

Predicting Sepsis in ICU Patients

Earlier intervention for
better health outcomes

Problem Identification

Context

Sepsis is a leading cause of death in US hospital patients

Solution Space

Classification of pre and non sepsis patients

Success Criteria

Accurate classification of pre and non sepsis patients in test set

Data Source

Hourly data from 40,336 ICU patients in 2 hospitals

Problem Statement: Early intervention in sepsis patients can lead to better health outcomes. Is it possible to predict sepsis in ICU patients hours before clinical diagnosis?

Data Structure & Source

Time (Hours)	Vital Signs 1-8	Laboratory Values 9-34	Demographics 35-40	Sepsis Label 41
t_0	0
t_1	1
....	0
t_n	0

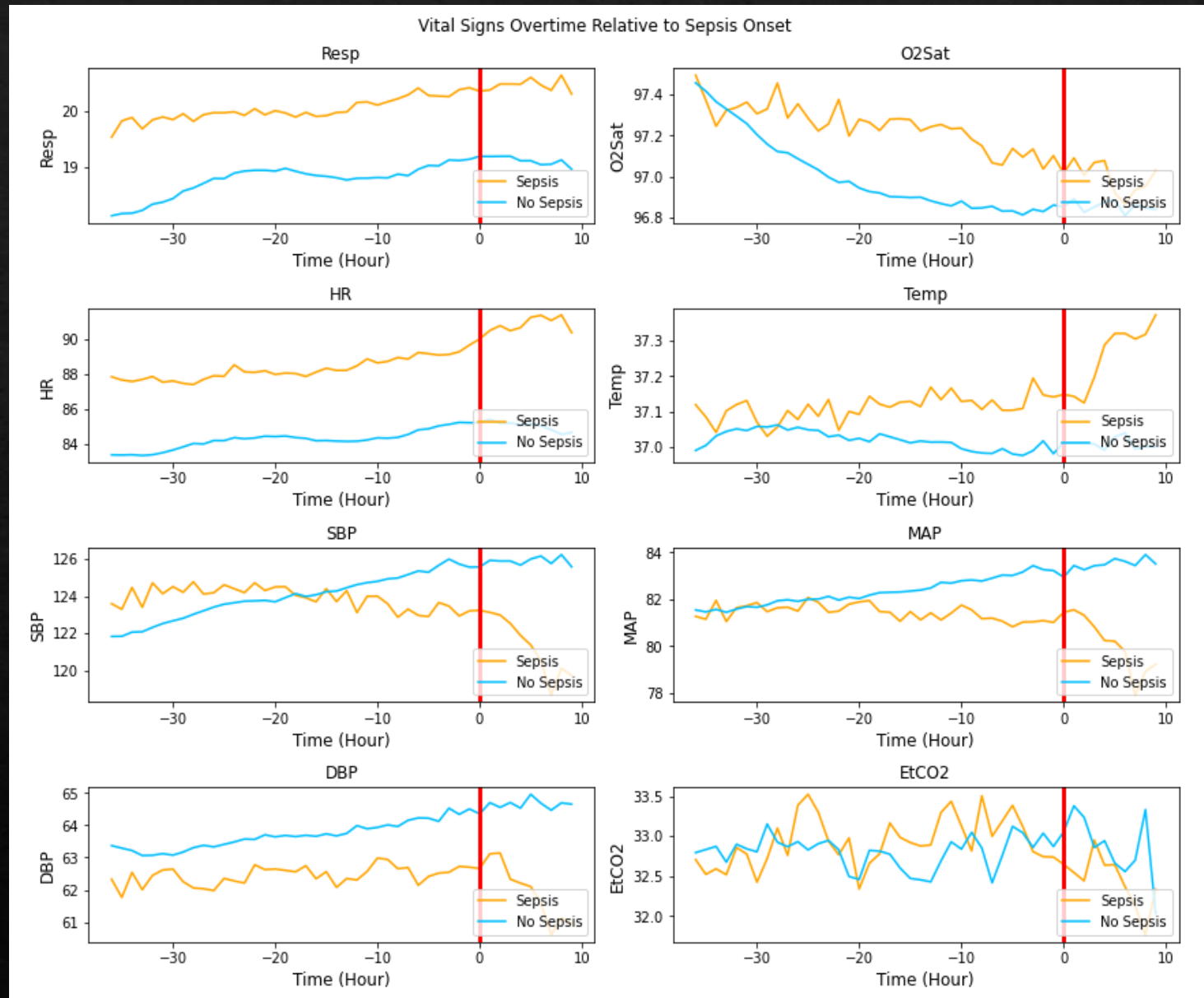
Data made available by Physionet Computing in Cardiology
Challenge 2019

