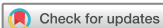inducible factor-2α antagonist in patients with previously treated advanced clear cell renal cell carcinoma. *J Clin Oncol* 2018;36: 867–874.

15. Cheng X, Prange-Barczynska M, Fielding JW, Zhang M, Burrell AL, Lima JDCC, *et al*. Marked and rapid effects of pharmacological HIF-2α antagonism on hypoxic ventilatory control. *J Clin Invest* 2020; 130:2237–2251.

Check for updates

Machine Learning Classifier Models: The Future for Acute Respiratory Distress Syndrome Phenotyping?

Acute respiratory distress syndrome (ARDS) is a heterogeneous clinical syndrome characterized by severe respiratory failure requiring mechanical ventilatory support for which there is no therapy and in which mortality remains approximately 40% (1, 2). A recent important advance has been the subclassification of ARDS into subphenotypes that have potential prognostic and/or therapeutic significance. Calfee and colleagues reported that an approach called latent class analysis (LCA) could identify one-third of patients with ARDS with a "hyperinflammatory" phenotype, with the remainder having a "hypoinflammatory" pattern (3). The hyperinflammatory phenotype had higher levels of proinflammatory biomarkers and poorer outcomes, including fewer ventilator-free and organ failure–free days and higher mortality. Biological plausibility to these subphenotypes has been suggested by differential responses to treatment (3, 4). A recent LCA of a large negative randomized controlled trial of simvastatin in ARDS demonstrated potential benefit in the hyperinflammatory group (5). Other approaches to ARDS phenotyping exist, including a "physiologic phenotyping" approach (based on the focal versus diffuse distribution of lung infiltrates) (6) and transcriptomics-based approaches (7).

Classifying patients into hyperinflammatory and hypoinflammatory subgroups in real time would enable rapid implementation of a precision medicine approach. Key impediments to the real-time use of these approaches are that nonstandard assays for Il-6 and sTNFR1 are required and that LCAs are computationally complex.

Computational modeling approaches are increasingly being applied in medicine. Machine learning (ML) is when a computer algorithm is trained with known input and output values to predict outcomes from novel data with similar input characteristics. In this issue of the *Journal*, Sinha and colleagues (pp. 996–1004) present data demonstrating the potential of an ML approach using readily available clinical data to identify ARDS subphenotypes (8).

ML Classification and ARDS

Sinha and colleagues examined whether an ML approach, specifically a variant of the gradient-boosting machine (GBM) algorithm (9), could accurately identify inflammatory subtypes of ARDS (8). They used standard clinical and laboratory parameters to categorize patients from three prior clinical trials into inflammatory subtypes. This data were divided into subsets using 10-fold cross-validation to train and optimize the settings (hyperparameters) of the ML algorithm with the objective of building a model for the task of categorizing entries into the hypoinflammatory and hyperinflammatory subphenotypes. The performance of the optimized GBM classifier model was evaluated on a fourth separate dataset by comparing the classes it assigned to the "gold-standard" LCA classes. For the hypoinflammatory class, the GBM model (with a probability threshold of 0.5) gave the correct answer (assuming LCA is correct) in 98% of cases (460 of 468), but for the hyperinflammatory class, it was only correct in 63% of cases (175 of 277). The combined accuracy for both classes was 85%.

The probability threshold at which cases are assigned to each subphenotype can be varied, which, in essence, moves borderline cases into the hypoinflammatory or the hyperinflammatory class. For example, moving to a threshold of 0.3, the GBM model was correct in 93% of cases for the hypoinflammatory class and 78% of cases for the hyperinflammatory class, with an overall accuracy of 87%. This changing threshold depended on the etiology of ARDS, particularly sepsis-related ARDS, indicating that clinical judgment may still be required in the implementation and interpretation of these algorithms. Although the authors highlight the performance of their method as quantified using the area under the curve (AUC), which provides a valid and useful measurement of the performance of a probabilistic classifier, classification accuracy is more amenable to direct clinical interpretation.

The authors showed that these subphenotypes could be predicted using in-hospital data generated up to admission to an ICU with ARDS. They also demonstrated that inflammatory biomarkers and 90-day mortality were higher in the hyperinflammatory phenotype as classified by the model. Findings for treatment interactions for statins, fluid management, and positive end-expiratory pressure strategy in ARDS for the ML-derived classes were consistent with those seen previously for LCA-derived classes. An alternative GBM model built only with readily available laboratory and vital sign variables achieved a comparable performance (AUC = 0.94) with one using all variables (AUC = 0.95) and was able to identify similar treatment interactions. We note that Calfee and colleagues took care to keep training data separate from their evaluation data by holding out a cohort as the evaluation group.

When selecting parameters to improve model performance, they appropriately performed cross-validation within the training data.

There are some inherent limitations to the approach, as the authors acknowledge. Most significantly, the databases used in this study came from NHBLI ARDS Network clinical trials, so the results might not generalize to real life data extracted from electronic medical records or observational cohorts.

ML: The Future of Personalized Medicine?

