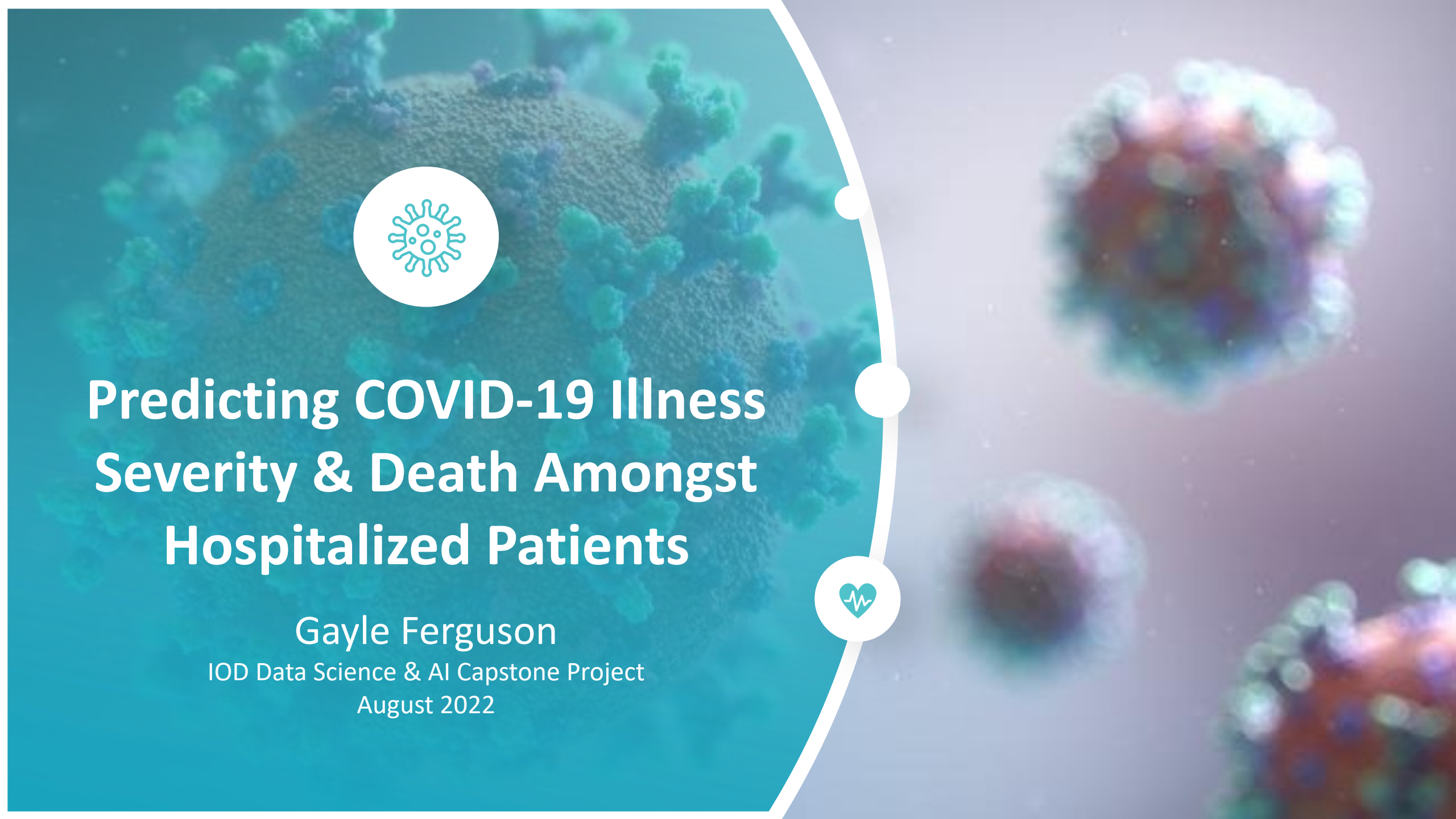


Predicting COVID-19 Illness Severity & Death Amongst Hospitalized Patients

Gayle Ferguson

IOD Data Science & AI Capstone Project

August 2022

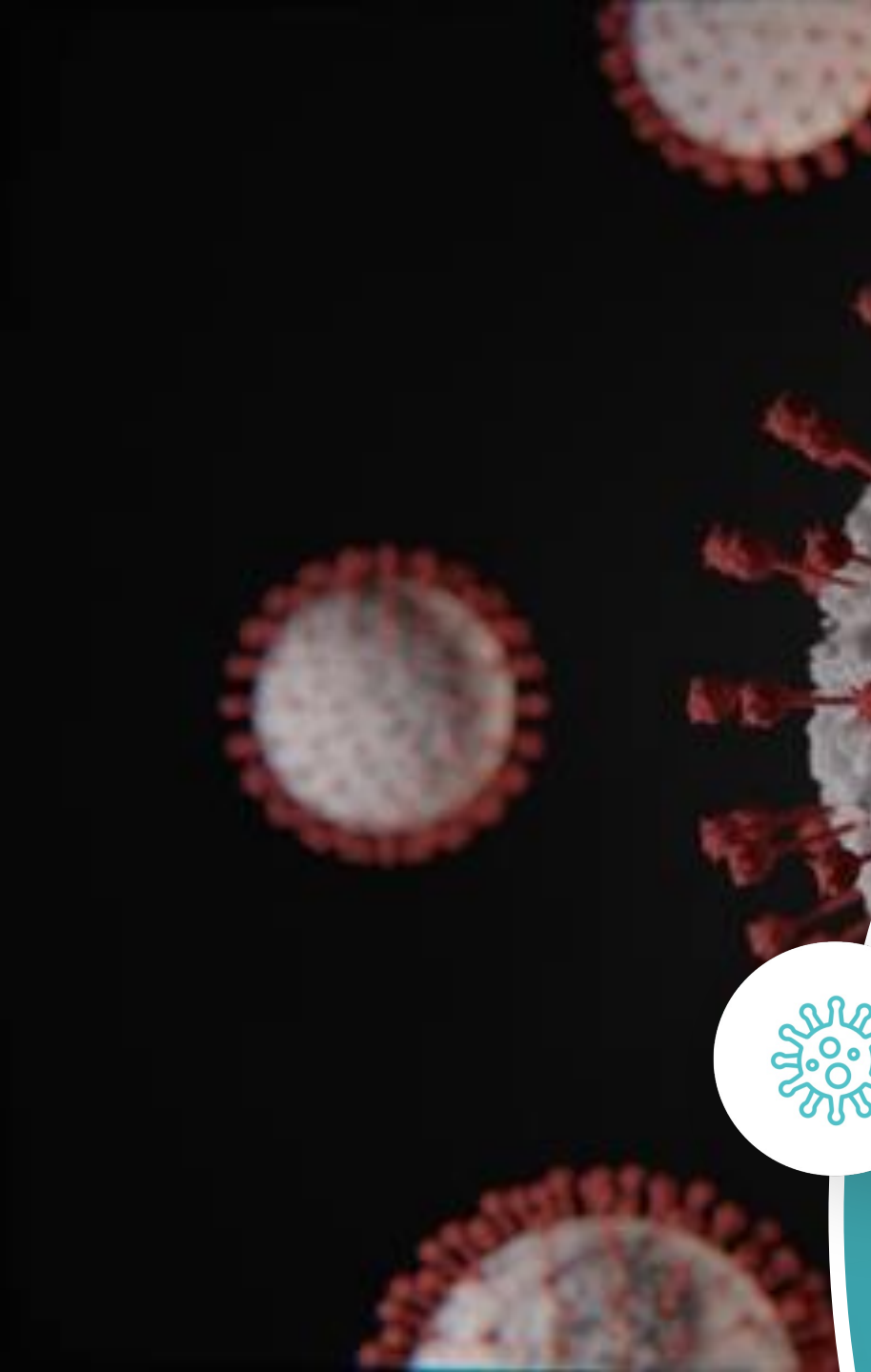
Goal: To predict COVID-19 severe illness and/or mortality from demographics, comorbidity and admission laboratory tests

COVID-19 is responsible for a 6.43M deaths worldwide since early 2020 and, as new variants emerge and spread, will continue to be a healthcare problem.

For most people, the disease is ‘mild’ and does not require hospitalisation.

However, in a significant portion of patients the disease is severe and fatal.

