Applied Analytics: Machine Learning Pipeline

Prediction of Sepsis, Severe Sepsis, and Septic Shock for ICU patients with chronic diseases

Carnegie Mellon University
Information System Management
Apr 27, 2018
Ya Ting Chang; Michelle Hsu

Dataset

MIMIC II

	Feature
Demographics	age, sex, white, asian, black, hispanic, others, unknown
Chronic diseases	diabetes, kidney, lung
Vital signs	blood.pressure, heart.rate, respiratory.rate, oxygen.saturation, temperature
Outcome	Sepsis, SeverSepsis, SepticShock

Explorative Data Analysis

Imbalanced Data

	Sepsis	Severe Sepsis	Septic Shock
0	97.6 %	86.8 %	93.4 %
1	2.44 %	13.2 %	6.65 %

SMOTE

	Sepsis	Severe Sepsis	Septic Shock
0	73.2 %	57.1 %	72.4 %
1	26.8 %	42.9 %	27.6 %

Missing Value

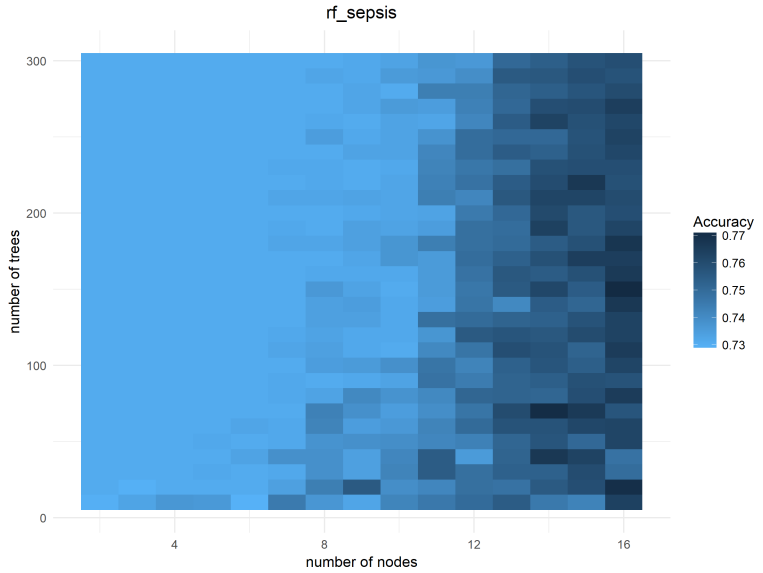
	Missing %
sex	0.366 %
age	0.122 %
blood.pressure	7.99 %
heart.rate	4.82 %
respiratory.rate	5.06 %
oxygen.saturation	5.06 %
temperature	6.52 %

