Artificial Intelligence
in the ICU

Vitaly Herasevich
Professor of Anesthesiology and Medicine
Department of Anesthesiology and Perioperative Medicine
Mayo Clinic
Rochester, Minnesota, USA
vitaly@mayo.edu

Daniel Diedrich
Associate Professor of Anesthesiology
Department of Anesthesiology and Perioperative Medicine
Mayo Clinic
Rochester, Minnesota, USA
diedrich.daniel@mayo.edu

Brian Pickering
Professor of Anesthesiology
Department of Anesthesiology and Perioperative Medicine
Mayo Clinic
Rochester, Minnesota, USA
pickering.brian@mayo.edu

Why Artificial Intelligence is Not Fixing the Problem of Sepsis in the Hospital

The use of Artificial Intelligence in detecting sepsis, AI prediction/detection models and how these healthcare tools need to complement clinical expertise.

The potential of Artificial Intelligence (AI) in healthcare has generated significant excitement and discussion. However, it's important to distinguish between the theoretical promise and the current state of evidence-based applications. AI has the capacity to drive a paradigm shift in healthcare, but its real-world impact is still being explored and refined. One of the driving factors behind the AI revolution in healthcare is the increasing availability of clinical data, largely attributed to the adoption of Electronic Medical Records (EMRs). EMRs have transformed the way patient data is stored, accessed, and analysed, providing a rich source of information that can be leveraged for various AI applications.

Adoption of Electronic Medical Records was slow. Mayo Clinic in 2005 was only one of 0.1% of USA hospitals with a fully digitised medical record (HIMSS Stage 7 criteria). That advantage allowed our institution to develop one of the first severe sepsis and septic shock electronic surveillance programme in 2006. This was especially important considering the ongoing challenges associated with sepsis management, particularly in hospitals with limited resources. These results represented a major step forward in leveraging EMR data but were far from exceptional, with sensitivity of 48%, specificity of 86%, and a positive predictive value of 32% (Herasevich et al. 2008). Tuning and optimisation of that algorithm over time resulted in improved performance with a sensitivity of 80% and a specificity of 96% (Harrison et al. 2015). When implemented in practice, the developed sepsis sniffer demonstrated a sensitivity of 79.9%, specificity of 76.9%, positive predictive value (PPV) of 27.9%, and negative predictive value (NPV) of 97.2%, which is similar in performance to other systems (Lipatov et al. 2022), highlighting the feasibility of such surveillance tools in the context of EMRs and sepsis management. Although this study didn't demonstrate changes in bundle compliance or hospital mortality, our experience from early EMR adoption to the development of advanced sepsis surveillance systems underscores the iterative nature of healthcare technology development and implementation.

In the intervening years, the adoption of EMRs has spread across the country and has dramatically increased the availability of clinical data, which may be used for research and development of novel informatics tools and the application of AI.

Last year, we published (ICU Management & Practice, Volume 22 - Issue 2, 2022) a manuscript which highlighted the lessons learned from a decade of studying sepsis surveillance and a possible path forward. In this manuscript, we discuss the use of AI in the detection of sepsis.

The concept of prediction in healthcare, especially in terms of disease onset and outcomes, has been a longstanding interest among physicians and other healthcare practitioners. This interest can be traced back to the time of Hippocrates and his famous aphorism "Primum non nocere" (First, do no harm), and can underscore the importance of predicting disease trajectories in an effort to provide effective treatments and thus minimise harm. Sepsis is a particularly challenging condition when it comes to prediction. Its non-specific early symptoms can often lead to delays in recognition and treatment, which in turn can result in poor patient outcomes. The introduction of the concept of Systemic Inflammatory Response Syndrome (SIRS) was a step toward recognising the broader signs of an inflammatory response in sepsis, but this broad definition has introduced challenges for developers trying to develop specific and accurate prediction algorithms.

Approximately 87% of sepsis cases originate outside of the hospital (Rhee et al. 2017), and this emphasises the critical role of the Emergency Department (ED) in the initial diagnosis and management of this condition. Much effort has been placed on devising an accurate sepsis prediction score for ED providers. Different diagnostic criteria for sepsis, such as the Sequential (Sepsis-Related) Organ Failure Assessment (qSOFA) score and
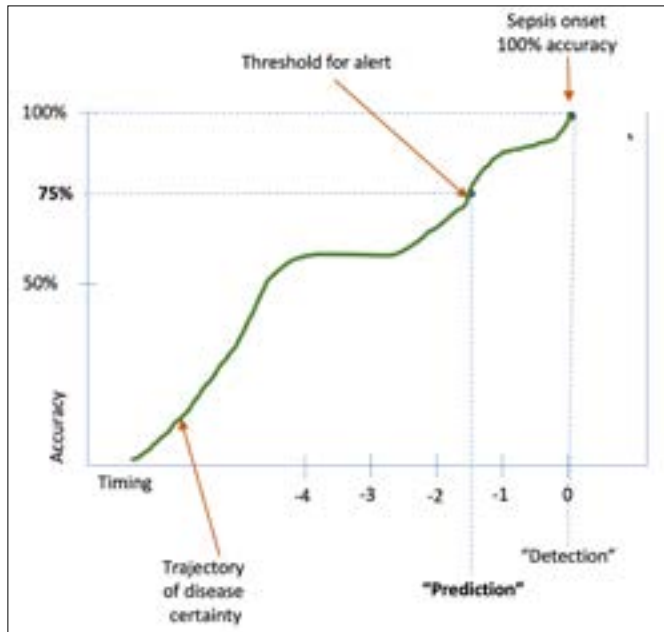
Figure 1. Time "zero" is the onset of disease (sepsis) when that could be predicted/detected with 100% accuracy. Everything earlier is prediction – with different levels of accuracy. The trajectory of disease certainty could have very different shapes.