Determining which patients are at high risk of severe illness or mortality is essential for clinical decision-making and ensuring sometimes scarce resources are prioritised for those who need them most.



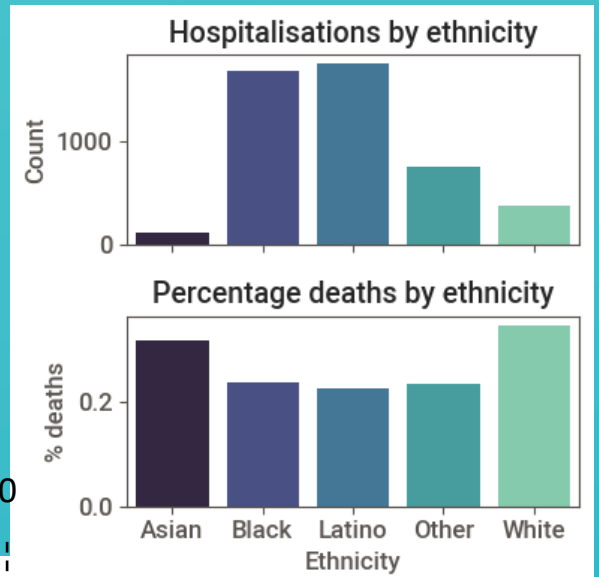
The Dataset

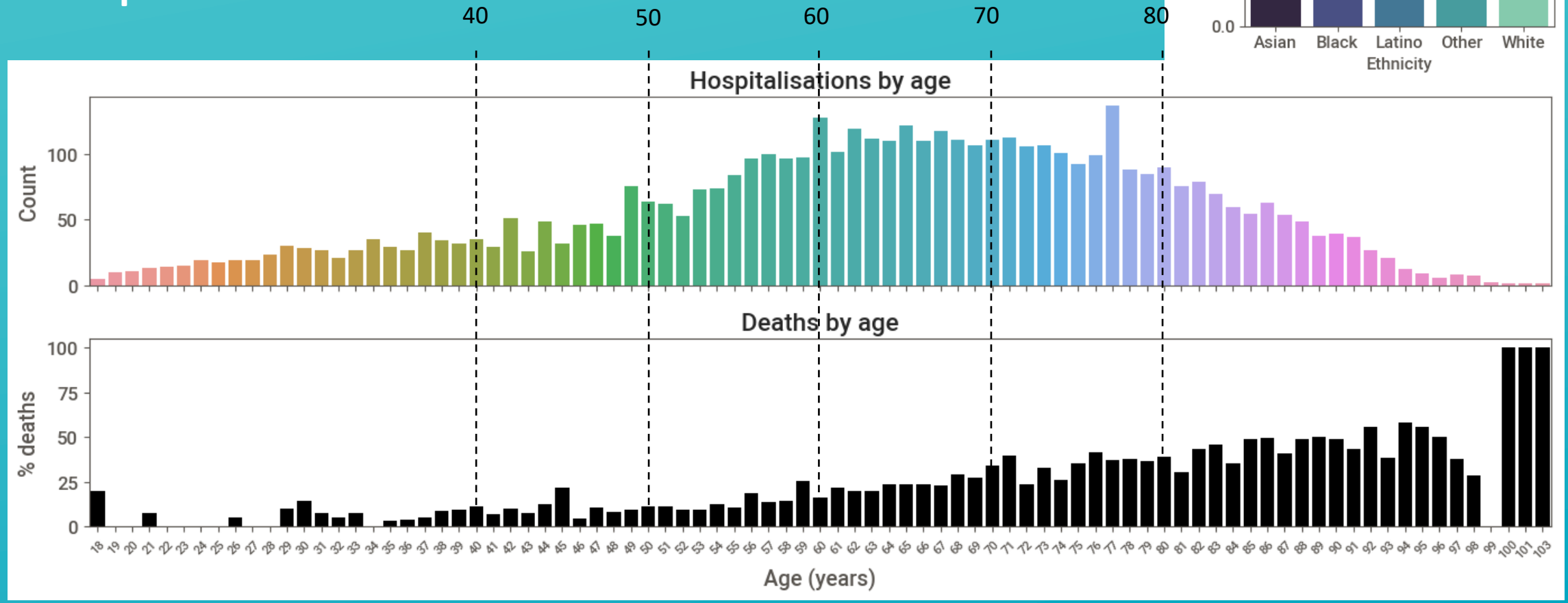
- 4711 instances
  - 1st and 2nd 3-weeks of the pandemic
  - Clinical Looking Glass [CLG]; Streamline Health, Atlanta, Georgia
- 82 features:
  - age, ethnicity, length of stay (9)
  - comorbidities(16)
  - admission test results (57)
- Targets (SL, classification):
  - death (0/1)
  - severity (0-11)

Source: https://www.kaggle.com/datasets/harshwalia/mortality-risk-clinincal-data-of-covid19-patients

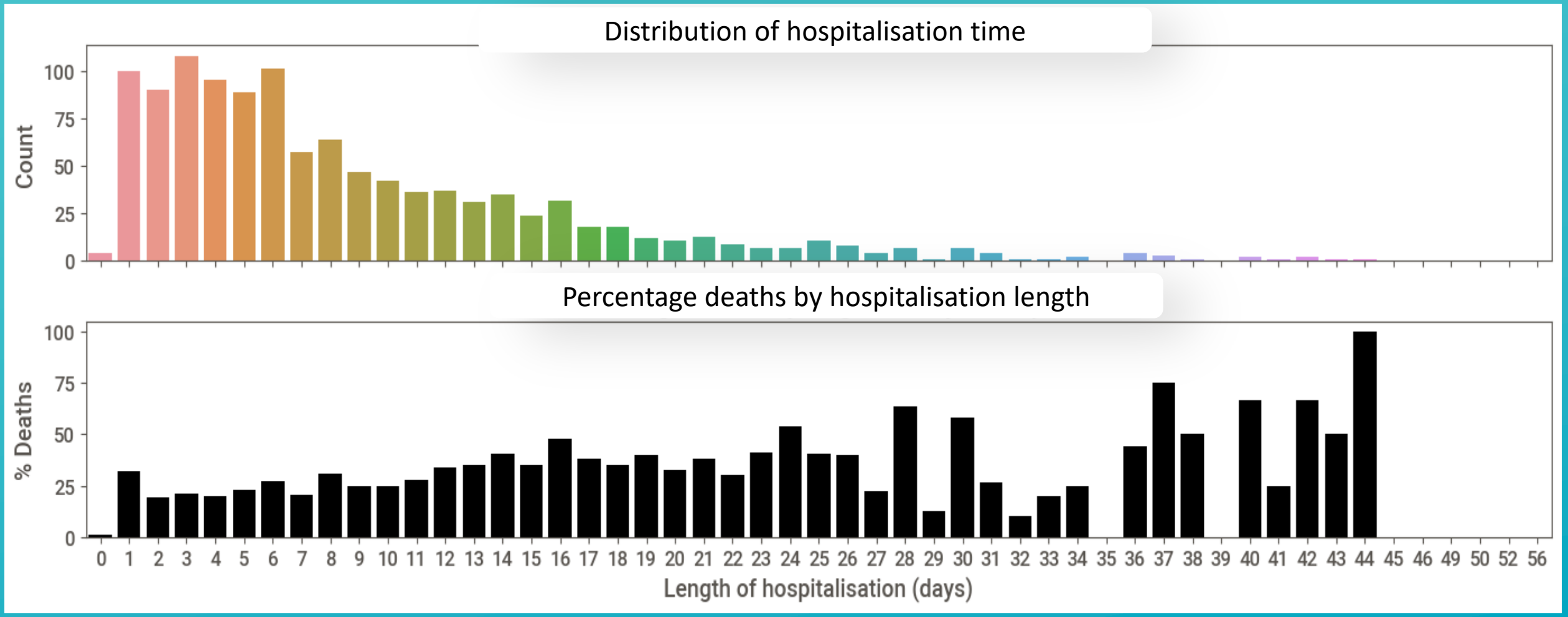
Demographic Insights:

- Hospitalised individuals were predominantly over 50
- Deaths were disproportionately among older people
- Hospitalizations were disproportionately Black or Latino, though a greater proportion of White and Asian patients died

