Leveraging Machine Learning to Predict Sepsis Mortality in Bangladesh

Kaden Bunch MBA [1], Nidhi Kadakia MD [2], Siraj Amanullah MD MPH[2], Gazi Md. Salahuddin Mamun MBBS, MPH[3], Shamsun Nahar Shaima MBBS, MPH[3], Abu Sayem Mirza Md. Hasibur Rahman MBBS, MPH[3], Alicia Genisca MD, Atin Jindal MD[4], Monira Sarmin MBBS, MCPS [3], Farzana Afroze MBBS, FCPS [3], Md. Tanveer Faruk MBBS, MPH[3], Abu Sadat Md. Sayeem, MBBS[3], Tahmeed Ahmed, MBBS, PhD[3], Mohammod Jobayer Chisti MBBS, MMed, PhD [3], Adam C. Levine MD, MPH, FACEP[2], Stephanie Chow Garbern MD, MPH, DTM&H [2]

[1] The Warren Alpert Medical School of Brown University, Providence, RI, USA
[2] Department of Emergency Medicine, The Warren Alpert Medical School of Brown University, Providence, RI, USA
[3] International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh
[4] Division of Hospital Medicine, Lifespan Health System, Providence, RI, USA

BROWN
Alpert Medical School

Introduction

Sepsis is the leading cause of child death globally, with low- and middle-income countries (LMICs) bearing the largest burden. LMICs often have limited prognostic tools and critical care capacity, leading to delayed recognition of advanced sepsis (Figure. 1).

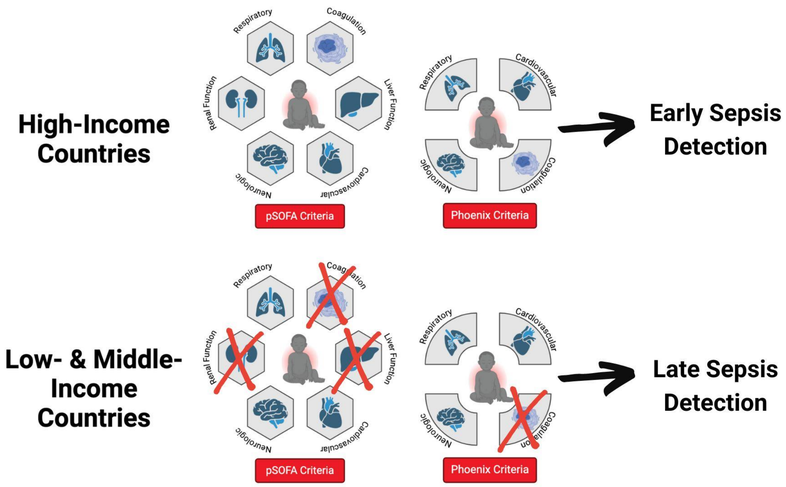

Figure 1. Comparison of HICs and LMICs in the speed of recognition of sepsis due to limited use of pSOFA and Phoenix Criteria.

Traditional sepsis scoring systems, such as the pSOFA and Phoenix criteria, are used to evaluate the severity of organ dysfunction and diagnose sepsis. While extremely valuable, these tests require extensive lab tests that are only sometimes available. This study aimed to evaluate machine learning (ML) models using readily accessible clinical data throughout Bangladesh to predict mortality in pediatric sepsis following the initial 24 hours post-suspicion of sepsis.

Objectives

- We hypothesize that sepsis prediction models based on widely available clinical measurements can match the accuracy of pSOFA and Phoenix criteria in resource-constrained settings
- In this study, we developed and evaluated ML models using readily accessible clinical data to predict mortality in pediatric sepsis.

Methods

Sample Collection: Data was collected from a prospective observational study of children (**N = 96**) with suspected sepsis admitted to the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr, b). Upon enrollment, study staff collected demographic, clinical, and laboratory tests, and upon study termination, mortality.

Variable Determination: To minimize overfitting, dimensionality reduction techniques were employed, and features were selected based on their significance in four distinct models (Figure 2).

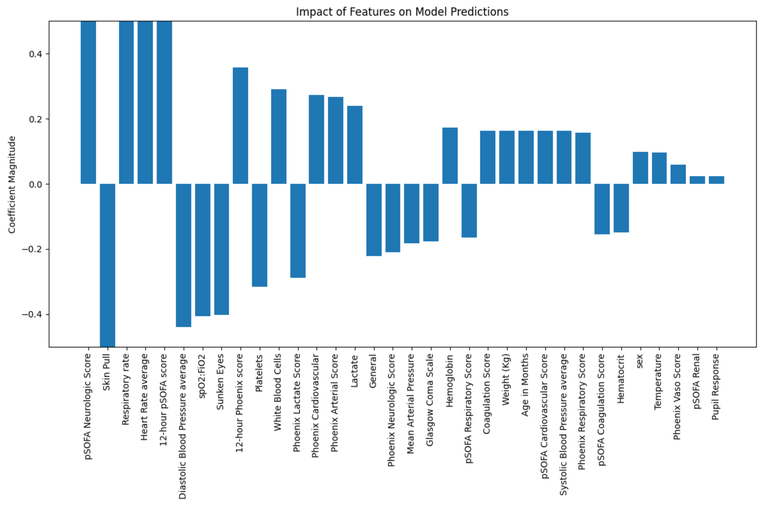

Figure 2. A bar chart illustrating the coefficient magnitudes of all features reflective of their mortality prediction, highlighting significant predictors.

