Predicting Neurological Outcome at Discharge in Patients with Traumatic Brain Injury

Health Informatics, Spring ‘21

Louis A. Gomez

Introduction

Traumatic Brain Injury (TBI) is a severe condition due to a sudden trauma such as a bump, blow or jolt to the brain which causes disruption of the normal brain functioning[1].

To measure the severity of TBI and neurological outcomes after discharge, the Glasgow Coma Scale (GCS) has been studied

1. A. Peterson, L. Xu, J. Daugherty, and M. Breiding, "Surveillance Report of Traumatic Brain Injury related Emergency Department Visits, Hospitalizations, and Deaths -United States 2014," Center for Disease Control and Prevention, 2019.

Introduction: Glasgow Coma Scale

The GCS is a behavioral assessment comprised of three scales: motor, verbal and eye response.

Motor subcomponents (mGCS) is a proxy for neurological function[2]

Eye Opening Response

- Spontaneous--open with blinking at baseline **4 points**
- To verbal stimuli, command, speech **3 points**
- To pain only (not applied to face) **2 points**
- No response **1 point**

Verbal Response

- Oriented **5 points**
- Confused conversation, but able to answer questions **4 points**
- Inappropriate words **3 points**
- Incomprehensible speech **2 points**
- No response **1 point**

Motor Response

- Obeys commands for movement 6 points
- Purposeful movement to painful stimulus 5 points
- Withdraws in response to pain 4 points
- Flexion in response to pain (decorticate posturing) 3 points
- Extension response in response to pain (decerebrate posturing) 2 points
- No response 1 point

M. Balestreri et al., “Predictive value of Glasgow coma scale after brain trauma: change in trend over the past ten years,” p. 2.

Related Work

CRASH – Corticosteroid Randomization After Significant Head Injury

IMPACT – International Mission for Prognosis and Analysis of Clinical Trails in TBI

Research Question

Can patient data from the first two days in the ICU can be used to predict neurological outcome at discharge?

Datasets

For this class project, I used static, vital signs and lab test results data from all patients admitted to the ICU with an ICD9 coding of TBI

ICD9 Codes: 800,801,850,851,852,853,854

Inclusion Criteria: Patients with an ICD 9 coding of TBI

Exclusion Criteria:
(i) Age < 18 yrs, (ii) ICU stay duration < 48 hours, (iii) A motor GCS score is not recorded at discharge

Datasets: Glasgow Coma Scale

Based on exclusion criteria, we have 452 unique ICU admissions for TBI with mGCS at discharge

Large data imbalance – (score of 6 has 75% of labels)

Scores	Categories	Labels per category
1	No response	23
2	Extension response in pain	6
3	Flexion in response to pain	8
4	Withdraws in response to pain	25
5	Purposeful movement to localized pain	51
6	Obeys commands	339

Methodology: Data Processing

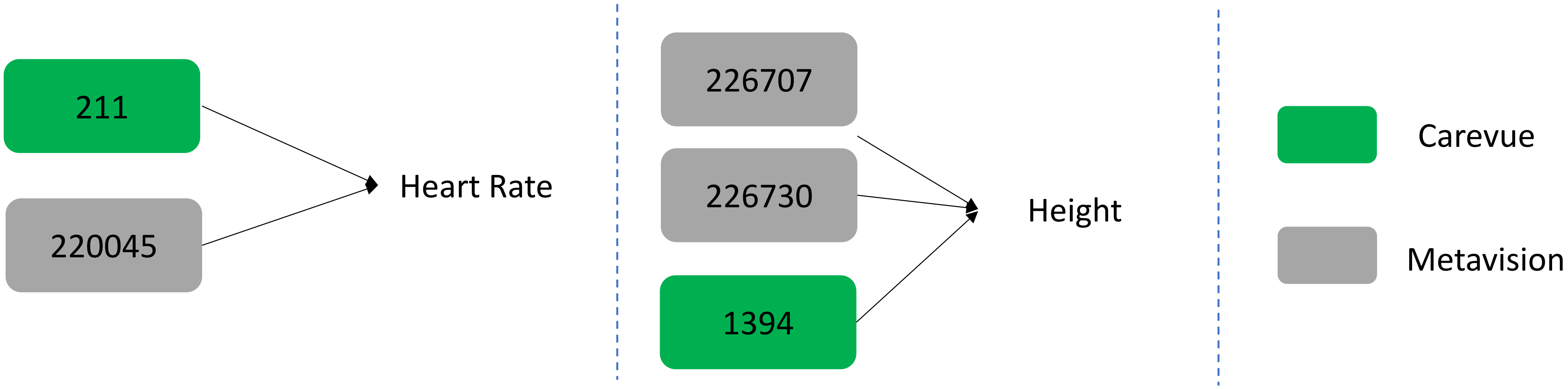
To process data, I followed the standardized approach implemented in MIMIC Extract[3]

