Developing Machine Learning Models to Predict for Pediatric Mechanical Ventilation-Associated Events (PedVAE) in Patients

Varunikaa TL[1], Bryan Ho[1], Yushan Zou[1], Rita Tsapatsis[1], Dheeksha Sudhakar[1], Katie J. O'Conor MD[3], Timothy Ruchti Ph.D.[2], Jules Bergmann MD, MHS[2], James Fackler MD[3], Casey O. Taylor PhD[1], Joseph L. Greenstein PhD[1]

1. Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD, USA
2. Nihon Kohden Digital Health Solutions LLC, Irvine, CA, USA
3. Johns Hopkins Medical Institutes (JHMI), Department of Anesthesiology and Critical Care Medicine, Baltimore, MD, USA

TEAM HAWK

Introduction

- Mechanical Ventilation is a life-saving intervention in pediatric intensive care units (PICUs), but it can lead to serious complications
- Pediatric ventilator associated events (PedVAEs) are linked to increased length of stay, mortality, and resource use.
- Current VAE models are primarily based on adult data, which may not translate well to pediatric patients
- There is a critical need for pediatric-specific tools to predict PedVAEs and support early clinical decision-making.
- This project uses high-resolution EHR and physiological time-series data from over 1,200 pediatric patients.
- We apply advanced machine learning techniques to develop predictive models tailored to the pediatric population.

Objective

To develop machine learning models that accurately predict pediatric ventilator associated events (PedVAEs) using electronic health records (EHR) and physiological time series data to enable early detection and assist with clinical decision making in the PICU.

Data Overview

Dataset: 1,237 patients admitted to the Johns Hopkins PICU who were mechanically ventilated between July 2016 and July 2022.

Available Features: Vitals, Respiratory Panels, Ventilation Settings, and Lab Values

Figure 1 details how we selected for our cohort. Our preprocessing steps included data smoothing, missing data handling, and feature engineering.