the Systemic Inflammatory Response Syndrome (SIRS) criteria, have been subject to various studies and evaluations in the ED setting, leading to sometimes conflicting results. One study conducted in two European clinical teaching hospitals in the Netherlands (Mignot-Evers et al.) found that the qSOFA score performed as well as or better than the SIRS criteria for identifying culture-positive sepsis and predicting in-hospital mortality and ICU admission, suggesting that the qSOFA score might be a valuable tool in the ED for stratifying patients' risk and informing clinical decisions. This finding aligns with the growing emphasis on the qSOFA score as a tool for quickly assessing patients at risk of sepsis-related organ dysfunction. A separate study published one year earlier (Gando et al. 2020) found that SIRS criteria had better performance for predicting infection than qSOFA scores in the ED, highlighting the complexity of sepsis diagnosis. It's

possible that different patient populations, settings, and factors influence the performance of these criteria. This variability underscores the importance of considering multiple factors, including the specific patient population and the clinical context, when evaluating and applying diagnostic criteria.

These differing results highlight the need for ongoing research and validation of sepsis diagnostic criteria, especially in the ED setting where early and accurate diagnosis is crucial. Additionally, it's important to recognise that clinical assessment and judgment play a significant role alongside these diagnostic tools. The decision-making process should be guided by a combination of clinical experience, available evidence, and the specific needs of each patient. The concept of certainty and accuracy, as well as the practical implications of using AI prediction models, are key considerations when applying these models to real-world healthcare scenarios.

In the context of AI and predictive modelling, the terms "prediction" and "detection" can be seen as points along a continuum of certainty and accuracy. Detection implies a high degree of certainty and accuracy, often approaching 100%. In contrast, prediction involves a range of probabilities or likelihoods of an event occurring, indicating varying levels of certainty **(Figure 1)**. AI experts often quote explainability as the key to usefulness in clinical practice. We would argue that this is less important for acceptability and meaningfulness than the distinction between prediction and detection. In practical terms, for a clinician, the question of when to act boils down to risk versus benefit. AI prediction/detection models in healthcare are tools that should complement clinical expertise. For an AI to be useful, they have to add something to the decision-makers' mental model. They need to reduce cognitive load by parsing data from large volumes of clinical data or to detect patterns and signals in multidimensional data that are difficult for individual clinicians to see in the moment of decision-making. Striking the right balance between accuracy, interpretability, and clinical utility is key. As the field continues to evolve, interdisciplinary collaboration between AI experts and healthcare professionals will be essential to ensure the meaningful integration of these

models into the clinical setting.

A critical consideration in the application of AI prediction models in healthcare is the trade-off between accuracy and practical utility. Predicting with 95% accuracy five minutes before the onset of sepsis has very limited practical utility. The same applies to a 12-hour prediction with 25% accuracy. Recent prospective validation of the AL/ML sepsis prediction model from a commercial EMR vendor failed to identify 67% of patients with sepsis and generated an alert for 18% of all hospitalised patients (Wong et al. 2021). Determining what constitutes an acceptable level of accuracy and how early predictions need to be made for meaningful clinical impact is a complex challenge that involves balancing various factors.

1. Acceptable Level of Accuracy: Different settings will have different acceptable levels of accuracy. Sensitivity is important in the home environment through the ED, where the consequences of a missed diagnosis could be devastating. Balancing this against the risks of overtreatment or false positive alert fatigue must be determined with all stakeholders, which will be essential in striking the right balance.

2. Lead Time: Early detection is valuable, but the lead time for predictions must be balanced with accuracy. Predicting an event too far in advance with limited accuracy might not be acceptable. The lead time needed for interventions to meaningfully impact the clinical condition should be used to guide the development of prediction models.

3. Clinical Workflow: The integration of prediction alerts into clinical workflows is vital. If alerts disrupt workflows or lead to alert fatigue, their utility diminishes. Alerts should be timely, actionable, and integrated into the existing care process.

4. Specificity and Sensitivity: It's important to assess both sensitivity (true positive rate) and specificity (true negative rate) of a prediction model. An overly sensitive model might produce numerous false positives, while an overly specific model could miss true positives.