Clinical Groupings → Unit Conversions → Data Filtering → Data Aggregation

S. Wang, M. B. A. McDermott, G. Chauhan, M. C. Hughes, T. Naumann, and M. Ghassemi, “MIMIC-Extract: A Data Extraction, Preprocessing, and Representation Pipeline for MIMIC-III,” Proc. ACM Conf. Health Inference Learn., pp. 222–235, Apr. 2020,

Clinical Groupings → Unit Conversions → Data Filtering → Data Aggregation

This involves combining the itemids that represent the same information at a high level.



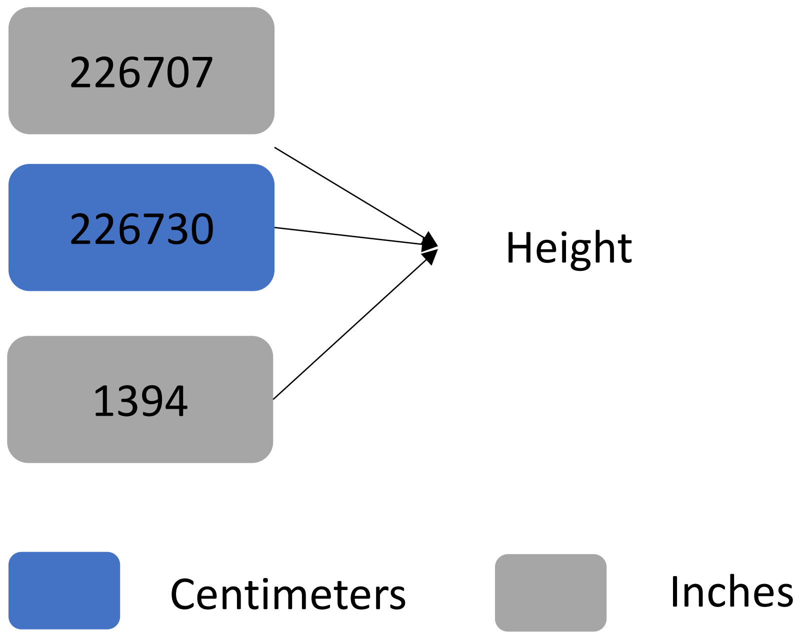
Clinical Groupings → Unit Conversions → Data Filtering → Data Aggregation

This involves converting similar itemids (defined in the previous step) to the same units.

