Budget Constrained Machine Learning for Early Prediction of Adverse Outcomes for COVID-19 Patients

DSI Workshop on Translational AI for Healthcare
March 18, 2021

Sam Nguyen & Ryan Chan
Machine Learning Engineers
Computational Engineering Division / Machine Learning Group

Lawrence Livermore National Laboratory

LLNL Laboratory Directed Research and Development (LDRD)
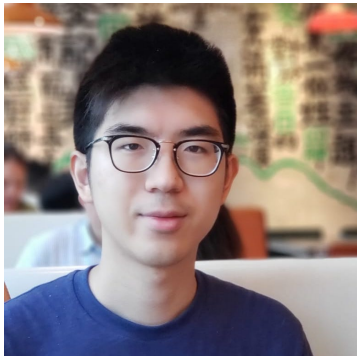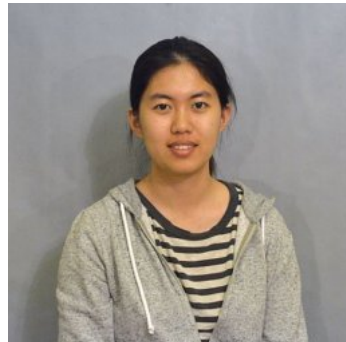## *Temporal Multimodal Learning*

Braden Soper

Jose Cadena

Sam Nguyen

Ryan Chan

Boya Zhang


Priyadip Ray

External Collaborators

The Collaboration with Univ. of Toledo and ProMedica

- Expertise in applying Machine Learning to Electronic Health Records

- Access to unique HIPAA compliant computing cluster for sensitive data

- Subject matter, data and clinical expertise

- Access to unique dataset consisting of EHRs of all patients admitted to ProMedica facilities for a COVID-19 infection
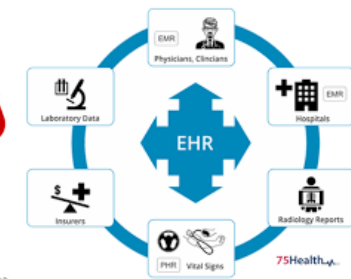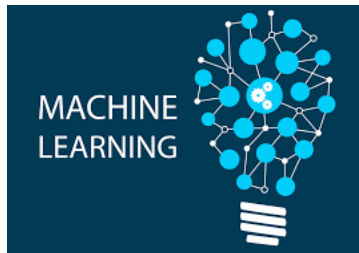
The Data

- ProMedica is the largest health care system in northwest Ohio and southeastern Michigan

- EHRs of all patients admitted to ProMedica facilities for a COVID-19 infection (positive nasopharyngeal PCR test for SARS-CoV-2)

- Data collected between 3/20/2020 and 12/29/2020

- Total of 1312 patients after various preprocessing requirements

Demographics	**Comorbidities**	**Vital Measurements**	**Lab results**
• Age • Gender • Race • Body Mass Index • …	• Asthma • Chronic Kidney Disease • Hypertension • Diabetes • …	• Respiratory Rate • %O2 Saturation • Blood pressure • Temperature • …	• Procalcitonin • LDH • D-dimer • Hemoglobin • …

Budget Constrained Machine Learning for COVID-19

- There have been a plethora of publications on **machine learning applications the COVID-19 pandemic**

- Many of these methods have focused on **static supervised learning methods** for predicting adverse outcomes

> *Identify various risk factors, e.g., obesity or age, that are highly correlated with adverse outcomes*

- Little work has been done on exploring the tradeoffs between budgets and machine learning utility

> ## *Our Motivation*
>
> - *It has been reported that overusing of labwork etc.: **TODO (from paper)***
> - *A better understanding of tradeoffs between budgets and ML utility would lead to better healthcare quality and efficiency*

The Task

Demographics

- Age
- Gender
- Race
- Body Mass Index
- …

Vital Measurements

- Respiratory Rate
- %O2 Saturation
- Blood pressure
- Temperature
- …

Comorbidities

- Asthma
- Chronic Kidney Disease
- Hypertension
- Diabetes
- …

Lab results

- Procalcitonin
- LDH
- D-dimer
- Hemoglobin
- …

The Task

Demographics

- Age
- Gender
- Race
- Body Mass Index
- ...

Vital Measurements

- Respiratory Rate
- %O2 Saturation
- Blood pressure
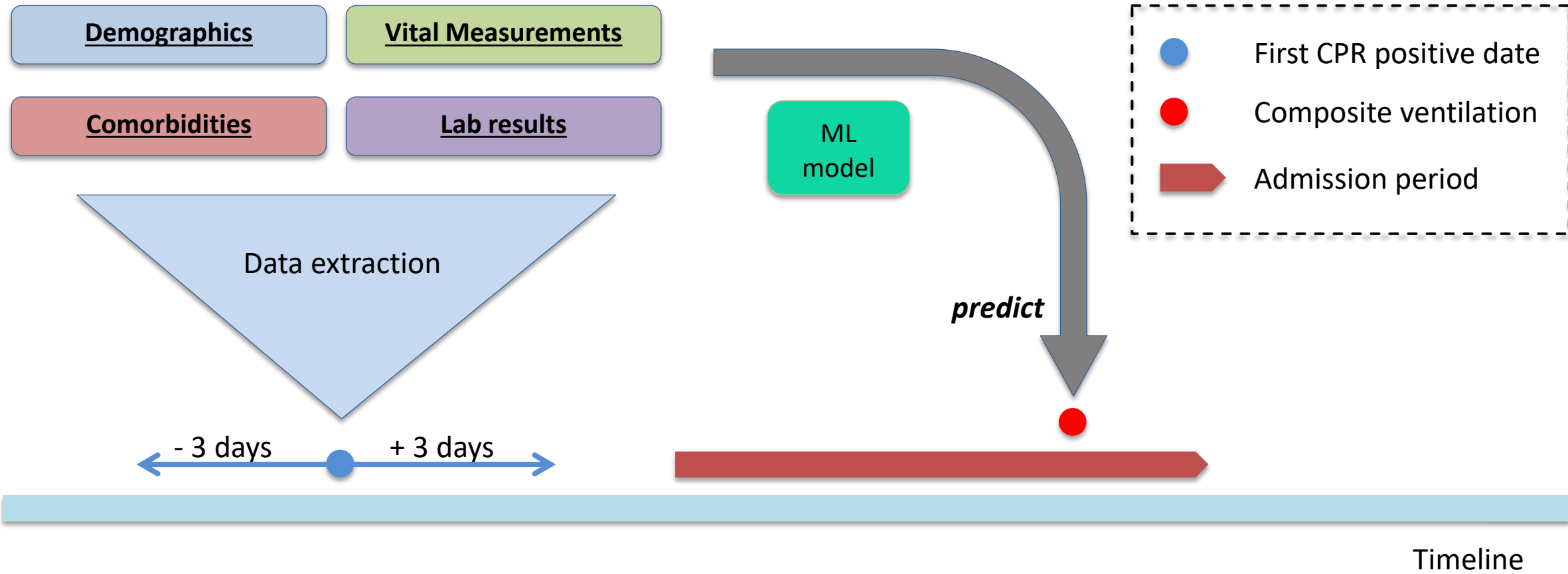- Temperature
- ...

Comorbidities

- Asthma
- Chronic Kidney Disease
- Hypertension
- Diabetes
- ...

Lab results

- Procalcitonin
- LDH
- D-dimer
- Hemoglobin
- ...