Second classifier column, pre-sepsis, added

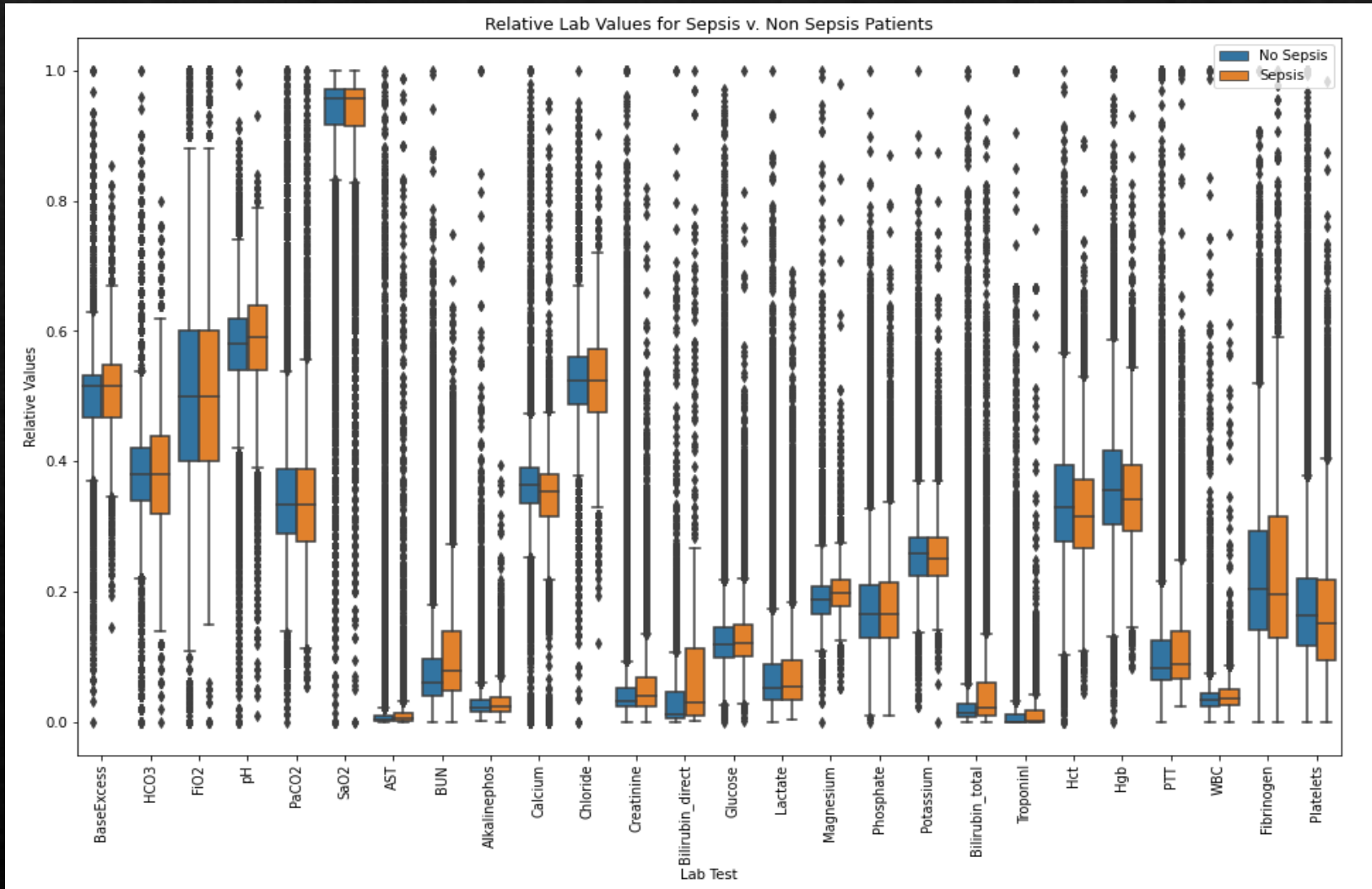
Sepsis Prevalence

Of **40,336** patients available in the data set, **7.27%** develop sepsis at some point during their hospital stay.

