

## EDITORIAL



# Utilizing machine learning in predictive modeling: what's next?

© The Author(s), under exclusive licence to Springer Nature Limited 2022

*Bone Marrow Transplantation* (2022) 57:699–700; <https://doi.org/10.1038/s41409-022-01622-9>

Machine learning (ML), a branch of artificial intelligence (AI), is increasingly being used to create predictive and prognostic healthcare models. Integrating such technologies in medicine might bring exciting changes in improving our risk stratification abilities and, therefore, help in both therapeutic and preventative measures. We have previously shown that the number of AI-related abstracts submitted to major hematology and bone marrow/stem cell transplant conferences between 2010 and 2017 increased around eight times [1]. Despite the increase in the number of articles, integrating ML/AI tools in medical practice continues to be slow due to significant challenges, including the need for high-quality big data.

Several examples in the literature have utilized ML in predicting outcomes, for instance, Nazha et al. [2] created a personalized model to risk-stratify patients with myelodysplastic syndromes. The model was built based on data from 1471 patients incorporating both genetic and clinical data. The algorithm was later validated using an external cohort (i.e., data were not used in building the model). The model performance was better than the current used MDS prognostic tools illustrating the possible impact of ML in optimizing predictive models' performance.

In this issue, Lee et al. [3] utilized supervised ML to create a prediction model for the risk of developing hepatic veno-occlusive disease/sinusoidal obstruction syndrome (VOD/SOS) and early death after transplant. VOD/SOS is a life-threatening disease after allogeneic hematopoietic cell transplantation (allo-HCT), with a high mortality rate, particularly if it is severe [4]. Thus, predicting the occurrence of VOD/SOS and factors contributing to it can provide insight into preventative measures (particularly if adjustable pre-transplant factors) and improve outcomes.

The authors included data from more than 2500 allo-HCT recipients with 20 selected features (14 immutable features and 6 adjustable features). The incidence of VOD/SOS was 3.4% (87 patients in 2572), and 49 patients (1.9%) developed clinically fatal severe to very severe VOD/SOS. Given the class imbalance, the authors used Synthetic Minority Over-sampling Technique. They created three models for VOD/SOS, severe to very severe VOD/SOS, and early death. They applied multiple algorithms to the data, including Naïve Bayes, Adaboost, logistic regression, Random Forest, and Extreme gradient boosting (XGBoost). XGBoost was the algorithm that achieved the best performance. Subsequently, the algorithm was validated using the k-fold cross-validation method, with an area under the curve (AUC) of 0.750 for all VOD/SOS, 0.778 for severe to very severe VOD/SOS, and 0.738 for early death. The article provided SHapley Additive exPlanations to help explain the contribution of different factors to the model, this showed that the most contributory factors included gender



(male), busulfan dose, age (older), FEV1 and disease risk index in case of VOD/SOS; however, haploidentical donor and history of liver dysfunction were additional contributors to the development of severe VOD/SOS.

This study provides interesting results and tackles a significant post-allo-HCT complication using data from a large number of patients. In 2018, the Center for International Blood & Marrow Transplant Research (CIBMTR) published a tool for risk scores using pre-transplant data to predict VOD/SOS [5]. They used the CIBMTR database with more than 13,000 patients. The predicting risk score model was validated using the random split method (i.e., the sample was randomly divided into two groups; one used for model creation and the other for validation) and had a c-statistics of 0.76 and was created using logistic regression. Factors associated with a higher risk of VOD included age (younger), low-performance status, disease status at transplant, and others. Lee et al. [3] applied the CIBMTR model to their data and found to have an AUC of 0.546, which was underperforming compared to all the other models.

The comparison above illustrates two particularly important points. The first is that adequately performing prediction models can be created using traditional regression models without the essential need for ML strategies, as can be seen by the studies utilizing the original CIBMTR data. Secondly, creating a model even when using a large dataset does not necessarily mean that it can be applied to different datasets (i.e., not generalizable to a different time and/or location). This study, for example, was done in a single institution, which has a specific population of patients and practices, some of which might not be commonly used in other institutions or countries (e.g., VOD/SOS prophylaxis used) and thus external validity of the dataset is limited.

Current ML literature in medicine is limited in several aspects, some of which have been highlighted by Lee et al. [3] ML algorithms remain a “black box,” and more efforts should be made to make it explainable to practitioners. Commonly, ML algorithms suboptimally report accuracy, use less preferred validation methodologies, lack calibration, and oversimplify clinical questions by dichotomizing outcomes [6–8]. In addition, advancing the type of ML research from “proof of concept” retrospective studies to prospective and randomized studies is needed to demonstrate clinical utility and outcomes improvement [9]. Finally, ML tools benefit in solving specific clinical problems, and researchers should use those tools appropriately as standard regression methods can still provide the ability to create predictive models and are preferred in certain situations. This article provided an insight into the possible uses and limitations of ML tools and is a step forward in the expansion of ML use to help improve outcomes. Future research and application should utilize ML predictive power and invest in large datasets in order to translate those applications to clinical practice.

Received: 9 February 2022 Revised: 12 February 2022 Accepted: 14 February 2022  
Published online: 15 March 2022

Ibrahim N. Muhsen <sup>1</sup> and Shahrukh K. Hashmi <sup>2,3,4</sup> ✉

<sup>1</sup>Department of Medicine, Houston Methodist Hospital, Houston, TX, USA. <sup>2</sup>Department of Medicine, Sheikh Shakhbout Medical City, Abu Dhabi, UAE. <sup>3</sup>Medical Affairs, Khalifa University, Abu Dhabi, UAE.

<sup>4</sup>Division of Hematology, Department of Medicine, Mayo Clinic, Rochester, MN, USA. ✉email: hashmi.shahrukh@mayo.edu

## REFERENCES

- Muhsen IN, ElHassan T, Hashmi SK. Artificial intelligence approaches in hematopoietic cell transplantation: a review of the current status and future directions. *Turk J Haematol.* 2018;35:152–7.
- Nazha A, Komrokji R, Meggendorfer M, Jia X, Radakovich N, Shreve J, et al. Personalized prediction model to risk stratify patients with myelodysplastic syndromes. *J Clin Oncol.* 2021;39:3737–46.
- Lee S, Lee E, Park SS, Park MS, Jung J, Min GJ, et al. Prediction and recommendation by machine learning through repetitive internal validation for hepatic veno-occlusive disease/sinusoidal obstruction syndrome and early death after allogeneic hematopoietic cell transplantation. *Bone Marrow Transplant.* 2022:1–9. Online ahead of print
- Coppell JA, Richardson PG, Soiffer R, Martin PL, Kernan NA, Chen A, et al. Hepatic veno-occlusive disease following stem cell transplantation: incidence, clinical course, and outcome. *Biol Blood Marrow Transplant.* 2010;16:157.
- Strouse C, Zhang Y, Zhang M-J, Digilio A, Pasquini M, Horowitz MM, et al. Risk score for the development of veno-occlusive disease after allogeneic hematopoietic cell transplant. *Biol Blood Marrow Transplant.* 2018;24:2072–80. <https://doi.org/10.1016/j.bbmt.2018.06.013>
- Salah HT, Muhsen IN, Salama ME, Owaidah T, Hashmi SK. Machine learning applications in the diagnosis of leukemia: current trends and future directions. *Int J Lab Hematol.* 2019;41:717–25.
- Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016;69:245–7.
- Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med.* 2019;17:230.
- Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17:195.

## AUTHOR CONTRIBUTIONS

INM and SKH wrote the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Shahrukh K. Hashmi.

**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.