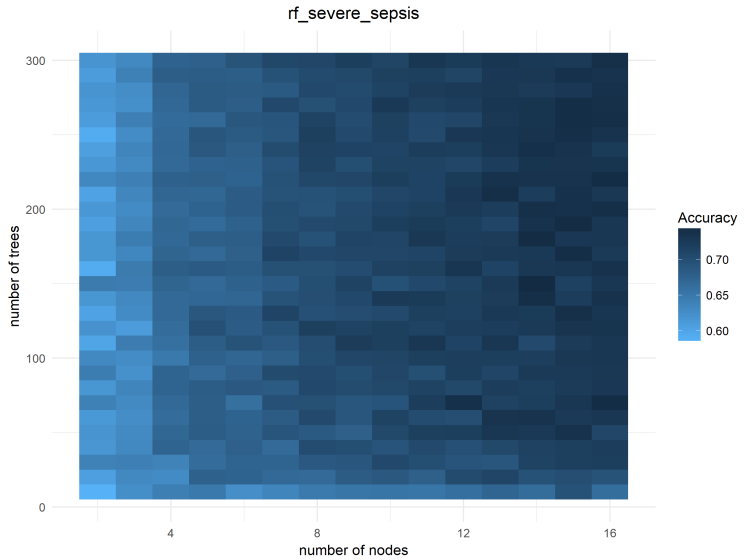
Modeling

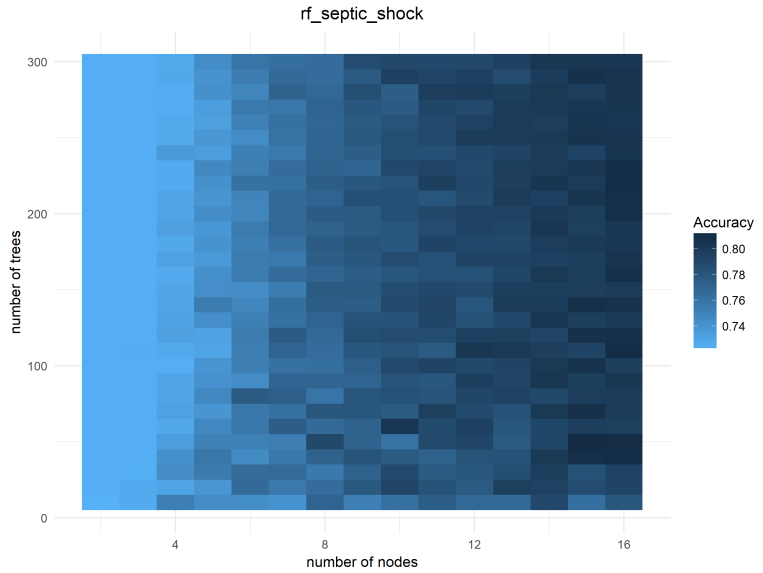
- **Naïve Bayes: baseline**
- **Random Forest: #trees, #nodes**
- **Gradient Boosting: #depths, #iterations**