MENTION MISSING DATA

The Task

Baseline Results

Model	Outcome	AP mean (std)	AUC mean (std)
XGBoost	Composite Ventilation	0.379 (0.038)	0.723 (0.038)
Gaussian Process	Composite Ventilation	0.357 (0.037)	0.713 (0.023)
Logistic classifier	Composite Ventilation	0.361 (0.035)	0.717 (0.013)
XGBoost	Mortality	0.354 (0.065)	0.802 (0.029)
Gaussian Process	Mortality	0.338 (0.055)	0.819 (0.012)
Logistic classifier	Mortality	0.328 (0.067)	0.805 (0.029)

Baseline Results

| Model	Outcome	AP mean (std)	AUC mean (std)
XGBoost	Composite Ventilation	0.379 (0.038)	0.723 (0.038)
Gaussian Process	Composite Ventilation	0.357 (0.037)	0.713 (0.023)
Logistic classifier	Composite Ventilation	0.361 (0.035)	0.717 (0.013)
XGBoost	Mortality	0.354 (0.065)	0.802 (0.029)
Gaussian Process	Mortality	0.338 (0.055)	0.819 (0.012)
Logistic classifier	Mortality	0.328 (0.067)	0.805 (0.029)

MENTION HOW AP IS BETTER SUITED

Looking Into the Black Box: SHAP
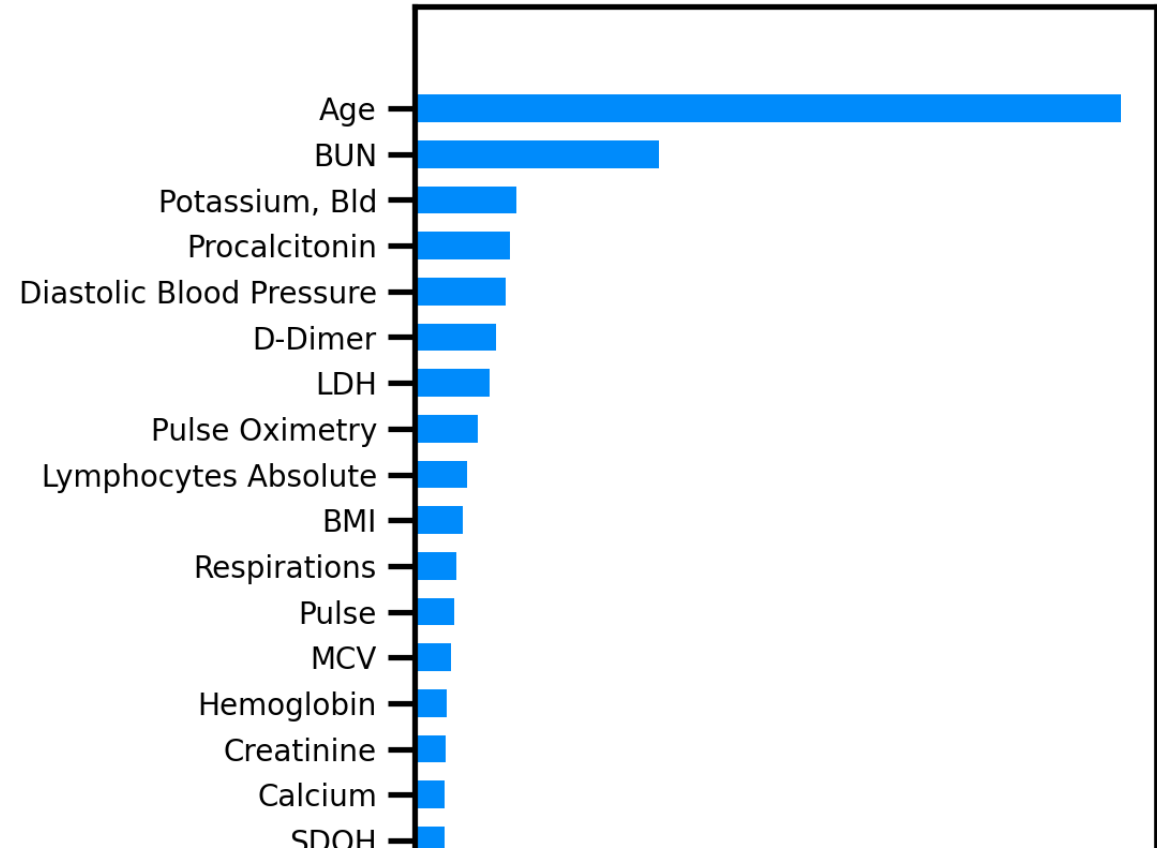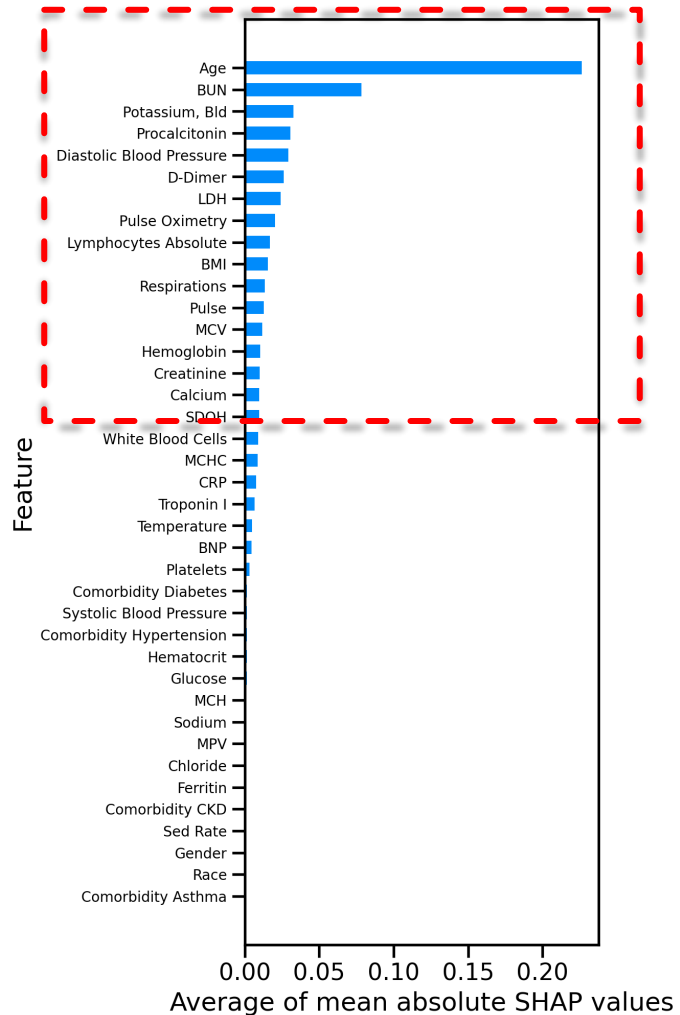
TALK ABOUT SHAP

