Artificial Intelligence
in the ICU

Frederic Michard
MiCo
Vallamand, Switzerland
frederic.michard@bluewin.ch
@MichardFrederic
michardconsulting.com

Filipe A Gonzalez
Intensive Care Unit
Hospital Garcia de Orta, Almada
University of Lisboa
Portugal
filipeandregonzalez@gmail.com

Patrick Schoettker
Department of Anesthesiology
CHUV
University Hospital Lausanne
Switzerland
patrick.schoettker@chuv.ch

Artificial Intelligence in Anaesthesia and Critical Care - Temptations and Pitfalls

Big data, Artificial Intelligence (AI) and machine learning are buzzwords. In this article, we briefly discuss what they mean for anaesthesiologists and intensivists, focusing on existing clinical applications.

We are today able to collect and store a considerable amount of patient-related data. These "big" data are typically part of Electronic Medical Record (EMR) systems and usually combine demographic, clinical and biological information. They may also contain images (e.g. ultrasound cardiac images) and physiologic waveforms. These data can be analysed with simple descriptive methods to report basic information regarding patient characteristics and outcomes such as hospital mortality, morbidity, and length of stay. This approach, useful for benchmarking and research, does not require artificial intelligence (AI).

A step further in the data analysis process consists of using machine learning (ML) algorithms (a subfield of AI), which have been trained to detect specific patterns of disease states or adverse events. As of today, most ML innovations approved for medical use have been developed in the field of imaging (radiology and pathology). It is indeed relatively easy to train an algorithm with a large database of images so that it becomes capable of detecting abnormalities that could be missed by a medical trainee or a seasoned but distracted clinician. In this respect, many ML algorithms have been designed to analyse chest x-rays and CT scans and to suggest a diagnosis (e.g., tracheal tube not correctly positioned on the chest x-ray of a mechanically ventilated patient or CT scan images suggestive of COVID-19 in a patient with ARDS). Recently, ML algorithms have also been implemented into ultrasound machines to facilitate and automate point-of-care echocardiographic evaluations (Nabi et al. 2019).
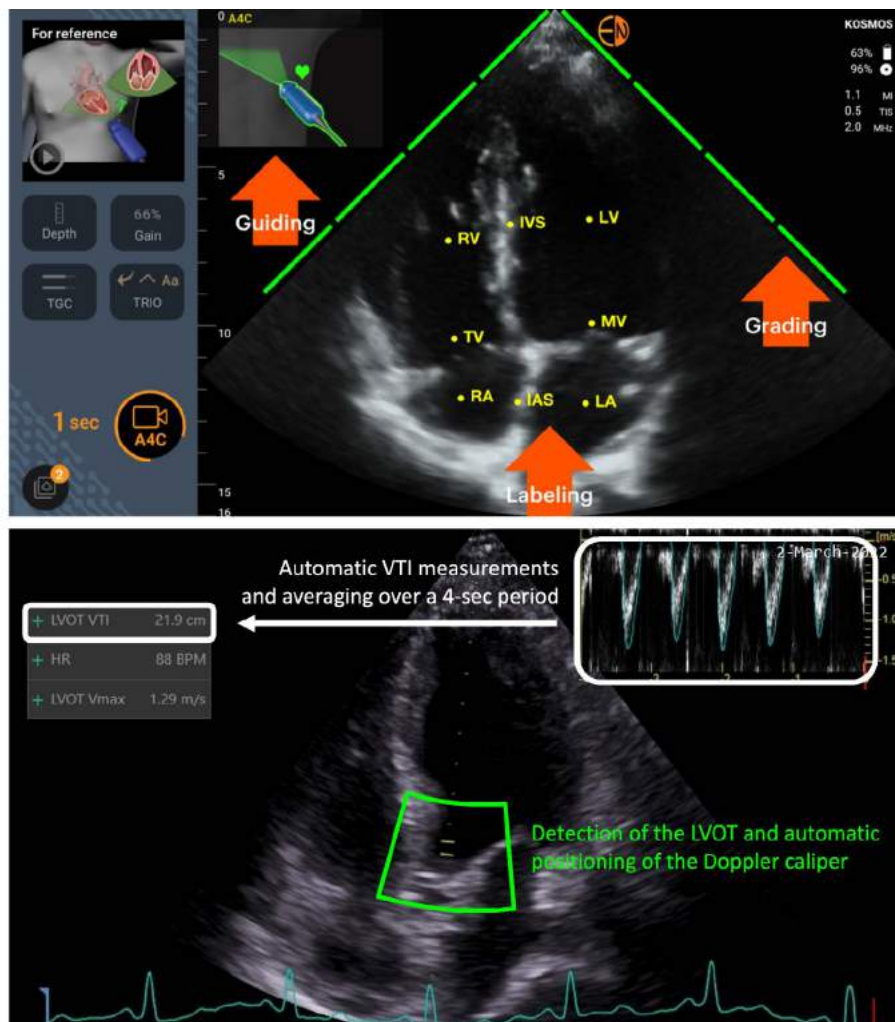
AI and Point-of-Care Echocardiography