Weight (kg, lbs., oz) → kilograms

Height (inches, cm) → Centimeters

Temperature (Celsius, Fahrenheit) → Celsius



Clinical Groupings → Unit Conversions → Data Filtering → Data Aggregation

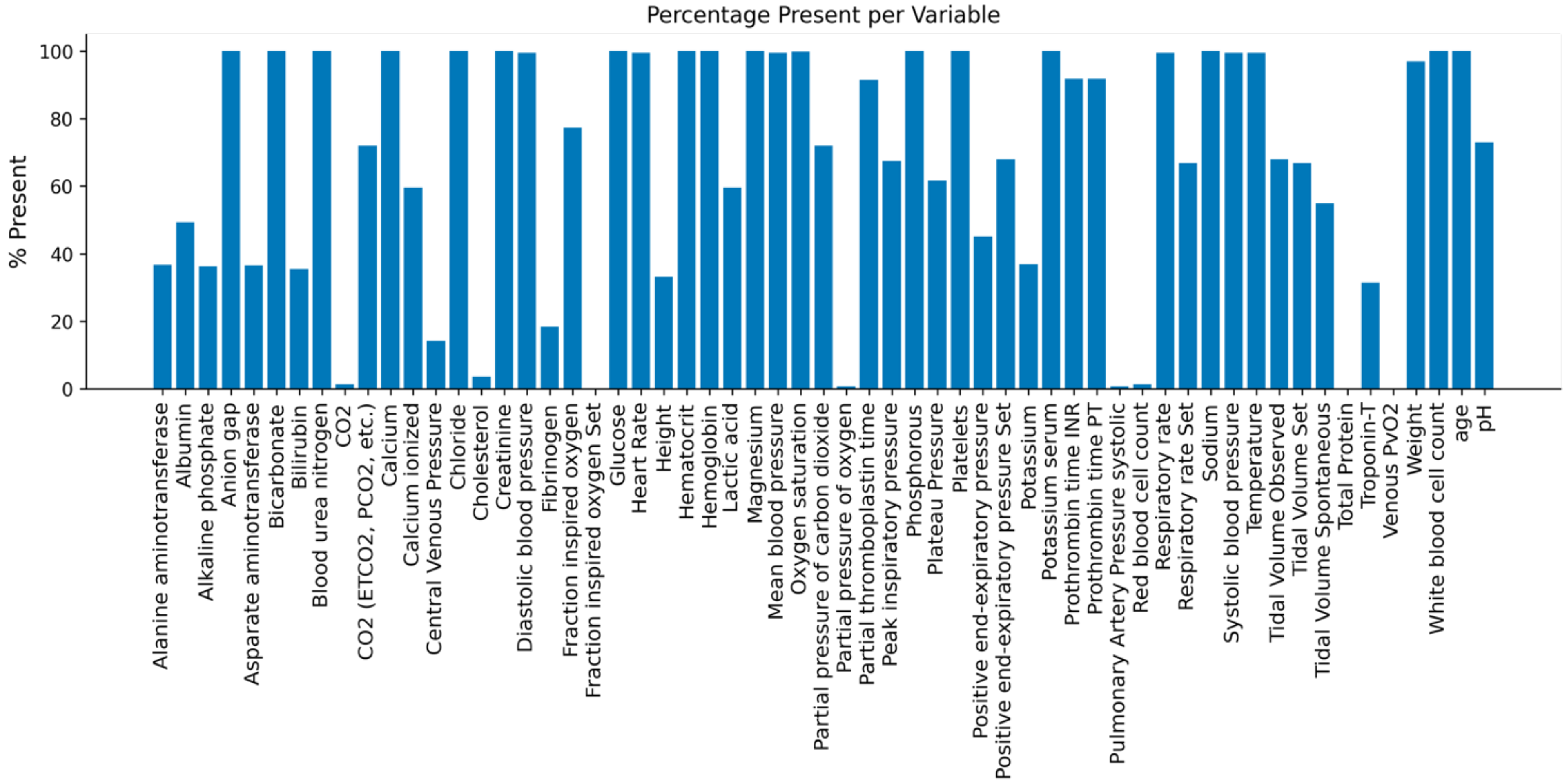
This process used recommended variable ranges and filters them.

For variable values higher (or lower) than the valid range were imputed using previous values

For variables values beyond the valid ranges (i.e outliers) were replaced with nans

Clinical Groupings → Unit Conversions → Data Filtering → Data Aggregation

I computed the average of all variables values over the two days of data for all admissions



Bar chart showing the percentage of data measured in each clinical grouping

Methodology: Data Processing

Dropped all features who had data for less than 50% of patients. 59 -> 40 features

Mean Imputation and normalization

Classification: Tasks

Classification tasks

Binary Classification Task
Non-command following (1,2,3,4,5) vs Command following (6)

339 sample in command following compared to 113 in non-command following

Classification: models

Classification Models:

Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosted Trees (XGBoost)

Training:

Split dataset into a 70/30 stratified split. Due to imbalanced dataset, added higher class weights to minority class. Default model parameters

Classification: Evaluations

Binary task

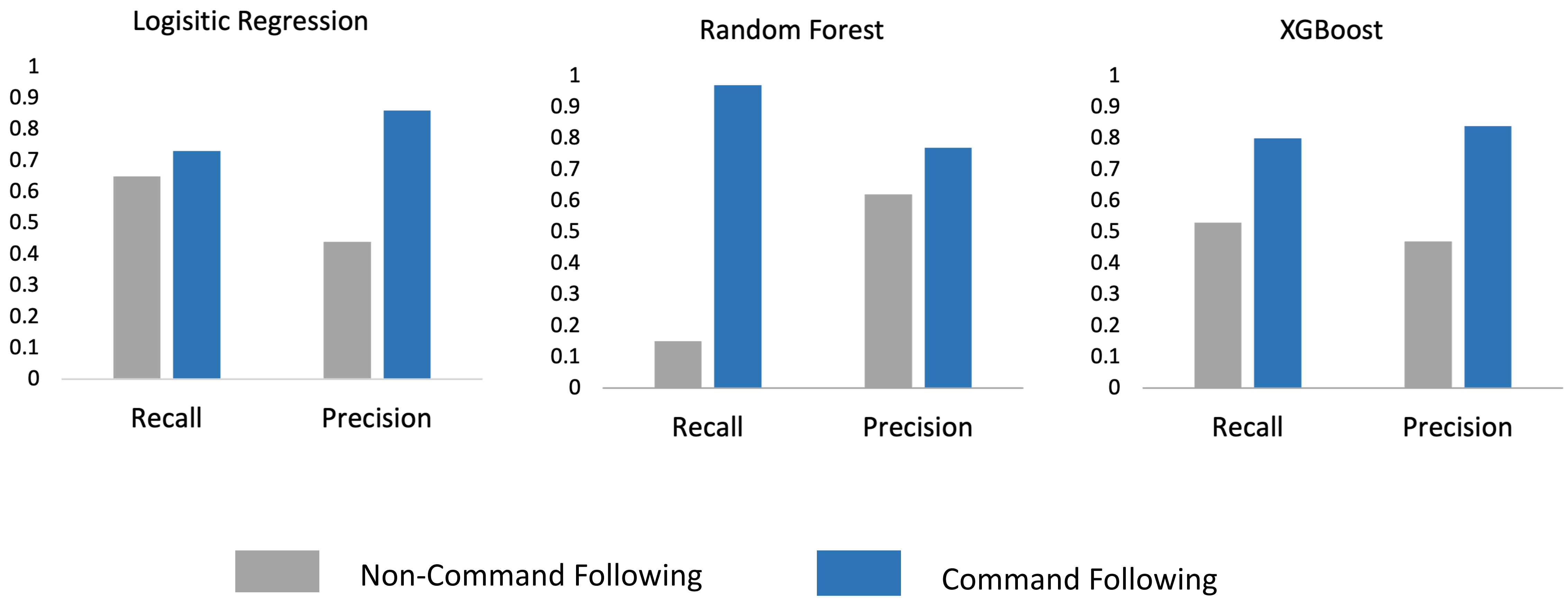
I report Precision, recall and F1-score.

