

"Can a machine think?" was the interrogative proposed by Alan Turing in 1950. Turing proposed what is now known has the Turing test, to assess whether AI was evolved enough to be indistinguishable from a human being. In the test, a computer and a human being are placed in two closed rooms, and an observer outside should guess which of the two is the computer and which is the human. If the human fails its guess, the computer has passed the Turing test. The academic debate on the concept of AI flourished in the following debate, with the term "artificial intelligence" being used the first time at the Dartmouth Conference, held in 1956 in New Hampshire. However, despite decades of discussion and argumentation, no real consensus on the definition of AI exits. The term generally refers to the ability of a machine to show "cognitive" capabilities, including the ability to learn, and to perform inference and deduction.

AI applied to the medical field may refer to machine learning to the ability to understand and produce natural language (also known as natural language processing [NLP]) or the capability to visualize and recognize objects (computer vision).

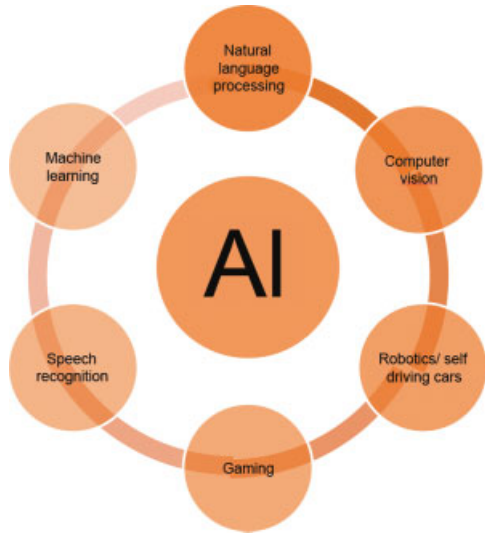

Fig. 1 The different facets of artificial intelligence.

Some of the key aspects encompassed by AI are reported in ►**Fig. 1**.

How Does a Machine Learn?: The Scope of Machine Learning

Machine learning techniques are algorithms designed to analyze very large datasets. The concept behind machine learning is to allow computers to learn without the need to program specific tasks that is without a human to understand, supervise, and interpret all steps of data analysis.

Machine learning models are trained on datasets that contain huge amounts of raw data and are based on numbers, or images, or texts. The advantage of machine learning models compared with standard analysis model is that the best algorithm is automatically tuned by the machine learning process, without being coded step by step by human interaction. In the simplest machine learning process, the following three key aspects are cooperating to yield a results: datasets (which contains the raw data), algorithms (which interpret the data), and selected features (the variables selected in the dataset to be used in the analysis; ►**Fig. 2**).

Features are used to define key aspects of the process in supervised learning and need human judgment and interaction. Unsupervised models, conversely, are built to interpret information without previous selection of features.

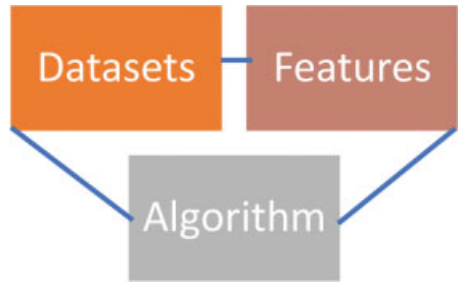

Fig. 2 The three aspects of machine learning.

Clinical Applications of Artificial Intelligence

In a study by De Fauw et al, AI algorithms employed convoluted neural networks (NN; a type of unsupervised machine learning which is specifically useful for evaluation of images) to analyze retinal pattern to detect retinal disease. This process is particularly complex, as it faces several difficulties such as variations in quality of image, variation in the technical process of acquisition, and variations in patient characteristic. To overcome these challenges, the authors used a two-step approach which is paradigmatic in computer vision, by means of a segmentation network, the model trained to map each voxel (a single data point of a regularly spaced three-dimensional grid) into one of the possible tissue types, according to predefined categories (by anatomy and pathology classification and including image artifacts). Subsequently, another neural network was trained to learn to analyze the segmentation map to provide diagnosis and referral decision. The two NNs were able to achieve very good performance in detection of retinal disease.[6]