The potential for ML to advance critical care medicine is increasingly recognized, such as in the prediction of ARDS development (8, 10, 11). It will be important to the potential and the limitations of ML algorithms, to recognize that because they are based on extracting rules and correlations that are implicit in data, their performance is limited by the characteristics of that data. The application of any ML approach is hampered by the lack of a gold-standard method to either diagnose or classify ARDS (8, 11, 12). It is also important to continue to improve our knowledge of the underlying biology and pathophysiologic mechanics, as this will further inform approaches to develop personalized medicines for patients with ARDS.

As we adopt more ML approaches to data interpretation, we should ask whether ML should be the new standard approach to data interpretation. Although traditional statistical approaches, such as multivariate logistic regression, have been widely adopted, ML approaches benefit from being able to discover more complex multifaceted nonlinear relationships between patient data and disease states. For example, ML has been demonstrated to outperform multivariate logistic regression analysis in predicting the risk of lower gastrointestinal rebleeds (13). However, although nonlinear ML models have greater explanatory power than linear statistical methods, the same nonlinearity properties can result in very poor predictions if models are used in situations that extrapolate far beyond the scope of the training data and can also give rise to "black box" models that do not support clinical insight. These challenges can be addressed through the careful choice of appropriate ML algorithms, the conscientious and rigorous evaluation of models, and the application of techniques from the emerging field of explainable artificial intelligence.

Sinha and colleagues are to be congratulated for advancing the field of ML in medicine and advancing the prospect of personalized medicines for ARDS by demonstrating a compelling application of ML in identifying ARDS subphenotypes from readily available clinical data with the potential for near-bedside deployment. ∎

Bairbre McNicholas, B.Sc., M.B. B.Ch., M.R.C.P.I., Ph.D., F.J.F.I.C.M.I.
Department of Anaesthesia and Intensive Care Medicine
Galway University Hospitals
Galway, Ireland

Michael G. Madden, Ph.D.
College of Science and Engineering
National University of Ireland
Galway, Ireland

John G. Laffey, B.Sc., M.B. B.Ch. B.A.O., M.D., M.A., F.F.A.R.C.S.I.
Department of Anaesthesia and Intensive Care Medicine
Galway University Hospitals
Galway, Ireland
and
School of Medicine
and
Regenerative Medicine Institute at the CURAM Centre for Medical Devices
National University of Ireland
Galway, Ireland

ORCID IDs: 0000-0001-9524-4021 (B.M.); 0000-0002-1246-9573 (J.G.L.).

References

1. McNicholas BA, Rooney GM, Laffey JG. Lessons to learn from epidemiologic studies in ARDS. *Curr Opin Crit Care* 2018;24:41–48.
2. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, *et al*.; LUNG SAFE Investigators; ESICM Trials Group. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA* 2016;315:788–800.
3. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA; NHLBI ARDS Network. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014;2:611–620.
4. Famous KR, Delucchi K, Ware LB, Kangelaris KN, Liu KD, Thompson BT, *et al*.; ARDS Network. Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. *Am J Respir Crit Care Med* 2017;195:331–338.
5. Calfee CS, Delucchi KL, Sinha P, Matthay MA, Hackett J, Shankar-Hari M, *et al*.; Irish Critical Care Trials Group. Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. *Lancet Respir Med* 2018;6:691–698.
6. Constantin JM, Jabaudon M, Lefrant JY, Jaber S, Quenot JP, Langeron O, *et al*.; AZUREA Network. Personalised mechanical ventilation tailored to lung morphology versus low positive end-expiratory pressure for patients with acute respiratory distress syndrome in France (the LIVE study): a multicentre, single-blind, randomised controlled trial. *Lancet Respir Med* 2019;7:870–880.
7. Bos LD, Schouten LR, van Vught LA, Wiewel MA, Ong DSY, Cremer O, *et al*.; MARS consortium. Identification and validation of distinct biological phenotypes in patients with acute respiratory distress syndrome by cluster analysis. *Thorax* 2017;72:876–883.
8. Sinha P, Churpek MM, Calfee CS. Machine learning classifier models can identify acute respiratory distress syndrome phenotypes using readily available clinical data. *Am J Respir Crit Care Med* 2020;202:996–1004.
9. Chen T, Guestrin C. (2016) XGBoost: a scalable tree boosting system. Presented at the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Proceedings. August 13–17, 2016, pp. 785–794.
10. Ding XF, Li JB, Liang HY, Wang ZY, Jiao TT, Liu Z, *et al*. Predictive model for acute respiratory distress syndrome events in ICU patients in China using machine learning algorithms: a secondary analysis of a cohort study. *J Transl Med* 2019;17:326.
11. Yang P, Wu T, Yu M, Chen F, Wang C, Yuan J, *et al*. A new method for identifying the acute respiratory distress syndrome disease based on noninvasive physiological parameters. *PLoS One* 2020;15:e0226962.
12. Reamaroon N, Sjoding MW, Lin K, Iwashyna TJ, Najarian K. Accounting for label uncertainty in machine learning for detection of acute respiratory distress syndrome. *IEEE J Biomed Health Inform* 2019;23:407–415.
13. Ayaru L, Ypsilantis PP, Nanapragasam A, Choi RC, Thillanathan A, Min-Ho L, *et al*. Prediction of outcome in acute lower gastrointestinal bleeding using gradient boosting. *PLoS One* 2015;10:e0132485.