Artificial Intelligence and the Risk for Intuition Decline in Clinical Medicine

Andres Duarte-Rojo, MD, PhD[1] and Ervin Sejdic, PhD[2]

Abstract: Artificial intelligence (AI) is revolutionizing big data analytics. In this issue of *The American Journal of Gastroenterology*, Ahn et al. introduce the AI-cirrhosis-electrocardiogram score that can grade the electrophysiologic cardiac changes present in patients with cirrhosis. Apart from showing excellent accuracy to identify cirrhosis, the AI-cirrhosis-electrocardiogram algorithm identified a biological gradient and signal reversibility after transplantation. Clinical applicability needs to be determined. Some concerns inherent to the use of AI are discussed, including the need to verify that the quality of data used for machine training is optimal. The black box nature of AI-identified associations is discussed, along with the lack of pathophysiologic coherence allowing intuitive medical reasoning.

Am J Gastroenterol 2022;117:401–402. https://doi.org/10.14309/ajg.0000000000001618

Big data analytics is the result of decades of technological innovation. As our personal and professional virtual storage grew from megabytes to gigabytes, and then to terabytes, the access to research data files containing thousands to millions of observations also grew exponentially. The Health Information Technology for Economic and Clinical Health Act, part of the American Recovery and Reinvestment Act of 2009, was created to incentivize hospitals and medical groups to implement the use of electronic health records. Because of such policy and the collaboration between physicians and computer scientists, access to big data was simplified. Nowadays most medical practices can process internal (or publicly available) data sets for the generation of knowledge or operations management.

Since the most basic machine learning description by Alan Turing in the late 1940s, artificial intelligence (AI) has drastically evolved, and it currently dominates most of our lives. Heuristic algorithms provide results for internet searches and recommend the music we listen to, the movies we watch, and the news that we read daily. Given the inherent difficulties in managing large databases, AI provides the perfect framework for unbiased and systematic big data analysis. The medical applications of AI are constantly evolving in research clinimetrics while taking over human-based decisions in systems' operations or the diagnosis and treatment of diseases. For example, electronic health record–derived databases processed through machine learning algorithms can facilitate patient care allocation and novel biomarkers research or support genomics to discover disease associations or contain microbial outbreaks (1–3).

In this issue, Ahn et al. introduce the AI-cirrhosis-electrocardiogram (ECG) (ACE) score, developed to identify electrocardiogram (ECG) signaling related to patients with cirrhosis. The authors trained a convoluted neural network (CNN) model with ECGs from 5,212 liver transplant candidates with decompensated cirrhosis and age-matched/sex-matched controls without liver disease (1:4 ratio). The primary outcome was the ability of the CNN to identify whether the ECG corresponded to a patient with cirrhosis and if the score changed with the severity of liver disease. Secondarily, authors examined ACE score changes after liver transplantation. The full model using standard 10-second, 12-lead ECGs showed excellent accuracy with an area under the curve of 0.908 for the identification of cirrhosis; results were unchanged when using only the first 2 seconds from the ECG. An ACE score of $\geq$0.17 showed a sensitivity of 85% and specificity of 83% for the detection of cirrhosis (higher scores in decompensated vs compensated disease). Interestingly, the ACE and model for end stage liver disease with sodium correction (MELD-Na) scores increased in parallel up to a MELD-Na of 20 (ACE score as high as 0.713 with MELD-Na 16–20), and in the 574 patients with post-transplant ECGs, there was a significant drop in the ACE score (from 0.762 to 0.183). A sensitivity analysis expanding case-control matching to comorbidities (e.g., cardiovascular disease) did not alter results.

Although the clinical relevance of these findings is still unclear, such investigations open the door to alternative methods of identifying cirrhosis, which could potentially be of particular value to patients with nonalcoholic steatohepatitis where the noninvasive assessment of cirrhosis carries some challenges or in transplant candidates given the risk of heart disease associated with liver transplantation. The method used, CNN, is a deep learning-based method more frequently featured in recent medical literature. Although often perceived as a synonym of machine learning, deep learning carries some nuances that ought to be clarified. For example, unlike traditional machine learning algorithms that require researchers to decide which features to use, deep learning methods typically determine their own features that provide the best discrimination among multiple object classes (4,5). The main caveat of these methods, however, relies on the participant selection process and the inherent biases introduced by the investigators when "training the machine."

