

Ményssa Cherifa<sup>1,2,3</sup> , Romain Pirracchio<sup>3,4,5</sup>

## What every intensivist should know about Big Data and targeted machine learning in the intensive care unit

*O que todo intensivista deveria saber sobre Big Data e aprendizado da máquina na unidade de terapia intensiva*

1. Institut National de la Santé et de la Recherche Médicale - Paris, France.
2. Université de Paris - Paris, France.
3. The ACTERREA Research Group - Paris, France.
4. Department of Anesthesia and Perioperative Medicine, Zuckerberg San Francisco General Hospital, University of California San Francisco - California, United States.
5. Department of Anesthesia and Intensive Care Medicine, Hôpital Européen Georges Pompidou - Paris, France.

### The increasing importance of Big Data in healthcare

The conjunction of increasingly available access to big medical data and of substantial progress in machine learning (ML) and artificial intelligence (AI) has created new, unforeseen opportunities for data science in healthcare. Big Data is described as having at least three distinct characteristics, volume, velocity, and variety, but in regard to healthcare, it also includes variability and value.<sup>(1)</sup> Therefore, it is very challenging to extract any useful information from Big Data using traditional statistical methods.<sup>(2)</sup> Big Data analytics has immense potential for improving quality of care, helping physicians and nurses to make more personalized clinical decisions, reducing waste and errors and possibly reducing the cost of care.<sup>(3)</sup> Anticipating organ dysfunction before it occurs can be extremely helpful to (i) make better and more tailored therapeutic decisions and (ii) in some instances, prevent the occurrence of organ failure by appropriately adjusting the therapeutics upfront. Additionally, the ability to predict any upcoming deterioration can be very helpful to assist clinical leadership in proactively allocating human resources. Malak et al. recently proposed a multiagent risk management architecture based on Big Data and analytics in order to create a collaborative and real-time environment to manage neonates with critical conditions at the neonatal intensive care unit (ICU).<sup>(4)</sup>

### Sources of healthcare Big Data

The “data revolution” in healthcare and, ultimately, in critical care depends on the ability to stream and store a large amount of information in a protected and encrypted central repository. Electronic medical records, bedside monitors, drug delivery devices, ventilators or dialysis machines are continuously generating data. It is becoming possible to combine these data with laboratory test results, procedures, caregiver notes, imaging reports, and, ultimately, outcomes, including long-term functional and behavioral outcomes. For instance, the Mayo Clinic has developed such a data warehouse, called the (Multidisciplinary Epidemiology and Translational Research in Intensive Care Data Mart (METRIC)),<sup>(5)</sup> while the Beth Israel Deaconess Medical Center (BIDMC) has launched a similar large database comprising deidentified health-related data associated with over 40,000 patients who stayed in critical care units, the MIMIC-III.<sup>(6)</sup> These two databases are openly available for scientific research purposes. Increasingly more ICUs, medical centers and even large-scale health networks are developing solutions to store and analyze patient data and benchmarks with different systems and organizations.<sup>(7)</sup>

**Conflicts of interest:** None.

Submitted on February 1<sup>st</sup>, 2019  
Accepted on May 5, 2019

**Corresponding author:**

Romain Pirracchio  
Department of Anesthesia and Perioperative  
Medicine  
Zuckerberg San Francisco General Hospital  
University of California San Francisco  
1001 Potrero avenue  
CA94110 San Francisco  
United States  
E-mail: romain.pirracchio@ucsf.edu  
DOI: 10.5935/0103-507X.20190069