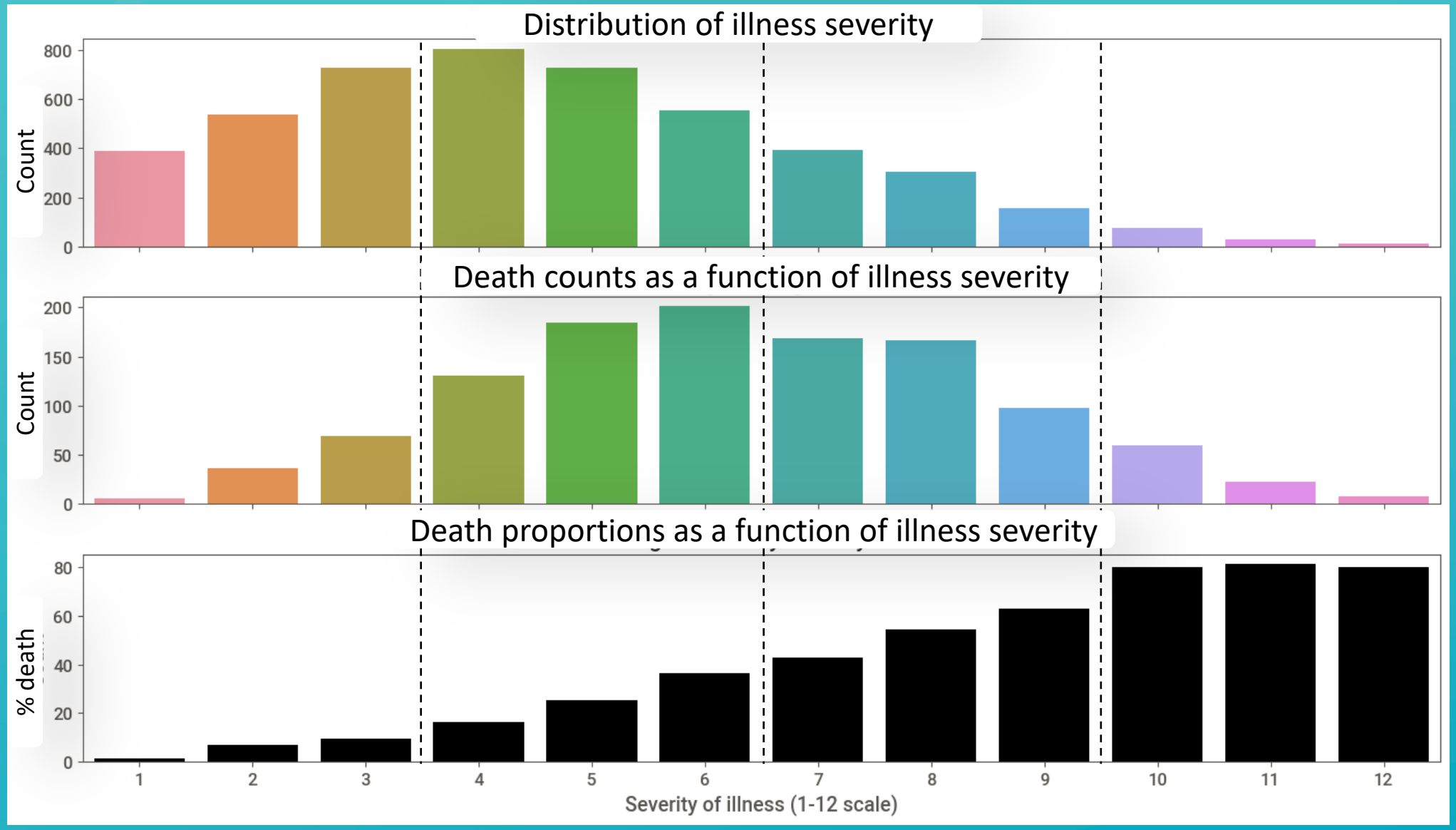


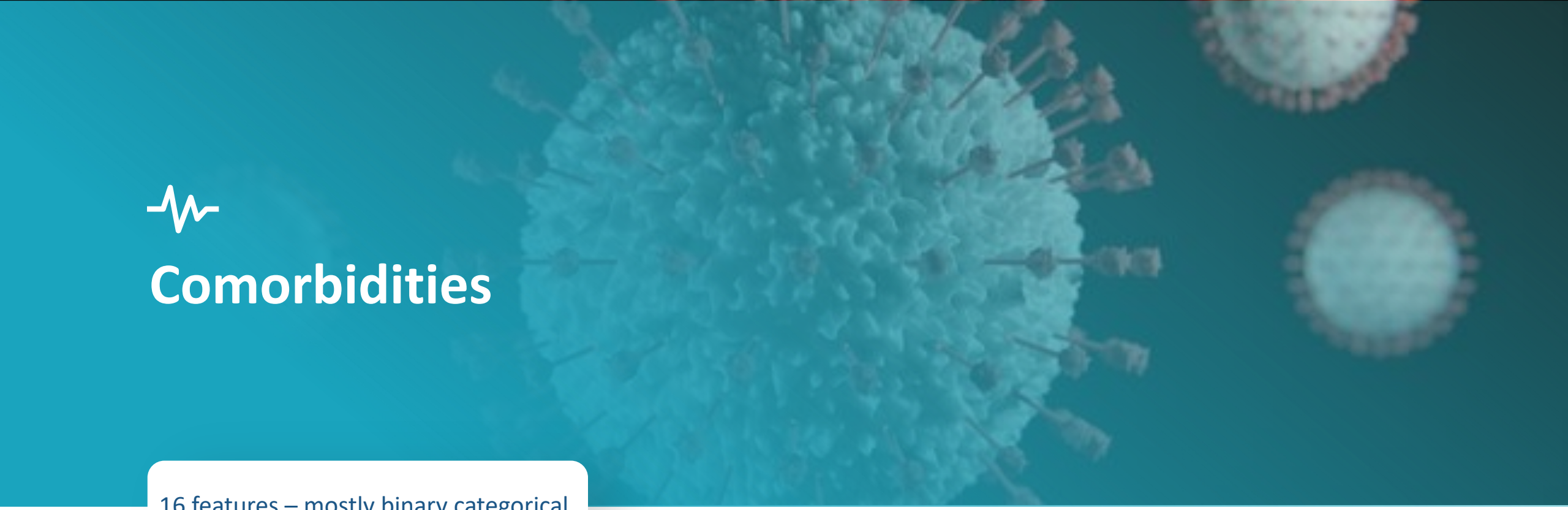


- Most hospital stays were less than 2 weeks.
- A greater proportion of deaths occurred amongst those who stayed over a month in hospital.



% Death is clearly correlated with severity of illness





Comorbidities

16 features – mostly binary categorical

1. Myocardial Infarction (MI)	(0/1)	9. Renal Disease	(0/1)
2. Peripheral Vascular Disease (PVD)	(0/1/2)	10. All CNS	(0/1)
3. Congestive Heart Failure (CHF)	(0/1)	11. Pure CNS	(0/1)
4. Cardiovascular Disease (CVD)	(0/1)	12. Stroke	(0/1)
5. Dementia	(0/1)	13. Seizure	(0/1)
6. Chronic obstructive pulmonary disease (COPD)	(0/1)	14. Old - Syncope	(0/1)
7. Diabetes mellitus complicated	(0/1)	15. Old – Other Neurological Conditions	(0/1)
8. Diabetes mellitus simple	(0/1)	16. Other Brain Lesions	(0/1)

Clinical Tests

Continuous measures

1. O_2 Saturation
2. Temperature
3. Mean Arterial Pressure (MAP)
4. D-dimer
5. Platelet count
6. International Normalized Ratio (INR)
7. Blood Urea Nitrogen (BUN)
8. Creatinine
9. Sodium
10. Glucose
11. Aspartate aminotransferase (AST)
12. Alanine transaminase (ALT)
13. White Blood Cell (WBC) counts
14. Lymphocytes
15. Interleukin-6 (IL6)
16. Ferritin
17. C-Reactive Protein
18. Procalcitonin
19. Troponin

