Neurocrit Care (2022) 37:S157–S159
https://doi.org/10.1007/s12028-022-01558-4

NEUROCRITICAL CARE SOCIETY

EDITORIAL

Navigating the Ocean of Big Data in Neurocritical Care

Rajat Dhar[1*] and Geert Meyfroidt[2]

Artificial intelligence (AI) will inevitably infiltrate and increasingly influence clinical practice and research in neurocritical care. The question is no longer whether or even when this will happen but how we respond to the opportunities and threats of exploring this terra incognita: some clinicians may watch anxiously from the shore as the AI armada sets sail for uncharted territory, whereas those who are more adventurous have elected to embark as "captains of big data" and determine the trajectory and destination of this scientific journey. Navigating through this still murky ocean of big data requires knowledge of evolving statistical and data analytic techniques that look to many like black magic. In reality, there is no mystical conjuring involved: machine learning techniques derive rules and reveal patterns in data by leveraging statistical analyses at an enormous scale and by combining variables in nonlinear and multidimensional ways. Just as with other statistical approaches, they cannot predict the future or make definitive causal inferences. It is imperative that physicians possess at least a modest understanding of the possibilities and limitations of AI to thrive in this technological future. This special issue of *Neurocritical Care* aims to guide the reader through this fascinating frontier of research. We invited several experts to review, debate, and provide perspectives on key aspects of these techniques and their potential to impact our field. We also include several original research submissions that harness big data and AI approaches to advance our understanding of common and vexing challenges in neurocritical care.

*Correspondence: dharr@wustl.edu
[1] Division of Neurocritical Care, Department of Neurology, Washington University School of Medicine in St. Louis, Campus Box 8111, 660 S Euclid AvenueSaint Louis, St. Louis, MO 63139, USA
Full list of author information is available at the end of the article

A few key themes emerge that are worth highlighting. Overall, the goals of big data and the neurocritical care community seem aligned: we all agree that bringing a more personalized, precision-oriented approach to the care of our patients is a long-sought and worthy aspiration [1]. For too long our field has been driven by uniform and often arbitrary physiologic targets for managing blood pressure, intracranial pressure, and other key parameters—derived from epidemiological associations with outcome—without the capacity for cogent individualization [2]. This is an inherent limitation of epidemiological data, further reinforced by the limits of the human mind: it is impossible for us absorb and integrate the avalanche of data available on our immeasurably complex patients and find patterns that would allow for a more targeted approach. The core philosophy of medical big data is to collect, curate, analyze, understand, and then harness the variability in data to identify biomarkers for risk stratification and individualization of targets and to provide actionable early warning systems. AI models can integrate dynamic data streams at high temporal resolution to make accurate risk predictions [3]. Building on this philosophy, a promising method has been developed to classify and predict delayed cerebral ischemia after subarachnoid hemorrhage by using a complex time series analysis algorithm called "time-varying temporal signal angle measurement" [4]. Such methods, if proven effective in a clinical setting, may shift our focus from a reactive approach to a proactive approach, harnessing multiple dimensions of data to prevent brain injury and target proactive interventions.

One of the key motivations for big data is to understand the immense heterogeneity in patient responses to brain injury. Such variability can be problematic and has led to the failure of many clinical trials that seek to apply a single intervention to a broad group of patients. Two articles in this special issue seek to leverage data to

Springer

dissect endotypes, which are patient subgroups that share biologic characteristics and might respond in a more uniform way. Azad et al. [5] propose a data-driven endotype approach for traumatic brain injury. The PREcision Care in cardiac arrest: Influence of Cooling duration on Efficacy in Cardiac Arrest Patients (PRECICECAP) investigators apply a similar data-driven philosophy to cardiac arrest [6] to "better define patterns of injury and tailor acute interventions" as a potential way to improve patient outcomes. Indeed, such an approach has great promise in resolving some of the challenges we face in managing complex patients and might lead to "smarter" enrollment in clinical trials [7]. In addition, the review by Podell and colleagues [8] discuss how capturing and analyzing physiologic data can be leveraged for real-time assessment of autonomic nervous system dysfunction after brain injury in ways and at scales invisible to the bedside clinician.