## Machine learning for predictive analytics and decision support in the intensive care unit

Because “Big Data includes heterogeneous, multispectral, incomplete and imprecise observations derived from different sources”,<sup>(8)</sup> the development of appropriate analytics and inference is needed. Machine learning, which is the component of AI that allows computers to make data-driven choices and predictions, is now considered as the solution of choice to harness big medical data.<sup>(9)</sup> Obviously, ML has the ability to model complex relationships between large explanatory features and desired outputs, such as patient outcomes. ML algorithms are usually divided into different categories: parametric vs. nonparametric methods, supervised vs. unsupervised algorithms, and unique vs. ensemble algorithms (Figure 1). Supervised learning algorithms are used to uncover the relationship between potential explanatory features and one or more known target outcomes. They are commonly applied in critical care for the prediction of clinical events, such as the prediction of ICU mortality.<sup>(10)</sup> In unsupervised learning algorithms, there is no specific targeted outcome; the goal is essentially to dig deep into the data structure in order to identify the correlation between features and create clusters of characteristics. These algorithms are currently mainly

used in precision medicine, in which the goal is to uncover subgroups of patients who share similar clinical or molecular characteristics.<sup>(11)</sup>

### Perspectives

The Food and Drug Administration (FDA) describes precision medicine as providing “the right patient with the right drug at the right dose at the right time”.<sup>(12)</sup> With the development of new ML algorithms, it is becoming feasible in the foreseeable future to analyze in real-time gigantic amounts of data directly streamed from the bedside in order to provide more personalized and relevant predictions. This field of stream analytics in which data are collected and used sequentially to update the current predicted algorithms is referred to as online ML.<sup>(7)</sup> Such an automated technology that is deployable bedside is the path for the ultimate goal of precision medicine. Thus, the next challenges are to create real-time support tools for personalized decision-making, allowing the clinician to better adapt his therapy for patients in critical situations. This current approach, called prescriptive analytics, refers to the prediction of treatment effects at the patient level. A statistical approach derived from causal inference methods may be used to estimate the benefit of treatment at the individual level rather than the population level. The