Multi-label Classification (3 labels) → 3 Binary Classification Tasks

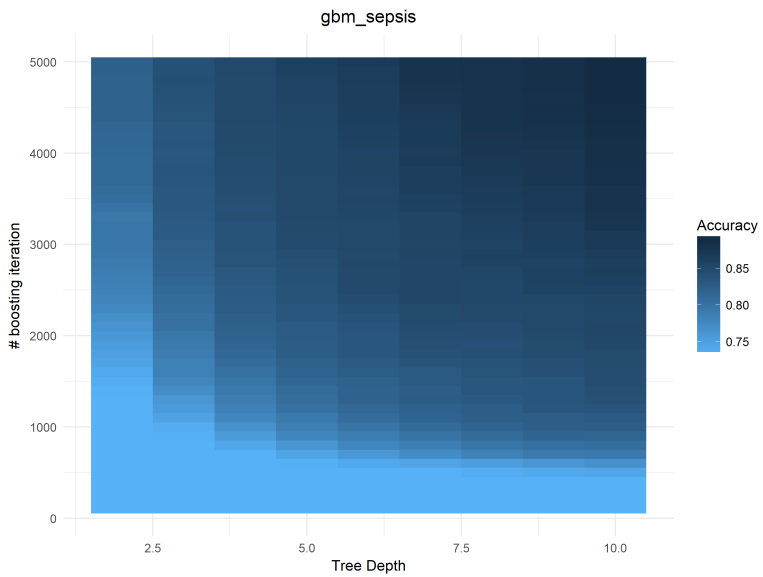
Random Forest: #trees, #nodes

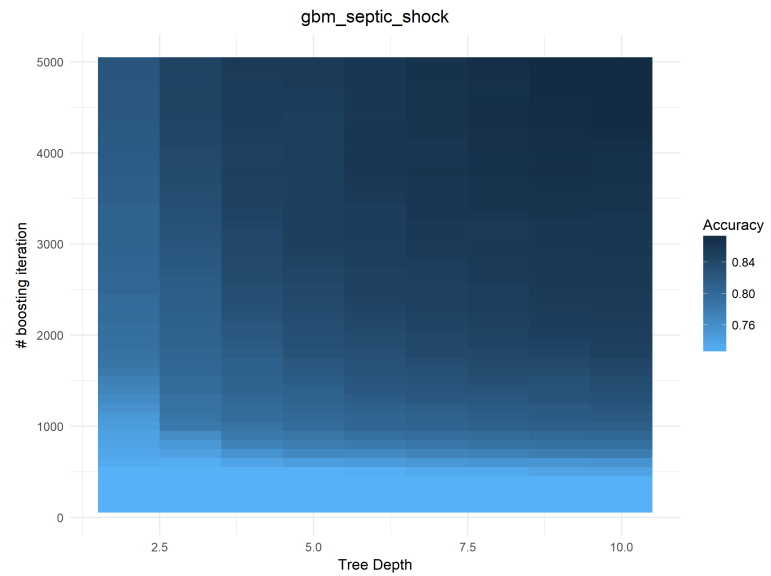