Feature Set Creation: Four distinct feature sets were created;

- Model 1 (**Clinical Variables**): WBCs, platelets, lactate, age, temperature, RR, HR, BP, spO2:FiO2, and GCS
- Model 2 (**Clinical Variables without Lab Values**): age, temperature, RR, HR, BP, spO2:FiO2, and GCS
- Model 3 (**Phoenix Criteria** by individual components): Respiratory, cardiovascular, coagulation, and neurologic components
- Model 4 (**pSOFA criteria** by individual components): Respiratory, Cardiovascular, Coagulation, Renal, and Neurologic components

ML Classifiers Employed: Data was trained and tested on the following seven ML classification models: Decision tree, Random forest, Support Vector Machine (SVM), Kernel SVM, Naïve Bayes, K nearest neighbors (KNN), and logistic regression.

Results

Table 1. Comparative performance of predictive models using different feature sets, including Clinical Variables (Model 1), Non-Lab Variables (Model 2), Phoenix Scores (Model 3), and pSOFA scores (Model 4). The metrics evaluated are Accuracy, F1 Score, and AUROC.

	Clinical Variables	Non-Lab Variables	Phoenix Scores	pSOFA Scores
Decision Tree				
Accuracy	0.829 [0.794 - 0.863]	0.852[0.805 - 0.898]	0.861[0.819 - 0.903]	0.842[0.785 - 0.894]
F1 Score	0.620[0.534 - 0.704]	0.668[0.550 - 0.782]	0.535[0.364 - 0.697]	0.593[0.434 - 0.741]
AUROC	0.788[0.725 - 0.848]	0.807[0.725 - 0.886]	0.693[0.556 - 0.814]	0.800[0.714 - 0.877]
Random Forest				
Accuracy	0.838[0.801 - 0.875]	0.829[0.800 - 0.861]	0.870[0.816 - 0.921]	0.851[0.807 - 0.895]
F1 Score	0.528[0.390 - 0.658]	0.409[0.250 - 0.562]	0.661[0.517 - 0.796]	0.606[0.472 - 0.734]
AUROC	0.871[0.819 - 0.922]	0.800[0.722 - 0.872]	0.809[0.694 - 0.911]	0.835[0.771 - 0.896]
SVM				
Accuracy	0.865[0.825 - 0.904]	0.874[0.823 - 0.923]	0.879[0.832 - 0.926]	0.875[0.834 - 0.916]
F1 Score	0.700[0.603 - 0.793]	0.652[0.486 - 0.804]	0.629[0.461 - 0.783]	0.679[0.550 - 0.788]
AUROC	0.904[0.861 - 0.943]	0.900[0.841 - 0.953]	0.862[0.803 - 0.919]	0.915[0.882 - 0.947]
Naïve Bayes				
Accuracy	0.851[0.813 - 0.890]	0.847[0.811 - 0.882]	0.843[0.806 - 0.882]	0.820[0.786 - 0.857]
F1 Score	0.651[0.548 - 0.757]	0.638[0.543 - 0.732]	0.636[0.541 - 0.733]	0.260[0.116 - 0.416]
AUROC	0.888[0.821 - 0.945]	0.869[0.794 - 0.938]	0.875[0.821 - 0.928]	0.907[0.865 - 0.943]
Logistic Regression				
Accuracy	**0.871[0.836 - 0.908]**	**0.879[0.830 - 0.926]**	**0.898[0.858 - 0.938]**	**0.866[0.837 - 0.896]**
F1 Score	**0.674[0.590 - 0.765]**	**0.669[0.499 - 0.818]**	**0.720[0.618 - 0.825]**	**0.667[0.587 - 0.749]**
AUROC	**0.918[0.880 - 0.953]**	**0.915[0.865 - 0.962]**	**0.894[0.849 - 0.938]**	**0.925[0.895 - 0.954]**
Kernel SVM				
Accuracy	0.829[0.791 - 0.869]	0.815[0.772 - 0.861]	0.875[0.836 - 0.917]	0.816[0.778 - 0.856]
F1 Score	0.393[0.238 - 0.553]	0.340[0.202 - 0.478]	0.627[0.493 - 0.758]	0.486[0.338 - 0.626]
AUROC	0.891[0.831 - 0.944]	0.855[0.789 - 0.917]	0.845[0.759 - 0.921]	0.839[0.801 - 0.880]
KNN				
Accuracy	0.853[0.805 - 0.894]	0.819[0.773 - 0.867]	0.848[0.805 - 0.893]	0.824[0.783 - 0.866]
F1 Score	0.515[0.362 - 0.662]	0.415[0.265 - 0.569]	0.546[0.398 - 0.688]	0.434[0.262 - 0.602]
AUROC	0.895[0.844 - 0.941]	0.789[0.722 - 0.857]	0.836[0.758 - 0.908]	0.811[0.756 - 0.863]

- Logistic Regression demonstrated high predictive performance with AUROC of 0.918 (CI: 0.880–0.953) for clinical variables and 0.915 (CI: 0.865-0.962) for clinical variables without labs.
- The Phoenix and pSOFA criteria were broken into individual sub-scores as variables, achieving an AUROC of 0.894 (CI: 0.849-0.938) and 0.925 (CI: 0.895-0.954).

Summary

We have shown that predictive modeling can perform as well as, or even better than, traditional sepsis scoring systems in resource-limited settings. Future research should focus on enhancing these models to improve their practical use. By employing machine learning models, we can improve the timely identification and treatment of pediatric sepsis, leading to a reduction in mortality in low—and middle-income countries.

Acknowledgments

This work was supported by the Emerging Infectious Disease and HIV Scholars Program at Brown University (NIH Grant *2R25AI140490)*