Looking Into the Black Box: SHAP
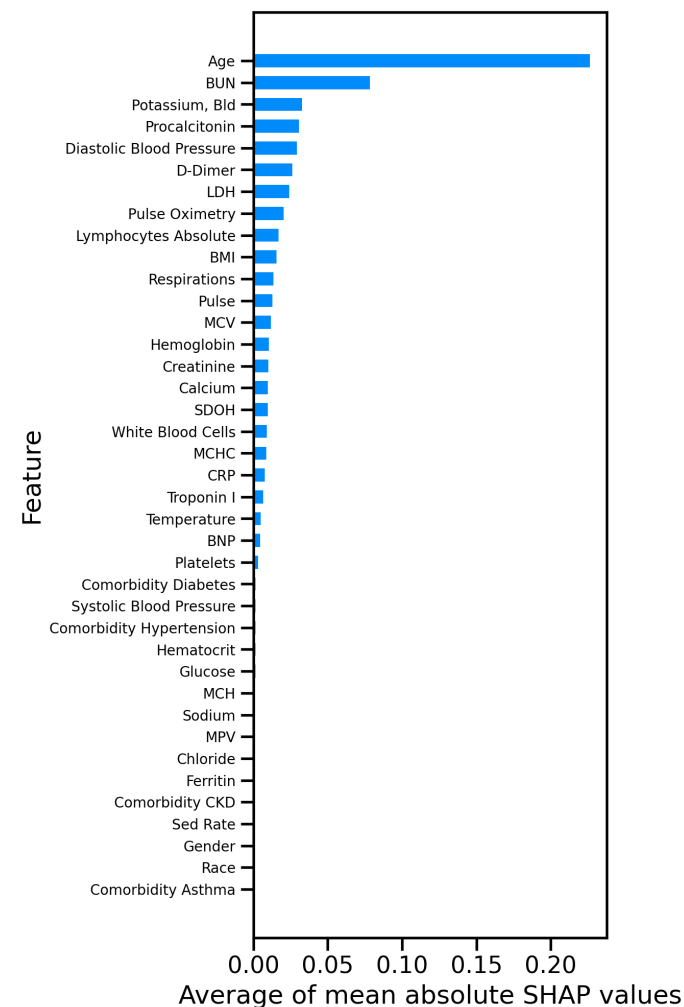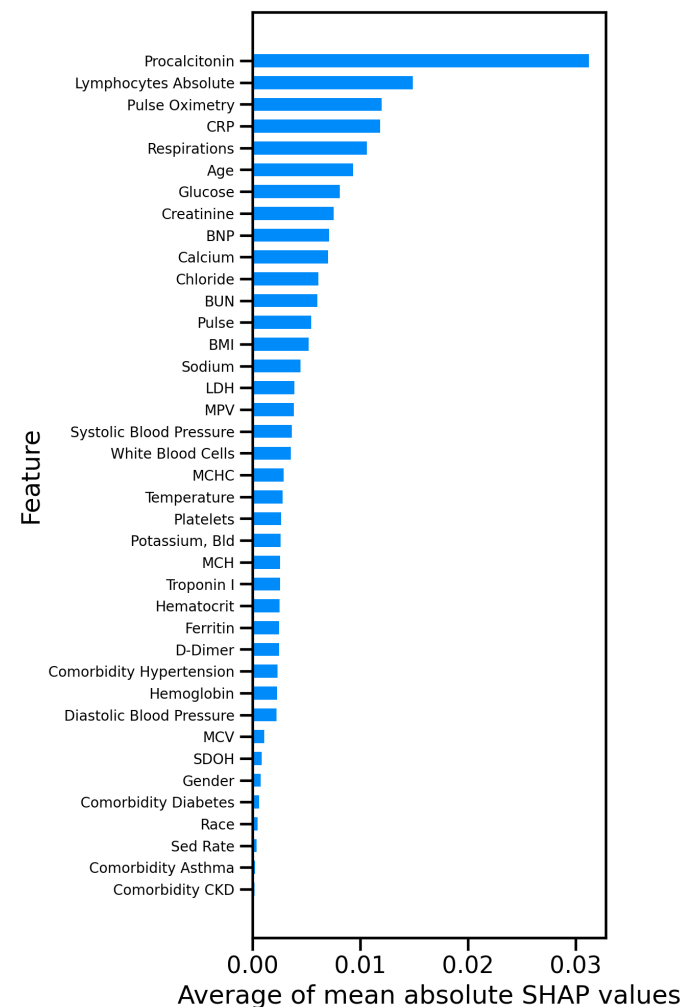
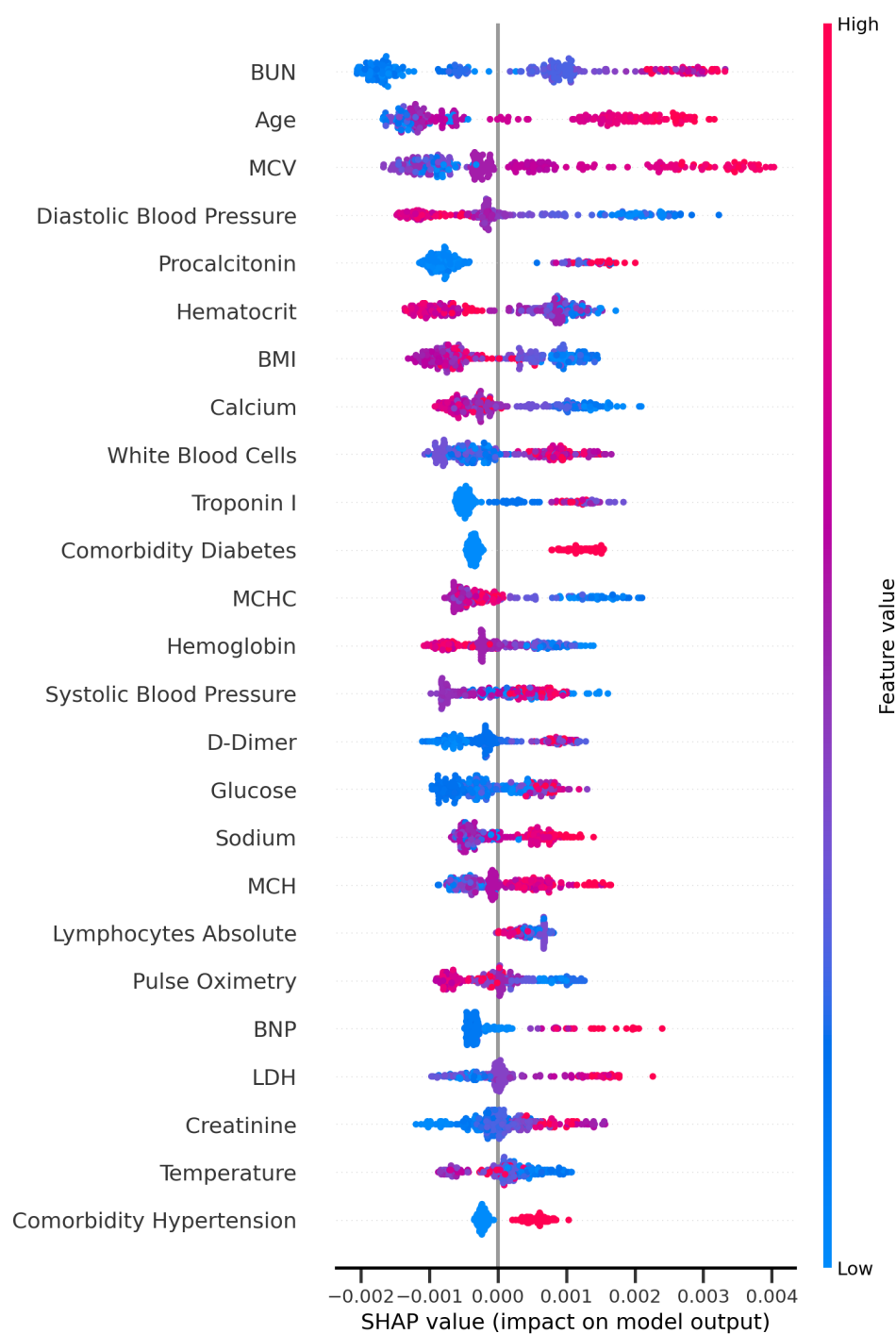# Looking Into the Black Box: SHAP for Composite Ventilation