Categorical data

20. O_2 Saturation < 94% (0/1)
21. Temperature > 38 °C (0/1)
22. MAP < 70 mmHg (0/1)
23. D-Dimer > 3 mg/ml (0/1)
24. Platelet score (0/1/2)
25. INR > 1.2 (0/1)
26. BUN > 30 mg/dL (0/1)
27. Creatinine Score (0/1/2)
28. Sodium < 139 or > 155 mEq/L (0/1)
29. Glucose < 60 or > 500 mg/dL (0/1)
30. AST > 40 units/L (0/1)
31. ALT > 40 units/L (0/1)
32. WBC < 1.8 or > 4.8 × 10^9/L (0/1)
33. Lymphocytes < 1 per nL (0/1)
34. IL6 > 150 pg/ml (0/1)
35. Ferritin > 300 ng/mL (0/1)
36. C-Reactive Protein > 1 mg/L (0/1)
37. Procalcitonin > 0.1 ng/ml (0/1)
38. Troponin > 0.1 ng/ml (0/1)

Pearsons Correlation

with ‘Death’

	corr. coefficient
Severity	0.431134
MAP<70	0.392984
Age	0.291806
AgeScore	0.286053
CrctProtein	0.244456
BUN	0.229488
Procalcitonin>0.1	0.211338
Procalcitonin>2	0.211109
O2Sat<90	0.207451
BUN>30	0.207235
CrtnScore	0.204768

with ‘Severity’

	corr. coefficient
AgeScore	0.631260
Age	0.612887
CrtnScore	0.531299
BUN	0.498741
BUN>30	0.475376
DDimer>3	0.465335
CrctProtein>10	0.461412
Death	0.431134
DDimer	0.428977
INR>1.2	0.421522
CrctProtein	0.407313
Procalcitonin>0.1	0.385839
Creatinine	0.378075
O2Sat<94	0.365804

AST>50	0.332478
MAP<70	0.323764
O2Sat<90	0.319387
CrctProtein>20	0.315406
Ferritin>1000	0.304264
Troponin>0.1	0.301974
Procalcitonin>2	0.295090
DDimer>17	0.285762
Sodium130<>145	0.261948
BUN>100	0.248345
IL6>60	0.239553
Sodium>155	0.232112
PltsScore	0.231550
Procalcitonin	0.222266
AST>100	0.220158
Sodium	0.216708
Lympho<1	0.202721

Workflow

1. Data Cleaning

- Identification of mismatches between columns and other errors
- Replacement of ‘0’ values with mode or median values

2. Feature Engineering

- Adjustment of thresholds for some existing features
- Creation of new binary features based on published normal ranges for blood test data
- Reduced the ‘severity’ (target) categories from 12 to 4

3.

Dimension reduction

PCA → **Dataset 1 (60 PC’s)**

Feature selection

Recursive Feature Elimination (RFE) → **Dataset 2 (Top 10)**, **Dataset 3 (Top 40)**

Select K Best (SKB) → **Dataset 4 (Top 10)**, **Dataset 5 (Top 40)**

4.

Models:

K-Nearest Neighbours
Naïve Bayes
Logistic Regression
SVC
Random Forest
XGBoost

Stacking Classifier

Results

Predicting 'Severity'

Logistic Regression, XGBoost and SVC models predicted 'severity' class with 94-95% accuracy

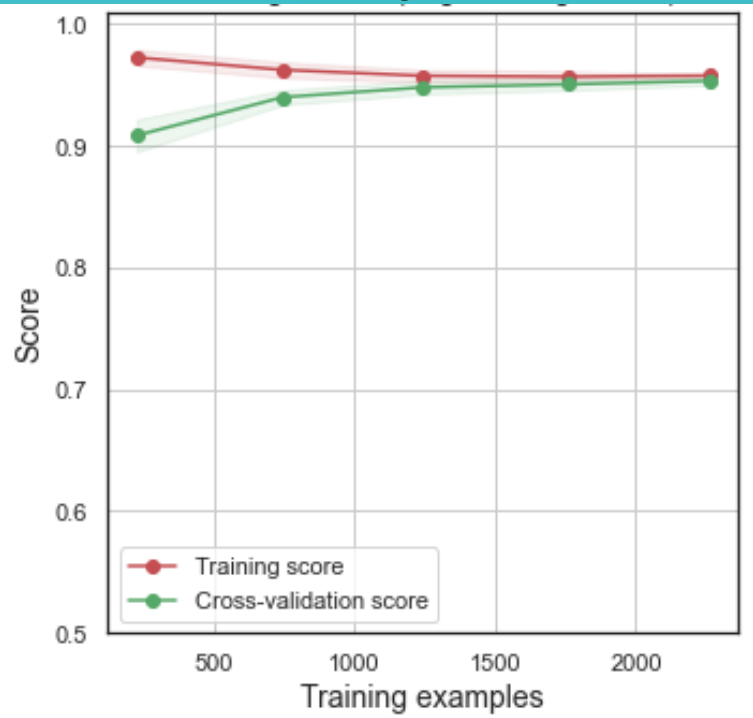
	Dataset 1 (60 PC's)	Dataset 2 (Top 10 - RFE)	Dataset 3 (Top 40 - RFE)	Dataset 4 (Top 10)	Dataset 5 (Top 40)
K-Nearest Neighbours	0.610403	0.92569	0.820594	0.795117	0.798301
Gaussian Naive Bayes	0.632696	0.767516	0.686837	0.695329	0.606157
Bernoulli Naive Bayes	0.692144	0.800425	0.767516	0.729299	0.719745
Logistic Regression	0.942675	0.955414	0.947983	0.825902	0.944798
SVC	0.941614	0.943737	0.94586	0.819533	0.943737
Random Forest	0.799363	0.912951	0.882166	0.799363	0.869427
XGBoost	0.842887	0.951168	0.917197	0.805732	0.915074
Stacking Classifier 1 (RF)	0.798301	0.933121	0.825902	0.79724	0.800425
Stacking Classifier 2 (XGB)	0.854565	0.934183	0.867304	0.79724	0.821656
Stacking Classifier 3 (LR)	0.840764	0.937367	0.888535	0.792994	0.87155
Stacking Classifier 4 (SVC)	0.804671	0.928875	0.869427	0.79724	0.854565
Stacking Classifier 5 (BNB)	0.436306	0.436306	0.436306	0.436306	0.436306
Stacking Classifier 6 (KNN)	0.868365	0.93949	0.91189	0.798301	0.883227
Stacking Classifier 7 (KMeans)	0.868365	0.93949	0.91189	0.798301	0.883227

Evaluation

- Fine-tuning produced models with >95% accuracy
- Convergence of training and cross-validation scores over training samples of increasing size indicates that the models are **not** over-fitting