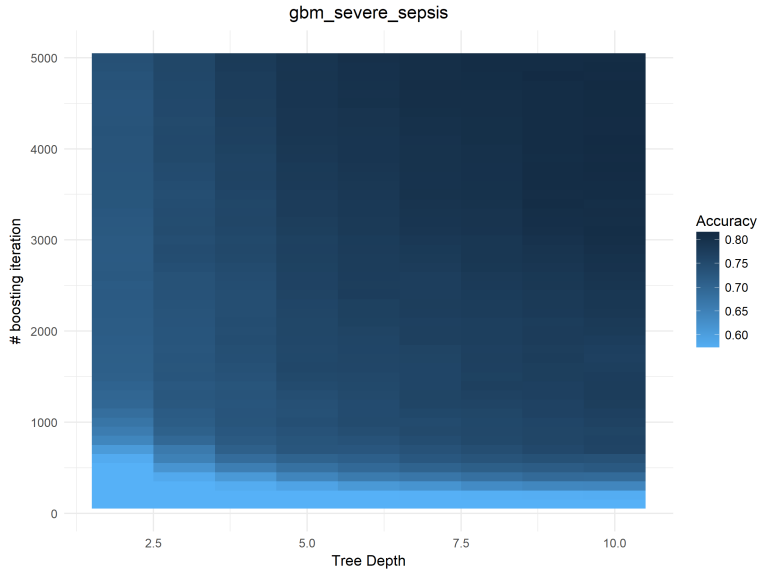




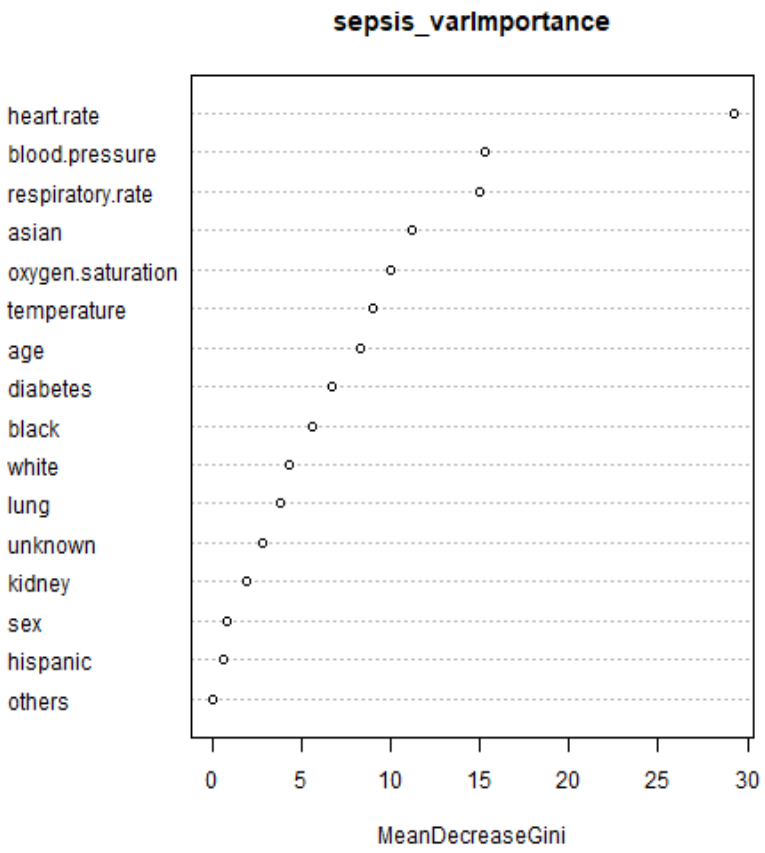
Gradient Boosting: #depths, #iterations

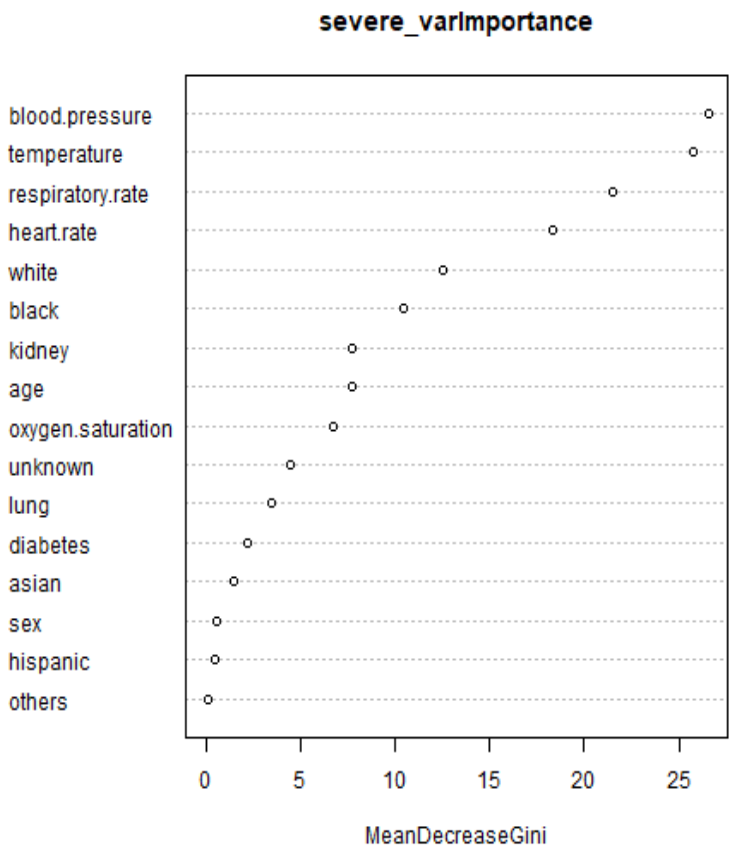


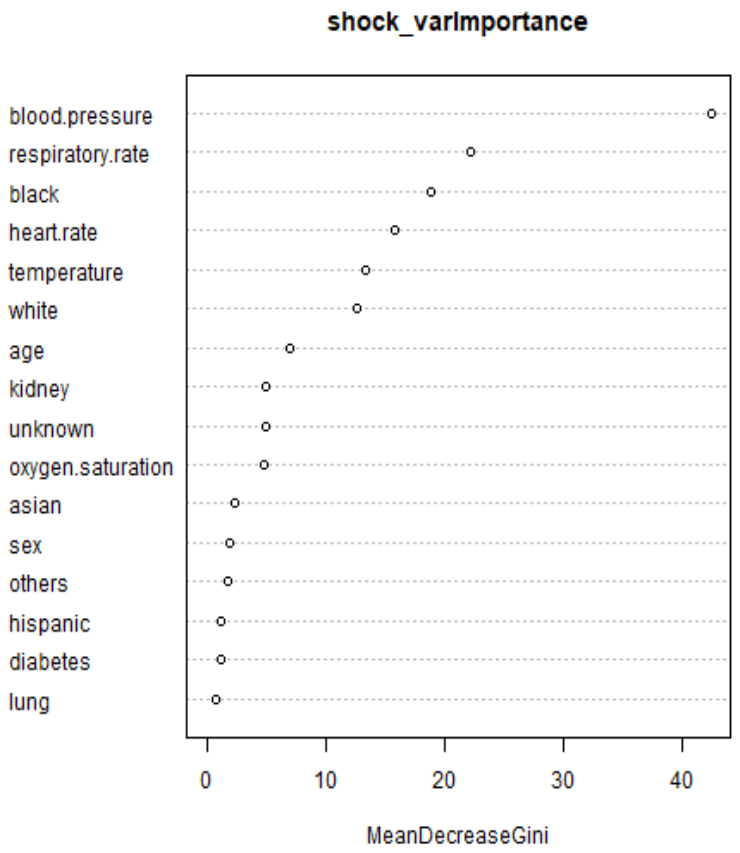