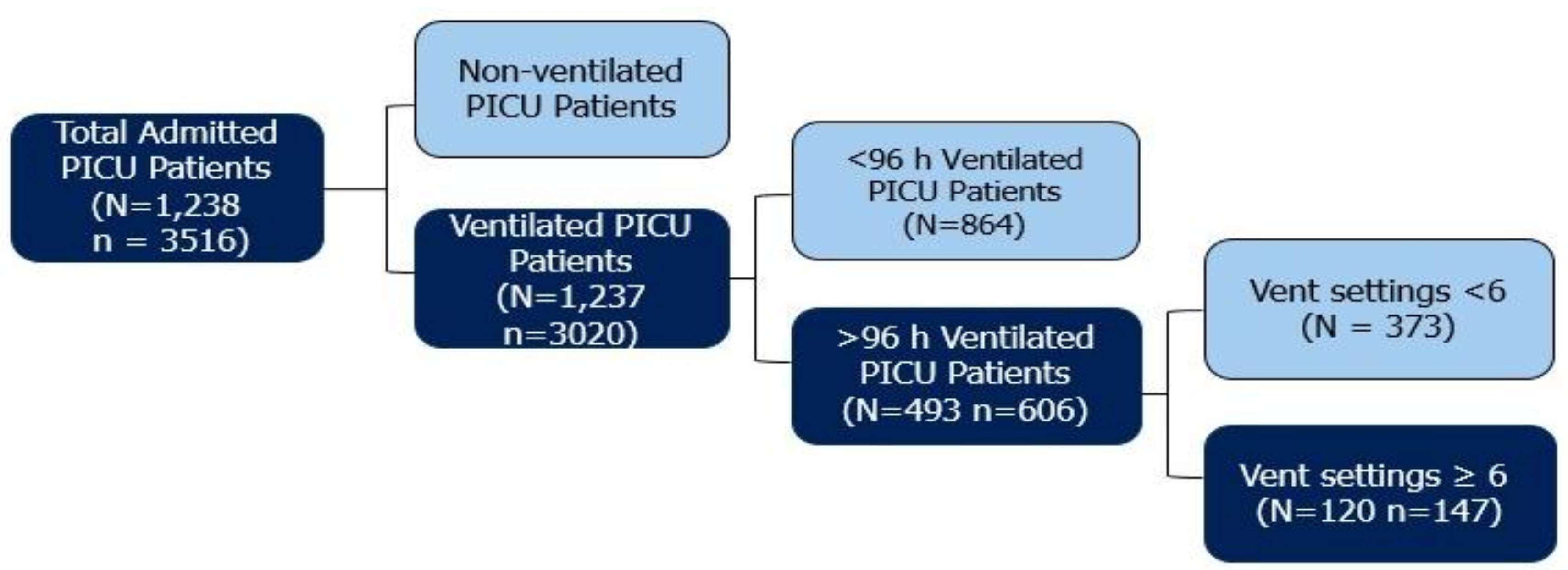


Figure 1. Cohort Identification. Patient selection by ventilation status, duration, and settings. (N – number of unique patients; n – number of encounters)

PedVAE Definition Refinement

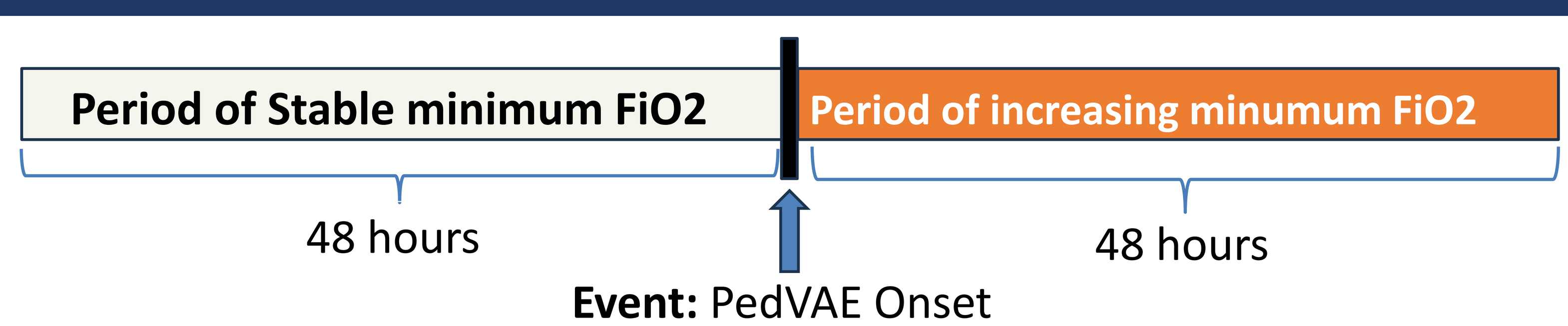


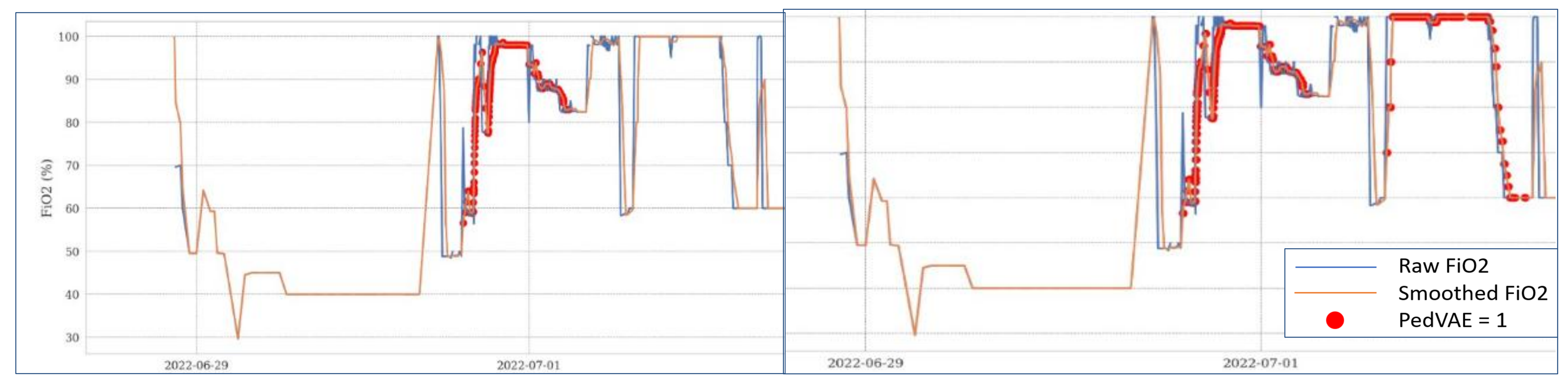
Figure 2. Standard PedVAE Definition. PedVAE definition of an increase in the minimum daily FiO2 by at least 0.25 from the baseline, sustained for two or more consecutive days after a period of stability ≥ 2 days