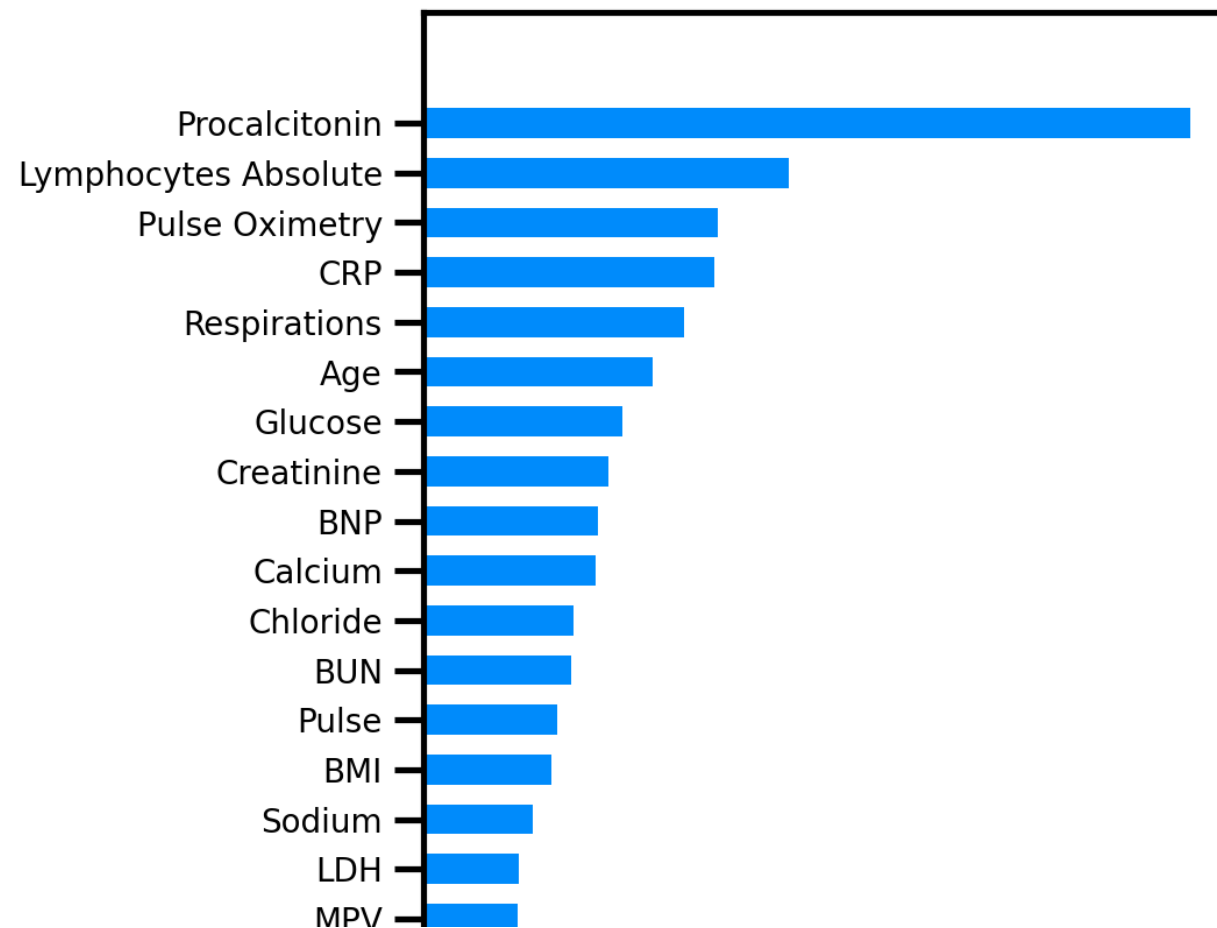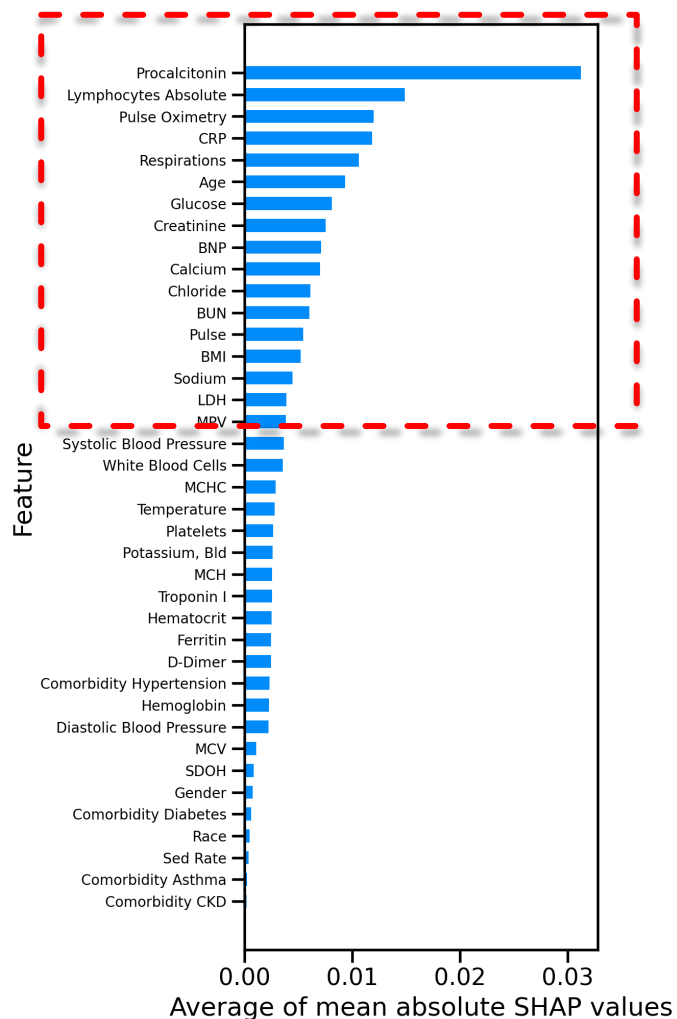
Looking Into the Black Box: SHAP for Mortality
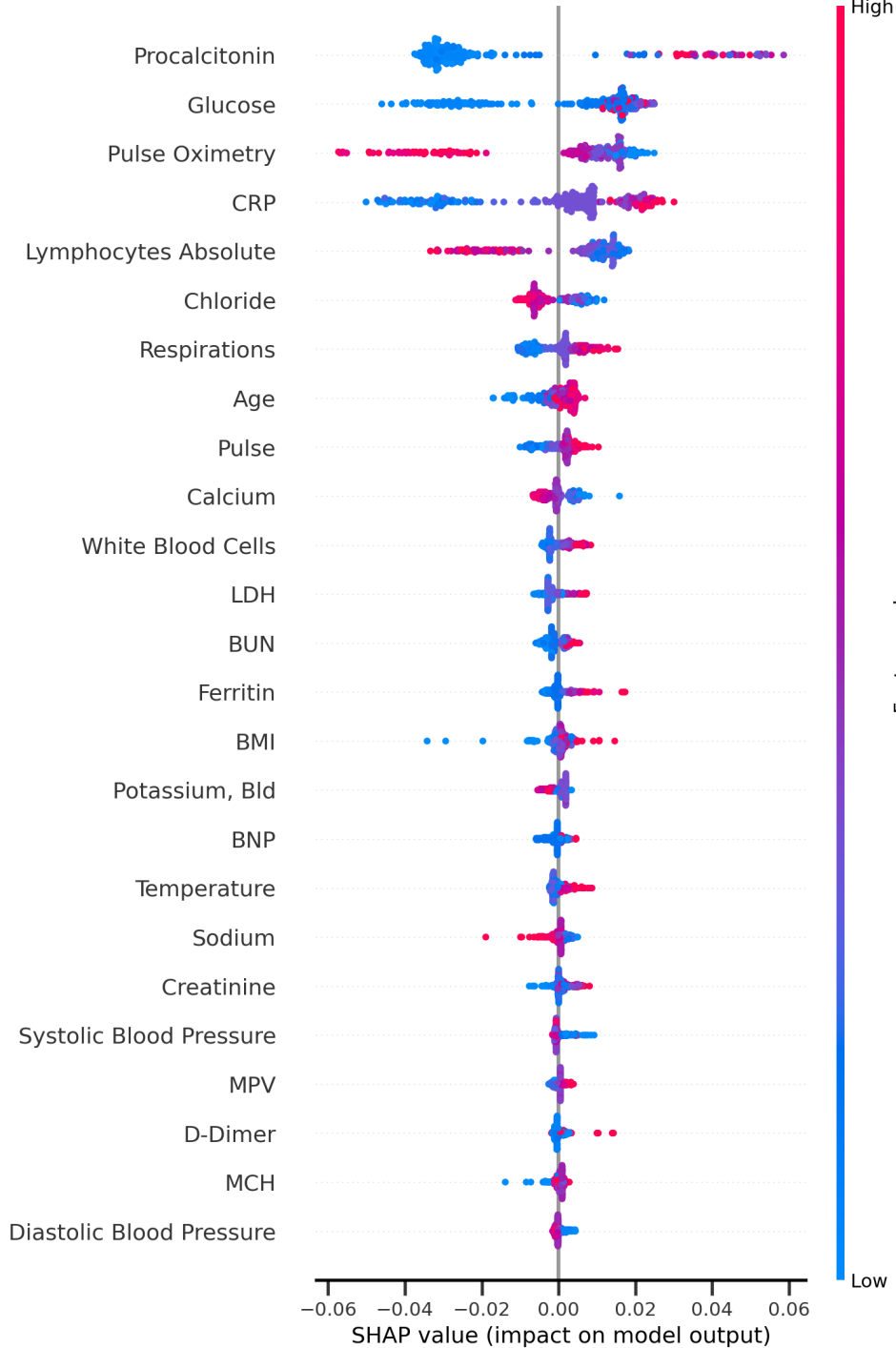
Looking Into the Black Box: SHAP