Several ML algorithms have been trained to recognise heart images and guide users to hold and position their transthoracic probe correctly. Such algorithms are also able to grade image quality and label heart structures. An example is displayed in **Figure 1.** Some ML algorithms can take echocardiographic measurements automatically. For instance, the autoVTI algorithm can recognise a 5-chamber apical view of the heart, automatically position the pulse wave Doppler caliper in the left ventricular outflow tract and record the sub-aortic velocity time integral (VTI) over a short time window **(Figure 1)**. A recent clinical evaluation suggests that the autoVTI algorithm may help trainees to be as efficient as echocardiography experts in estimating VTI, stroke volume (SV ~ VTI x Pi) and cardiac output using ultrasounds (Gonzalez et al. 2022). Several ML algorithms have also been developed for the automatic estimation of left ventricular ejection fraction (LVEF). Comparison studies suggest they may enable novices to measure LVEF as accurately and with better reproducibility than experts taking manual measurements (Varudo et al. 2022). Other ultrasound algorithms have been designed to predict fluid responsiveness in mechanically ventilated patients from the automatic quantification of the inferior vena cava respiratory variation or to detect pulmonary oedema from the automatic quantification of lung B lines.

In summary, the value of ML algorithms to help novices perform point-of-care echocardiographic evaluations has been documented in several clinical studies. However, given the fact that the proportion of intensivists trained to perform echocardiography is increasing sharply, whether AI innovations are necessary to increase the number and quality of ultrasound haemodynamic evaluations remains to be established.

AI and Continuous Blood Pressure Monitoring

In the search for cuffless and continuous blood pressure monitoring techniques, ML algorithms have been proposed to estimate blood pressure and its changes from the

Figure 1. Examples of machine learning algorithms designed for point-of-care echocardiography. **Top** - Auto guiding to obtain an apical 4-chamber view of the heart, auto-grading to ensure optimal image quality, and auto-labelling. From EchoNous (Redmond, WA, USA), with permission. **Bottom** - Automatic detection of the apical 5-chamber view of the heart, automatic positioning of the Doppler caliper in the left ventricular outflow tract (LVOT) and automatic measurements of subaortic velocity-time integral (VTI). From GE Healthcare (Chicago, IL, USA) with permission.

analysis of photoplethysmographic (PPG) waveforms. Historically, PPG waveforms were recorded by medical-grade pulse oximeters, but they are today frequently obtained from smartwatches, adhesive patches, optical bracelets, rings or smartphone cameras (Festo et al. 2023). A few of these devices, mainly designed for the detection or follow-up of patients with chronic hypertension, have been cleared for medical use. Recent independent clinical evaluations suggest they may not always be able to detect the physiologic night-time dipping nor therapeutic changes in blood pressure (Tan et al. 2023). As a matter of fact, these devices require frequent recalibrations and carry the potential to track changes in blood pressure over short time periods rather than measure absolute numbers (Ghamri et al. 2020). Interestingly, this would not be an obstacle to their use during surgery, in ICU patients or even in hospital wards to detect hypotensive and hypertensive episodes and trigger intermittent blood pressure spot-checks with a reference clinical method (e.g.,

the oscillometric brachial cuff method). In these settings, the reference method would be used not only to confirm changes in blood pressure but also to recalibrate the algorithm.

AI to Forecast Clinical Deterioration

As mentioned above, ML algorithms can detect specific patterns of overt disease states. They can also be trained to detect patterns associated with pre-disease states or patterns observed before the occurrence of specific adverse events.

For instance, multiple ML algorithms have been developed to create scores (e.g., eCART or HAVEN scores) predicting severe adverse events in patients hospitalised in regular hospital wards. Several studies have shown these AI-derived scores are able to predict ICU admission, cardiac arrest, and death with an area under the curve (AUC) around 0.8-0.9 (as a reminder, a random guess would be associated with an AUC of 0.5 and a perfect prediction with an AUC of 1.0). However, their predictive value is frequently only slightly higher when not simply comparable to what is possible to achieve with existing scores such as the modified early warning score (MEWS) or the national early warning score (NEWS) - both scores which are easy to calculate from vital sign spot-checks (Bartkowiak et al. 2019).