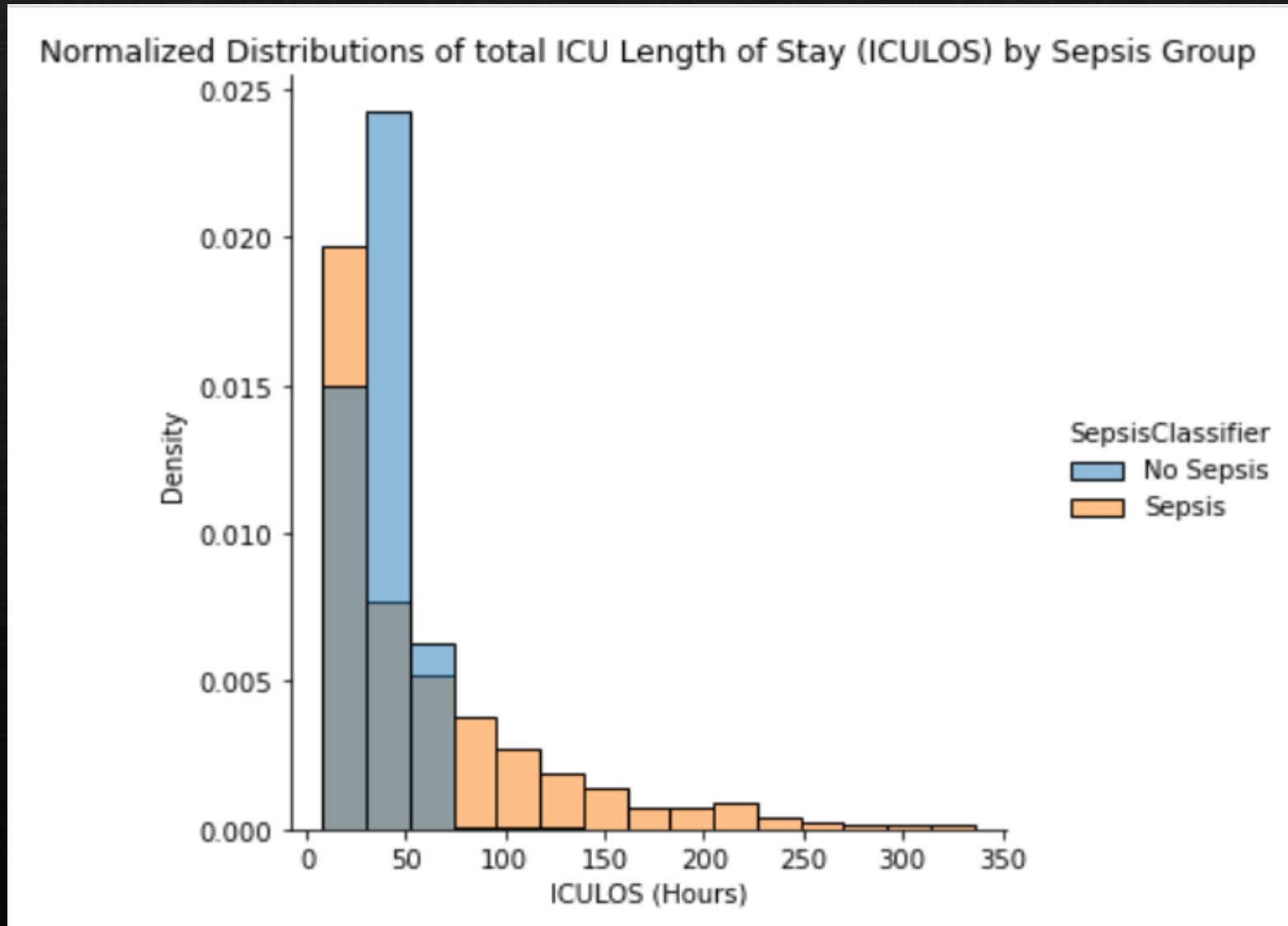
Exploratory Data Analysis: Vital Signs



Exploratory Data Analysis: Normalized Lab Values



Exploratory Data Analysis: ICU Length of Stay



Feature Engineering

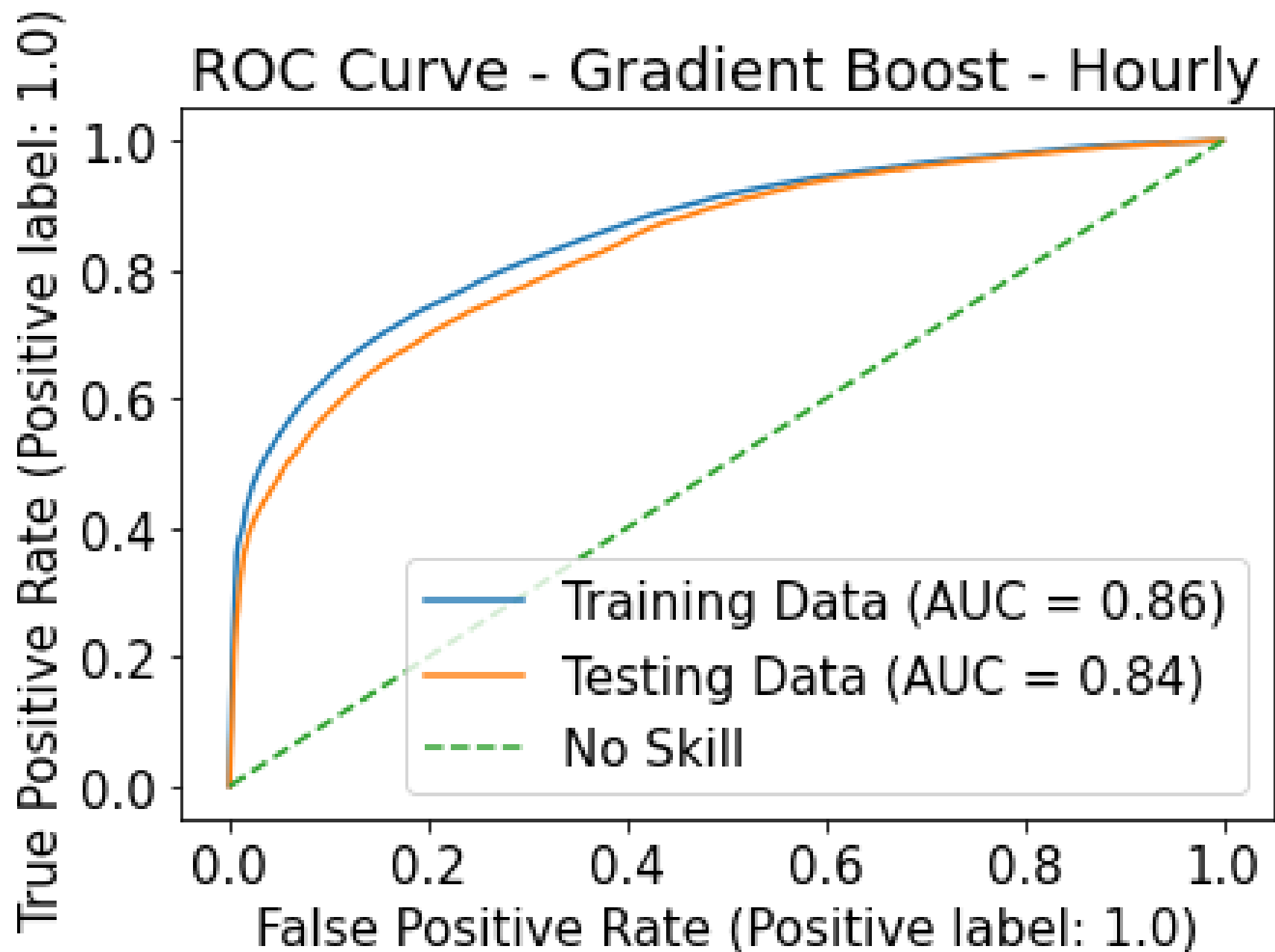
- ◆ Interpolation & forward filling
 - ◆ Lab & Vital Sign Data
- ◆ Changes in vital sign column
 - ◆ Past 1, 2, 3 Hours
- ◆ Lab value indicator, forward filled

	d3_HR	d2_HR	d1_HR	HR
0	0.0	0.0	0.0	97.0
1	0.0	0.0	0.0	97.0
2	0.0	-8.0	-8.0	89.0
3	-7.0	-7.0	1.0	90.0
4	6.0	14.0	13.0	103.0
5	21.0	20.0	7.0	110.0
6	18.0	5.0	-2.0	108.0
7	3.0	-4.0	-2.0	106.0
8	-6.0	-4.0	-2.0	104.0
9	-6.0	-4.0	-2.0	102.0
10	-2.0	0.0	2.0	104.0