Definition Pipeline:

- Smooth the raw data by the mean value within 1 hour window
- Compute minimum FiO2 for the stable and increasing period
- Compare and define whether current time point has PedVAE or not, label each time point

Definition Refinement:

- Change FiO2 threshold from 0.25 to 02
- Try different combination of time periods: 48h+48h, 48h+24h, 24h+24h, and 24h + 12h
- PedVAE identification before and after definition refinement:



Before Refinement
After Refinement

Figure 3. PedVAE identification. Identification of PedVAE events both before and after definition refinement with 48 hours of stability followed by 24 hrs of increased FiO2.

Results

Model	Data before the PedVAE event			
	Accuracy	Precision	F1 Score	AUC ROC
LR[a]	0.7007	0.7258	0.6000	0.6780
RF[b]	0.7483	0.7108	0.7613	0.8002
XGB[c]	**0.7755**	**0.7671**	**0.7724**	**0.8276**
SVM[d]	0.6463	0.5962	0.6716	0.6780

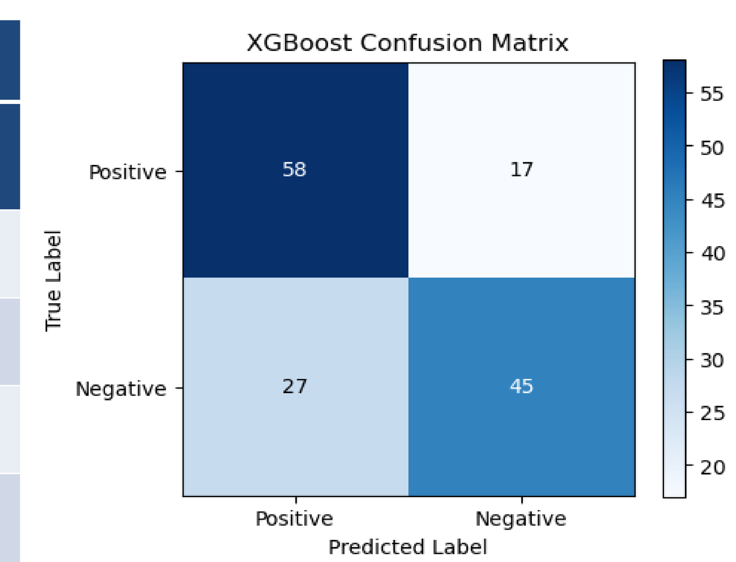


Figure 4. Comparison of Machine Learning Model Performance in Cross-Validation. A) Table comparing model accuracy, precision, F1 scores, and AUC ROC. a - Logistic Regression, b - Random Forest, c - XGBoost, d - SVM. B) Confusion matrix of the best performing model, XGBoost, comparing predicted label vs. true label normalized by true label.