Budget Constrained Machine Learning

Group	Cost
DCM (Demographic and comorbidities: Age, Sex, Gender, Race, Respiratory Rate, BMI, Temperature, Systolic/Diastolic Blood Pressure, Pulse Oximetry, FIO2, O2 Sat, SPO2, Asthma, Chronic Kidney Disease, Diabetes, Hypertension)	0
BMP (Basic Metabolic Profile: Sodium, Chloride, Glucose, GFR MDRD Non Af Amer, Creatinine, BUN, Calcium, Potassium, Bld)	1
CBC (Complete Blood Count: Hemoglobin, Lymphocyte, MCH, Hematocrit, White Blood Cells, Platelets, MCHC, MCV, MPV)	1
D-Dimer	2
LDH	2
Sed rate	2
CRP	2
BNP	3
Troponin	3
Procalcitonin	3
Ferritin	3

Table 3: Pre-defined cost structure for each group of features (may be modified based on the healthcare facility) based on a relative cost score where 0 indicates clinical features that are least expensive and 3 indicates clinical features that are most expensive.

Budget Constrained Machine Learning

Group	Cost
DCM (Demographic and comorbidities: Age, Sex, Gender, Race, Respiratory Rate, BMI, Temperature, Systolic/Diastolic Blood Pressure, Pulse Oximetry, FIO2, O2 Sat, SPO2, Asthma, Chronic Kidney Disease, Diabetes, Hypertension)	0
BMP (Basic Metabolic Profile: Sodium, Chloride, Glucose, GFR MDRD Non Af Amer, Creatinine, BUN, Calcium, Potassium, Bld)	1
CBC (Complete Blood Count: Hemoglobin, Lymphocyte, MCH, Hematocrit, White Blood Cells, Platelets, MCHC, MCV, MPV)	1
D-Dimer	2
LDH	2
Sed rate	2
CRP	2
BNP	3
Troponin	3
Procalcitonin	3
Ferritin	3

This cost structure can be modified

Table 3: Pre-defined cost structure for each group of features (may be modified based on the healthcare facility) based on a relative cost score where 0 indicates clinical features that are least expensive and 3 indicates clinical features that are most expensive.
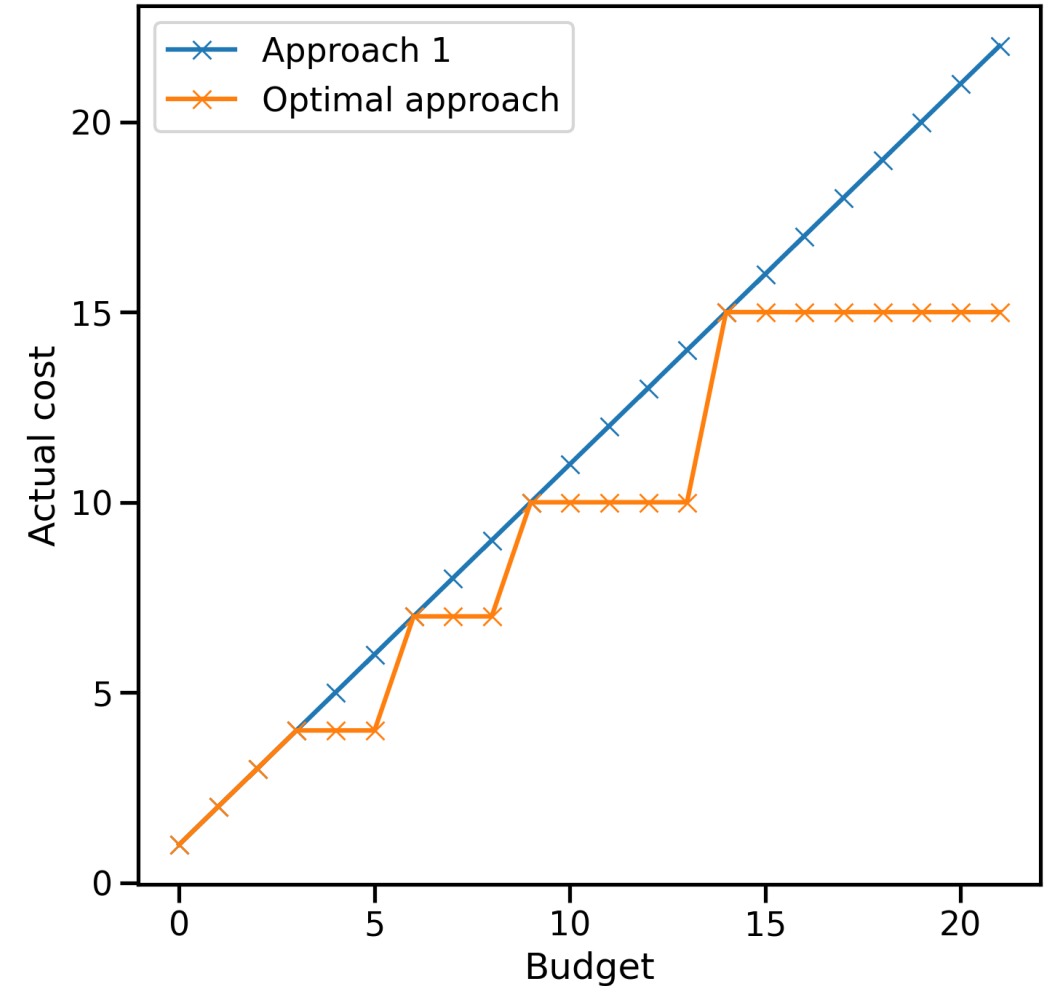
Budget Constrained Machine Learning