Random Forest Variable Importance

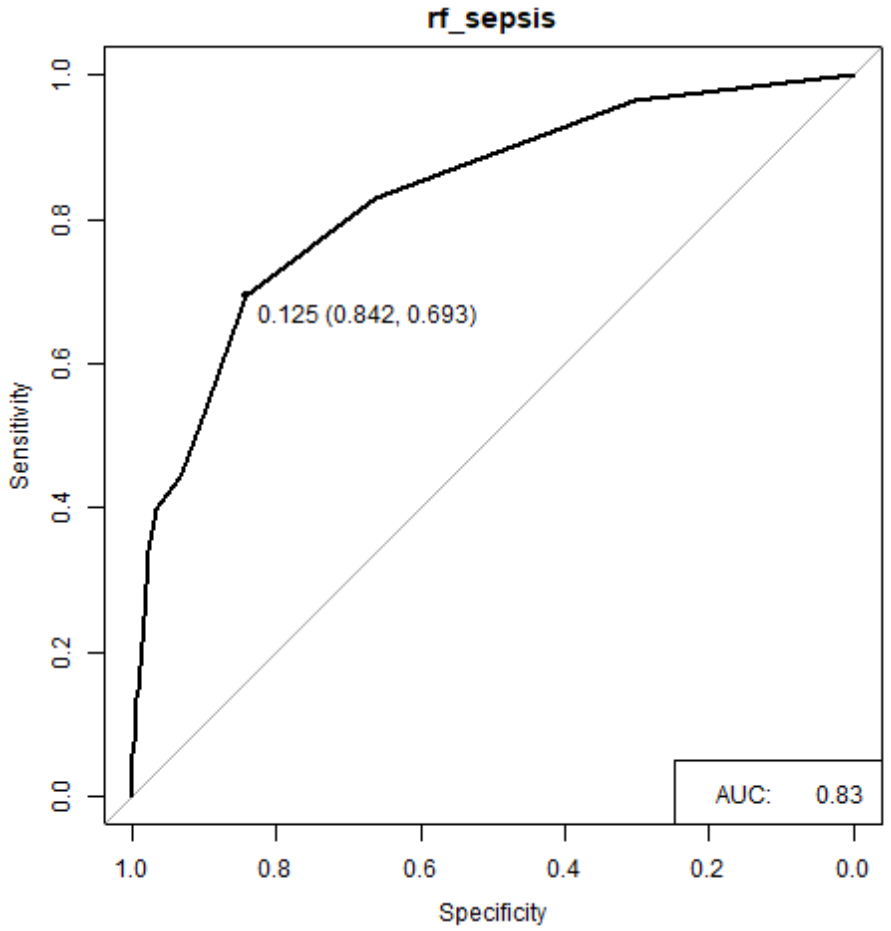




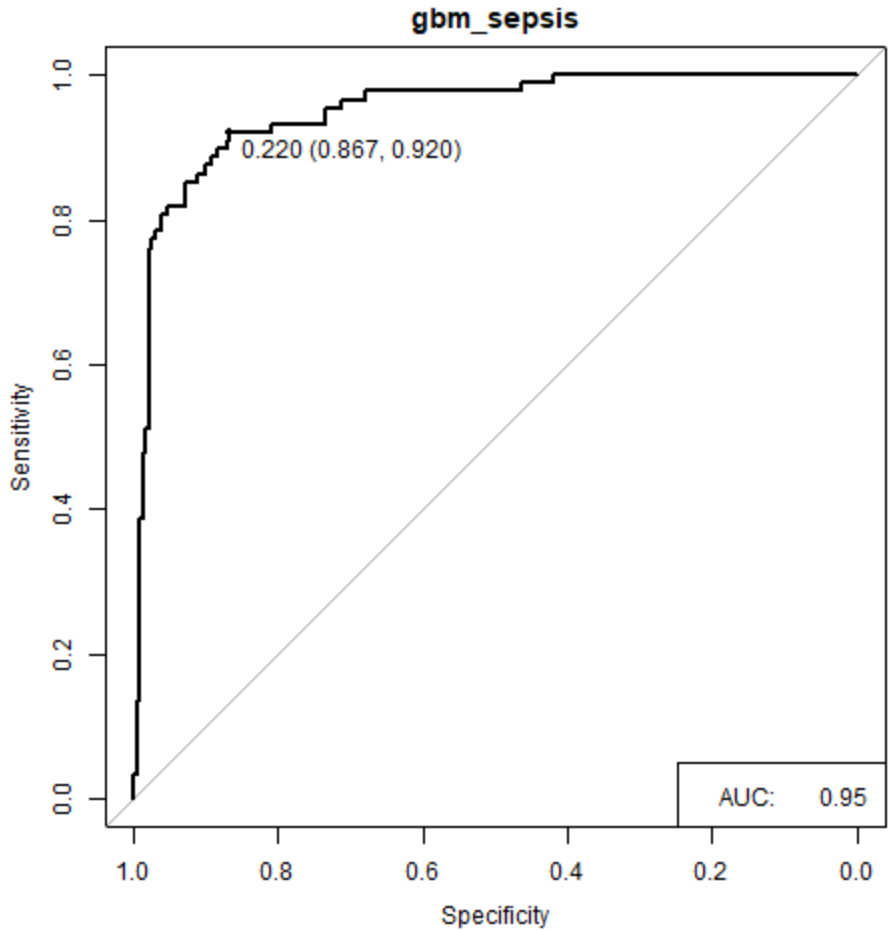


Sepsis – Model Performance

Random Forest

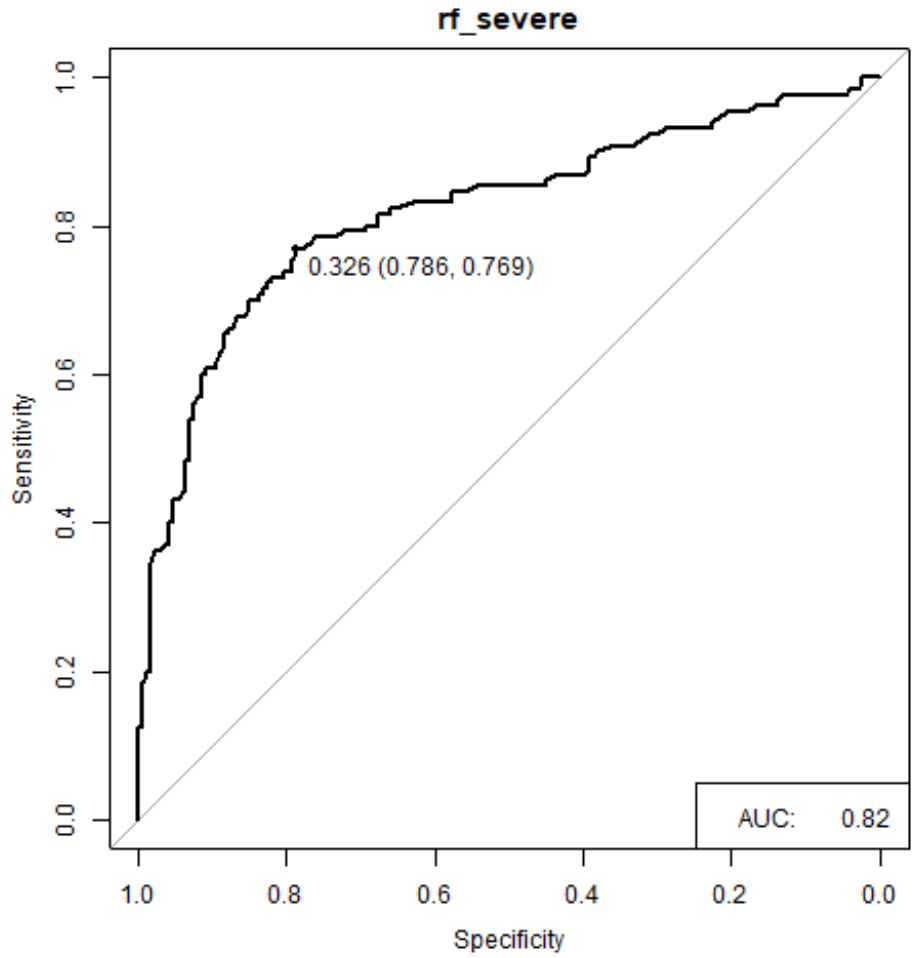