5. Prospective Validation: A model's performance in real-world clinical scenarios might differ from its performance in

controlled research settings. Prospective evaluation against gold standard clinical evaluation is essential prior to more widespread implementation.

6. Population Variability: Patient populations can vary, and models should ideally be trained and validated on diverse patient cohorts to ensure generalisability.

7. Continuous Improvement: AI models should undergo continuous improvement based on feedback and real-world performance. Feedback loops that enable refining the model's accuracy and clinical impact are essential. In a related topic, post-market surveillance and reporting should be included with any model deployment. This will ensure that unintended cases of harm resulting from model deployment are picked up early.

Does this mean AI/ML methods are not useful in sepsis prediction? The key to their success lies in developing intelligent and context-aware systems that go beyond simple associative models based on available Electronic Medical Record (EMR) data. While challenges exist, smarter approaches can harness the power of AI to improve sepsis prediction and patient outcomes. Here are some considerations for developing effective AI-driven sepsis prediction systems that we have learned from our experience of building these alerts for over 20 years;

Feature Engineering: Instead of relying solely on raw EMR data, effective sepsis prediction models can benefit from careful feature engineering. This involves partnering with clinicians and selecting relevant patient variables, incorporating time-series data, and considering the mechanisms of sepsis progression. Future generations of AIs (Large language models or generative AI) may have access to such large quantities of data and incorporate powerful new analytics approaches to achieve mechanistic insight without the need for feature engineering, but for now, this is a step we advocate.

Multimodal Data Integration: AI models can be enhanced by integrating multiple data sources beyond EMRs, such as laboratory results, vital signs, imaging data, novel sensors, computer vision, work context, and patient demographics. This broader dataset could be useful in improving the performance of algorithms in real-world clinical situations.

Time-Series Analysis: Sepsis often exhibits dynamic changes over time. Advanced AI methods, like time-series analysis and recurrent neural networks, can capture temporal patterns and trends, allowing for more accurate predictions.

Clinical Context: Incorporating clinical context, such as patient history, co-morbidities, and clinical guidelines, can enhance the predictive power of AI models. This extends to the work setting (home versus ED versus ICU). Context-aware models can calibrate to the operating conditions and offer more meaningful predictions that align with actual clinical scenarios.

Multi-model approach: Combining predictions that take advantage of Boolean logic, multiple AI models or algorithms (ensemble approaches) can improve accuracy and reduce the impact of individual model weaknesses.

Interpretability: Developing models that provide not just predictions but also explanations for those predictions can be useful for stakeholder buy-in, building trust and facilitating shared decision-making.

Continuous Learning: AI models should be designed for continuous learning, adapting to changes in patient populations and healthcare practices over time. A mechanism for automatically capturing clinical insights, health system and patient population outcomes and making these available as training data for the model should be included in the implementation environment. This will facilitate the realisation of a learning health system.

Real-Time Integration: For early sepsis detection, real-time integration with clinical workflows and rapid response systems is essential. This ensures timely interventions and avoids delays in care delivery.

Clinical Validation: Rigorous clinical validation in diverse settings is crucial to demonstrate the effectiveness and reliability of AI-driven sepsis prediction systems.

Human-Machine Collaboration: AI should augment, not replace, clinical expertise. The goal should be to develop models that are implemented in a way that promotes collaboration between AI systems and healthcare professionals.

Taken together with advances in monitoring, data access, computing power and sensor miniaturisation, there is a very high likelihood in the near future that AI-powered clinical digital assistance will be available and used in healthcare settings.

Conflict of Interest

None.

References

Gando S, Shiraishi A, Abe T et al. (2020) Sepsis Prognostication in Intensive Care Unit and Emergency Room (SPICE) (JAAM SPICE) Study Group. The SIRS criteria have better performance for predicting infection than qSOFA scores in the emergency department. Sci Rep. 10(1):8095.

Harrison AM, Thongprayoon C, Kashyap R et al. (2015) Developing the surveillance algorithm for detection of failure to recognize and treat severe sepsis. Mayo Clin Proc. 90(2):166-75.

Herasevich V, Afessa B, Chute CG, Gajic O (2008) Designing and testing computer-based screening engine for severe sepsis/septic shock. AMIA Annu Symp Proc. 966.

Lipatov K, Daniels CE, Park JG et al. (2022) Implementation and evaluation of sepsis surveillance and decision support in medical ICU and emergency department. Am J Emerg Med. 51:378-383.

Mignot-Evers L, Raaijmakers V, Buunk G et al. (2021) Comparison of SIRS criteria and qSOFA score for identifying culture-positive sepsis in the emergency department: a prospective cross-sectional multicentre study. BMJ Open. 11(6):e041024.

Rhee C, Dantes R, Epstein L et al. (2017) Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. JAMA. 318(13):1241-1249.

Wong A, Otles E, Donnelly JP et al. (2021) External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med. 1;181(8):1065-1070.