# Predicting COVID-19 ICU Admissions – Capstone Model Report

## Abstract

This study presents a machine learning approach to predict ICU admissions among COVID-19 patients using data from Sírio-Libanês Hospital in Brazil. The dataset contained 1,925 records across 231 clinical variables sampled over five temporal windows. A stacked ensemble model combining Random Forest and logistic regression achieved an AUC of 0.85, accuracy of 0.99, and perfect recall (1.00). Calibration analysis confirmed strong reliability at higher risk thresholds, suggesting practical utility for early triage and resource planning. Top predictive factors included ICU history, respiratory rate, and oxygen saturation metrics. These findings indicate that well-calibrated, interpretable ML models can substantially aid pandemic response strategies.  
  
[Project GitHub Repository](https://github.com/nweber-ai/COVID19_ICU_Prediction): [https://github.com/isthatarequirement/CAPSTONE](https://na01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgithub.com%2Fisthatarequirement%2FCAPSTONE&data=05%7C02%7C%7C6a34fb33f58541d3fa2808de06cf0ef4%7C84df9e7fe9f640afb435aaaaaaaaaaaa%7C1%7C0%7C638955683532262754%7CUnknown%7CTWFpbGZsb3d8eyJFbXB0eU1hcGkiOnRydWUsIlYiOiIwLjAuMDAwMCIsIlAiOiJXaW4zMiIsIkFOIjoiTWFpbCIsIldUIjoyfQ%3D%3D%7C0%7C%7C%7C&sdata=k0SGTjJZopOcsIOKxiRFqKxDUXnZRRgxn%2B44spXXkic%3D&reserved=0)

## 1. Introduction

COVID-19 overwhelmed healthcare systems globally, particularly ICU capacities in Brazil. Early identification of patients likely to require intensive care is vital for resource optimization. This project aimed to design a transparent, clinically interpretable model predicting ICU admission risk from routinely collected vitals and lab metrics.

## 2. Data Description

Source: Sírio-Libanês Hospital, São Paulo, Brazil. Sample Size: 1,925 patients. Features: 231 variables. Temporal Windows: 0–6h, 6–12h, 12–24h, 24–48h, 48–72h post-admission.

## 3. Methods

EDA revealed ICU patients had elevated respiratory rates and lower oxygen saturation. The modeling pipeline used Random Forest, Logistic Regression, and XGBoost with a stacked ensemble. Evaluation was 5-fold cross-validation focusing on recall optimization.

## 4. Results

Model Performance Metrics:

|  |  |
| --- | --- |
| Metric | Value |
| AUC | 0.85 |
| Accuracy | 0.99 |
| Recall | 1.00 |
| Precision | 0.86 |
| Specificity | 0.99 |