NLP is still at the beginning of its development but its potential is huge. NLP is designed to allow the extraction of concepts from the natural language used by doctors in clinical charts to communicate with colleagues and health care professional.[7] NLP is still limited by different levels of development according to different languages and countries, but in the near future, it will provide fundamental understanding in the richest resource of medical knowledge such as the medical narrative included in clinical charts.[8]

In intensive care medicine, the most used AI algorithms are those based on machine learning, as critical care is a good source for large dataset of numeric data derived by complex continuous monitoring and by continuous therapies.

►**Fig. 3** illustrates the complexity of data generated by a single ICU admission, starting from baseline patient data, to the level of organ failure at the debut of critical illness, and encompassing all therapies and biological and clinical data derived from the entire ICU admission.

A machine learning strategy for prognostication considering the entire ICU admission could have clear advantage over commonly employed scores such as SOFA (Sequential Organ Failure Assessment) and APACHE-II (Acute Physiology And Chronic Health Evaluation II), as it would reach a higher performance and represent a step forward toward personalized-medicine compared with standard scores.[9,10]

The most common machine learning models which have been used in critical care are described in the next sections

Classification versus Regression Models

In general terms, classification models are generally used to automatically predict whether an object is part of a category or another. The aim of the model is thus to define a mapping function able to assign a patient to a discrete output variable (e.g., a "labelled" feature), using data from some input variables.

Conversely, regression models are generally used to predict a quantity, that is, how much the blood pressure while rise or fall according after the administration of a medication.

