Machine Learning-Based Decision Tree for Identifying and Grading Severity of ICANS Based on Neurological Adverse Events in Patients Treated with CD19 CAR T-Cell

Alexandru Socolov, MBAn[1] | Penelope Lafeuille, MS[1] | Sheila Diamond, MS, CGC[1] | Eric Yang, PhD[1] | Francesco Maura, MD[2] | Jacob Aptekar, MD, PhD[1] | Esther Nie, MD, PhD[3]

1 Medidata, a Dassault Systèmes company, New York, NY, 10014 | 2 Sylvester Comprehensive Cancer Center. University of Miami, Miami, FL, 33136 | 3 Department of Neurology, Stanford University, Stanford, CA, 94304

INTRODUCTION

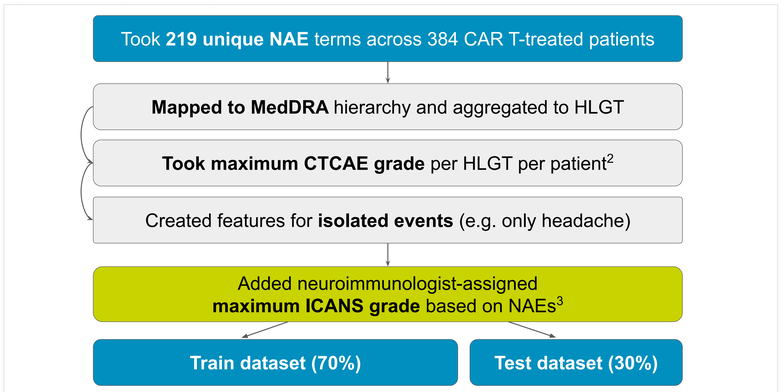
- One of the major challenges associated with Chimeric Antigen Receptor T-cell (CAR T) therapy is the development of immune effector cell-associated neurotoxicity syndrome (ICANS). This study uses machine learning in combination with clinical insight to build **a decision tree algorithm that accurately identifies and grades ICANS severity** in patients treated with CD19 CAR T therapy based on the neurological adverse events (NAEs).

- This approach allows us to **study the driving factors behind ICANS** as well as **perform retrospective ICANS analyses** in patient groups where the ICANS grades were not previously available.

METHODS

- We conducted a retrospective analysis of **384 patients who exhibited NAEs** after being treated with **CD19-targeted CAR T** therapy in the pooled clinical trial data from the Medidata Enterprise Data Store. Majority of patients had various types of Lymphoma while 9% had B-cell ALL[1]. Of note, this dataset excluded patients without NAEs; those patients were assumed to not experience ICANS. Data was processed according to Figure 1.

Figure 1. Data processing steps*: from verbatim NAE terms to train/test datasets*



- The final dataset included the **CTCAE severity of the NAEs mapped to MedDRA groups as features and the expert-labeled ICANS grades as the target variable.** The data was then split into train (70%) and test (30%) sets stratifying by study and ICANS grade.

- We fit a **decision tree** while performing 5-fold hyperparameter grid search on the train set. The tree with the highest area under the Receiver Operating Characteristic curve one-vs-one (**ROCAUC OVO**) was selected and evaluated on the test set using an array of classification performance metrics.

1 - 350 patients had some type of Lymphoma: majority DLBCL, some MCL, FL, B-cell Lymphomas. 34 patients had B-cell ALL.
2 - Except for the MedDRA term "Neurological Disorders NEC" which was broken down to the High Level Term (HLT) level due to the large size of this group (73 out of 219 terms)
3 - Neurological Adverse Events recorded in the first 28 days post infusion during the clinical trial along with their CTCAE toxicity and symptom onset day were displayed to the neuroimmunologist.

RESULTS

Figure 2. Decision tree of the ICANS grading model

At each decision step, the distribution of the MedDRA group is presented. Each final leaf contains a color break-down of true ICANS grades together with the (majority) predicted class. Decision tree plotted with Dtreeviz Python package.

