Separating the Wheat from the Chaff

Zvia Agur PhD

Institute for Medical BioMathematics (IMBM), Bene Ataroth, Israel

KEY WORDS: coronavirus disease 2019 (COVID-19), data science, machine learning, prediction model, prognostic model

IMAJ 2022; 24: 705–707

The coronavirus disease 2019 (COVID-19) pandemic has had a profound impact on our world and has cost millions their lives. It has disrupted economies and education systems and has taken away means of support from masses of people around the world. No wonder this pandemic is like a black hole, drawing in all resources and all expertise. In the scientific arena, the pandemic has created a tremendous opportunity for new and exciting synergies between different disciplines. One of the most prominent synergies in the fight against the COVID-19 pandemic uses machine learning to diagnose and prognosticate the disease.

Machine learning is responsible for some of the most sensational technological advancements in modern times, including self-driving vehicles, for example, or the discovery of hundreds of exoplanets–planets that orbit stars other than the sun. Machine learning algorithms automatically build a computational model that uses sample data, also known as *training data*, to make decisions without being explicitly programmed to make those decisions. This property renders machine learning especially attractive when medicine faces a global outbreak of a fast-spreading new disease caused by an unfamiliar virus that threatens to inflict damage of biblical dimensions. The enormous gap between the almost non-existent knowledge about the disease, on the one hand, and the urgency in finding efficient solutions to it, on the other, underscores the potential value of a method enabling the prediction of processes, such as personal disease progression, with no prior knowledge on the driving forces underlying these processes. Indeed, since the disease outbreak, machine learning has been the backbone of thousands of publications suggesting models for the diagnosis or prognosis of people with COVID-19.

Machine learning traces its roots to the 1950s when Arthur Samuel of IBM developed a computer program for playing checkers. He coined the term *Machine Learning* for mechanisms he designed, which allowed his program to improve on its own [1]. But machine learning remained a niche area for decades, taking off only in the 21st century when increasing computing power and gigantic amounts of data converged to finally take full advantage of machine learning algorithms, which require massive data and fast processing speed to be useful. Yet, until recently, the contribution of this field to healthcare was limited. The COVID-19 pandemic has changed this, providing the impetus for the increasing willingness of physicians to join forces with data scientists in the quest for solutions for the long list of unknowns of the current crisis.

The downside of this exciting development is the need to implement the new synergy straightaway, whereas fruitful collaboration depends on thorough interdisciplinary understanding, which demands time and effort: the data scientists should understand the crucial needs of the physicians, and their practical limitations, while the physicians should be able to evaluate the quality and the feasibility of applying the proposed machine learning tools. Unfortunately, most of the machine learning-based prediction models for COVID-19 published thus far, are fraught with faults in both the methodology itself, the suitability of the data used for model development, the validation of model accuracy, and the applicability to the clinic [2-5].

Take, for example, the work by Yadaw and colleagues [6] from the Icahn School of Medicine at Mount Sinai, New York, USA. Yadaw and colleagues presented machine learning models predicting mortality during medical encounters of unspecified duration in patients with COVID-19, who had been admitted to the Mount Sinai Health System in the New York City area. The researchers highlighted a model they developed, which was based on three features: patient's age, minimum oxygen saturation throughout their medical encounter, and type of patient encounter (inpatient, outpatient, or telehealth visits). They used a relatively large patient dataset for model development (n=3841). The number of patients who died (n=313) seems appropriate for the statistical analysis [7], and high accuracy is achieved in model validation (area under the curve [AUC] of 0.91). The authors suggested using this model in clinical settings to guide the management and prognostication of patients affected by the COVID-19 disease.

However, the experienced reader may not be convinced by the proposition of Yadaw and colleagues. In their article [6], the authors mentioned some of the caveats hampering the clinical use of

the model, notably, insufficient external validation of its accuracy. But the unmentioned methodological problems in the work seem to be insurmountable. Essentially, the highlighted model predicts death using measurements collected throughout the entire encounter of the patient with the health system, with no specific moment at which the prediction is generated and tested.