[1]Center for Liver Diseases and Starzl Transplantation Institute, University of Pittsburgh and UPMC; [2]Swanson School of Engineering, University of Pittsburgh.
Correspondence: Andres Duarte-Rojo, MD, PhD. E-mail: duarterojoa@upmc.edu.
Received November 30, 2021; accepted December 27, 2021; published online January 13, 2022

Alvan Feinstein warned us of the inherent risk of meta-analyses, where the quality of the studies could not be substituted by quantity ("garbage in, garbage out") and where results could not be generalized as a consequence of the aggregation of larger and more varied populations (6). The black box nature of AI is, in principle, no different to this. Thus, it is not possible to know which components of ECG were incorporated into the ACE score. Patients with advanced liver disease can present with ECG abnormalities, such as prolonged QT intervals, QRS/T wave abnormalities, reduced heart rate variability, chronotropic incompetence, persistent tachycardia, or bradycardia in relation to beta-blockade (7–9). However, it is unclear which of the above ECG characteristics, if any, the CNN used to distinguish patients with cirrhosis apart from controls. As such, the AI black box lacks the Aristotelian intuitive reasoning (i.e., factors governing pathophysiologic principles) physicians use to build their knowledge on, which helps better associate ill conditions and individualize therapeutics. Importantly, there was no difference in ACE scores among patients with the presence or absence of beta-blocker use, so at least we know the ACE model is applicable to patients with or without exposure to such class of medications.

Cross-pollination research with biomedical informaticians, computer engineers, and scientists is a new trend in gastroenterology/hepatology academia. Like in any other field of research, we need to remain accountable for the quality of the research through proper training of AI models by providing an unbiased real-world experience to algorithms. This way, the end product would be applicable and beneficial to patients. Our group developed a self-attention model to automatically quantify sarcopenia from routine computed tomography scans and is developing an algorithm to monitor adherence to home-based physical rehabilitation with personal activity trackers (unpublished data). We hope that such initiatives will find their niche in practice to further advance the care of patients with liver disease. Similarly, for the ACE score to become an added value in our armamentarium, it needs to find its sweet spot where the clinical utility will be unparalleled. In a few years, we could plausibly be prescribing aspirin plus statins to liver transplant recipients not reversing their ACE scores to reduce their risk for major adverse cardiovascular events. However, while we enjoy the road of AI research, we need to guarantee that good scientific intuition does not decline.

CONFLICTS OF INTEREST

Guarantor of the article: Andres Duarte-Rojo, MD, PhD.
Specific author contributions: A.D.-R.: conceptualization, drafting of the article and critical revisions. E.S.: drafting of the article and critical revisions.
Financial support: None to report.
Potential competing interests: A.D.-R. receives research grant support and has served as a consultant for Axcella Health. E.S. declares no potential conflicts of interest.

REFERENCES

1. Yang NI, Yeh CH, Tsai TH, et al. Artificial intelligence-assisted identification of genetic factors predisposing high-risk individuals to asymptomatic heart failure. Cells 2021:10:2430.
2. Sundermann AJ, Chen J, Kumar P, et al. Whole genome sequencing surveillance and machine learning of the electronic health record for enhanced healthcare outbreak detection. Clin Infect Dis 2021;7:e638–42.
3. Ge X, Raghu VK, Chrysanthis PK, et al. CausalMGM: An interactive web-based causal discovery tool. Nucleic Acids Res 2020;48:W597–W602.
4. Wainberg M, Merico D, Delong A, et al. Deep learning in biomedicine. Nat Biotechnol 2018;36:829–38.
5. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interf 2018;15: 20170387.
6. Feinstein AR. Meta-analysis: Statistical alchemy for the 21st century. J Clin Epidemiol 1995;48:71–9.
7. Jansen C, Chatterjee DA, Thomsen KL, et al. Significant reduction in heart rate variability is a feature of acute decompensation of cirrhosis and predicts 90-day mortality. Aliment Pharmacol Ther 2019;50:568–79.
8. Josefsson A, Fu M, Bjornsson E, et al. Prevalence of pre-transplant electrocardiographic abnormalities and post-transplant cardiac events in patients with liver cirrhosis. BMC Gastroenterol 2014;14:65.
9. Izzy M, VanWagner LB, Lin G, et al. Redefining cirrhotic cardiomyopathy for the modern era. Hepatology 2020;71:334–45.