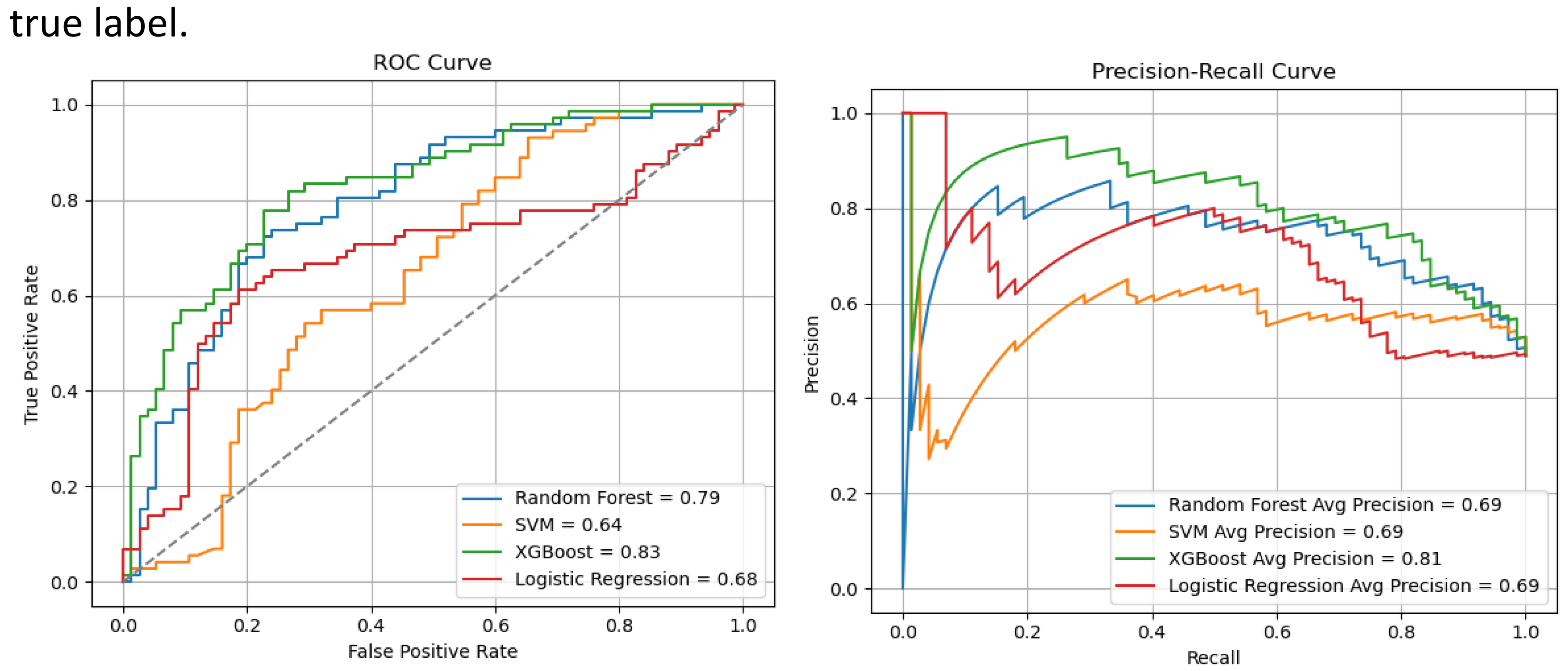