This situation raises questions about the actual prognostic value of the only time-varying model parameter: the minimum oxygen saturation, and about when and how the model should be used. As the predictive value of time-varying clinical parameters tends to increase when measured closer to the outcome, in this case death of the patient, it remains unclear how to interpret the reported performance measurement (i.e., AUC of 91%) vis-à-vis the time of measurement of this time-varying predictor) [3]. The mere definition of the minimum saturation as the lowest value of oxygen saturation over the entire encounter [6] implies that the prediction itself becomes immaterial at the time it is created when the patient is already dying or discharged. Furthermore, patients who did not die by the end of the study were considered as remaining alive. But since the death of these patients might have occurred after the study ended, the actual incidence of mortality could be underestimated, putting in doubt the value of the minimum oxygen saturation as a sole time-varying predictor [3]. A possible solution for such a conundrum may be to fix a short-term prediction scope, such as, "predict death in the coming 96 hours." But from the applicability point of view, the fixed follow-up window should be carefully determined to allow sufficient time for efficacious relief of the predicted fatal outcome (e.g., by corticosteroids [8]).

The result of Yadaw and colleagues [6] that the minimum oxygen saturation is responsible for the high predictive capacity of the model is striking also from another point of view. Even though the leading cause of death of critically ill patients with COVID-19 is a refractory respiratory failure (45%), more than half of the deceased patients died from other causes, such as cardiac arrest or hemorrhagic events [9]. Therefore, it is not clear how minimum oxygen saturation represents almost all the potentially deceased patients in [6]. Overestimation of the model accuracy is conceivable in this case, due to potential correlations between the consecutive measurements over time in the same patients.

The analysis of the work of Yadaw and colleagues [7] describes how some of the prerequisites for prediction models could become more helpful in the clinic. Better collaboration is necessary among researchers from different backgrounds, specialties, and institutes for determining the clinical need and for sharing patient data from COVID-19 studies and registries. Another issue is the requirement for external model validation, currently complicated by the incompatibility of recording systems in different hospitals. Consensual representation of the patient's follow-up and treatments is required to allow for external validation of prediction models and their subsequent generalization. Most important, in this context, is the necessity to adhere to unified sets of criteria for evaluating prediction models, such as the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) set of recommendations [10], or the Prediction Model Risk Of Bias Assessment Tool criteria (PROBAST), which enable accurate evaluation of the risk of bias and applicability of a prediction model [2]. Another important way to refine the prognostic model landscape is by a critical analysis of the diverse modeling efforts and by recommendations for their improvement.

At present, there is an urgent need to separate the wheat from the chaff and underline those predictive models that can become useful in the clinic. But how can one do this? The pandemic has created huge amounts of information, which the traditional method of academic reporting cannot encompass. As a result, atlases and catalogs, covering extensive disease-related data, acquire a special status these days. An example is the multi-omics blood atlas of immune profiles of patients with varying COVID-19 severity, back-to-back with the immune profiles of patients with influenza or sepsis, and healthy volunteers. This massive dataset, by more than 200 scientists from many research centers could aid future drug developers and designers of precision medicine modalities [11]. Another example is the COVID-19-related mortality dataset by Karlinsky and Kobak [12], which the authors used to compute the excess mortality in each country during the COVID-19 pandemic and identify the countries that have been substantially underreporting their COVID-19 deaths.

The review article by Shapiro and colleagues [13], from Tel Aviv Sourasky Medical Center, Israel, joins this new class of publications. They worked to separate the wheat from the chaff in the multitude of prognostic models for COVID-19 by cataloging and scrutinizing the major models for classifying patients at risk of deterioration. The authors discussed the tools at our disposal for critical model assessment and evaluated the clinical adequacy of the analyzed models. First, Shapiro and colleagues discussed scoring systems, both established scores and scores designed specifically for COVID-19 patients. Next, they listed and analyzed models that use machine learning to predict risk in COVID-19 patients. Shapiro and colleagues provided a comprehensive table of models and their main attributes, and added their point of view on the highlights and difficulties in each of the models. Their research can serve as a concise navigation map in the turbulent water of machine learning risk predictors for COVID-19.

CONCLUSIONS

Ultimately, prediction models should be objectively tested in prospective clin-

ical trials and evaluated to understand how they improve the clinical outcomes. These models may be used to better triage patients to an appropriate level of care, streamline resource allocation, improve care in times of hospital overload, and optimize the timing of disease-modifying treatment. Well-validated prediction models can empower care teams and healthcare administrators to make the right decisions under stress.