Budget	Total Cost	Optimal Set of Features	AP mean (std)
0	0	DCM	0.279 (0.049)
1	1	DCM, BMP	0.289 (0.016)
2	2	DCM, CRP	0.314 (0.071)
3	3	DCM, BMP, LDH	0.331 (0.034)
4	4	DCM, BMP, CBC, CRP	0.353 (0.045)
5	4	DCM, BMP, CBC, CRP	0.353 (0.045)
6	4	DCM, BMP, CBC, CRP	0.353 (0.045)
7	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
8	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
9	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
10	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
11	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
12	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
13	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
14	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
15	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
16	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
17	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
18	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
19	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
20	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
21	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
22	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)

Table 6: Optimal set of features at different budget levels for predicting need of ventilation. Total Costs is the sum of the cost of each feature group in the selected set. The cost of each feature group is listed in Table 3.

Budget Constrained Machine Learning

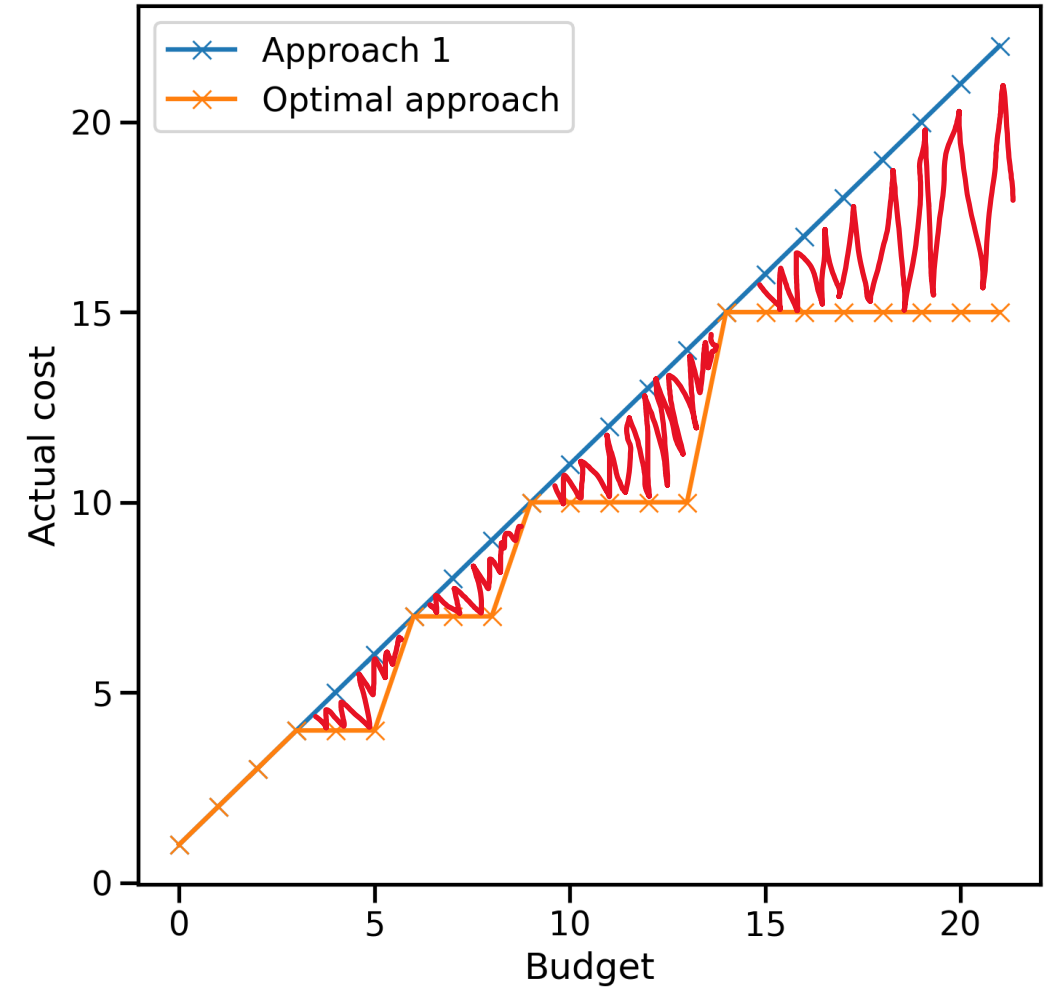| Budget	Total Cost	Optimal Set of Features	AP mean (std)
0	0	DCM	0.279 (0.049)
1	1	DCM, BMP	0.289 (0.016)
2	2	DCM, CRP	0.314 (0.071)
3	3	DCM, BMP, LDH	0.331 (0.034)
4	4	DCM, BMP, CBC, CRP	0.353 (0.045)
5	4	DCM, BMP, CBC, CRP	0.353 (0.045)
6	4	DCM, BMP, CBC, CRP	0.353 (0.045)
7	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
8	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
9	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
10	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
11	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
12	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
13	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
14	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
15	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
16	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
17	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
18	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
19	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
20	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
21	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
22	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)

Table 6: Optimal set of features at different budget levels for predicting need of ventilation. Total Costs is the sum of the cost of each feature group in the selected set. The cost of each feature group is listed in Table 3.