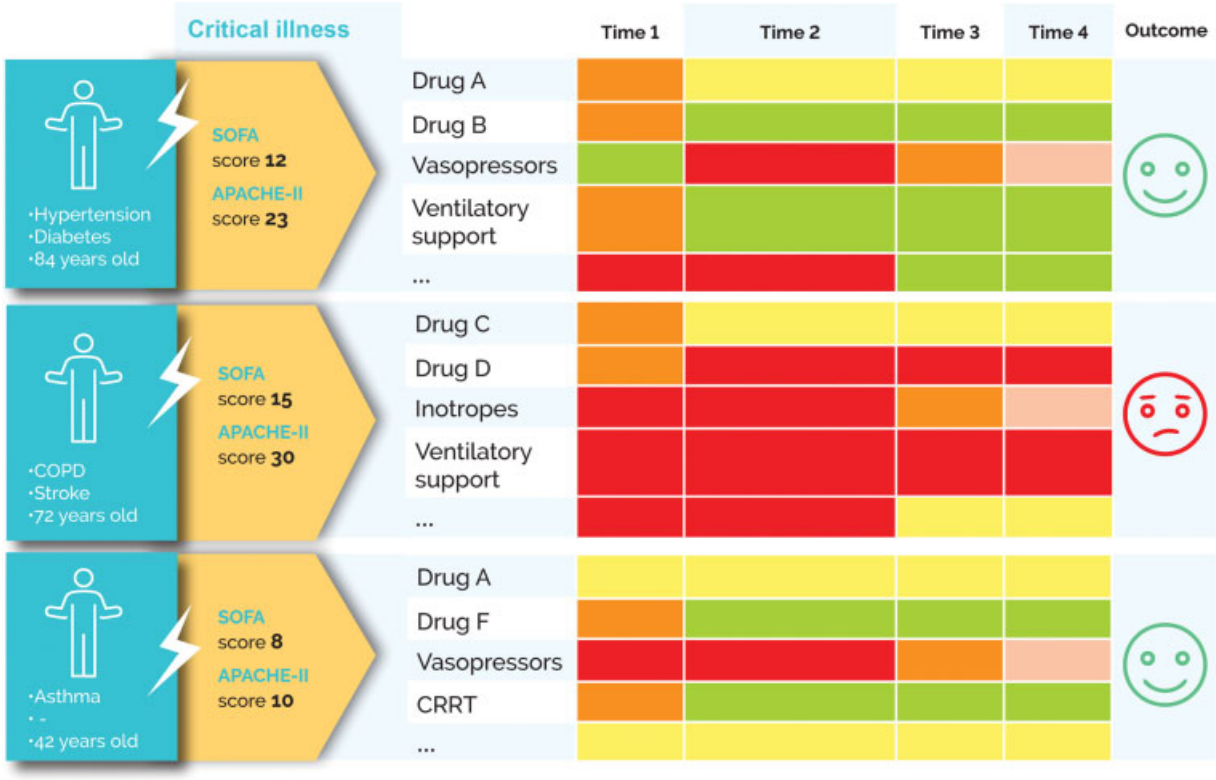

Fig. 3 The complexity and variability data generated by single ICU admissions. COPD, chronic obstructive pulmonary disease; ICU, intensive care unit.

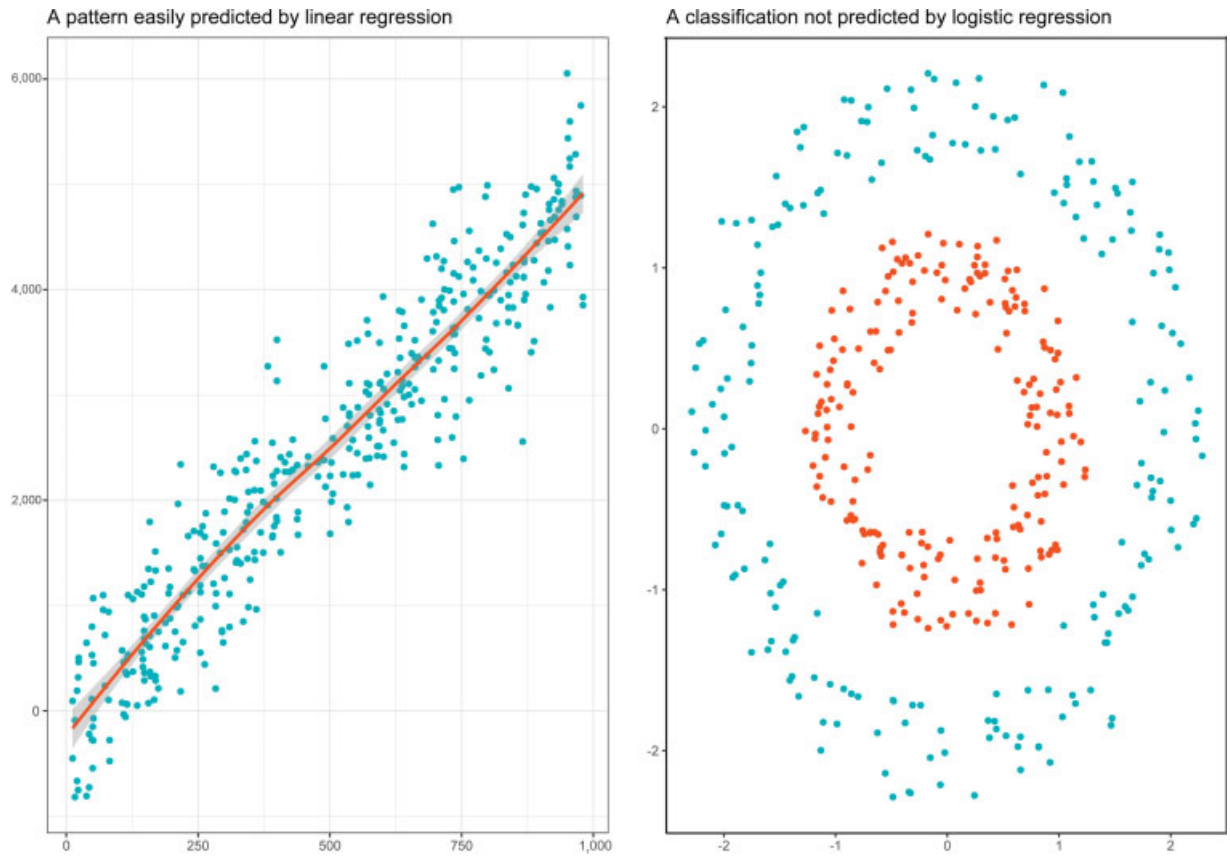

Fig. 4 (**A**) A simple pattern predicted by a linear model. (**B**) A clear pattern which could have a very poor performance using logistic regression.