Multiple attempts have been made to detect sepsis at an early stage, fasten therapeutic management and improve patient outcomes. As of today, the results of sepsis "sniffer" implementation programmes have been conflicting, with some reporting a decrease in time-to-antibiotic and in-hospital mortality (Shimabukuro et al. 2017), whereas others, including the recent evaluation of the EPIC system (widely used in the US), reported poor discrimination (AUC 0.63) and calibration in predicting the onset of sepsis (Wong et al. 2021). Another potential ML application is known as reinforcement learning. It enables the development of algorithms designed to provide dynamic therapeutic recommendations, which have been shown to be associated with improved organ func-

tion and/or survival (Komorowski 2018). Whether such prescriptive algorithms may be accepted by clinicians (particularly by experts in sepsis management) and may improve clinical outcomes remains unknown.

Machine learning algorithms have also been developed and proposed to predict postoperative morbidity and mortality, with reported AUCs that may exceed 0.9. However, this predictive value does not always overcome what is possible to achieve with simple scores such as the SORT score (Wong et al. 2018). Of note, the subjective prediction made by clinicians has been shown to be associated with an AUC of 0.89! (Wong et al. 2018). Therefore, whether there is a need for complex ML scores to predict postoperative outcomes remains debatable.

Machine learning algorithms have recently been proposed to predict haemodynamic instability and, more specifically, systemic hypotension. The hypotension prediction index (HPI) is a commercially available ML-derived score calculated from the analysis of the arterial pressure waveform. It has been shown to forecast intraoperative hypotension 5-15 minutes ahead with an AUC ranging between 0.75-0.95. However, recent publications have highlighted the fact that HPI is the mere reflection of the mean arterial pressure (MAP) and, as a result, that its predictive value may not be superior to MAP monitoring (Mulder et al. 2023).

In summary, the predictive value of machine learning algorithms is hardly disputable. However, the superiority over existing and simpler methods often remains to be determined, and the complexity/benefit and cost/benefit ratio may therefore be questioned.

The Pitfalls of Predictive Analytics

Predictive analytics is associated with at least four main limitations and/or pitfalls, which are summarised in **Figure 2**.

The first one is to believe that everything is predictable. As highlighted by Chen and Asch in a famous New Engl. J. Med. editorial (Chen and Asch 2017), "no amount of algorithmic finesse or computer power can squeeze out information that is not present". Google X, an Alphabet subsidiary, reported that its initiative to discover a biomarker for depression and anxiety in brainwave data fell short of its goal. Given the fact that they had almost unlimited resources and an army of top-level computer scientists working on this project, it is likely that brainwave data simply did not contain the predictive information they were looking for. In addition, some events are unpredictable by nature. As an example, which algorithm could predict hypotension related to surgical injury (e.g., vena cava injury during liver surgery) or the decision to deepen anaesthesia or sedation with a propofol bolus? During surgery and in ICUs, multiple external factors are susceptible to modify clinical trajectories in one direction or the other. When steady states do not exist, it becomes challenging to predict short-term clinical trajectories (Michard and Teboul 2019).

Secondly, poor data quality is one of the main factors holding up the big data revolution in healthcare (Dhindsa et al. 2018). This limitation is often summarised as "garbage in, garbage out". Indeed, one may use the best predictive algorithm, but if we feed it with wrong data, artefacts and/or damped physiologic waveforms, one may logically end up with wrong predictions.

Thirdly, it is paramount to understand that predicting does not necessarily mean preventing. When the prediction is not followed by one or more appropriate actions susceptible to modify the clinical trajectory, logically, nothing can be prevented. In the largest HPI randomised controlled

1. Everything is not predictable
- No amount of algorithmic finesse can squeeze out **information that is not present**
- In the OR and ICU, many **external factors** can modify short-term clinical trajectories

2. Garbage in, garbage out
- Poor **data quality** impacts the ability to forecast adverse events

3. Predicting is not preventing
- Forecasting must be followed by an **action plan**
- Clinicians may not **trust** and follow AI recommendations

4. Can we treat probabilities?
- Who would **take the risk** to give a treatment to patients who may not need it?

Figure 2. The four main pitfalls of predictive analytics

trial published so far (Maheshwari et al. 2020), anaesthesiologists who were alerted about the risk of hypotension failed to prevent hypotensive events. Interestingly, it appeared that most of them did not feel the need and/or the right to give fluid, vasopressors, or inotropes to patients who were still haemodynamically stable and only had a probability of becoming hypotensive. This finding is an excellent illustration of the reluctance of clinicians to trust and follow AI recommendations (Gaube et al. 2021).

Fourthly, there are risks associated with the treatment of probabilities. Therefore, one may hardly envision being proactive from a therapeutic standpoint. One may be proactive by performing bacteriological samples when predicting sepsis or by upgrading surveillance when predicting clinical deterioration (e.g., by offering continuous monitoring and/or ICU admission). There is no harm in doing so. There might be economic consequences, but no

harm to the patient. In contrast, giving antibiotics to a probability of sepsis or administering vasopressors to a probability of hypotension might be risky and is, therefore, questionable (Michard and Futier 2023). Who would accept receiving treatment with known side effects for a predicted disease or adverse event that may never occur? And who would be responsible in case of complications?