Budget Constrained Machine Learning

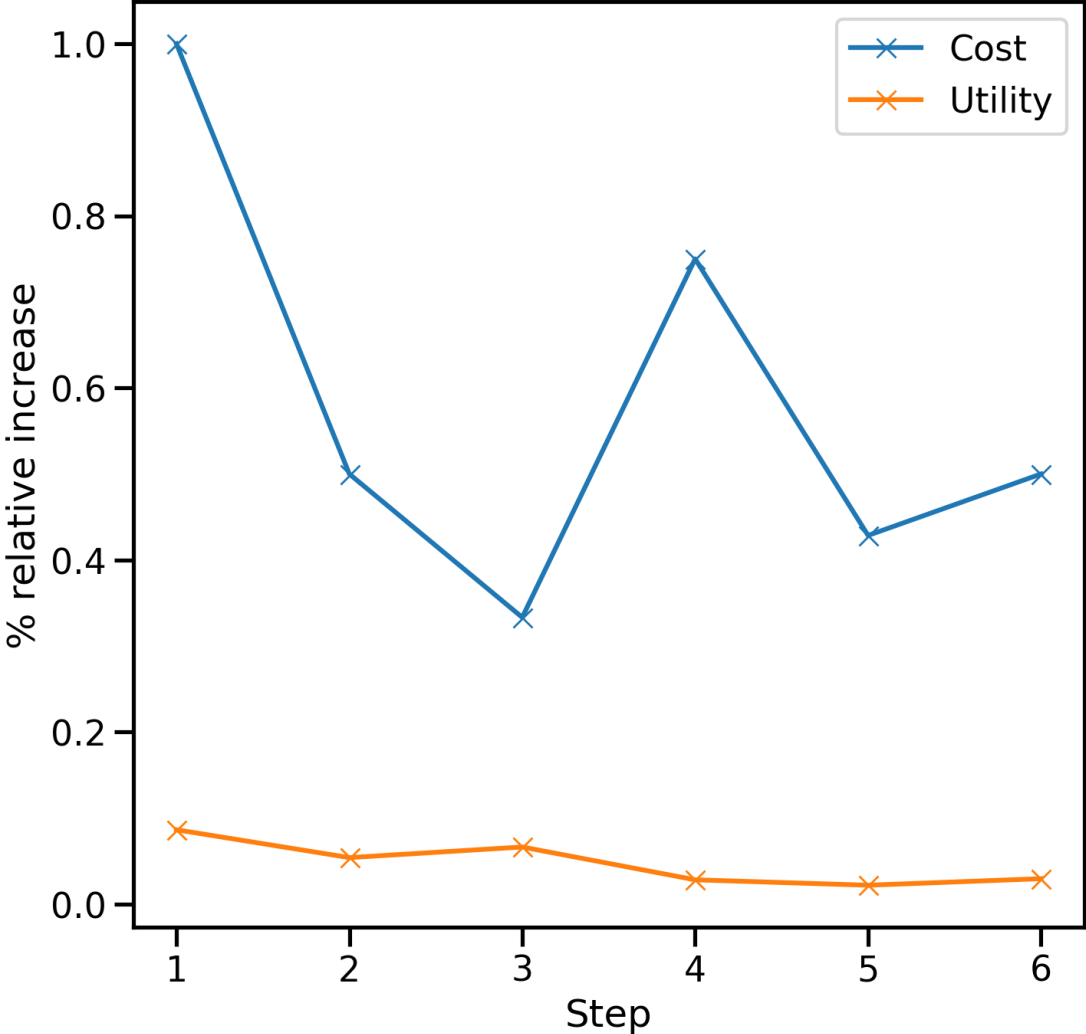| Budget	Total Cost	Optimal Set of Features	AP mean (std)
0	0	DCM	0.279 (0.049)
1	1	DCM, BMP	0.289 (0.016)
2	2	DCM, CRP	0.314 (0.071)
3	3	DCM, BMP, LDH	0.331 (0.034)
4	4	DCM, BMP, CBC, CRP	0.353 (0.045)
5	4	DCM, BMP, CBC, CRP	0.353 (0.045)
6	4	DCM, BMP, CBC, CRP	0.353 (0.045)
7	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
8	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
9	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
10	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
11	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
12	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
13	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
14	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
15	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
16	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
17	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
18	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
19	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
20	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
21	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
22	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)

Table 6: Optimal set of features at different budget levels for predicting need of ventilation. Total Costs is the sum of the cost of each feature group in the selected set. The cost of each feature group is listed in Table 3.

Budget Constrained Machine Learning

Budget	Total Cost	Optimal Set of Features	AP mean (std)
0	0	DCM	0.279 (0.049)
1	1	DCM, BMP	0.289 (0.016)
2	2	DCM, CRP	0.314 (0.071)
3	3	DCM, BMP, LDH	0.331 (0.034)
4	4	DCM, BMP, CBC, CRP	0.353 (0.045)
5	4	DCM, BMP, CBC, CRP	0.353 (0.045)
6	4	DCM, BMP, CBC, CRP	0.353 (0.045)
7	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
8	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
9	7	DCM, BMP, CBC, D-dimer, Procalcitonin	0.363 (0.051)
10	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
11	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
12	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
13	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
14	10	DCM, BMP, CBC, CRP, BNP, Troponin	0.371 (0.030)
15	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
16	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
17	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
18	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
19	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
20	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
21	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)
22	15	DCM, BMP, D-dimer, LDH, Sedrate, CRP, BNP, Procalcitonin	0.382 (0.047)

Table 6: Optimal set of features at different budget levels for predicting need of ventilation. Total Costs is the sum of the cost of each feature group in the selected set. The cost of each feature group is listed in Table 3.
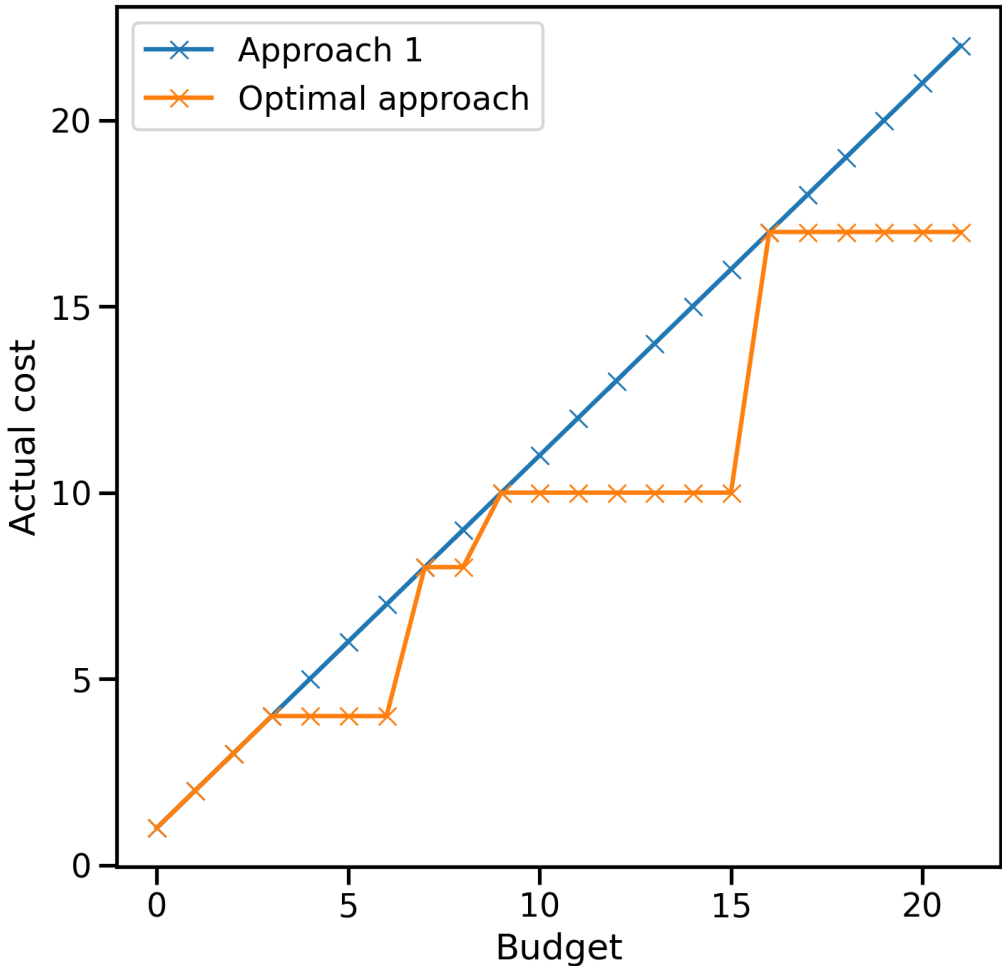
Budget Constrained Machine Learning

