Early Identification of the Need for CRRT in Critically Ill Children: A Machine Learning Approach

Mark Wainwright[1], Qingyang Li[1], Shina Menon[1], Alexander Doud[1], David Van De Sompele[1], Kin Vong[1], Vijayaraghavan Makkakode[1], Hillary Bourdrez[1]

[1] Seattle Children's Hospital, Seattle, WA, USA

Background

Continuous Renal Replacement Therapy (CRRT) is a life-saving intervention in critically ill children with acute kidney injury (AKI). At our institution approximately 1% of patients receive CRRT during their ICU stay. The decision to initiate CRRT may be based on a combination of factors including the patient's clinical status, underlying comorbidities, caregivers' concerns, and resource availability. We sought to develop a machine learning (ML) model using Electronic Health Records (EHR) data to support clinical decision making for earlier identification of children in the pediatric ICU (PICU) who require support with CRRT. The goal of the model is to identify patients at least one day before the clinical team makes the decision for CRRT for patients in the ICU.

Methods

This study was approved by the local Institutional Review Board. Seattle Children's Hospital admits over 2000 patients annually to its 39-bed medical-surgical PICU. We used data extracted from patients admitted to the PICU with a length of stay greater than 24 hours between 2008 and 2021. We selected candidate data elements based on clinical expertise and attention to the anticipated United States Core Data for Interoperability (USCDI) requirements. The structured data included demographics, vital signs, laboratory values, and fluid balance data. We restricted the analysis to the initial 12 days after ICU admission. To address the class imbalance problem of the data, we oversampled the data from the CRRT patient group and then selected data using a random 24-hour window preceding the outcome of interest for each new sample. By including test cases with data collected 1 day before the CRRT decision, this approach was intended to enable earlier identification of the need for CRRT. Only the initial CRRT encounter per patient was included in the modeling to prevent bias, even if multiple encounters occurred for a patient. We engineered features by vectorizing the patient state and then selected features using correlation-based feature selection (CFS) and information gain (IG) feature selection in combination with 5 methods of classification: random forest, logistic regression, Naïve Bayes (NB), support vector machine, and extreme gradient boosting. Data curation, analyses and development were conducted using Python (version 3.9).

Results

19457 PICU encounters (159 with at least one episode of CRRT) were identified and then stratified into training and test sets after data augmentation. The test set accounted for 15% of the data. Models constructed using IG contained 436 features versus 42 features using CFS. The top features were related to blood urea nitrogen, creatinine, platelets, coma score, and fluid balance. Models built with NB approach outperformed the other approaches. The NB model with IG feature selection achieved area under the ROC curve (AUROC) of 0.925, area under the precision-recall curve (AUPRC) of 0.813, accuracy of 0.88 and F1 score of 0.88. The NB model with CFS achieved comparable accuracy and F1-score but the performance was slightly better with 0.945 AUROC, 0.890 AUPRC.

Discussion

We present an interoperable decision support pipeline for the early identification of the need for CRRT in critically ill children. The model uses structured, time-series data from the EHR and ML algorithms to aid clinicians in making decisions about the need for CRRT. The study's limitations include being conducted in a single institution and not considering unstructured data, which may improve the model's performance. Future work will focus on external validation of the model with real-time data and the development of a user-friendly interface to facilitate its use in the clinical setting.