

COMMENT



Machine learning: remember the fundamentals

Kristyn S. Beam^{1,2} and John A. F. Zupancic^{1,2}✉

© The Author(s), under exclusive licence to the International Pediatric Research Foundation, Inc 2022

Pediatric Research (2023) 93:291–292; <https://doi.org/10.1038/s41390-022-02420-1>

Machine learning has become increasingly incorporated into our everyday lives. In medicine, technological strides in recent years have allowed these techniques to predict various aspects of care, including diagnoses and prognoses, through sophisticated analysis of data, more recently allowing the incorporation even of images into algorithms. In the field of neonatology, a variety of machine learning applications have been developed, including examples for illness severity,¹ retinopathy of prematurity,² sepsis,³ and neurodevelopmental outcomes.^{4,5}

In this issue of *Pediatric Research*, Baker and Kandasamy present a systematic review examining studies that use machine learning to predict neurodevelopmental outcomes in preterm infants.⁶ Machine learning is a field of artificial intelligence that utilizes computer algorithms to generate predictive models automatically from large datasets, without being explicitly programmed to a specific task. Baker and Kandasamy searched for studies published between 2010 and 2022 and identified 11 publications that met their eligibility criteria of using a machine learning method to examine or predict neurodevelopmental outcomes. Their review documents a high degree of variability in the data inputs and outputs of the studies, and notes that studies remained ambiguous about which features were most predictive of neurodevelopmental outcomes. They conclude that the variability and ambiguity are mostly due to a lack of data standardization, differences in defining the outcome of interest, and variation in the machine learning methods used. These issues are important considerations for how machine learning can be applied to various problems in the field of neonatology.

In this commentary, we discuss how the goals of machine learning models determine the type of model used and how the definition of outcomes can also affect our interpretation of models.

THE GOAL: PREDICTION VERSUS DESCRIPTION

Baker and Kandasamy describe various machine learning methods used either to describe or infer associations among factors related to neurodevelopmental outcomes or to predict a probability of neurodevelopmental outcomes. Inference models, or those describing features that are associated with an outcome, are common in medicine and comprise the majority of historical models, including the familiar linear and logistic regression models. Their frequent use is mostly a byproduct of the types of data that have been available in medicine and of long-standing computational limitations. Data that are collected from retro-

spective chart reviews, through secondary analyses of randomized controlled trial data, or through manual identification of fields in electronic health records, require a more parsimonious approach and lend themselves to inference, but omit large amounts of information that may be helpful for prediction. The benefit of this type of analysis is that we can usually understand on a basic and intuitive level how specific pieces of information are biologically related to the outcome of interest. Often it is explained in a way that makes sense as far as how we think about clinical practice.

Recently, a significant turning point has been reached through vastly more complex technological modeling capabilities that can incorporate larger amounts of data, broader types of data (for example, those derived directly from medical images), and more flexible organization of data. Such approaches allow more precise and accurate prediction models, and include random forests, classification models, and convolutional neural networks, which enable image analysis by comparing neighboring pixels to predict the next pixel. Although they have higher predictive utility, a challenge to utilizing such machine learning models is that we cannot always explain the mechanism through which these clinical predictors may be related to the predicted probability of the outcome; this is the “black box” analysis that is typically referenced in artificial intelligence methods. The result may be discomfort among clinicians due to their “inexplicability”; however, there are differing opinions among the larger field of machine learning as to how explainable models should be.⁷

It should be noted that categorizing analytic approaches reductively as either descriptive or predictive may lead to missing some of the more nuanced aspects of the machine learning model used. For example, the authors note that several studies employed a method called “backtracking” and “partial derivatives” to describe associations between clinical features and neurodevelopmental outcomes; however, these analyses in fact use prediction-based models including neural networks, random forests, and support vector machines to predict the probability of developing specific neurodevelopmental outcomes.

