Background Immune checkpoint inhibitors (ICI) have improved outcomes in several tumor types allowing subgroups of patients to have longer, higher quality lives. However, potential life-threatening immunotoxicities can arise in susceptible patients. Identifying patients at high risk of immunotoxicities alongside those responding well can help patients understand risk-benefit profile of the treatment, improve clinical trial cohort selection, and inform therapy selection in clinical settings. Herein, we introduce a machine learning (ML) framework that can accurately predict common immunotoxicities – hepatitis, colitis, and pneumonitis – alongside efficacy utilizing routinely collected Electronic Health Records (EHR) data.

Methods Our models rely on real-world EHR data of over 2,200 ICI-treated patients from Vanderbilt University Medical Center obtained prior to December 31, 2018. During the design of the predictive models, we set the prediction time point as the ICI initiation date for each patient. 1-year prediction time window was applied to create binary labels for the four prediction outcomes. Pneumonitis and colitis episodes were manually curated to establish the labels. The hepatitis label was defined to be 1 if any of the four liver enzymes exceeded three times the upper limit of normal. Overall survival served as a surrogate for efficacy. Structured data and clinician notes prior to ICI initiation were utilized to create features for the models. Feature engineering involved aggregating laboratory measurements over 60 and 120-day time windows. 1-year window was applied for other data types including ICD-10 codes, procedures, medication, and smoking history. In model development, patients were randomly partitioned into training (80%) and test (20%) sets for each outcome. An experiment involved a baseline and an alternative model, where the latter was selected if it demonstrated statistically significant superior performance based on outer loop results from a nested cross-validation process on the training set.

Results A random forest classifier was developed for each outcome. (Table 1) demonstrates performance results with 95% bootstrap confidence intervals on the test set. Overall, each model shows reasonably strong performance achieving an AUC between 0.72 and 0.76. (Table 2) contains the features used in the models.

Conclusions To our knowledge, this is the first ML solution that can assess the risk-benefit profile of ICI for patients, based on their medical history. Structured data and clinician notes prior to ICI initiation were utilized to create features for the models. Feature engineering involved aggregating laboratory measurements over 60 and 120-day time windows. 1-year window was applied for other data types including ICD-10 codes, procedures, medication, and smoking history. In model development, patients were randomly partitioned into training (80%) and test (20%) sets for each outcome. An experiment involved a baseline and an alternative model, where the latter was selected if it demonstrated statistically significant superior performance based on outer loop results from a nested cross-validation process on the training set.

Results A random forest classifier was developed for each outcome. (Table 1) demonstrates performance results with 95% bootstrap confidence intervals on the test set. Overall, each model shows reasonably strong performance achieving an AUC between 0.72 and 0.76. (Table 2) contains the features used in the models.

Conclusions To our knowledge, this is the first ML solution that can assess the risk-benefit profile of ICI for patients, based on their medical history. Structured data and clinician notes prior to ICI initiation were utilized to create features for the models. Feature engineering involved aggregating laboratory measurements over 60 and 120-day time windows. 1-year window was applied for other data types including ICD-10 codes, procedures, medication, and smoking history. In model development, patients were randomly partitioned into training (80%) and test (20%) sets for each outcome. An experiment involved a baseline and an alternative model, where the latter was selected if it demonstrated statistically significant superior performance based on outer loop results from a nested cross-validation process on the training set.

Ethics Approval The Vanderbilt University Medical Center Health Sciences #3 institutional review board approved this study, tracked as #211814. The IRB determined the study poses minimal risk to participants, and a waiver of consent was granted.

http://dx.doi.org/10.1136/jitc-2023-SITC2023.1300