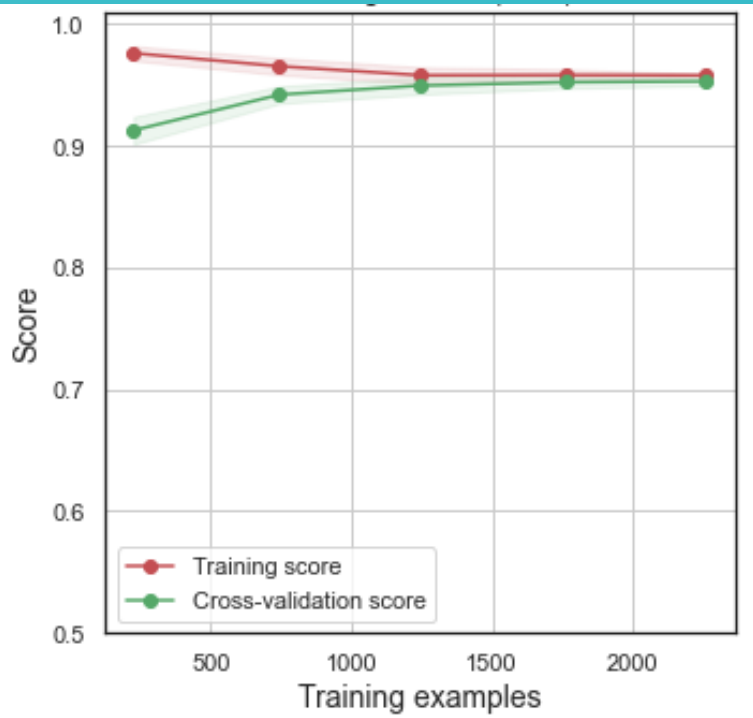
Logistic Regression

Accuracy 0.9554



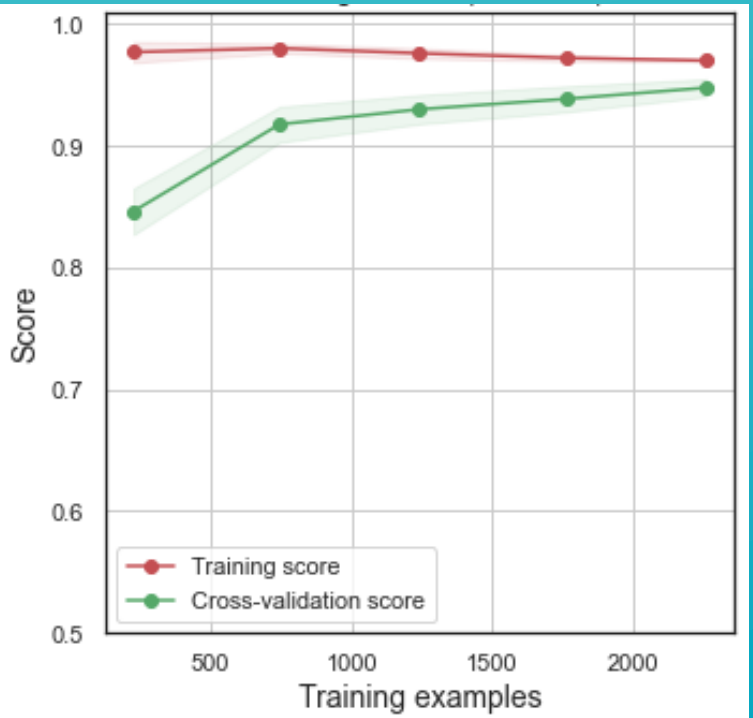
SVC

Accuracy 0.9544



XGBoost

Accuracy 0.9512



Evaluation

- LogReg and SVC models perform similarly: only 2 “misses” for Class 4 and 8 “misses” for Class 3
- XGBoost produces a greater number of “misses” for the more severe classes

LogReg

	predicted_severity class 1	predicted_severity class 2	predicted_severity class 3	predicted_severity class 4
actual_severity class 1	348	3	1	0
actual_severity class 2	17	391	4	0
actual_severity class 3	0	8	143	2
actual_severity class 4	0	0	2	24

SVC

	predicted_severity class 1	predicted_severity class 2	predicted_severity class 3	predicted_severity class 4
actual_severity class 1	342	9	1	0
actual_severity class 2	13	395	4	0
actual_severity class 3	0	8	144	1
actual_severity class 4	0	0	2	24

XGB

	predicted_severity class 1	predicted_severity class 2	predicted_severity class 3	predicted_severity class 4
actual_severity class 1	348	3	1	0
actual_severity class 2	17	392	3	0
actual_severity class 3	0	12	141	0
actual_severity class 4	0	0	8	18

Evaluation

The SVC model gave slightly better precision and recall for classes 3 and 4 than the other two models

TP/(TP+FP)
1- precision = “false alarms”

TP/(TP+FN)
1-recall = “misses”

SVC

	precision	recall	f1-score	support
severity class 1	0.96	0.97	0.97	352
severity class 2	0.96	0.96	0.96	412
severity class 3	0.95	0.94	0.95	153
severity class 4	0.96	0.92	0.94	26
accuracy			0.96	943
macro avg	0.96	0.95	0.95	943
weighted avg	0.96	0.96	0.96	943

Results

- **Baseline accuracy is 73%**

(*i.e.* we'd we correct 73% of the time by always predicting that the patient will survive!)

- **Random Forest and XGBoost models were the best performers, but only 9% better than the baseline at best**

Predicting 'Death'

	Dataset 1 (60 PC's)	Dataset 2 (Top 30 - RFE)	Dataset 3 (Top 60 - RFE)	Dataset 4 (Top 30 - SKB)
Gaussian Naive Bayes	0.746285	0.782378	0.754777	0.772824
Bernoulli Naive Bayes	0.780255	0.779193	0.738854	0.757962
Logistic Regression	0.780255	0.784501	0.785563	0.789809
SVC	0.794055	0.788747	0.792994	0.788747
Random Forest	0.800425	0.81741	0.819533	0.819533
XGBoost	0.802548	0.813163	0.821656	0.820594
Stacking Classifier 1 (RF)	0.802548	0.809979	0.814225	0.819533
Stacking Classifier 2 (XGB)	0.800425	0.81741	0.819533	0.819533
Stacking Classifier 3 (LogReg)	0.800425	0.81741	0.819533	0.819533

Evaluation of XGBoost Model

Predicting 'Death'

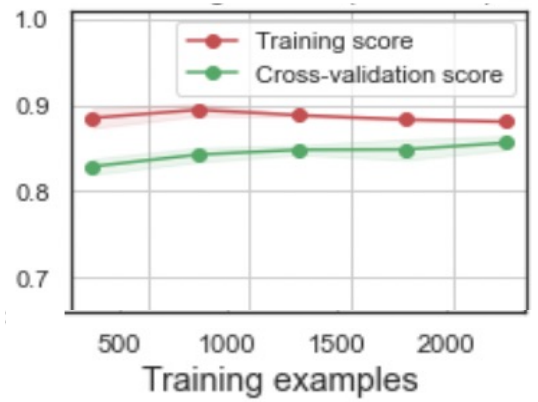
Dataset 3 (Top 60 - RFE)

X_test

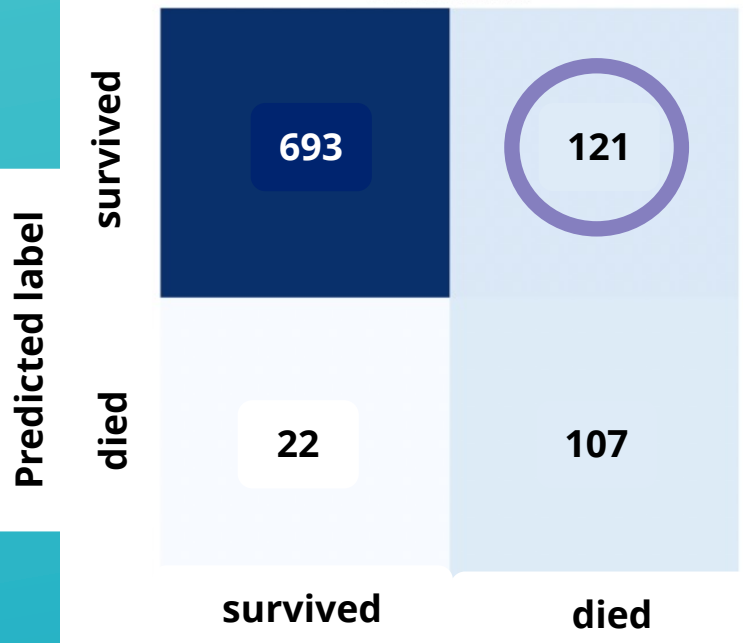
Accuracy:	**0.8484**
Precision:	**0.8295**
Recall:	**0.4693**

Recall is especially poor,
***i.e.* more than 50% of 'deaths' are wrongly predicted as 'survived'**