The most known models for classification and regression are linear and logistic regression. Linear regression uses a linear combination of features to model an outcome (►**Fig. 4A**), and is represented by the following function:

$$y = w_1x_1 + w_2x_2 + w_nx_n + b,$$

where y is the outcome, w_n is the slope of each x_n feature, and b is the y intercept.

Logistic regression, despite its name, is a classification learning algorithm. It uses a sigmoid function to assign an event a probability, which is by definition restricted between 0 and 1. It is one of the most popular classification algorithm used in medicine. However, in clinical medicine, we are rarely so lucky to have a simple pattern predicted by linear or logistic regression (►**Supplementary Material**, available in the online version). With an increasing complexity of scenarios related with increasing amount of data, and with better representation of reality, more complex models are needed to be able to interpret the data (►**Fig. 4B**).

Machine Learning Algorithms: Supervised Learning Models and Labeled Data

Supervised learning algorithms are commonly employed in intensive care medicine. The term supervised refers to the process of learning from labeled data. With training, the algorithm will search for patterns which best correlate with outcome.

Labeling is the process that assign a data to a category or label it. In medicine, this process normally involves a human who maps clinical data against a known definition. The definition may identify a syndrome, such as sepsis or acute respiratory distress syndrome (ARDS), but also quantify a condition, such as severity of chronic obstructive pulmonary disease (COPD) or level of frailty.

Labeling is essential in supervised learning and is also the base for clinical scales used in medicine, where we use to define categories that are easily understandable by physicians and convey information on diagnosis or prognosis.

While being largely used for research purpose, clinical labeling is generally not integrated with EHR system. Even when health care professional categorizes a patient within a certain scale, the results are often reported in natural language within the clinical diary, thus losing the utility of the label beyond the clinical context.

Therefore, one of the most common strategies to derive clinical labels is to use diagnostic and procedure reimbursement codes. In several countries, the hospital admission process completes with the mapping of International Classification of Diseases, ninth revision (ICD-9) codes for reimbursement purposes.[11] However, limitations of ICD-9 coding are well known. In a large study including 161,529 patients, ICD-9 codes had poor accuracy in reporting pneumonia etiology, with good specificity but low sensitivity, (as low as 14% for some type of bacterial pneumonia).[12,13] Nonetheless, as data labeling represent a burden in terms of time and resources, as it implies the need of one or more experienced person, ICD-9

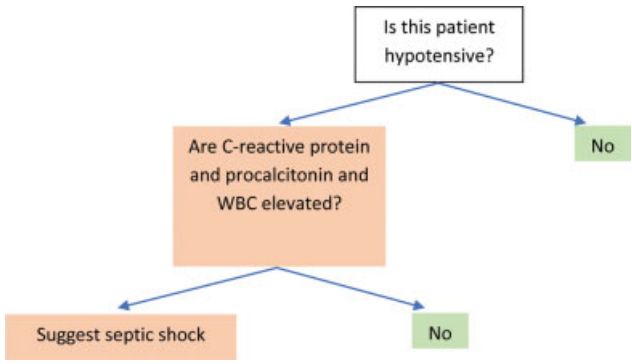

Fig. 5 A simple decision tree on sepsis. WBC, white blood cell.

codes are frequently used even in the most famous critical care databases, the Medical Information Mart for Intensive Care, version III (MIMIC-III) database.[14]

Among supervised models, we report here a brief description of decision trees, SVM, and random forest.