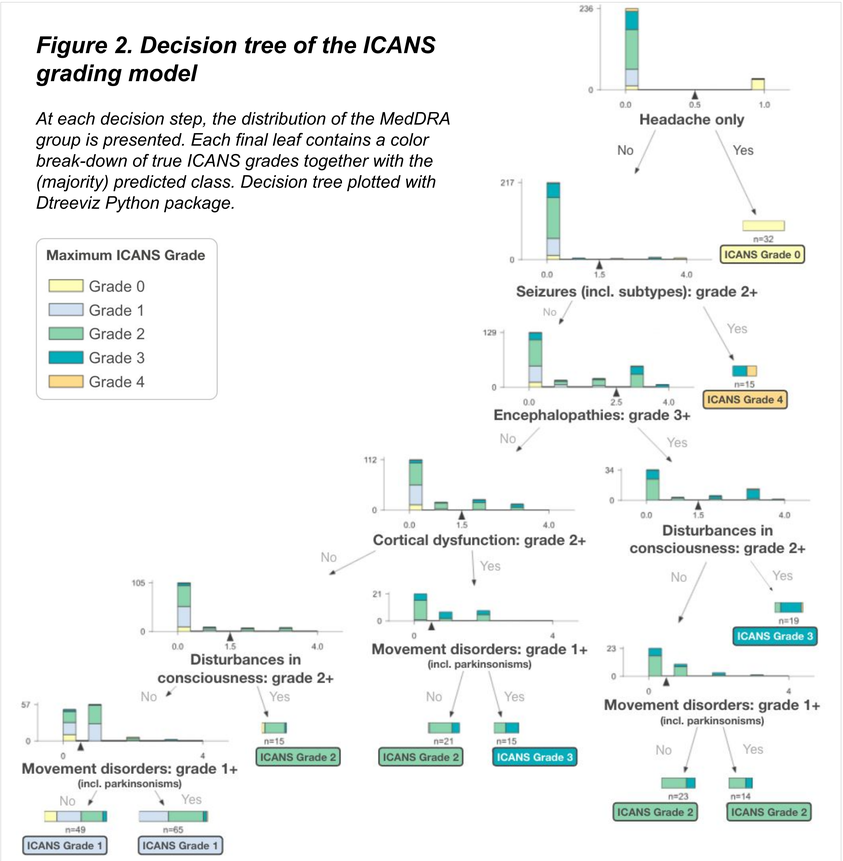
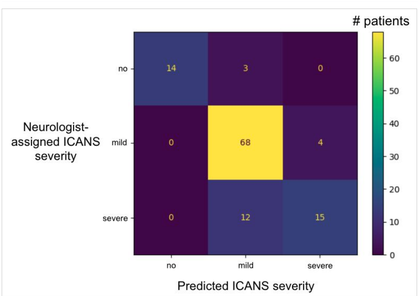


Figure 3A and 3B. Classification matrices *of neurologist-assigned versus predicted ICANS severity (3A) and ICANS grade (3B) in the out-of-sample patient cohort (n=116).*

3A: Predicting ICANS severity
(0 = no, 1-2 = mild, 3-4 = severe)



3B: Predicting exact ICANS grade
(from 0 to 4)

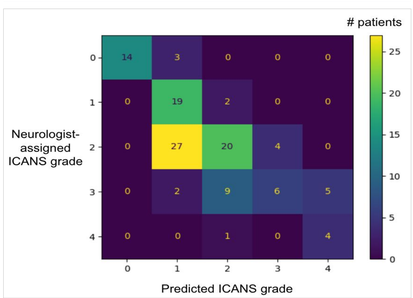


Table 1. Performance metrics *of the decision tree model*

Metric	Predicting ICANS severity (0 = no, 1-2 = mild, 3-4 = severe)	Predicting exact ICANS grade (from 0 to 4)
Balanced Accuracy	**0.775**	0.639
F1 score (weighted)	**0.829**	0.536
ROC AUC (weighted)	**0.935**	0.863
PR AUC (weighted)	**0.867**	0.582

- The training algorithm selected **six MedDRA groups of NAEs to be predictive of ICANS**: Headaches, Seizures (incl subtypes), Encephalopathies, Cortical dysfunction NEC, Disturbances in consciousness NEC, and Movement disorders (incl parkinsonism). See Figure 2 for full decision tree.

- **Majority of model errors are bound to +/- one grade** when comparing the neurologist-assigned versus predicted ICANS grades in the out-of-sample patient group (Figure 3A and 3B). Only 3% of patients are mislabeled by two grades. Among misclassified patients, NAEs atypical for ICANS arise like Syncope or Lethargy; the evaluation of these requires further interdisciplinary discussion. The system implementing this decision tree should build in a pause for whenever such terms are encountered.

- **None of the mild/severe ICANS cases were missed,** i.e. recall of predicting any ICANS (mild + severe) is 100%. The precision of predicting any ICANS is 97%. **When identifying severe ICANS,** precision and recall are 79% and 56%.

CONCLUSIONS

- Our machine learning-based decision tree **accurately predicts the severity and grade of ICANS in CD19 CAR T treated patients**, enabling retrospective analyses of ICANS patterns in populations where the syndrome had not been graded previously.

- This type of **clinical decision support tool** has the potential to be a powerful aid to physicians helping identify and manage toxicities earlier, especially to those **less familiar with ICANS**.

- **A single grading algorithm results in more uniform ICANS grades** across patient groups. In clinical research context, this is particularly important to enable consistent evaluation of cohorts.

- **Future directions**: further validation across **multiple physicians** is warranted to assess model's generalizability and refine its predictive capabilities. Additionally, we plan to continue studying ICANS or other neurological phenotypes that arise in high risk populations using different classes of immunotherapies other than CD19 CAR T.