THE OUTCOME: BINARY VS CONTINUOUS CLASSIFICATION

A second important issue highlighted in Baker and Kandasamy relates to how binary versus continuous expression of outcomes can change how we think about the predicted probabilities produced by machine learning models. In dichotomizing values within scoring systems, the cutoff at which to draw a binary classification can be

¹Department of Neonatology, Beth Israel Deaconess Medical Center, Boston, MA, USA. ²Department of Pediatrics, Harvard Medical School, Boston, MA, USA.

✉email: jzupanci@bidmc.harvard.edu

Received: 22 November 2022 Accepted: 30 November 2022

Published online: 22 December 2022

arbitrary and might shift patients into a “low-risk” or “high-risk” group on the basis of even one point, which may not be clinically significant. Several papers included in the Baker and Kandasamy review categorize infants as low or high risk of atypical neurodevelopment at 18–24 months, which might similarly reflect clinically insignificant differences in the underlying Bayley scores. Understanding how outcomes are expressed can thus ultimately change how we view the predictive value of a model.

Machine learning models can make a prediction on a continuous scale much easier by predicting more granular outcomes.⁸ An important example of this distinction in neonatology is bronchopulmonary dysplasia (BPD). There are two options when attempting to predict pulmonary outcome early in an infant’s life. The first is that we can predict whether the infant will or will not have BPD. This is undoubtedly clinically meaningful since not having the diagnosis is associated—in descriptive analyses—with lower morbidity, mortality, and resource utilization. However, the diagnosis of BPD includes a wide-ranging group of phenotypes that spans minimal low-flow oxygen via nasal cannula at 36 weeks postmenstrual age to tracheostomy throughout infancy, with associated increased mortality risk. Therefore, although our habitual approach of predicting “BPD” vs “no BPD” might have some clinical relevance, the use of more advanced machine learning techniques might allow the prediction of the particular level of respiratory support at a particular time. The more precise information available through use of continuous outcomes will potentially allow for testing and adoption of more customized interventions.⁹

CONCLUSIONS

Machine learning models are emerging in medicine and in neonatology. The availability of richer data sources and powerful computational platforms provides clinicians and researchers the ability to think about new ways to predict outcomes and new questions to ask of the data. By learning from large granular datasets of neonatal data, machine learning can effect a paradigm shift toward more precise and accurate outcome prediction. However, just as with traditional statistical techniques, users of machine learning approaches must consider not only the technical aspects of their work, but also fundamental issues such as modeling goals and outcome specification.

REFERENCES

1. Saria, S., Rajani, A. K., Gould, J., Koller, D., & Penn, A. A. Integration of early physiological responses predicts later illness severity in preterm infants. *Sci. Transl. Med.* **2**, 48ra65 (2010).
2. Brown, J. M. et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol.* **136**, 803 (2018).
3. Song, W. et al. A predictive model based on machine learning for the early detection of late-onset neonatal sepsis: development and observational study. *JMIR Med. Inform.* **8**, e15965 (2020).
4. He, L. et al. A multi-task, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Sci. Rep.* **10**, 15072 (2020).
5. Vassar, R. et al. Neonatal brain microstructure and machine-learning-based prediction of early language development in children born very preterm. *Pediatr. Neurol.* **108**, 86–92 (2020).
6. Baker, S. & Kandasamy, Y. Machine learning for understanding and predicting neurodevelopmental outcomes in premature infants: a systematic review. *Pediatr. Res. Current Issue* (2022).
7. Ghassemi, M., Oakden-Rayner, L. & Beam, A. L. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* **3**, e745–e750 (2021).
8. Halabi, S. S. et al. The RSNA pediatric bone age machine learning challenge. *Radiology* **290**, 498–503 (2019).
9. Royston, P., Altman, D. G. & Sauerbrei, W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat. Med.* **25**, 127–141 (2006).

AUTHOR CONTRIBUTIONS

Dr. Beam and Dr. Zupancic both conceived of the presented idea, contributed to the writing of the manuscript, and approved the final version.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to John A. F. Zupancic.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.