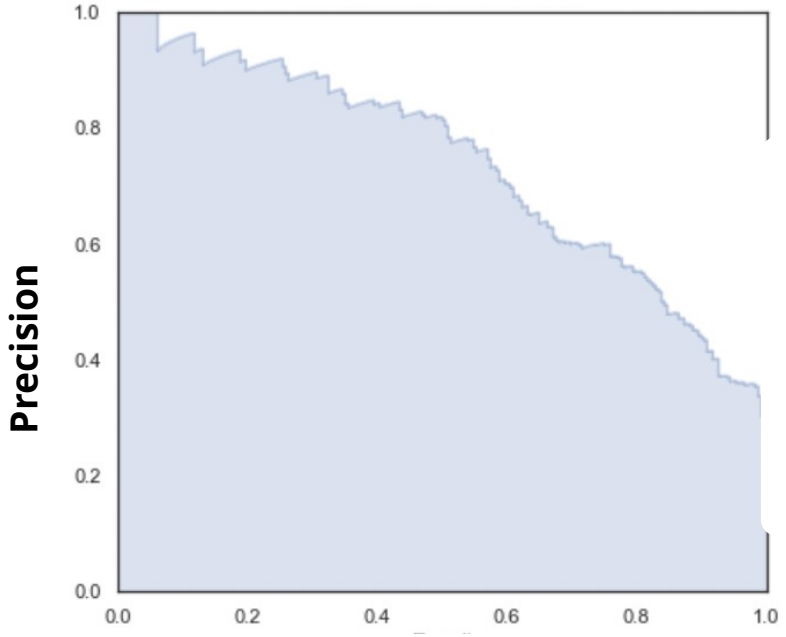
Learning Curves



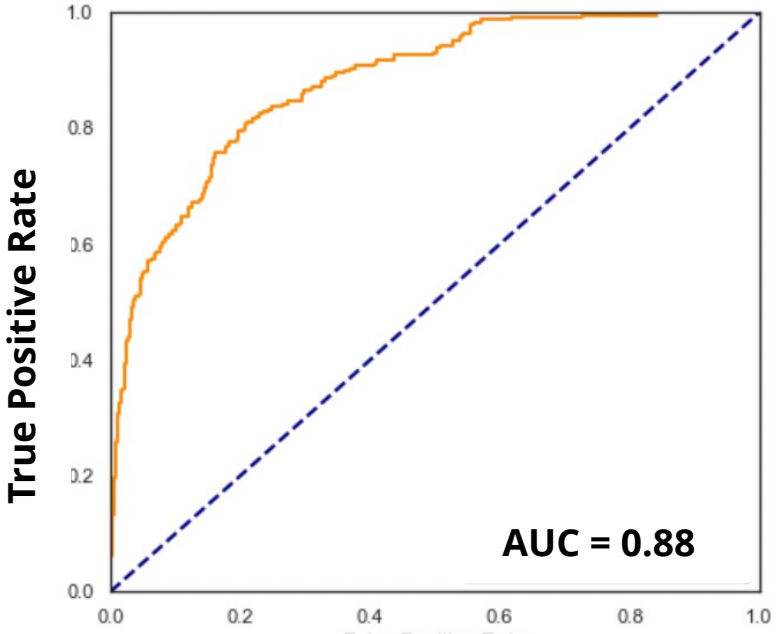
Confusion Matrix



2-Class Precision-Recall Curve



ROC Curve



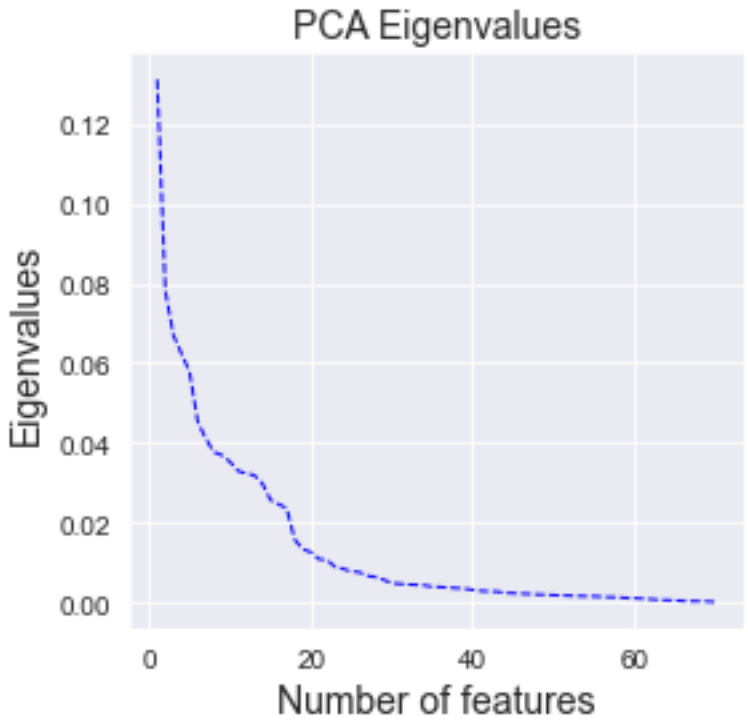
Conclusions

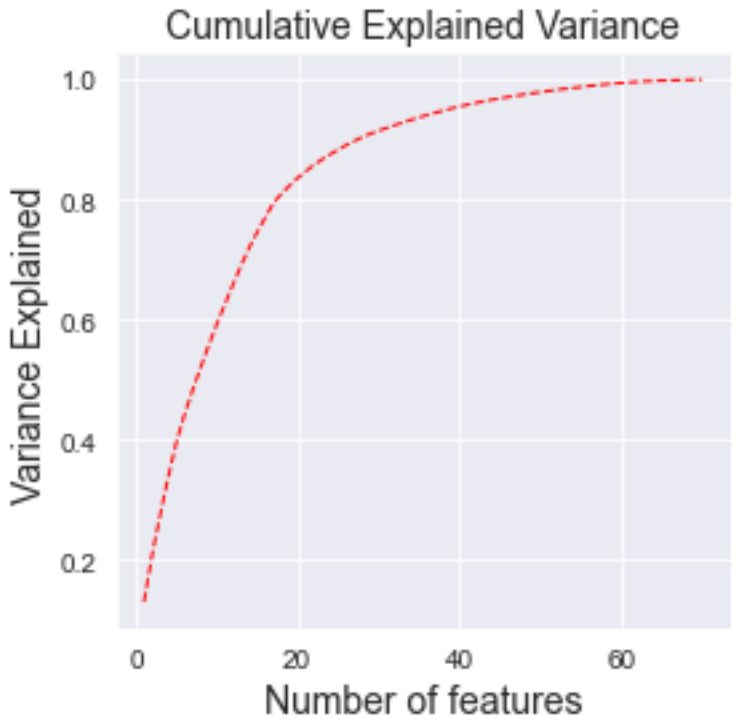
- SVC model can predict ‘severity’ of illness (4 classes) with > 95% accuracy and precision, and recall of 92%, and is easily scalable
- Predicting ‘death’ is less accurate, with > 50% “misses”; further data examination is needed to identify the cause of prediction error
- Notwithstanding, accurate prediction of severe illness will enable hospitals to triage the most vulnerable patients, preventing many deaths

Top 10 Predictors of ‘Severity’ (by RFE)
Mean Arterial Pressure < 70 mmHg
International Normalized Ratio (INR) > 1.2
C-reactive Protein > 10 mg/L
D-Dimer > 3 mg/mL
Platelet Score
Age Score
O_2 Saturation < 94%
Aspartate aminotransferase (AST) > 50 u/L
Creatinine Score
Temperature <= 36 ºC

The End

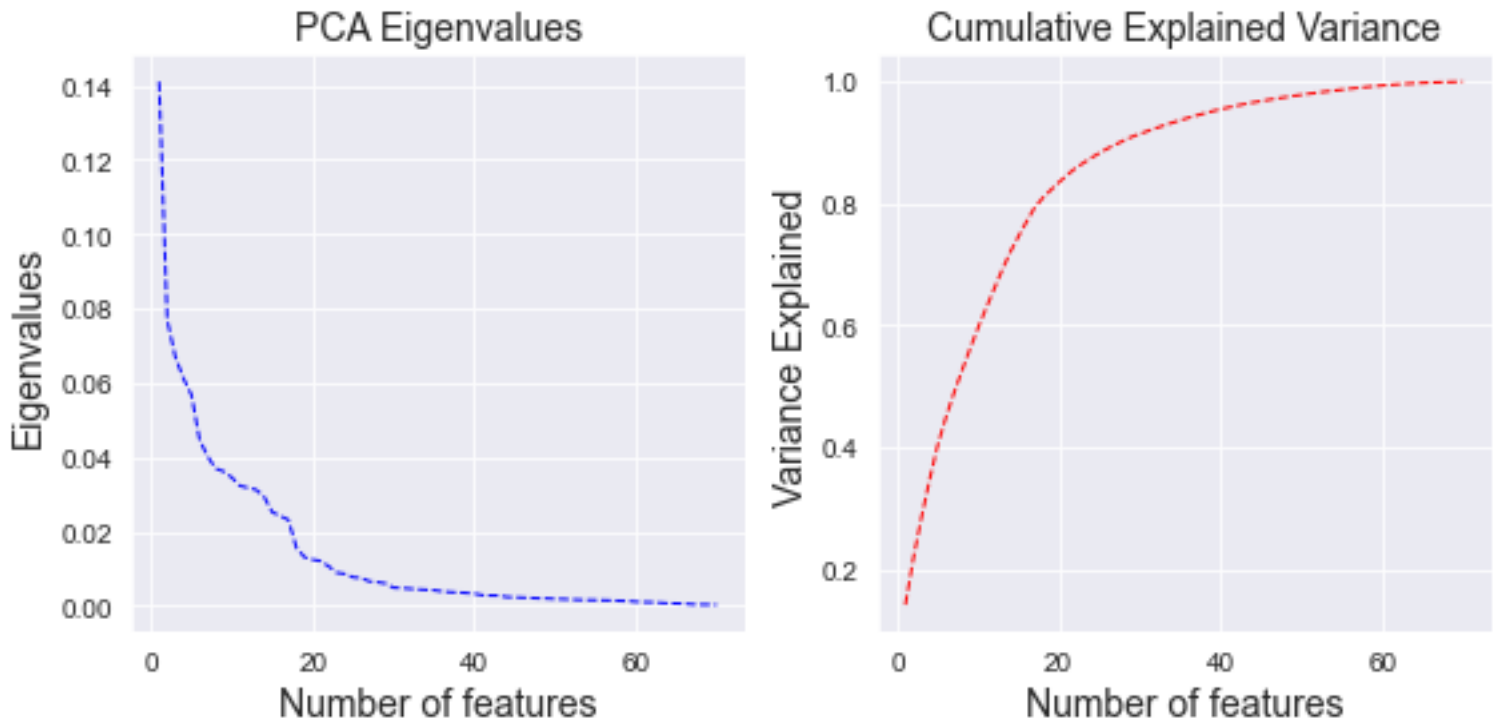
Explained Variance and Cumulative Explained Variance after dimension reduction by PCA