These analyses rely on collecting large volumes of high-dimensional and curated data, including imaging and brain/systemic physiological conditions, and extracting consistent injury signatures. One major challenge in bringing big data to bear in these ways is the need to collaborate across sites and bring together nations and professional societies. The insightful perspective provided by the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) authors highlights the challenges of multicenter data collection, including the immense workload of data curation and harmonization [9]. Moberg and colleagues [10] provide a useful perspective on the harmonization of physiologic data, stressing the need for standards for interoperability and clear metadata and labeling so we can make meaningful use of all the data available. The PRECICECAP investigators propose a platform to integrate physiologic data with electronic health record annotations for harmonization of data across sites and even modules and pipelines to facilitate analysis [6]. One key challenge to high-throughput phenotyping is abstracting outcomes from unstructured reports (such as clinical or radiology notes): natural language processing may help solve this scalability problem, as demonstrated in one study in this issue, in which it was employed to extract edema-related and hemorrhage-related outcomes from 2,289 radiology reports of patients with stroke [11]. In a similar vein, an imaging platform has recently been initiated to bring together computed tomography and magnetic resonance imaging for stroke and other brain injuries. Its goal is to facilitate high-throughput processing of brain images by using AI algorithms to measure phenotypes such as edema and hemorrhage from large cohorts of patients [12–14]. Dissecting the observed variability at a scale sufficient to find key pathways and mechanisms might allow scientists to begin evaluating the genetic basis for brain injury phenotypes [15]. Such collaborative data-driven work also aligns with the goals of the Curing Coma campaign, a concerted effort by the Neurocritical Care Society to advance the science of all disorders of consciousness [16]. This will require continued focus on building such pipelines, including encouraging collaborations, concrete coordinated strategies for collection, curation, and harmonization of multimodal data, statistical approaches to finding patters and endotypes, and enriched trials to test new therapies. [17].

In parallel, we need to recognize the persistent challenges and limitations of the big data approach. The allure and immense hype of AI, as Dr. Citerio points out in his editorial, has not yet been realized [18]. In fact, there remains a chasm between the promise and the reality. There are a growing number of publications expounding the predictive value of machine learning models, but there are only a precious few demonstrating that the implementation of such algorithms actually improves anything; few are even validated beyond a single population, ignoring diversity and potential biases [19]. Practitioners often struggle when reviewing the ever-expanding panoply of publications applying machine learning techniques to data analysis, not knowing how to critically review the methodological machinations behind the curtain of these algorithms. Gravesteijn and colleagues [20] provide a valuable review of machine learning's strengths and pitfalls, including the use of clustering to find endotypes (as proposed in the articles cited above) and the critical need for skepticism and validation of findings from machine learning studies. In addition, we must remember that an ethical and equity-minded perspective must be applied when collecting and implementing data-driven decision making in medicine [21]. It is possible (but not a given) that AI will assist us with the goal of expanding access of underserved communities to consistent, high-quality neurocritical care expertise, broadening equitable care in the face of inevitable shortages of neurointensivists [2]. However, gaining traction requires the building of clinician (and patient) trust, and to achieve this models must be transparent or at least provide some degree of interpretability. It is unlikely that we will accept a machine's decision to perform an invasive intervention on a patient without understanding what factors are driving this alert. A perspective on enhancing the interpretability of machine learning models is provided by Moss and colleagues [22].

In summary, we believe that without conscious planning and careful validation, there is a real risk that the potential of big data and AI will fail to meet the promise of transforming medicine. Early models have demonstrated a potential to tackle many of the current challenges in neurocritical care and are ready to move toward bedside implementation and prospective validation in clinical trials. Ultimately, we need more thoughtful science, and not necessarily more data, to solve some of the issues with big data models [23]. We hope that the perspectives of the publications in this special issue will begin a concerted conversation to that end.

Author details
[1] Division of Neurocritical Care, Department of Neurology, Washington University School of Medicine in St. Louis, Campus Box 8111, 660 S Euclid AvenueSaint Louis, St. Louis, MO 63139, USA. [2] Department and Laboratory of Intensive Care Medicine, University Hospitals Leuven, Louvain, Belgium.

