



The 65th ASH Annual Meeting Abstracts

POSTER ABSTRACTS

803. EMERGING TOOLS, TECHNIQUES AND ARTIFICIAL INTELLIGENCE IN HEMATOLOGY

Machine Learning-Based Decision Tree for Identifying and Grading Severity of Immune Effector Cell-Associated Neurotoxicity Syndrome (ICANS) Based on Neurological Adverse Events in Patients Treated with CD19 CAR-T

Alexandru Socolov, MBAn¹, Penelope Lafeuille¹, Sheila Diamond, MS, CGC², Eric Yang, PhD¹, Francesco Maura, MD³, Jacob Aptekar, MD PhD², Esther Nie, MD PhD⁴

¹ Medidata AI, Medidata, a Dassault Systemes Company, New York, NY

² Medidata AI, Medidata, a Dassault Systemes company, New York, NY

³ Sylvester Comprehensive Cancer Center, Myeloma Division, University of Miami, Miami, FL

⁴ Department of Neurology, Stanford University, Stanford, CA

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), which can vary in severity and requires prompt intervention. This study proposes using machine learning in combination with clinical insight to build a decision tree that accurately identifies and grades ICANS severity in patients treated with CD19 CAR-T based on the neurological adverse events (NAEs).

Methods

We conducted a retrospective analysis of 384 patients that exhibited NAEs after being treated with CD19-targeted CART-cell therapy in the pooled clinical trial data from the Medidata Enterprise Data Store. Of note, this dataset excludes patients without NAEs; these patients were assumed to not experience clinical ICANS.

Patient NAEs were reviewed and the highest ICANS grade was assigned by a neuroimmunologist, resulting in the following patient counts: grade 0 (n=61), 1 (n=69), 2 (n=166), 3 (n=75), 4 (n=13). The clinical expert graded each patient based on the NAEs recorded in the first 28 days post infusion during the clinical trial, their CTCAE toxicity and symptom onset day.

To process NAEs, we first mapped each of the 219 verbatim terms to the MedDRA hierarchy using a proprietary Medidata coding algorithm. We then aggregated the NAEs to the High Level Group Term (HLGT) level 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). The mapping of NAEs to MedDRA groups was reviewed together with the neuroimmunologist.

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. The model output was evaluated at the ICANS grade and severity (no/moderate/severe) level. The confusion matrices for the test set and the full decision tree were plotted.

Results

When binning the ICANS grade prediction by severity (grade 0 = no ICANS, grade 1-2 = moderate, grade 3-4 = severe), the model achieves a strong out-of-sample accuracy of 83.6% and ROCAUC OVO of 0.935. Regarding the individual ICANS grade prediction, the model yielded lower out-of-sample accuracy of 54.3%, with the majority of errors between grades 1 and 2. In spite of this, there was still reasonable discrimination based on the out-of-sample ROCAUC OVO values of 0.863.

The training algorithm selected six MedDRA groups of NAEs to make predictions with: Headaches, Seizures (incl subtypes), Encephalopathies, Cortical dysfunction NEC, Disturbances in consciousness NEC, and Movement disorders (incl parkinsonism). We observe that patients with only headaches (a separate feature in our dataset) are all labeled as having no ICANS while those with seizures of grade 2 and higher are classified as grade 4. Refer to Figure 1 for the full decision tree.

Comparing the neurologist-assigned versus predicted ICANS grades in the out-of-sample patient group (Figure 2) revealed that the majority of errors were bound to +/- one grade. Only three out of 116 patients were mislabeled by two grades. Crucially, none of the grade 1+ patients were predicted by the model to have grade 0.

Conclusion

Our machine learning-based decision tree accurately predicts the severity and grade of ICANS in CD19 CAR-T treated patients, aiding in clinical decision-making as well as retrospective analyses of ICANS patients. This type of clinical decision support tool has the potential to be a powerful aid to physicians by providing them with earlier physiological insights and data-driven assessments to help them triage patients accordingly.

The model's performance highlights its potential clinical utility in improving patient care and outcomes. Further validation across multiple neurologists and prospective studies are warranted to assess its generalizability and refine its predictive capabilities.

Acknowledgments

We thank the developers of *Scikit-learn* and *Dtreeviz* Python packages for open-sourcing their code.

Disclosures Socolov: Medidata, a Dassault Systèmes company: Current Employment. **Lafeuille:** Medidata, a Dassault Systèmes company: Current Employment. **Diamond:** Medidata, a Dassault Systèmes company: Current Employment. **Yang:** Medidata, a Dassault Systèmes company: Current Employment. **Aptekar:** Medidata, a Dassault Systèmes company: Current Employment. **Nie:** Medidata, a Dassault Systèmes company: Consultancy.