Predicting ‘Severity’

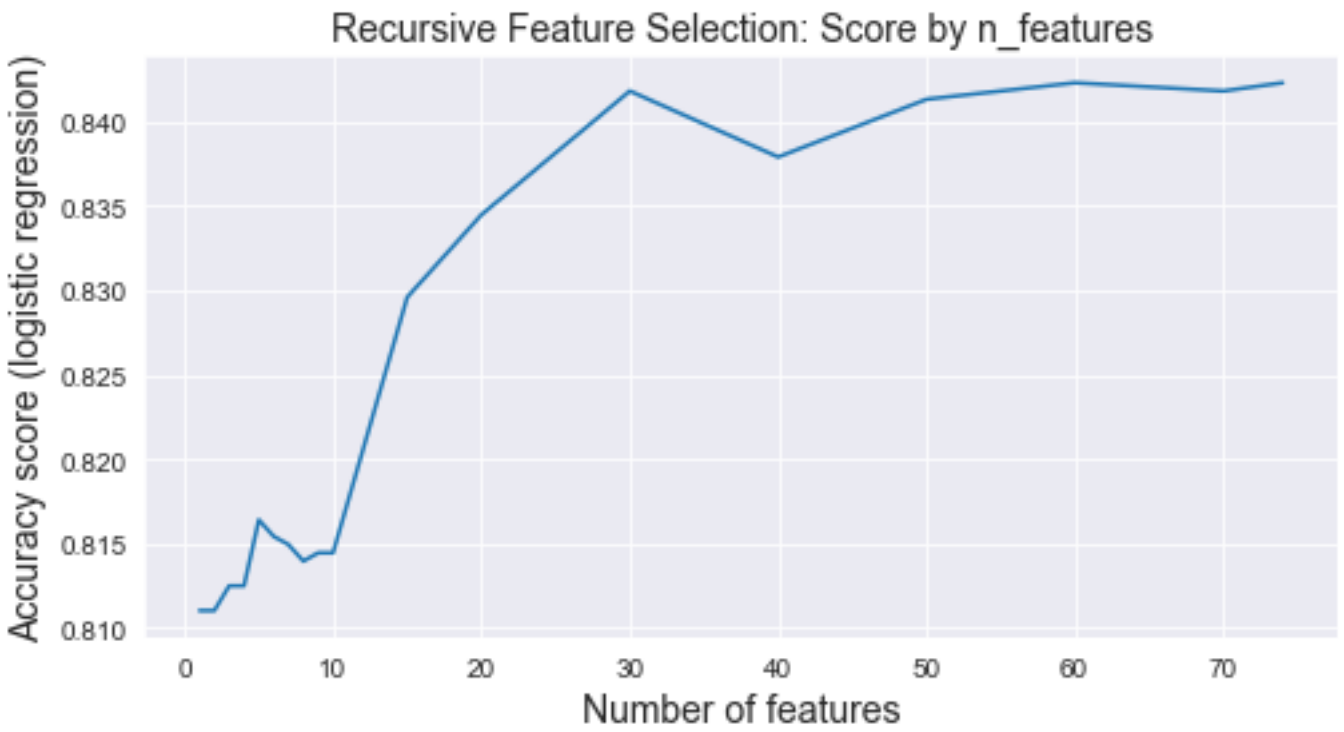
Explained Variance and Cumulative Explained Variance after dimension reduction by PCA



Predicting ‘Death’

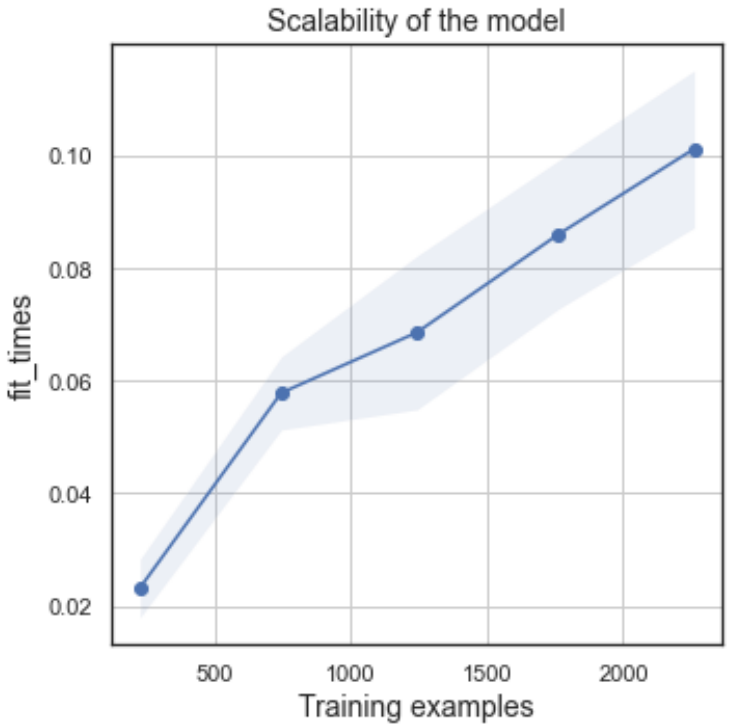
Recursive Feature Elimination for Predictors of Illness ‘Severity’

Recursive Feature Elimination for Predictors of ‘Death’

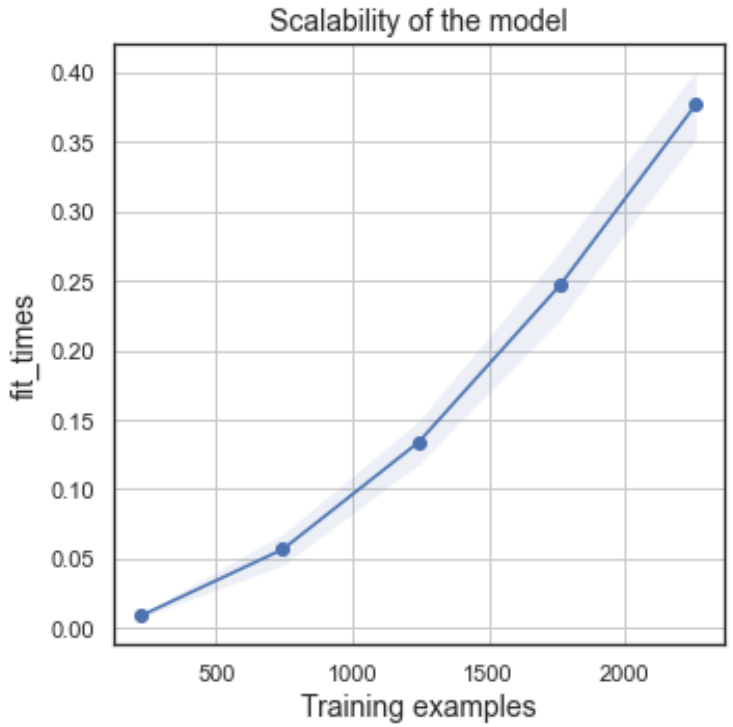


Scalability of the models for Predicting ‘Severity’