Decision Trees

A decision tree is a flowchart-like model which produces an outcome after processing input information through several decisional nodes which correspond to each tree node. The decision tree is depicted upside down with the root at the top (►**Fig. 5**). The first node is called root, while the end of the branches is called leaf. The advantage of decision tree is a clear representation of feature importance and relations.

Support Vector Machines

The goal of SVM model is to define a hyperplane in an N-dimensional space, which is able to classify data points when a line or a plane in a two- or three-dimensional space would not be useful. ►**Fig. 6** reports a visual simplification of a SVM that is readable, as hyperplanes of n-dimensions are very difficult to imagine for human beings. Support vectors are the data which influence orientation and position of the hyperplane which separate the data for classification. In SVM, by means of a function which transform the original space in a space of higher dimensionality, we may be able to classify every points.

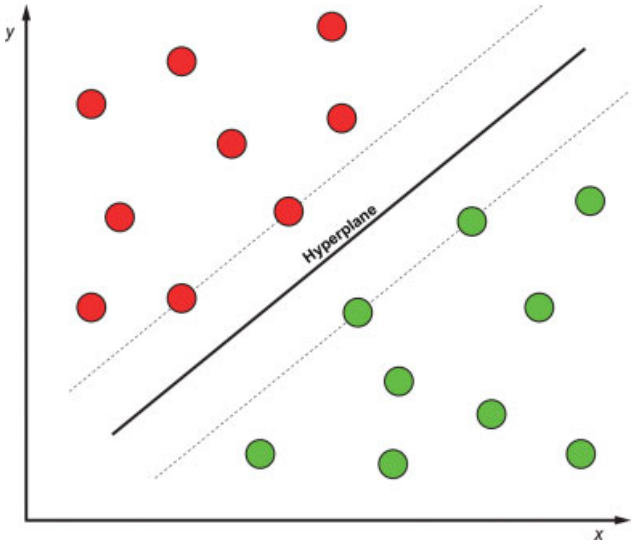

Fig. 6 Example of a support vector machine, with hyperplane and support vectors (dotted lines).

Random Forest

Random forests are based on a large number of individual decision trees that work in parallel. Each tree is different from the others, and classifies the outcome independently, and through a sort of democratic process the classification with the higher number of votes produces the final output. The idea behind the model is that the overall group of uncorrelated trees will produce better performance than any of the single tree (►**Fig. 7**).

Deep Learning: Neural Network

To improve processing of complex information, deep learning models try to replicate the structure of a human brain. They use nonlinear transformations to increase the level of abstraction, and differently from supervised models, deep models may be used without previous labeling or feature selection.

In its most simple form, a neural network is based on multilayer perceptrons that are a series of several layers of neurons. Every neuron holds a number from 0 to 1, and when the number inside the neuron is above a certain threshold, called activation number, the neuron will activate. Activation of some group of neurons from one neural layer bring activation to other neurons in the next layer, in a way that is similar to biological neurons. The last layer is the output layer, where the group of neurons activated more frequently will prevail in a single output, giving the final interpretation of the model. NN can start from only a few layers, and reach hundreds of layers of neurons. As the neural network increases in complexity and number of layers, the number of data to correctly train the algorithm increases parallelly. Deep learning has been used in image recognition and in computer vision, as well as in NLP.

Reinforcement Learning

Reinforcement learning is a third option in machine learning that is particularly useful in sequences of decision. When the algorithm is trained, each choice of the algorithm leads to a reward or penalty, in a trial-and-error game, which enables to solve a complex problem in an uncertain environment. The algorithm tries to maximize the reward and reduce the penalty. The programmer sets the rewards and penalties, but produces no suggestion on how to solve the problem.