Budget	Total Cost	Optimal Set of Features	AP mean (std)
0	0	DCM	0.243 (0.062)
1	1	DCM, BMP	0.285 (0.048)
2	2	DCM, LDH	0.311 (0.060)
3	3	DCM, BMP, LDH	0.317 (0.050)
4	4	DCM, LDH, CRP	0.340 (0.070)
5	4	DCM, LDH, CRP	0.340 (0.070)
6	4	DCM, LDH, CRP	0.340 (0.070)
7	4	DCM, LDH, CRP	0.340 (0.070)
8	8	DCM, BMP, LDH, CRP, Troponin	0.343 (0.079)
9	8	DCM, BMP, LDH, CRP, Troponin	0.343 (0.079)
10	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
11	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
12	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
13	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
14	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
15	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
16	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
17	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
18	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
19	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
20	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
21	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
22	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)

Table 5: Optimal set of features at different budget levels for predicting mortality. Total Costs is the sum of the cost of each feature group in the selected set. The cost of each feature group is listed in Table 3.

Budget Constrained Machine Learning

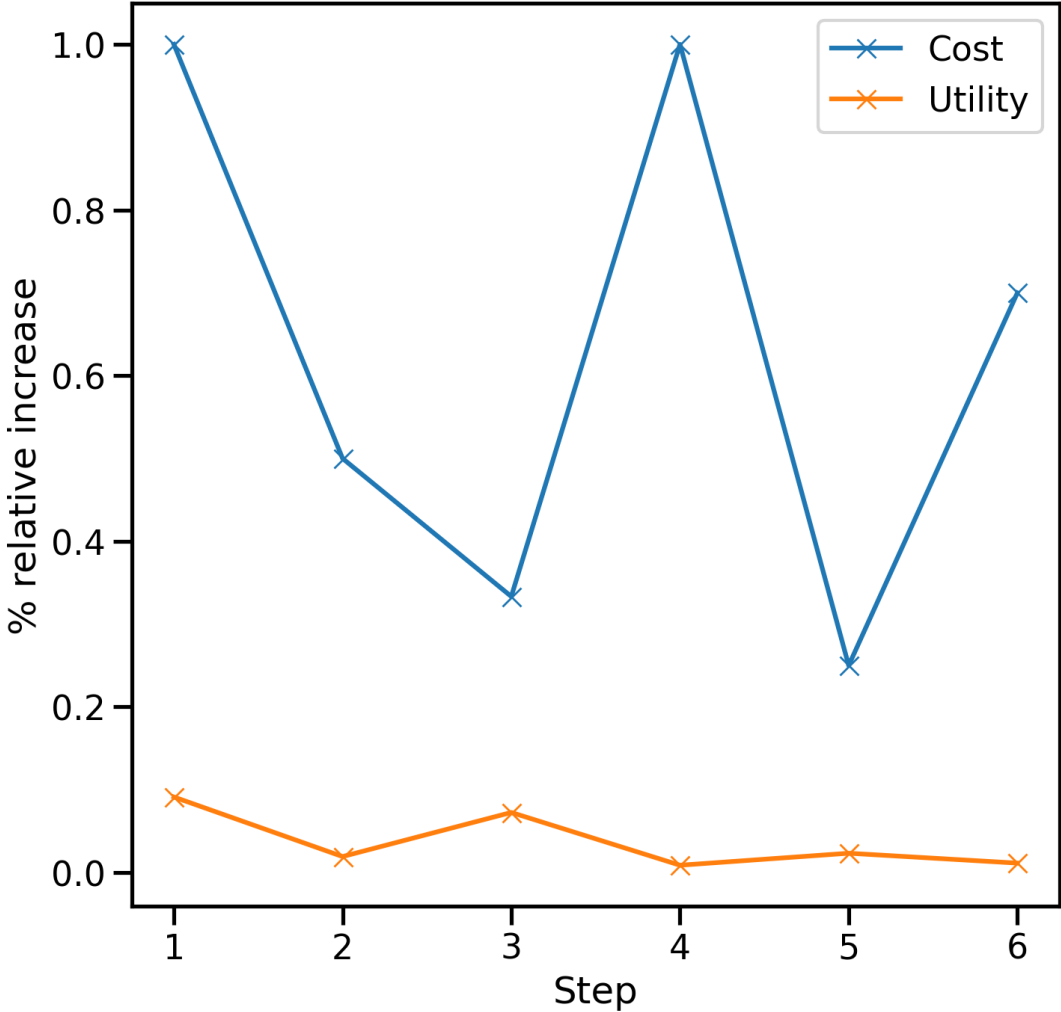| Budget	Total Cost	Optimal Set of Features	AP mean (std)
0	0	DCM	0.243 (0.062)
1	1	DCM, BMP	0.285 (0.048)
2	2	DCM, LDH	0.311 (0.060)
3	3	DCM, BMP, LDH	0.317 (0.050)
4	4	DCM, LDH, CRP	0.340 (0.070)
5	4	DCM, LDH, CRP	0.340 (0.070)
6	4	DCM, LDH, CRP	0.340 (0.070)
7	4	DCM, LDH, CRP	0.340 (0.070)
8	8	DCM, BMP, LDH, CRP, Troponin	0.343 (0.079)
9	8	DCM, BMP, LDH, CRP, Troponin	0.343 (0.079)
10	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
11	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
12	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
13	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
14	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
15	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
16	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
17	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
18	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
19	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
20	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
21	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
22	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)

Table 5: Optimal set of features at different budget levels for predicting mortality. Total Costs is the sum of the cost of each feature group in the selected set. The cost of each feature group is listed in Table 3.

Budget Constrained Machine Learning

Budget	Total Cost	Optimal Set of Features	AP mean (std)
0	0	DCM	0.243 (0.062)
1	1	DCM, BMP	0.285 (0.048)
2	2	DCM, LDH	0.311 (0.060)
3	3	DCM, BMP, LDH	0.317 (0.050)
4	4	DCM, LDH, CRP	0.340 (0.070)
5	4	DCM, LDH, CRP	0.340 (0.070)
6	4	DCM, LDH, CRP	0.340 (0.070)
7	4	DCM, LDH, CRP	0.340 (0.070)
8	8	DCM, BMP, LDH, CRP, Troponin	0.343 (0.079)
9	8	DCM, BMP, LDH, CRP, Troponin	0.343 (0.079)
10	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
11	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
12	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
13	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
14	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
15	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
16	10	DCM, BMP, LDH, Sedrate, CRP, Procalcitonin	0.351 (0.080)
17	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
18	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
19	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
20	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
21	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)
22	17	DCM, BMP, CBC, D-dimer, LDH, CRP, BNP, Procalcitonin, Ferritin	0.355 (0.075)

Table 5: Optimal set of features at different budget levels for predicting mortality. Total Costs is the sum of the cost of each feature group in the selected set. The cost of each feature group is listed in Table 3.

Lawrence Livermore National Laboratory