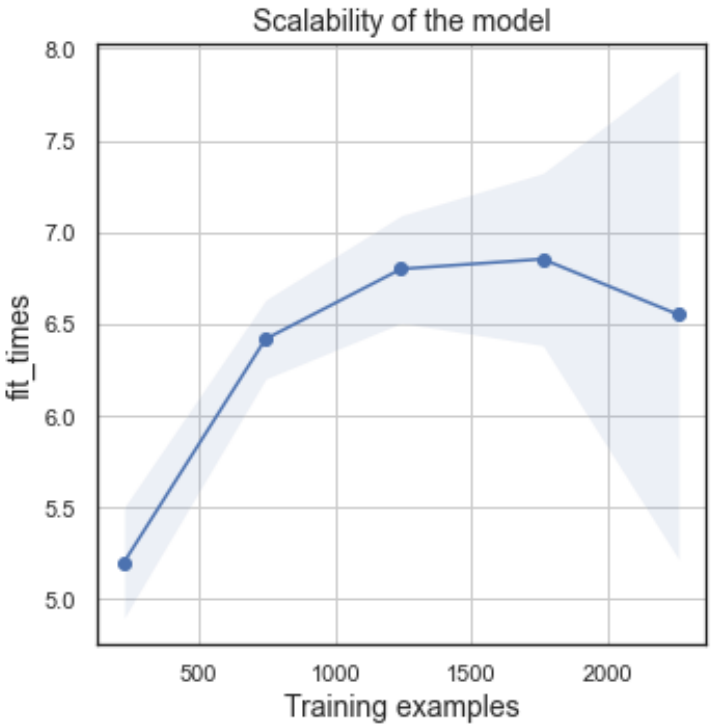
Logistic Regression



SVC

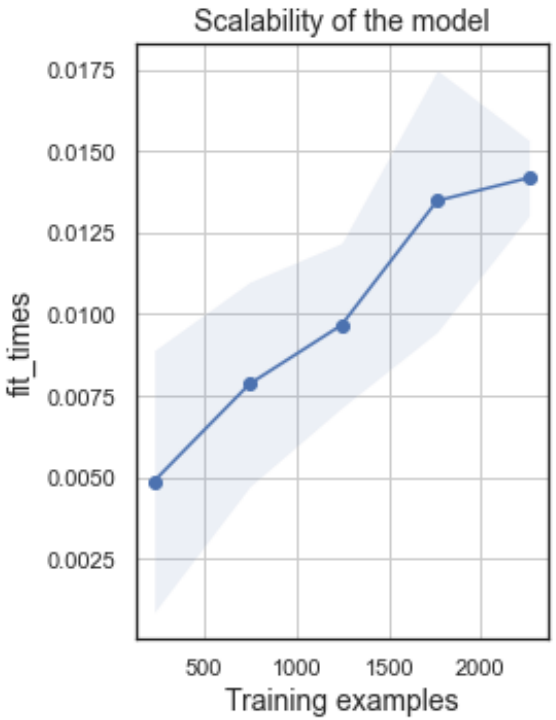


XGBoost

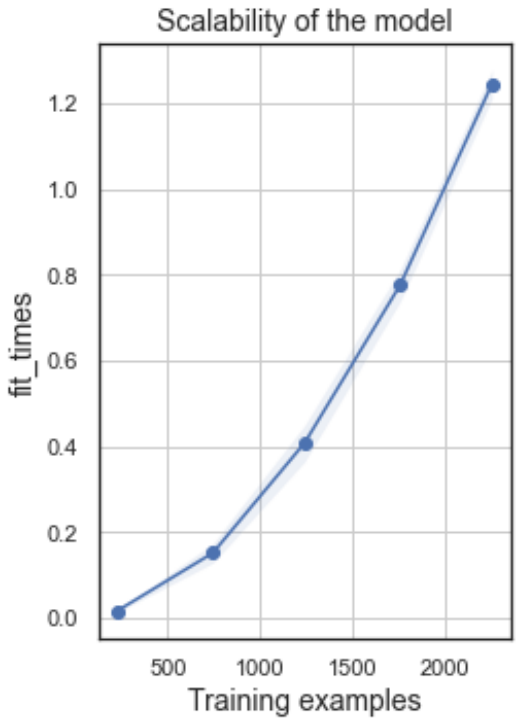


Scalability of the models for Predicting ‘Death’

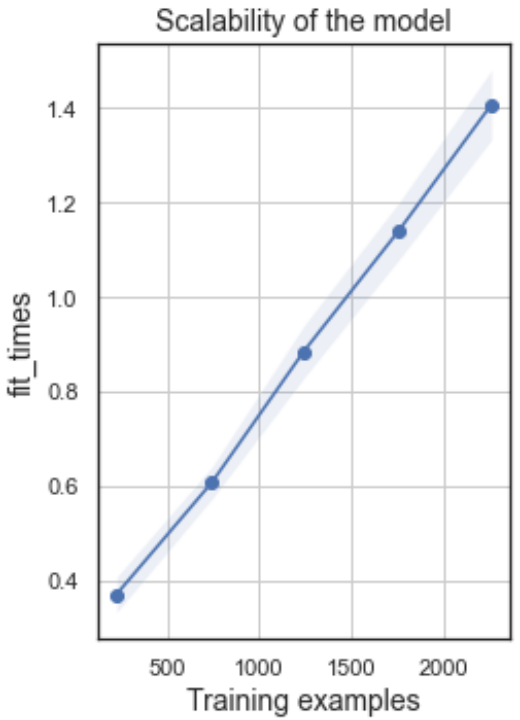
Logistic Regression



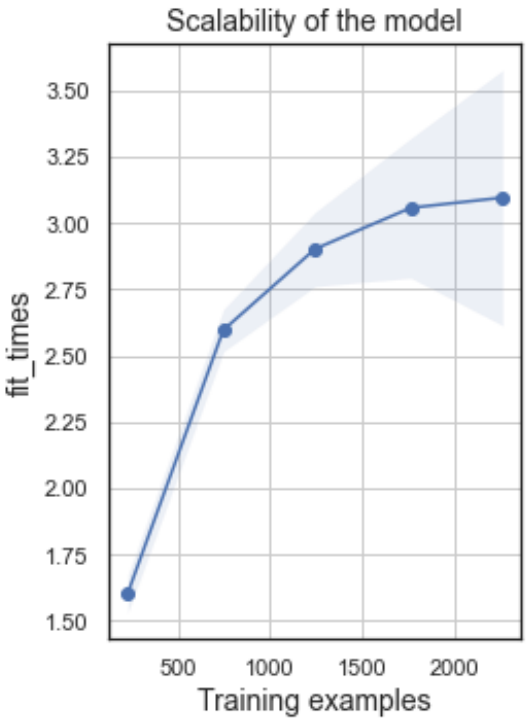
SVC



Random Forest



XGBoost



Top10 features for predicting illness severity by RFE (left) and SKB (right)

```
The Top 10 Features, in order, are:

'MAP<70'
'INR>1.2'
'CrctProtein>10'
'DDimer>3'
'PltsScore'
'AgeScore'
'O2Sat<94'
'AST_over50'
'CrtnScore'
'Temp<=36'
```

```
The Top 10 Features, in order, are:

'AgeScore'
'Age'
'CrtnScore'
'BUN'
'BUN_over30'
'DDimer_over3'
'CrctProtein_over10'
'INR_over1.2'
'DDimer'
'CrctProtein'
```

Top10 features for predicting illness severity (left) and death (right) by RFE

```
The Top 10 Features, in order, are:

'MAP<70'
'INR>1.2'
'CrctProtein>10'
'DDimer>3'
'PltsScore'
'AgeScore'
'O2Sat<94'
'AST_over50'
'CrtnScore'
'Temp<=36'
```

```
The Top 10 Features, in order, are:

'MAP<70'
'Stroke'
'O2Sat<90'
'OldSyncope'
'Age'
'DDimer>17'
'Asian'
'AllCNS'
'Troponin'
'OtherBrnLsn'
```