Hybrid Supervised and Unsupervised Machine Learning Model Improves Prediction of New-Onset Dialysis in Orthotopic Heart Transplant Patients

Roshan Mathi, MSc[1], Weston McDonald[1], Benjamin Usry[1], Khaled Shorbaji, MD[1], Brett Welch, MBA[1], Arman Kilic, MD[1]

1. *Division of Cardiothoracic Surgery, Department of Surgery, Medical University of South Carolina, Charleston, SC, USA*

Introduction

Machine learning (ML) approaches to clinical risk prediction have demonstrated improved performance over traditional statistical approaches. New-onset dialysis following isolated orthotopic heart transplantation (OHT) is an important contributor to mortality risk and quality of life in recipients, yet traditional approaches to predicting its occurrence have performed only modestly. This study sought to develop and evaluate a hybrid supervised and unsupervised ML approach to estimating the risk of new-onset dialysis following OHT.

Methods

This was a retrospective study of patients (n = 26,505) with no history of dialysis undergoing primary, isolated OHT between 2000 and 2022 as identified in the United Network for Organ Sharing (UNOS) database. The study cohort was split randomly into training (80%) and validation sets (20%) and under-sampling using K-Nearest Neighbors clustering was used to account for class imbalance. ML models, including a Multilayer Perceptron Classifier, a Support Vector Classifier, and a Logistic Regression Model were developed and compared. Sensitivity, accuracy, precision, and F_1 score were measured, and discriminatory performance was assessed using the area under the receiver-operating-characteristic curve (AUC).

Results

The rate of new dialysis was 11.3% following OHT with comparable rates in the training and validation cohorts. In the validation cohort, the support vector classifier had the highest AUC Score (0.774 (95% CI, 0.748 – 0.800), sensitivity (0.74), F_1 Score (0.72), accuracy (0.70), and precision (0.70). All other models yielded poorer performance by every metric. Predictive risk factors that were common across models included congenital heart disease, pre-transplant glomerular filtration rate, and human leukocyte antigen mismatch.

Conclusions

This study demonstrates that an ML approach using a support vector model was associated with improved sensitivity and discriminatory capability compared to other modeling approaches including logistic regression in predicting new-onset dialysis risk after OHT. The use of ML models in OHT may have utility in patient selection, modification of risk factors, and programmatic evaluation.