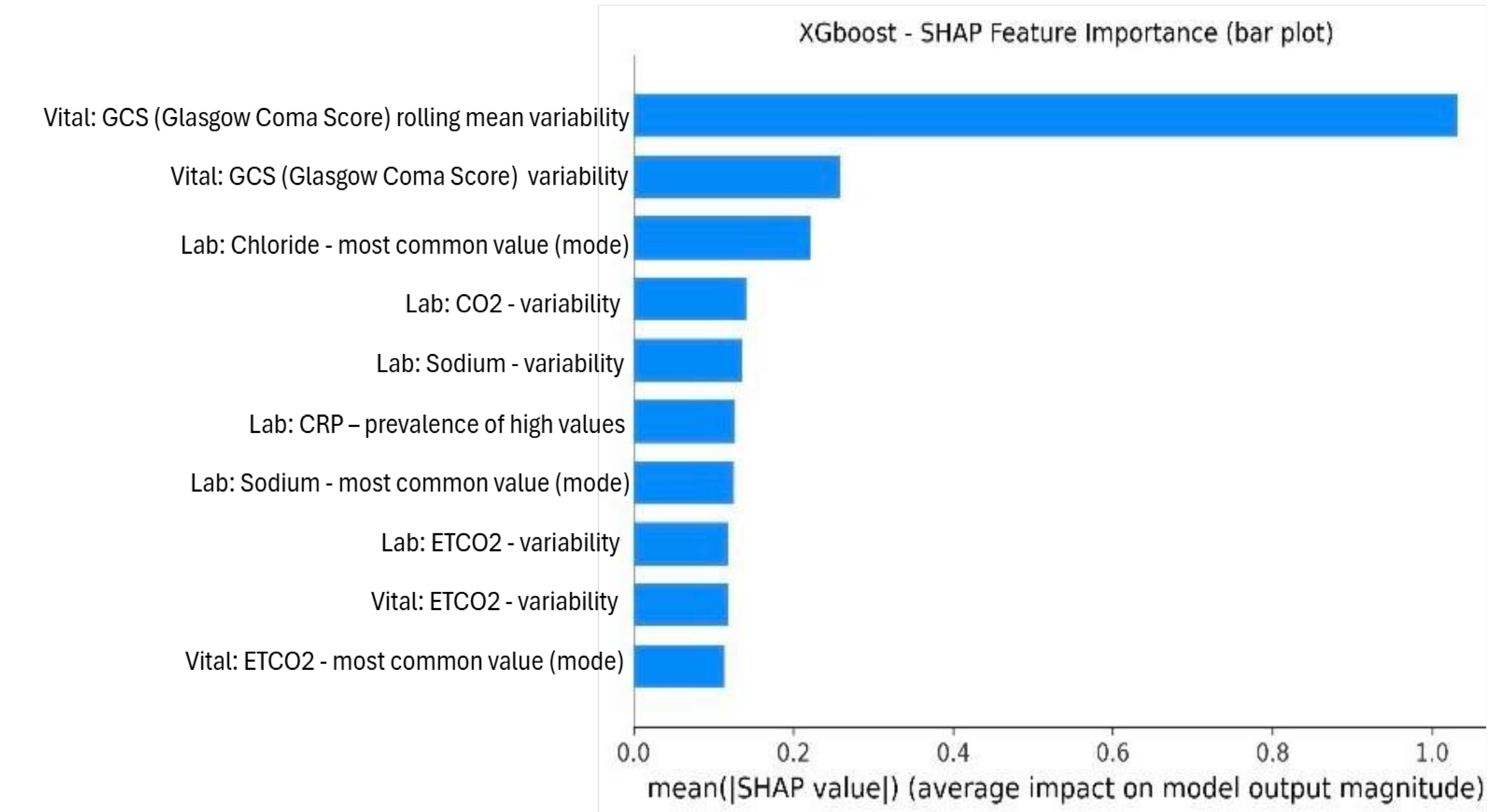