Gradient Boosting

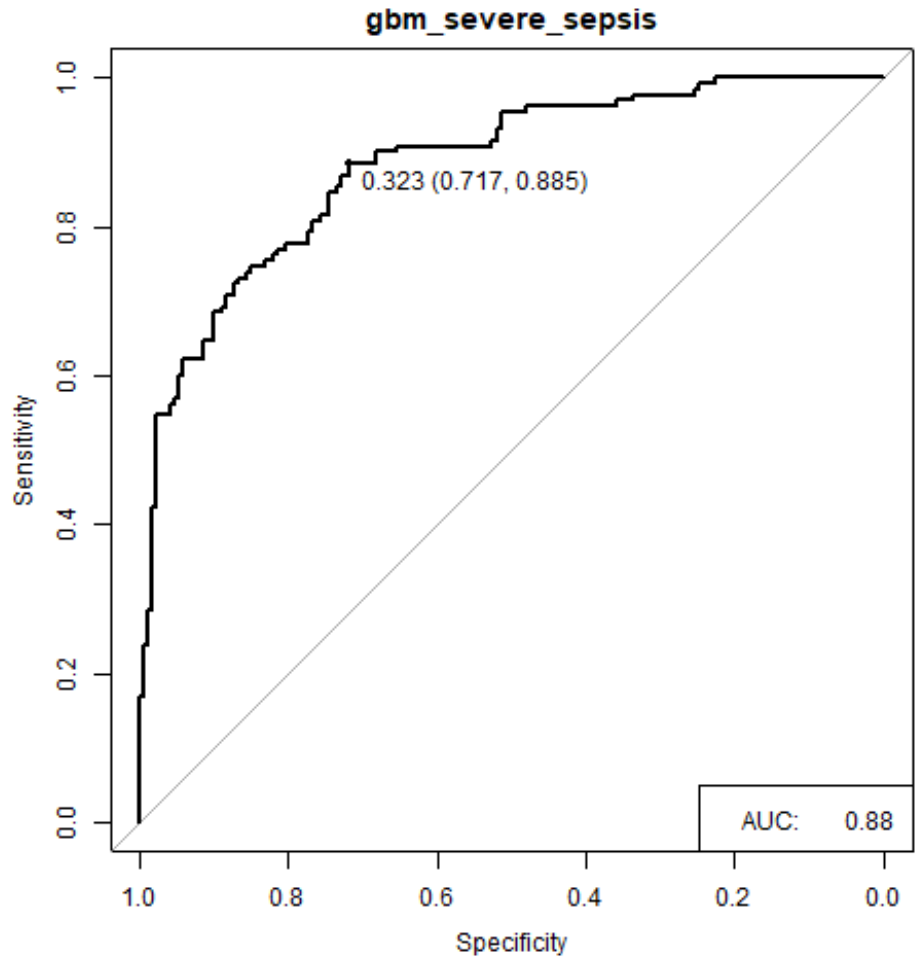


Severe Sepsis – Model Performance

Random Forest

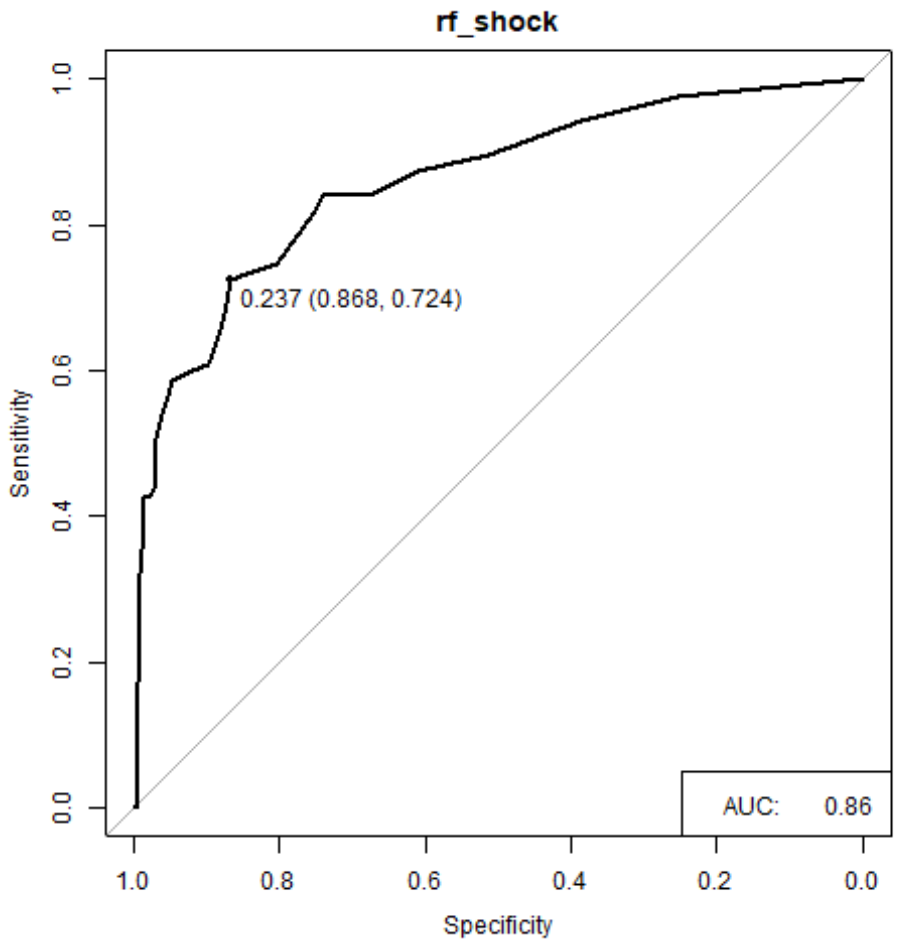


Gradient Boosting

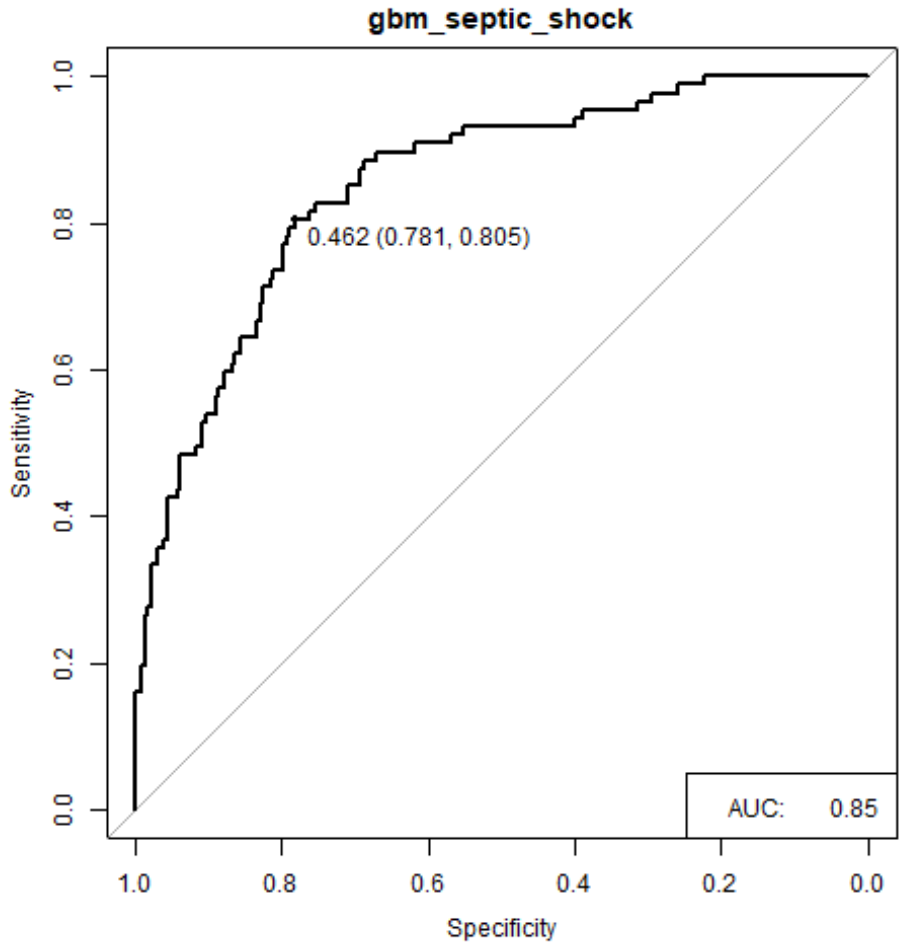


Septic Shock – Model Performance

Random Forest



Gradient Boosting



Model Comparison

Sepsis	Accuracy Rate	AUC
Naïve Bayes	0.78	0.84
Random Forest	0.77	0.84
Gradient Boosting	0.89	0.93

Severe Sepsis	Accuracy Rate	AUC
Naïve Bayes	0.68	0.72
Random Forest	0.73	0.83
Gradient Boosting	0.802	0.88

Septic Shock	Accuracy Rate	AUC
Naïve Bayes	0.82	0.80
Random Forest	0.82	0.85
Gradient Boosting	0.89	0.91