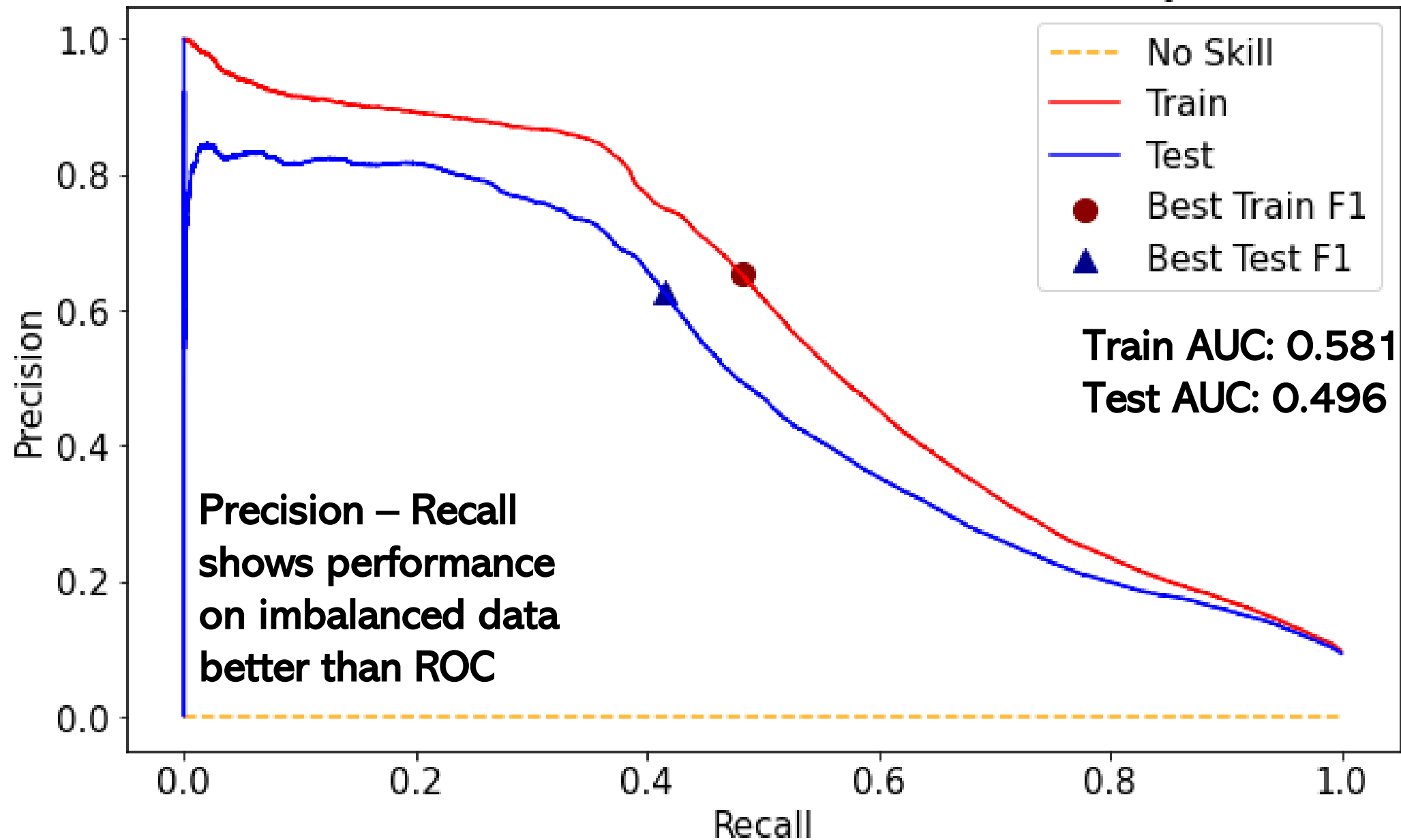
Final Model Results

Classification Scheme 1: Hourly Classification

- Hour by hour classification of a patient being in the pre-sepsis period or not.



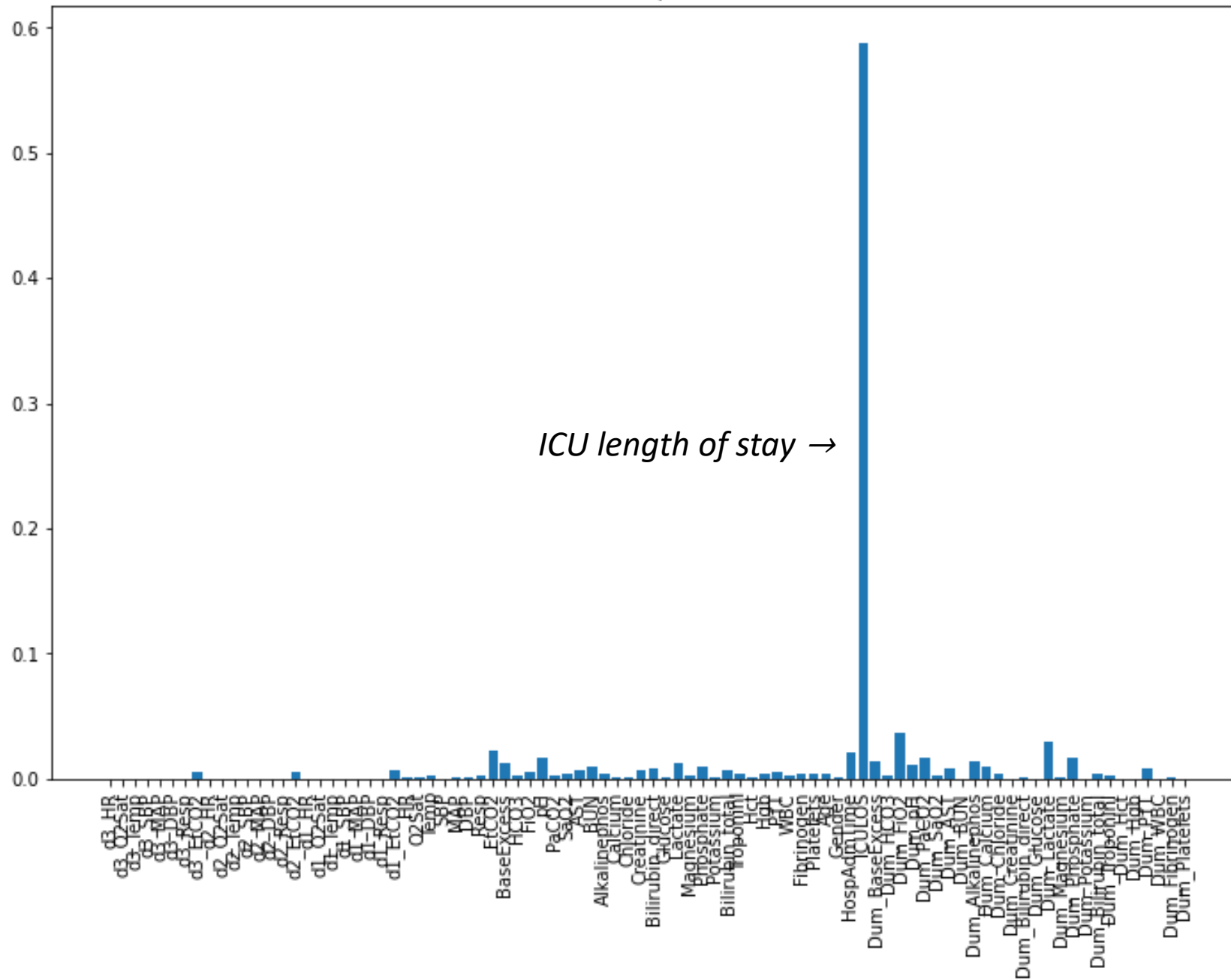
Precision- Recall - Gradient Boost - Hourly