Figure 5. Performance of XGBoost model with all features on held-out test set. Precision-Recall Curve, where AP indicates Average Precision.



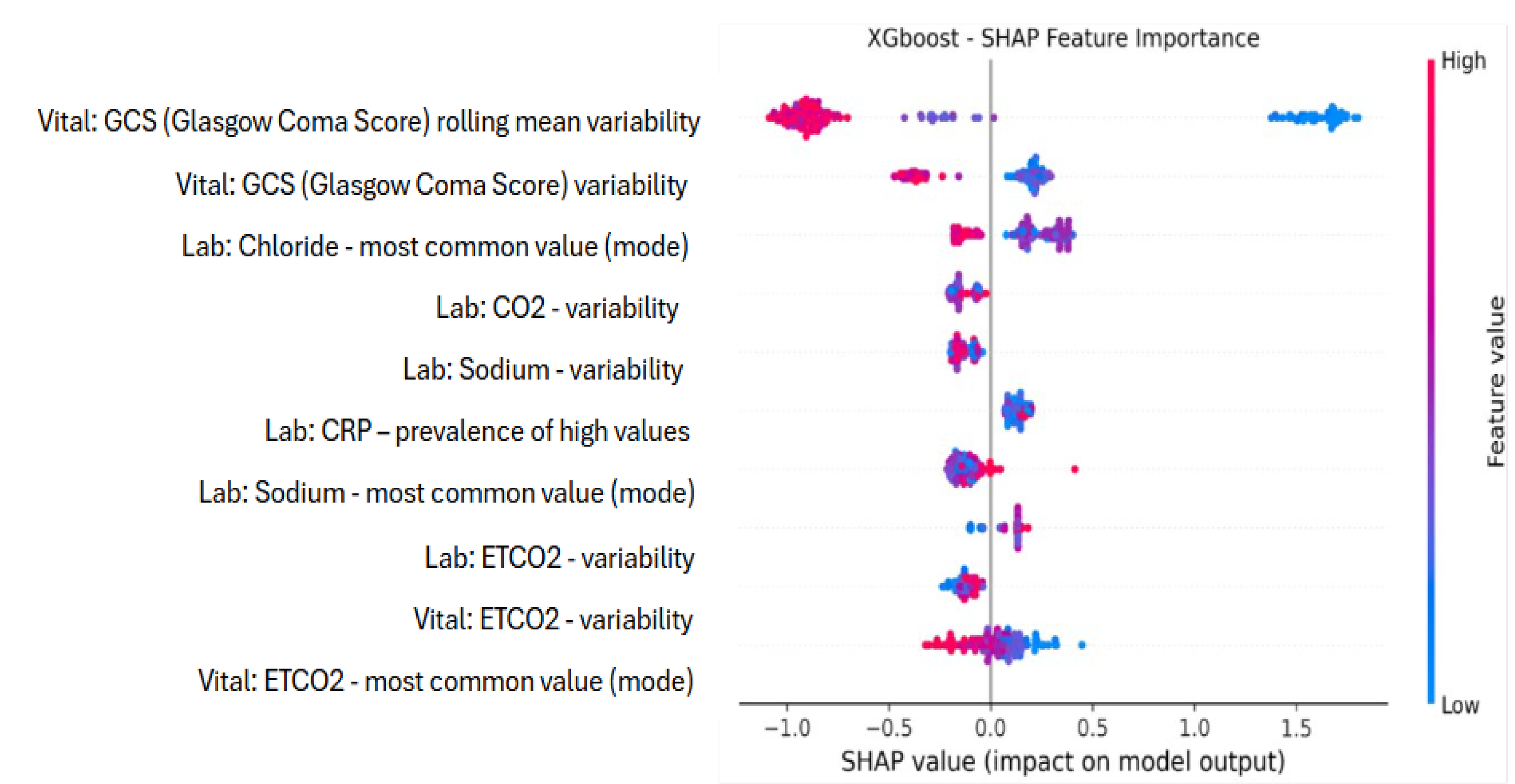


Figure 6. SHAP Analysis of Model Results. Shapley analysis for the final XGBoost model based on the training & validation data. Shows the impact of the top 10 most important features on PedVAE classification.

Conclusions & Future Direction

- Developed effective machine learning models to predict pediatric ventilator-associated events (PedVAEs) using detailed EHR and physiological data.
- XGBoost model demonstrated the best performance, showing strong potential for early detection in pediatric intensive care units (PICUs).
- Refined PedVAE definitions led to improved predictive accuracy.

Future directions:

- Validate models across various clinical settings and diverse patient populations to ensure generalizability and robustness.
- Integrate real-time predictive analytics into clinical decision support systems.
- Enhance clinicians' ability to intervene earlier, potentially reducing morbidity and improving outcomes for pediatric patients receiving mechanical ventilation.