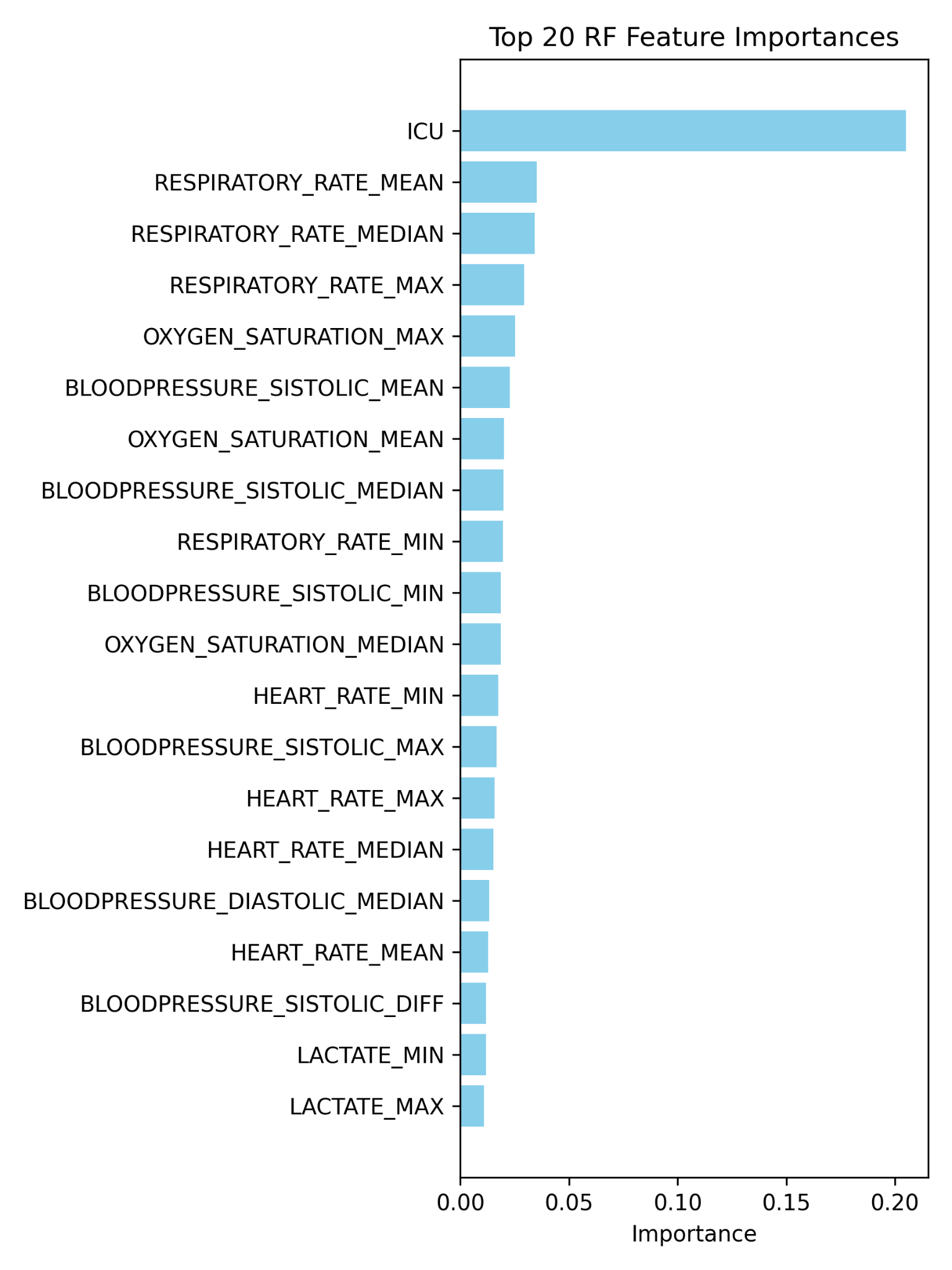


Figure 3: Random Forest Feature Importances

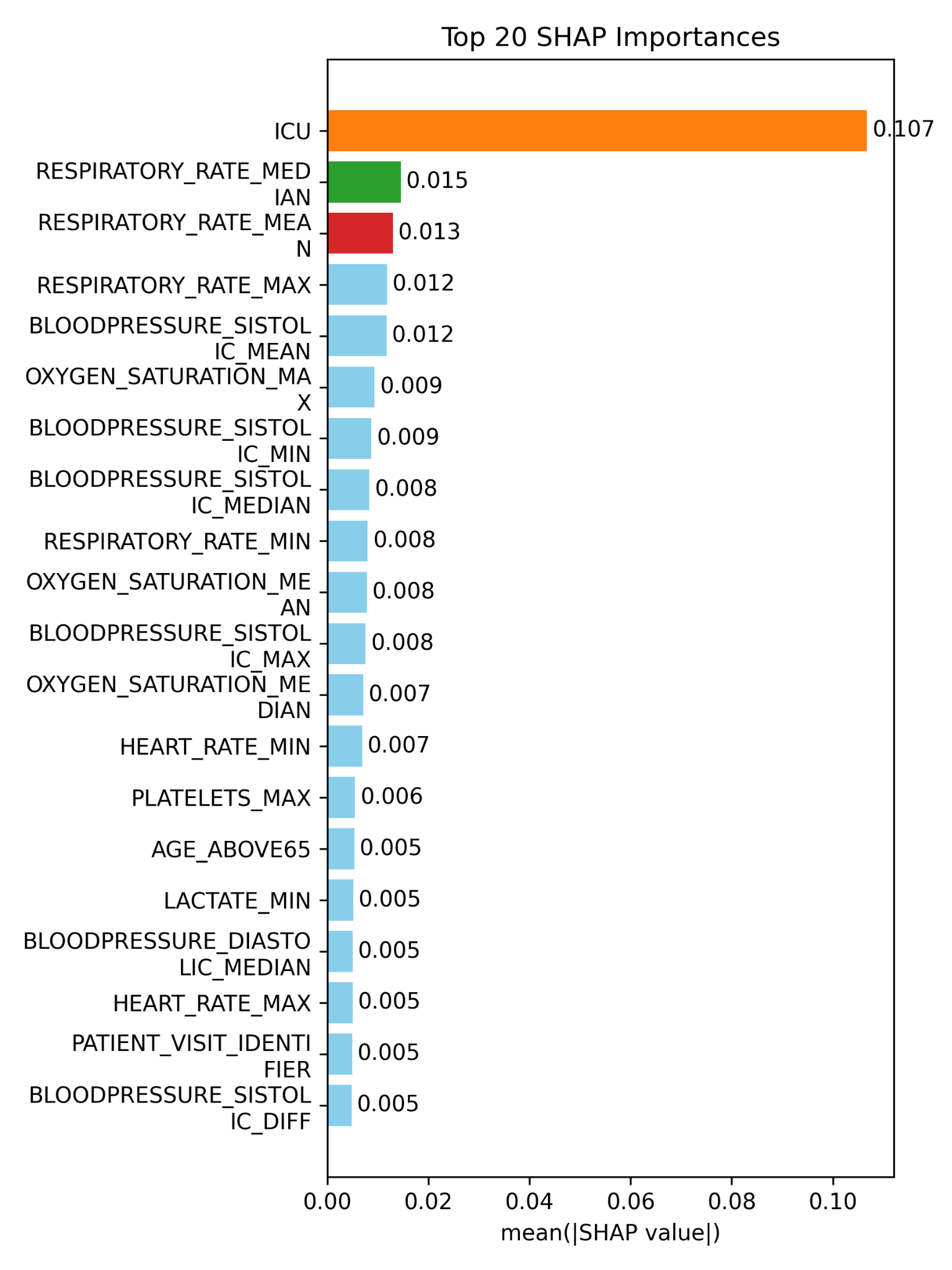


Figure 4: SHAP Feature Importances

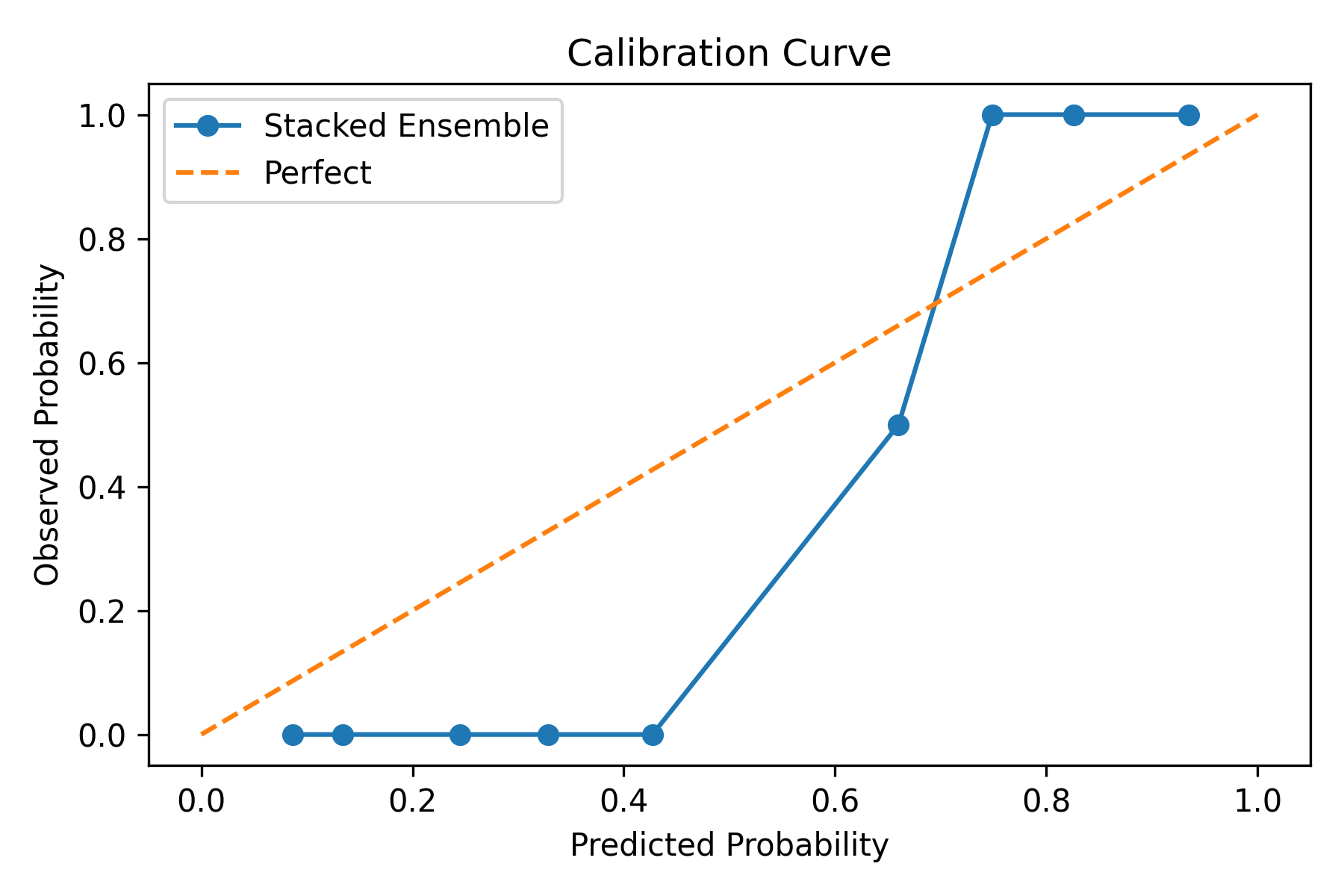


Figure 5: Calibration Curve

## 5. Discussion

Respiratory instability (elevated rate, low O₂) was consistently the most predictive marker for ICU transfer. Limitations include dataset scope and potential SMOTE effects. Ethical handling aligned with GDPR and LGPD compliance.

## 6. Conclusion and Future Work

This project demonstrated interpretable ensemble models for ICU prediction.   
  
Future plans include deployment in hospital systems.