Examples of Machine Learning in Critical Care

The increasing adoption of EHR in the ICU is prompting the diffusion of data science and machine learning in the critical care environment. Hemodynamic data from monitors, infusion data from infusion pumps, respiratory data from ventilators, are generating large amount of data which can be

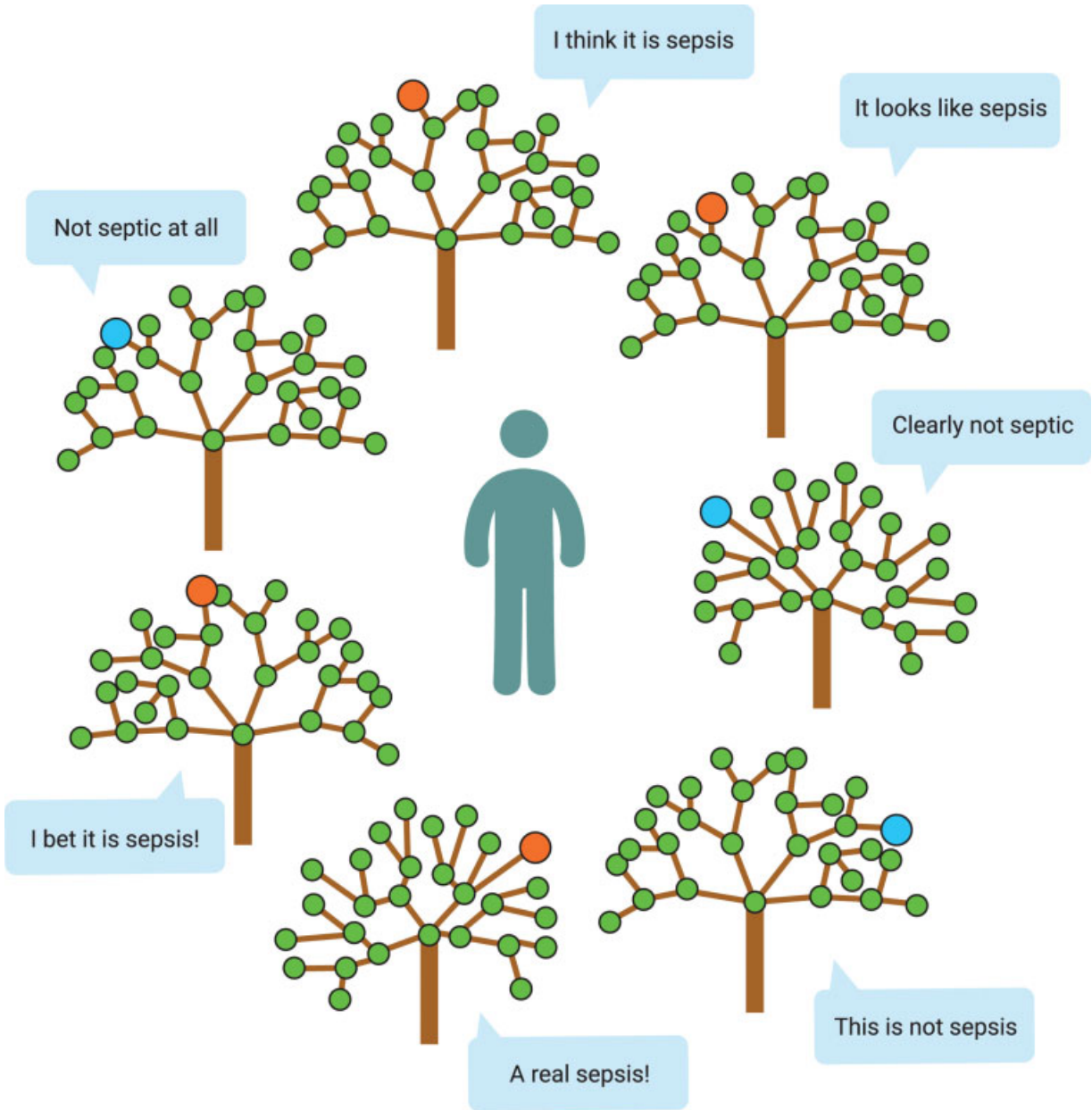

Fig. 7 An example random forest.

compared with other source of big data such as the omics (i.e., genomics or proteomics).