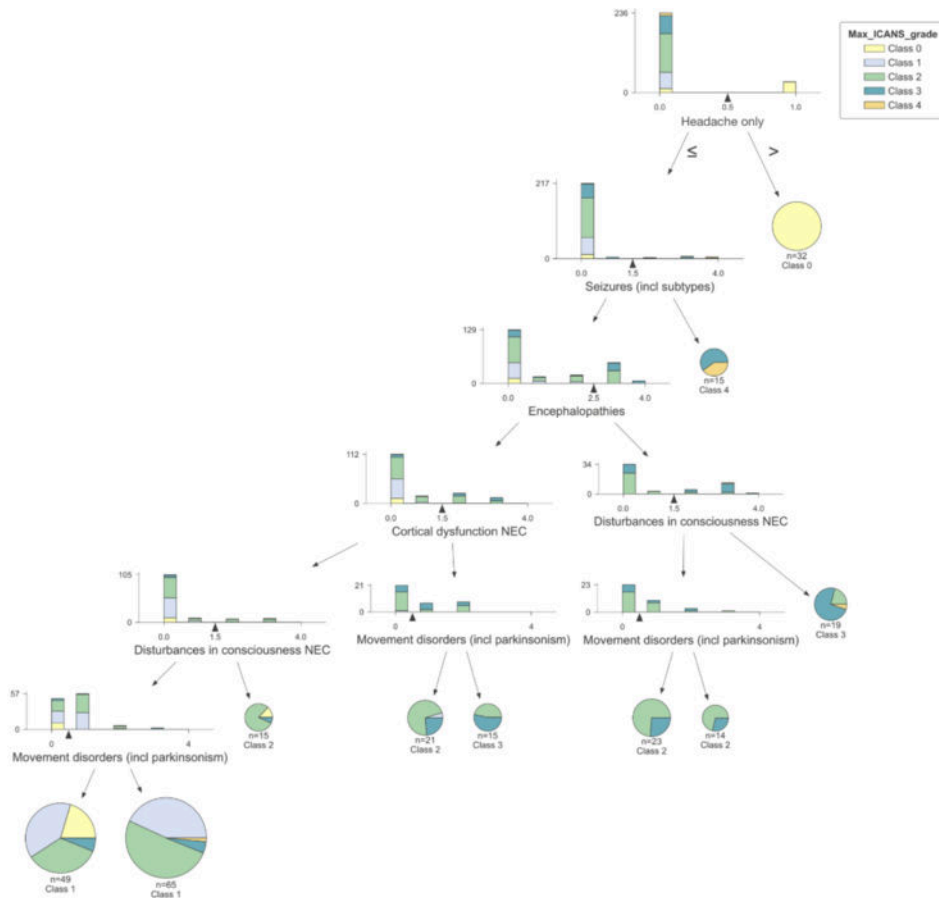


Figure 1. Decision tree of the ICANS grading model. At each decision step, the distribution of the MedDRA group is presented. Each final leaf contains a color break-down of true ICANS grades together with the (majority) predicted class. Training parameters for decision tree (scikit-learn library): max_depth: 6, min_samples_leaf: 14, criterion: gini, class_weight: balanced.

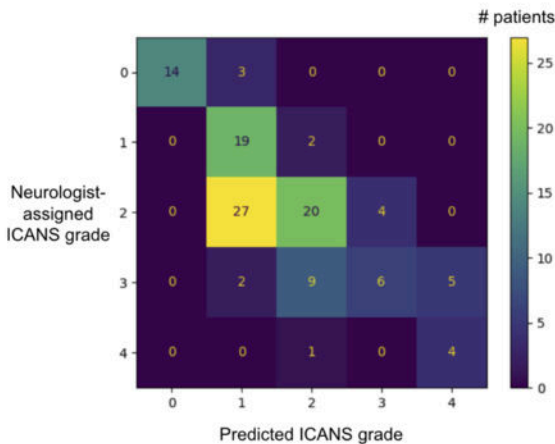


Figure 2. Classification matrix of neurologist-assigned versus predicted ICANS grades (via the below "confusion matrix") in the out-of-sample patient cohort. A model capable of perfectly replicating the neurologist-assigned ICANS grades would have non-zero # patients across the diagonal and zero everywhere else.

Figure 1

<https://doi.org/10.1182/blood-2023-182147>