Conclusion

Big data, AI, and, more specifically, machine learning algorithms are hot topics for medical journals and scientific events. For start-ups, they are also very useful keywords to raise funds. However, one may acknowledge that, as of today, and from a practical standpoint, the AI elephant gave birth to a mouse in the field of anaesthesiology and intensive care. Prospective clinical trials are indispensable not only to assess the safety of AI innovations but also to demonstrate superiority over exist-

ing and simpler methods. In the digital medicine era, whereas many medical students are eager to work on AI projects and to participate in datathons, it might be useful to remind them that "the immediate challenge to improving quality of care is not discovering new knowledge, but rather how to integrate what we already know into practice" (Urbach and Baxter 2005). Therefore, although we should keep our eyes and ears wide open for AI innovations, we should also continue to focus on basic initiatives (more nurses and doctors, better training with simulation, better compliance to existing guidelines, and better use of existing monitoring tools) that are known to improve patient outcomes and satisfaction.

Conflict of Interest

FM is the founder and managing director of MiCo, a consulting and research firm based in Switzerland. FAG and PS have nothing to disclose.

References

Bartkowiak B, Snyder AM, Benjamin A et al. (2019) Validating the electronic cardiac arrest risk triage (eCART) score for risk stratification of surgical inpatients in the postoperative setting: retrospective cohort study. Ann Surg. 269:1059-63.

Chen JH, Asch SM (2017) Machine learning and prediction in medicine - Beyond the peak of inflated expectations. N Engl J Med. 376:2507-2509.

Dhindsa K, Bhandari M, Sonnadara RR (2018) What's holding up the big data revolution in healthcare? BMJ. 363:k5357.

Festo C, Vannevel V, Ali H, et al. (2023) Accuracy of a smartphone application for blood pressure estimation in Bangladesh, South Africa, and Tanzania. npj Digit. Med.

Gaube S, Suresh H, Raue M, et al. (2021) Do as AI say: Susceptibility in deployment of clinical decision-aids. npj Digit. Med. 4:31.

Ghamri Y, Proença M, Hofmann G et al. (2020) Automated pulse oximeter waveform analysis to track changes in blood pressure during anesthesia induction: a proof-of-concept study. Anesth Analg. 130:1222-33.

Gonzalez FA, Varudo R, Leote J et al. (2022) The automation of sub-aortic velocity time integral measurements by transthoracic echocardiography: clinical evaluation of an

artificial intelligence-enabled tool in critically ill patients. Br J Anaesth. 129:e116-e119

Komorowski M, Celi LA, Badawi O et al. (2018) The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med. 24:1716-20.

Maheshwari K, Shimada T, Yang D et al. (2020) Hypotension prediction index for prevention of hypotension during moderate-to high-risk non-cardiac surgery. Anesthesiology. 133:1214-1222.

Michard F, Futier E (2023) Predicting intraoperative hypotension: from hope to hype and back to reality. Br J Anaesth.131: 199-201.

Michard F, Teboul JL (2019) Predictive analytics: beyond the buzz. Ann Intensive Care. 9:46.

Mulder MP, Harmannij-Markusse M, Donker DW et al. (2023) Is continuous intraoperative monitoring of mean arterial pressure as good as the hypotension prediction index algorithm? Anesthesiology. 138:657-658.

Nabi W, Bansal A, Xu B (2021) Applications of artificial intelligence and machine learning approaches in echocardiography. Echocardiography. 38:982-92.

Shimabukuro DW, Barton CW, Feldman MD et al. (2017) Effect of machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomized clinical trial. BMJ Open Respir Res. 4:e000234.

Tan I, Gnanenthiran SR, Chan J et al. (2023) Evaluation of the ability of a commercially available cuffless wearable device to track blood pressure changes. J Hypertension. 41:1003-10.

Urbach DR, Baxter NN (2005) Reducing variation in surgical care. BMJ. 330:1401-2.

Varudo R, Gonzalez FA, Leote J et al. (2022) Machine learning for the real-time assessment of left ventricular ejection fraction in critically ill patients: a bedside evaluation by novices and experts in echocardiography. Crit Care. 26:386.

Wong A, Otles E, Donnely JP et al. (2021) External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med. 181:1065-70.

Wong DJN, Harris S, Sahni A et al. (2020) Developing and validating subjective and objective risk-assessment measures for predicting mortality after major surgery: An international prospective cohort study. PLoS Med. 17:e1003253.