Komorowski et al developed a computational model through the use of reinforcement learning to dynamically suggest optimal treatments for adult patients in ICUs.[15]

The model was built and validated on the two largest datasets available in ICU: the (MIMIC-III) which was used as training set to develop the model, and the eICU Research Institute Database (eRI).[14] The model demonstrated that an AI clinician performed bettered, compared with clinicians, in selecting intravenous fluids and vasopressors. On average, the AI clinician recommended lower doses of fluids and higher doses of vasopressors compared with actual treatments. Moreover, patients receiving the dose more close to those suggested by AI clinician had the lowest mortality.[15]

In a study by Davoudi et al, pervasive monitoring and machine learning were used on 22 patients admitted to an ICU to continuously assess delirium and agitation.[16] Patient were labeled according to CAM (confusion assessment method)-ICU scale. Camera and accelerometers were employed to record facial expression and movements. Three accelerometers were placed on patient wrist, ankle, and arm to identify posture. A pretrained neural network was used for facial recognition and detection of expression through single elements.[17] This very nice study is one of the first study to assess patient emotions continuously, in patients with and without delirium. Moreover, this system offered the possibility to continuously analyze patient movements, circadian rhythm disruption, in an objective and reproducible way.

The Future

With an explosion in quantity and resolution of data from the critical care environment, machine learning models will become popular in intensive care research, and will provide deeper understanding of the complexity behind ICU care.

Beside clinical research purposes, the power of AI in intensive care may be unleashed by two further steps: clinical DSS systems and precision medicine.

A step toward precision medicine will be taken when data science in intensive care will allow to study subgroups of patients so homogeneous and distinct from other groups that they will represent the "prototype" of a single patient, having exactly the same age, same previous pathologies, same home medications, and same ICU admission reason and organ failure.[18]

DSS integrated in the EHR will be able to inform instantly the critical care team of variation in patient conditions, consequent changes in prognosis, and suggest further diagnostic or therapeutic actions.[19] Limitation of current example of DSS is that they are based on a low number of features and lack external validity, all factors that reduce their performance. In the future, DSS derived by machine learning algorithms built on thousands of features may constitute a major clinical advantage. For example, a DSS may use laboratory, clinical, hemodynamic waveform data, and microbiological data to inform the caring team of a possible new episode of sepsis in an ICU patient, hours before its clinical manifestation, suggesting further diagnostic and therapeutic steps, according to the analysis of microbiological data taken from every patient admitted in that ward during the previous 10 years.

This system will not replace the critical care team, physician and nurses will be always in charge of the patient, taking all decision, but will benefit from an increased level of information which could not be available using other methods.

Limitations of Artificial Intelligence

Despite its major advantages, the AI clinicians bear some important limitations. It would be exceedingly difficult for an algorithm to include, in its decision process, the large spectrum of opinions and believes which influence personal choices of every single patients. Cultural difference and diversity in perception of intensive care interventions are a fundamental part of the care of critically ill patients.[20,21]

Moreover, machine learning models may find spurious association and erroneously interpret them as real relationships between events. To avoid these kinds of misinterpretation, humans should always oversee the outcomes of machine learning process and verify their results.

A third limitation is related with source data, every machine learning model can be as good as the data underlying it. As above mentioned, AI models need very large amount of high-quality and high-resolution data; however, most of the clinical knowledge on patient care is transmitted by natural language within clinical charts or during oral transmissions between colleagues, and these are still precluded to AI algorithms. Differences between software systems, local protocols, and medical practice between different countries, different centers, or even within the same hospital may further impair the performance of AI models

All these aspects are limitations of AI. Eventually, despite its potency in classification of disease, stratification of patients, identification of the best treatment toward precision medicine, the AI clinician would never be able to sit with a patient, take his hand, guess the best words needed to communicate his/her medical condition, and discuss with her or with him the best treatment options.

A Plea for Standardized Data Labeling and Digital Data Sharing

The potential benefit of machine learning in clinical practice are directly related with the amount, quality, and resolution of the data.