Feature importance for each model.

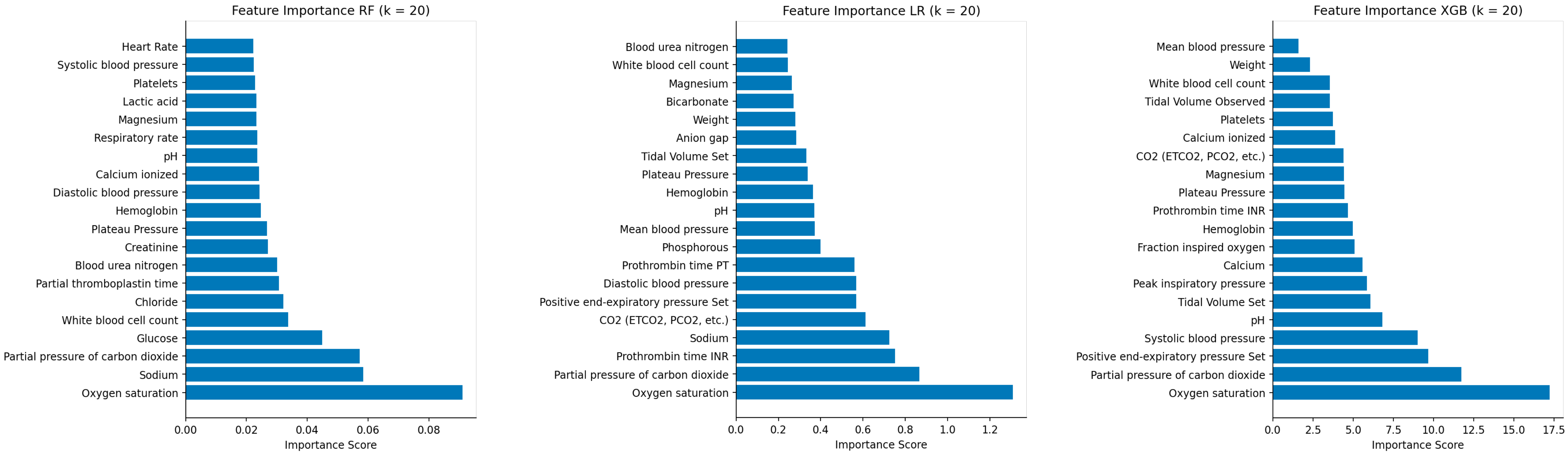
Results

Classification Models	**Evaluation Metrics**		
	Avg. Recall	Avg. Precision	Macro F1-Score
Logistic Regression	0.69	0.65	0.66
Random Forest	0.56	0.70	0.55
XGBoost	0.67	0.66	0.66

Results


Logisitic Regression
Random Forest
XGBoost
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Recall
Precision
Non-Command Following
Command Following

Results: Feature Importance (top 20)



Seven features in common across all models:
pH, white blood cell count, magnesium, partial pressure of carbon dioxide, oxygen saturation, plateau pressure, and hemoglobin

Limitations

Naïve simple mean imputation – future work could study impact of imputation of predictive performance

Small cohort of only 452 admissions – larger dataset is needed to learn a more general model

Imbalanced data approaches – other techniques to either oversample minority or under-sample majority classes

Future work could include other data modalities – clinical notes, drugs, and waveform data

Questions?