Publisher's Note

References
1. Shrestha GS, Suarez JI, Hemphill JC 3rd. Precision medicine in neurocritical care. JAMA Neurol. 2018;75(12):1463–4.
2. Hemphill JC 3rd. Pro: neurocritical care big data and AI: it's about expertise. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-021-01434-7.
3. Raj R, Luostarinen T, Pursiainen E, et al. Machine learning-based dynamic mortality prediction after traumatic brain injury. Sci Rep. 2019;9(1):17672.
4. Megjhani M, Weiss M, Kwon SB, et al. Vector angle analysis of multimodal neuromonitoring data for continuous prediction of delayed cerebral ischemia. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-022-01481-8.
5. Azad TD, Shah PP, Kim HB, Stevens RD. Endotypes and the Path to precision in moderate and severe traumatic brain injury. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-022-01475-6.
6. Elmer J, He Z, May T, et al. Precision Care in Cardiac Arrest: ICECAP (PRECICECAP) study protocol and informatics approach. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-022-01464-9.
7. Prescott HC, Calfee CS, Thompson BT, Angus DC, Liu VX. Toward smarter lumping and smarter splitting: rethinking strategies for sepsis and acute respiratory distress syndrome clinical trial design. Am J Respir Crit Care Med. 2016;194(2):147–55.
8. Podell J, Pergakis M, Yang S, et al. Leveraging continuous vital sign measurements for real-time assessment of autonomic nervous system dysfunction after brain injury: a narrative review of current and future applications. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-022-01491-6.
9. Maas AIR, Ercole A, De Keyser V, Menon DK, Steyerberg EW. Opportunities and challenges in high-quality contemporary data collection in traumatic brain injury: the CENTER-TBI experience. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-022-01471-w.
10. Moberg R, Moyer EJ, Olson D, Rosenthal E, Foreman B. Harmonization of physiological data in neurocritical care: challenges and a path forward. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-022-01524-0.
11. Miller MI, Orfanoudaki A, Cronin M, et al. Natural language processing of radiology reports to detect complications of ischemic stroke. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-022-01513-3.
12. Foroushani HM, Dhar R, Chen Y, et al. The Stroke Neuro-Imaging Phenotype Repository (SNIPR): an open data science platform for stroke research. Front Neuroinform. 2021;15: 597708.
13. Dhar R, Falcone GJ, Chen Y, et al. Deep learning for automated measurement of hemorrhage and perihematomal edema in supratentorial intracerebral hemorrhage. Stroke. 2020;51(2):648–51.
14. Dhar R, Hamzehloo A, Kumar A, et al. Hemispheric CSF volume ratio quantifies progression and severity of cerebral edema after acute hemispheric stroke. J Cereb Blood Flow Metab. 2021;41(11):2907–15.
15. Kirsch E, Szejko N, Falcone GJ. Genetic underpinnings of cerebral edema in acute brain injury: an opportunity for pathway discovery. Neurosci Lett. 2020;730: 135046.
16. Provencio JJ, Hemphill JC, Claassen J, et al. The Curing Coma Campaign: framing initial scientific challenges-proceedings of the first Curing Coma Campaign Scientific Advisory Council Meeting. Neurocrit Care. 2020;33(1):1–12.
17. Hammond FM, Katta-Charles S, Russell MB, et al. Research needs for prognostic modeling and trajectory analysis in patients with disorders of consciousness. Neurocrit Care. 2021;35(Suppl 1):55–67.
18. Citerio G. Big data and artificial intelligence for precision medicine in the neuro-ICU: bla, bla, bla. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-021-01427-6.
19. van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med. 2021;47(7):750–60.
20. Gravesteijn BY, Steyerberg EW, Lingsma HF. Modern learning from big data in critical care: primum non nocere. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-022-01510-6.
21. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. N Engl J Med. 2018;378(11):981–3.
22. Moss L, Corsar D, Shaw M, Piper I, Hawthorne C. Demystifying the black box: the importance of interpretability of predictive models in neurocritical care. Neurocrit Care. 2022. https://doi.org/10.1007/s12028-022-01504-4.
23. Telenti A. Do we need more data or more science in data science? Forbes. 2020. https://www.forbes.com/sites/forbestechcouncil/2020/02/20/do-we-need-more-data-or-more-science-in-data-science/?sh=6164b84b6cc6. Accessed 9 Jun 2022.