Nowadays, even with the increasing diffusion of EHR system in the ICUs, the background structure of the data are so different between different ICUs, to be very difficult to merge, even when different ICUs are using the same software and often even when considering ICUs from the same hospital.

Moreover, barriers related with country-specific privacy regulation, ethical committee regulations are legislative differences may further impair sharing and merging of data.

This is a lost opportunity, for patients, clinical researchers, doctors, and nurses from the entire intensive care community.

Conclusion

We advocate the diffusion of a common structure for ICU EHR databases, based on a common nucleus of core data, and a common dictionary for labeling of core features, which may be in future used to merge large amount of data between different centers. Core data and labeling could be chosen through a consensus process.

We also propose to stakeholders and legislators to take further steps to simplify the diffusion of anonymized data between different centers in different countries and allowing the creation of large datasets, which could grant improvements in clinical care while protecting patients' rights.

Conflict of Interest
None declared.

References

1 Bhatt A. Evolution of clinical research: a history before and beyond James Lind. Perspect Clin Res 2010;1(01):6–10
2 Atkins D, Best D, Briss PA; GRADE Working Group, et al. Grading quality of evidence and strength of recommendations. BMJ 2004; 328(7454):1490–1494
3 Fleming PS, Koletsi D, Ioannidis JPAA, Pandis N. High quality of the evidence for medical and other health-related interventions was uncommon in Cochrane systematic reviews. J Clin Epidemiol 2016;78:34–42
4 Vincent JL, Marini JJ, Pesenti A. Do trials that report a neutral or negative treatment effect improve the care of critically ill patients? No. Intensive Care Med 2018;44(11):1989–1991

5 Iwashyna TJ, Burke JF, Sussman JB, Prescott HC, Hayward RA, Angus DC. Implications of heterogeneity of treatment effect for reporting and analysis of randomized trials in critical care. Am J Respir Crit Care Med 2015;192(09):1045–1051
6 De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018;24(09):1342–1350
7 Fu S, Chen D, He H, et al. Clinical concept extraction: a methodology review. J Biomed Inform 2020;109:103526
8 Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. JMIR Med Inform 2020;8(03):e17984
9 Vincent JL, de Mendonça A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. Crit Care Med 1998;26(11):1793–1800
10 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985;13 (10):818–829
11 Bouza C, Lopez-Cuadrado T, Amate-Blanco JM. Use of explicit ICD9-CM codes to identify adult severe sepsis: impacts on epidemiological estimates. Crit Care 2016;20(01):313
12 Higgins TL, Deshpande A, Zilberberg MD, et al. Assessment of the accuracy of using ICD-9 diagnosis codes to identify pneumonia etiology in patients hospitalized with pneumonia. JAMA Netw Open 2020;3(07):e207750
13 Sarrazin MS, Rosenthal GE. Finding pure and simple truths with administrative data. JAMA 2012;307(13):1433–1435
14 Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3:160035
15 Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018;24(11):1716–1720
16 Davoudi A, Malhotra KR, Shickel B, et al. Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. Sci Rep 2019;9(01):8020
17 Amos B, Ludwiczuk B, Satyanarayanan M. OpenFace: a general-purpose face recognition library with mobile applications. Available at: http://elijah.cs.cmu.edu/DOCS/CMU-CS-16-118.pdf. Accessed August 10, 2020
18 Patel SK, George B, Rai V. Artificial intelligence to decode cancer mechanism: beyond patient stratification for precision oncology. Front Pharmacol 2020;11:1177
19 Liu S, See KC, Ngiam KY, Celi LA, Sun X, Feng M. Reinforcement learning for clinical decision support in critical care: comprehensive review. J Med Internet Res 2020;22(07):e18477
20 Aslakson RA, Wyskiel R, Thornton I, et al. Nurse-perceived barriers to effective communication regarding prognosis and optimal end-of-life care for surgical ICU patients: a qualitative exploration. J Palliat Med 2012;15(08):910–915
21 Levin PD, Sprung CL. Cultural differences at the end of life. Crit Care Med 2003;31(05):S354–S357