Correspondence

Dr. Z. Agur
Institute for Medical BioMathematics (IMBM), Bene Ataroth *6099100*, Israel
Phone: (972-3) 973-3075/4068
Fax: (972-3) 973-3410
email: agur@imbm.org

References

1. Foote KD. A brief history of machine learning: Dataversity. [Available from https: //www.dataversity.net/a-brief-history-of-machine-learning/].
2. Miller JL, Tada M, Goto M, et al. Prediction models for severe manifestations and mortality due to COVID-19: a systematic review. *Acad Emerg Med* 2022; 29 (2): 206-16.
3. Leeuwenberg AM, Schuit E. Prediction models for COVID-19 clinical decision making. *Lancet Digit Health* 2020; 2 (10): 496-7.
4. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020; 369: 1328.
5. Bachtiger P, Peters NS, Walsh SL. Machine learning for COVID-19-asking the right questions. *Lancet Digit Health* 2020; 2 (8): 391-2.
6. Yadaw AS, Li YC, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *Lancet Digit Health* 2020; 2 (10): 516-25.
7. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; 368: 441.
8. Bartoletti M, Marconi L, Scudeller L, et al. Efficacy of corticosteroid treatment for hospitalized patients with severe COVID-19: a multicentre study. *Clin Microbiol Infect* 2021; 27 (1): 105-11.
9. Contou D, Cally R, Sarfati F, Desaint P, Fraisse M, Plantefeve G. Causes and timing of death in critically ill COVID-19 patients. *Crit Care* 2021; 25 (1): 79.
10. Goodacre S, Thomas B, Sutton L, et al. Derivation and validation of a clinical severity score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort study. *PLoS One* 2021; 16 (1): e0245840.
11. Consortium C-M-oBA. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell* 2022; 185 (5): 916-38 e58.
12. Karlinsky A, Kobak D. The World Mortality Dataset: tracking excess mortality across countries during the COVID-19 pandemic. *medRxiv* 2021; (10): e69336.
13. Shapiro M, Yavne Y, Shepshelovich D. Predicting which patients are at risk for clinical deterioration in COVID-19: a review of the current models in use. *IMAJ* 2022; 24 (11): 699-704.

Capsule

STING-induced regulatory B cells compromise NK function in cancer immunity

An immunosuppressive tumor microenvironment is a major obstacle in the control of pancreatic and other solid cancers. Agonists of the stimulator of interferon genes (STING) protein trigger inflammatory innate immune responses to potentially overcome tumor immunosuppression. Although these agonists hold promise as potential cancer therapies, tumor resistance to STING monotherapy has emerged in clinical trials and the mechanism(s) is unclear. **Li** et al. showed that the administration of five distinct STING agonists, including cGAMP, results in an expansion of human and mouse interleukin (IL)-35+ regulatory B cells in pancreatic cancer. Mechanistically, cGAMP drives expression of IL-35 by B cells in an IRF3-dependent but type I interferon-independent manner. In several preclinical cancer models, the loss of STING signaling in B cells increases tumor control. Furthermore, anti-IL-35 blockade or genetic ablation of IL-35 in B cells also reduces tumor growth. Unexpectedly, the STING–IL-35 axis in B cells reduces proliferation of natural killer (NK) cells and attenuates the NK-driven anti-tumor response. These findings reveal an intrinsic barrier to systemic STING agonist monotherapy and provide a combinatorial strategy to overcome immunosuppression in tumors.

Nature 2022; 610: 373
Eitan Israeli

Capsule

Endosomal changes tied to disease

Frontotemporal dementia and amyotrophic lateral sclerosis share key genetics and pathology, but the connection between different known facets of their disease biology is not always clear. **Shao** et al. discovered an interplay between the disease-associated genes *C9orf72* and *TBK1*. Large repeats of glycine-alanine, which are produced by an expansion in *C9orf72*, sequestered TBK1 into inclusions, inhibiting its function and impairing the downstream endosomal pathway. A mutation in *TBK1* worsened these defects, enhancing disease phenotypes in mice. Remarkably, the disruption of the endosomal pathway also proved sufficient to induce the aggregation of TAR-DNA binding protein 43 (TDP-43), a key driver of degeneration in these diseases.

Science 2022; 378: 94
Eitan Israeli