Confusion Matrix: Hourly

Actual	Predicted					
	0	1	0	1	0	1
0	82%	17.9%	97.5%	2.5%	98.4%	1.6%
1	32.1%	67%	58.5%	41.5%	62.9%	37%
Threshold	10%		21.6%		30%	

ICU length of stay →

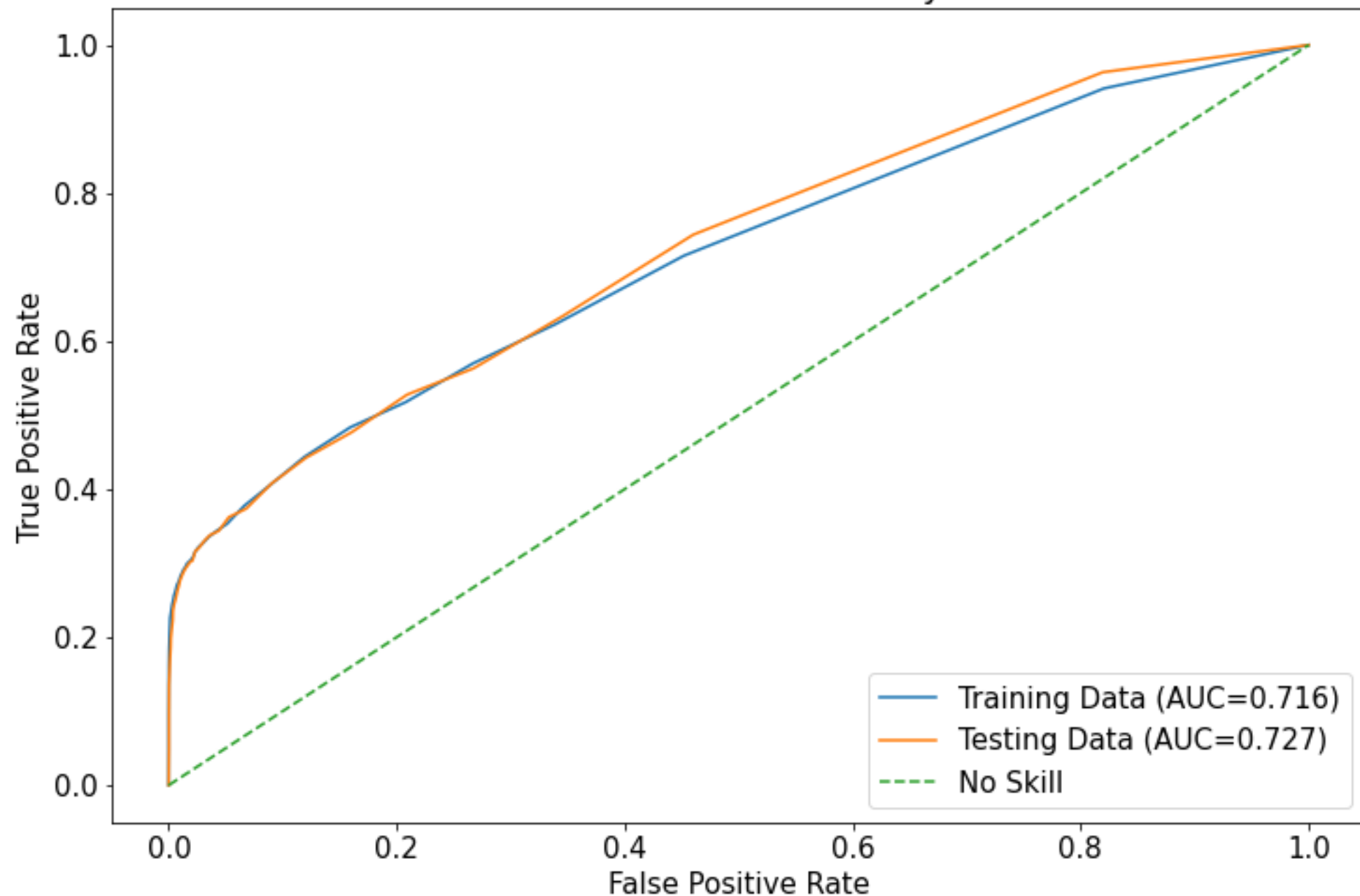


Final Model Results

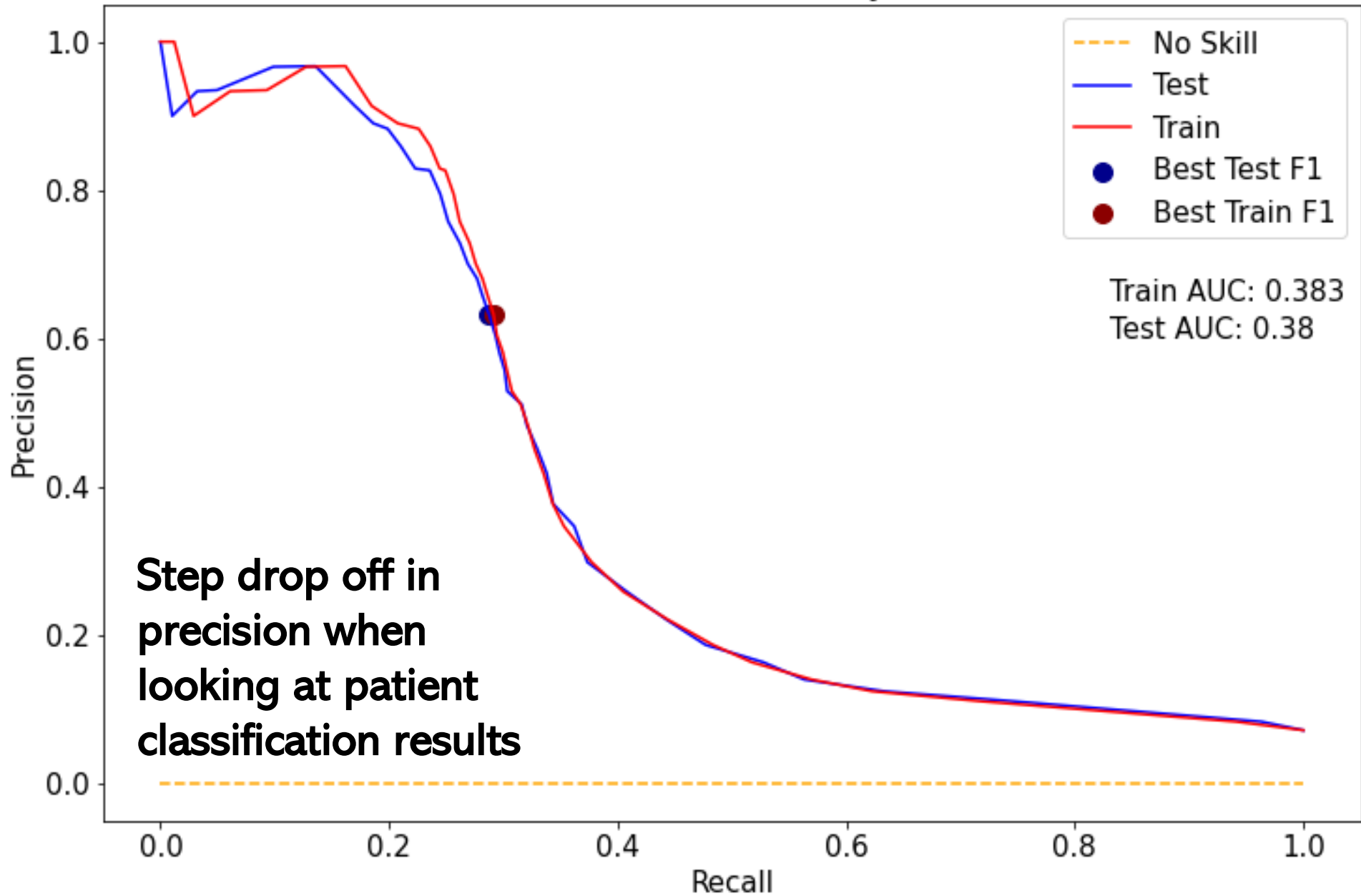
Classification Scheme 2: By-Patient Classification

- Non-sepsis patient classified with any positive hourly predictions, or a sepsis patient with positive prediction at any point in pre-sepsis period : 1
- Non-sepsis patient with no positive predictions, or a sepsis patient with positive prediction only during/after sepsis or not at all: 0

