Machine Learning for Healthcare 2020 – Clinical Abstract Track

Development of Machine Learning Models for Early Prediction of Clinical Deterioration in Pediatric Inpatients

Zohaib Shaikh MD candidate[1,2], Daniel Witt MIDS, MD candidate[2,3], Tong Shen MS[4], William Ratliff MBA[2], Harvey Shi[2], Michael Gao[2], Marshall Nichols MS[2], Mark Sendak MD MPP[2], Suresh Balu MBA MS[2], Karen Osborne BSN RN[5], Karan Kumar MD MS[6], Kimberly Jackson MD[6], Andrew McCrary MD MS[7], Jennifer Li MD MHS[7]

[1]Duke University School of Medicine, [2]Duke Institute for Health Innovation, [3]Mayo Clinic Alix School of Medicine, [4]Department of Biomedical Engineering, Duke University, [5]Duke University Health System, [6]Division of Pediatric Critical Care Medicine, Duke University, [7]Division of Pediatric Cardiology, Duke University

Background.

Approximately 50,000 children die each year in the United States, and more than half die in inpatient hospital settings. Hospitalized children can decompensate quickly, and predicting which child might decompensate is often difficult. Several pediatric early warning systems (PEWS) have been developed and studied for early detection of deterioration with wide-ranging performance. Most PEWS currently in use almost exclusively use a patient's dynamic features, such as vital signs and appearance, although some studies have found static features, such as medical history, to be powerful predictors of deterioration. Although scores incorporating both static and dynamic features have improved performance, the increased number of features to be assessed by bedside nurses can be time-consuming and resource intensive. Duke University's Department of Pediatrics and the Duke Institute for Health Innovation (DIHI) formed a transdisciplinary team to develop a machine learning model to evaluate an extensive number of clinical features and accurately predict a pediatric inpatient's risk of deterioration.

Methods.

Data were collected from 17,630 inpatient encounters for 10,388 unique pediatric patients (defined as <18 years of age at hospital admission or <25 years of age and on a pediatric service) at a large academic health system from October 2014 to August 2018. Encounters limited to only the labor and delivery and/or neonatal units were excluded. The deterioration outcome was defined as an unplanned transfer to the intensive care unit (ICU) or inpatient mortality. Planned admissions to the ICU from the emergency department or the operating room and direct transfers from labor and delivery were excluded. A total of 542 predictive features were built from patients' age, sex, comorbidities, and prior inpatient encounters at the time of admission, and vitals, lab results, orders, and medication administrations during the encounter. Features with numerical values were processed by creating event flags, 24-hour rolling mean, minimum, and maximum values, and hourly differences in values. Non-numerical elements were mapped into features according to representations of clinical severity. The models are designed to generate hourly predictions of the risk of an unplanned transfer to the ICU over the subsequent 24 hours and mortality over the subsequent 48 hours. Models were trained using light gradient boosting machine (LGBM), lasso-penalized logistic regression (LR), and random forest (RF) methods. Models were evaluated on the accuracy of hourly predictions using area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) and compared to the current institutional standard of care, the Duke PEWS (D-PEWS).

Results.

Of the 17,630 encounters, 6.0% experienced the deterioration outcome (1,022 encounters with unanticipated ICU transfer, 108 culminating in inpatient mortality, and 81 with both events). In most encounters, ICU transfer was observed to occur soon after admission, with 25% of encounters experiencing a transfer in the first 22 hours, 50% in the first 101 hours, and 75% in the first 536 hours (mean 510 ± 959 hours). The LGBM model performed best in predicting the deterioration outcome, with an AUROC of 0.847 (95% CI, 0.840-0.854) and AUPRC of 0.082 (95% CI, 0.076-0.090), compared to the RF (AUROC: 0.814 [95% CI, 0.806-0.822]; AUPRC: 0.067 [95% CI, 0.061-0.075]) and the LR models (AUROC: 0.812 [95% CI, 0.804-0.822]; AUPRC: 0.071 [95% CI, 0.065-0.078]) and D-PEWS (AUROC: 0.690 [95% CI, 0.686-0.693]; AUPRC: 0.066 [95% CI, 0.063-0.069]).

Conclusion.

Machine learning models utilizing a wide range of clinical data elements exhibit improved performance in predicting early clinical deterioration in pediatric inpatients compared to the current institutional standard of care, the D-PEWS. The transdisciplinary team aims to pilot the LGBM model within pediatric inpatient units in late 2020 to support proactive patient assessment by rapid response teams while collecting prospective data. Integration of this model into the rapid response workflow may improve care and mitigate morbidity and mortality for children at high risk of deterioration.