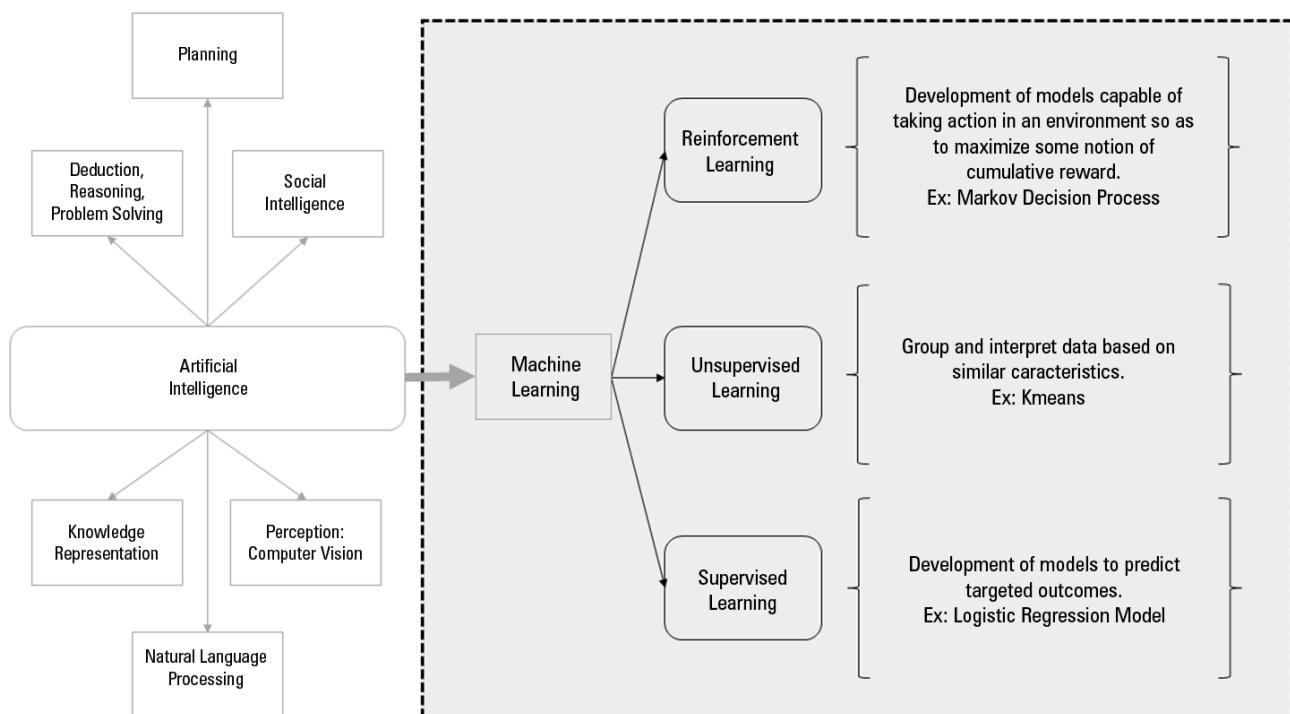


Figure 1 - Artificial intelligence and different types of machine learning algorithms.

definition and estimation of such parameters will allow, if coupled with Big Data, to support the clinician in his decisions by highlighting optimal therapeutic choice strategies. Komorowski et al. developed a computational model able to dynamically suggest optimal treatments for adult patients with sepsis in the ICU.<sup>(13)</sup>

### Current limitations and conclusions

One needs to acknowledge the existence of limitations that will need to be overcome in order to allow for targeted ML to become a reality in the future.<sup>(14)</sup> First, while ICUs are now generating gigabytes of data each day,

only a small fraction is currently accessible for research purposes.<sup>(15)</sup> Second, important questions remain about how best to leverage big medical data and ML in the ICU. Randomized controlled trials will be needed to demonstrate the benefit of predictive and prescriptive analytics in critically ill patients. However, considering recent advances, big medical data and ML offer a unique opportunity to dramatically change our paradigm from the era of evidence-based medicine in which therapeutic decisions are essentially based on population-level shreds of evidence to a new era of optimal and personalized clinical decision support.

---

## REFERENCES

- Kalbandi I, Anuradha J. A brief introduction on big data 5Vs characteristics and hadoop technology. *Proc Comput Sci*. 2015;48:319-24.
- Sanchez-Pinto LN, Luo Y, Churpek MM. Big data and data science in critical care. *Chest*. 2018;154(5):1239-48.
- Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. *Int J Med Inform*. 2018;114:57-65.
- Malak JS, Safdari R, Zeraati H, Nayeri FS, Mohammadzadeh N, Farajollah SS. An agent based architecture for high-risk neonate management at neonatal intensive care unit. *Electron Physician*. 2018;10(1):6193-200. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29588819>
- Herasevich V, Pickering BW, Dong Y, Peters SG, Gajic O. Informatics infrastructure for syndrome surveillance, decision support, reporting, and modeling of critical illness. *Mayo Clin Proc*. 2010;85(3):247-54.
- Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035.
- Pirracchio R, Cohen MJ, Malenica I, Cohen J, Chambaz A, Cannesson M, Lee C, Resche-Rigon M, Hubbard A; ACTERREA Research Group. Big data and targeted machine learning in action to assist medical decision in the ICU. *Anaesth Crit Care Pain Med*. 2019;38(4):377-84.
- Dinov ID. Volume and value of big healthcare data. *J Med Stat inform*. 2016;4. pii: 3.
- Gambus P, Shafer SL. Artificial intelligence for everyone. *Anesthesiology*. 2018;128(3):431-3.
- Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *Lancet Respir Med*. 2015;3(1):42-52.
- Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. *Sci Transl Med*. 2015;7(287):287ra71.
- Food and Drug Administration (FDA). U.S. Department of Health and Human Services. Paving the way for personalized medicine: FDA's Role in a New Era of Medical Product Development [Internet]. Maryland: FDA; 2013 [cited 2019 Jan 14]. Available from: <https://www.fda.gov/oc/2013/01/10-28-13-Personalized-Medicine.pdf>
- 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):1716-20.
- Naidus E, Celi LA. Big data in healthcare: are we close to it? *Rev Bras Ter Intensiva*. 2016;28(1):8-10.
- Celi LA, Mark RG, Stone DJ, Montgomery RA. "Big data" in the intensive care unit: Closing the data loop. *Am J Respir Crit Care Med*. 2013;187(11):1157-60.