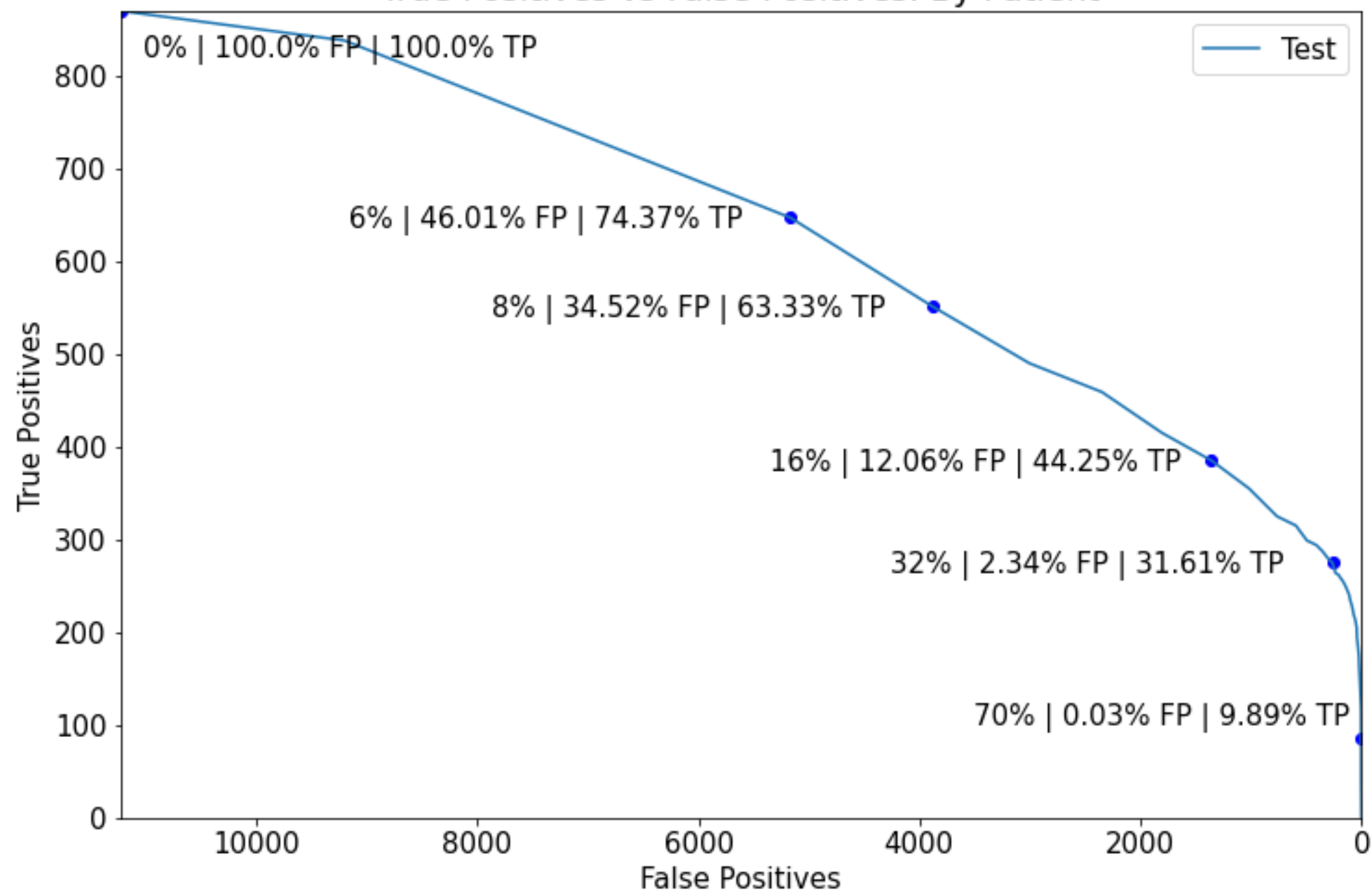
ROC Curve - Gradient Boost: By Patient



Precision-Recall Curve: By Patient



True Positives vs False Positives: By Patient



Confusion Matrix: Patient Level

Actual	Predicted					
	0	1	0	1	0	1
0	18.1%	81.9%	76%	24%	97.9%	2.1%
1	3.7%	96.3%	45.9%	54.1%	69.7%	30.3%
Threshold	4%		11%		34%	

Use Cases

The by-patient classification schema is used because it more accurately reflects how the model will perform across all patients, regardless of ICU length of stay. It is used for the following scenarios.

Sepsis Concern Eliminator

4% | TPR .96 | FPR .81

Indicates if a patient is
very unlikely to develop
sepsis

Elevated Sepsis Intervention Warning

11% | TPR .54 | FPR .24

Sepsis risk elevated;
patient should be
monitored closely

Critical Sepsis Intervention Warning

34% | TPR .30 | FPR .02

Sepsis eminent; clinician
should treat for sepsis

Conclusions & Future Directions

- ◆ Sepsis is known to manifest very differently in different people
- ◆ Nonetheless, the selected gradient boost model had skill in distinguishing groups
- ◆ The elevated sepsis intervention use case is the most feasible, because it could augment clinician's decisions, while not being relied upon too greatly
- ◆ For actual clinical use, a model that could predict sepsis onset with greater precision and within a specific time frame is needed
- ◆ Recurrent neural net models have seen more success in this, as they can process sequences of data^[1]

[1] Liu, L., Wu, H., Wang, Z., Lieu, Z, Zhang, M. Early Prediction of Sepsis From Clinical Data via Heterogeneous Event Aggregation. arXiv.org. 2019 Oct 12; 1910.06792v1.