Predicting Death with COVID-19

Sunna Jo

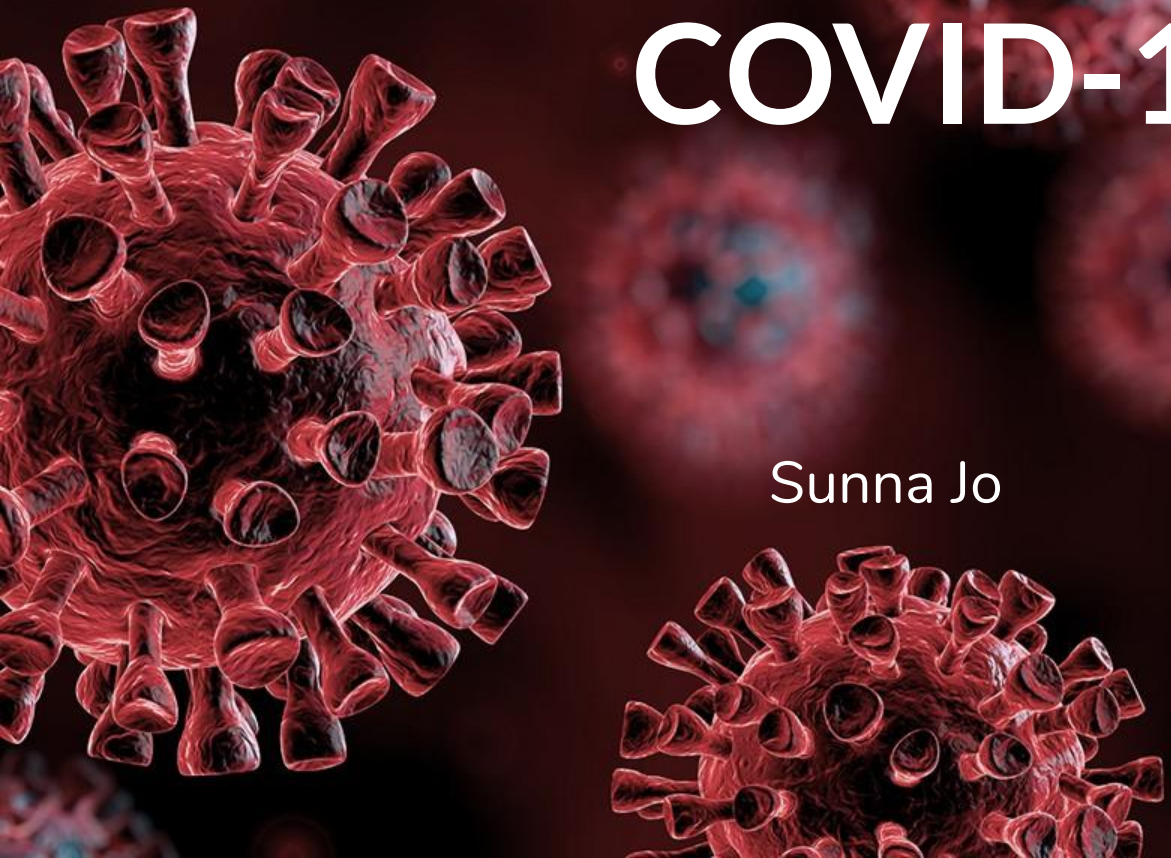


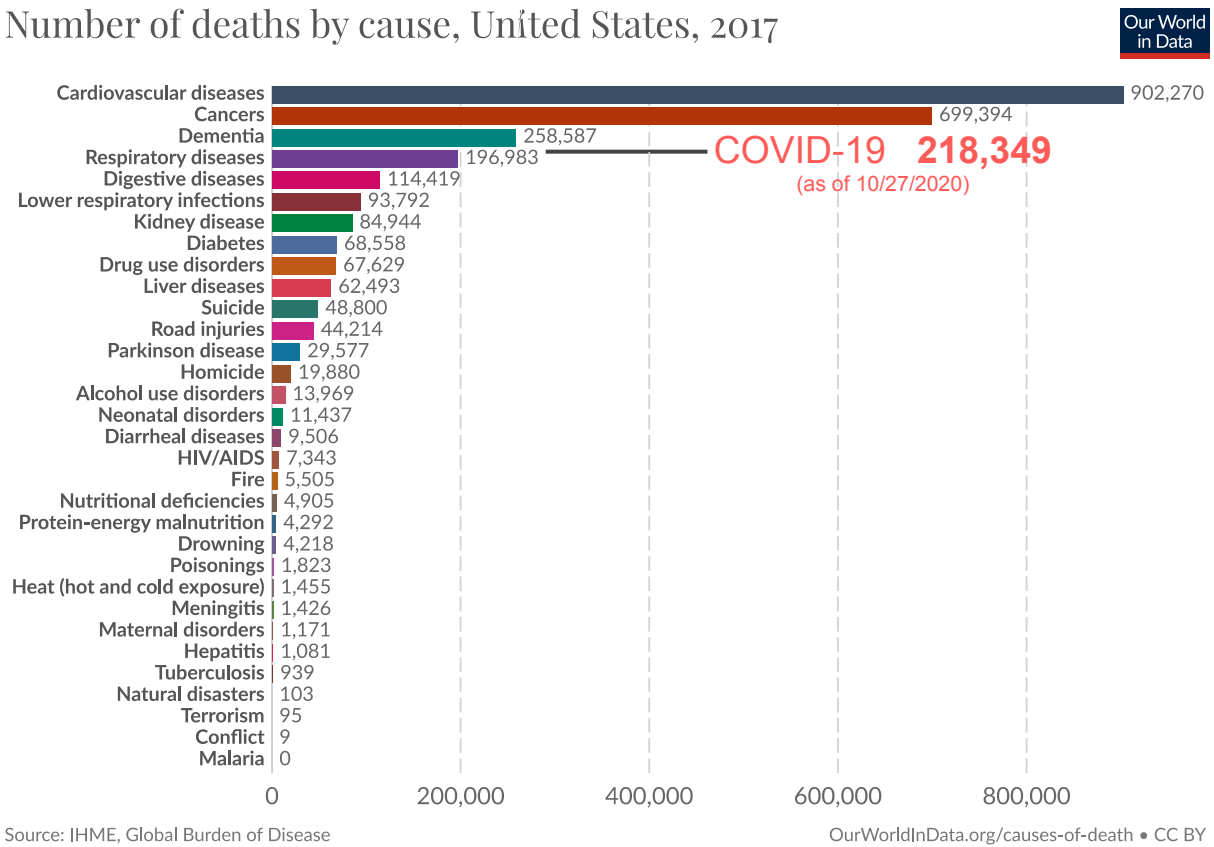
Image source: https://www.tmj4.com/news/coronavirus

Overview

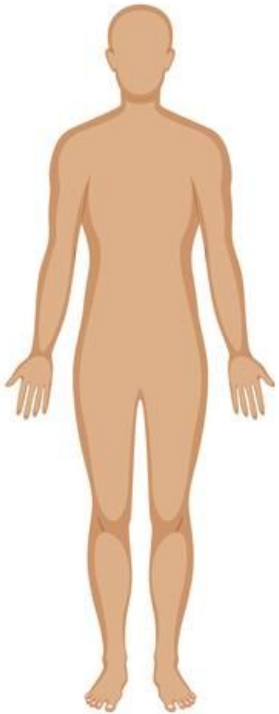
- Background & Objective
- Process
- Final Model
- Conclusions
- Application
- Future Work

Image source: https://www.kansashealthsystem.com/patient-visitor/covid19-update

How bad is it?


Number of deaths by cause, United States, 2017
Our World in Data
Cardiovascular diseases 902,270
Cancers 699,394
Dementia 258,587
Respiratory diseases 196,983
COVID-19 218,349
(as of 10/27/2020)
Digestive diseases 114,419
Lower respiratory infections 93,792
Kidney disease 84,944
Diabetes 68,558
Drug use disorders 67,629
Liver diseases 62,493
Suicide 48,800
Road injuries 44,214
Parkinson disease 29,577
Homicide 19,880
Alcohol use disorders 13,969
Neonatal disorders 11,437
Diarrheal diseases 9,506
HIV/AIDS 7,343
Fire 5,505
Nutritional deficiencies 4,905
Protein-energy malnutrition 4,292
Drowning 4,218
Poisonings 1,823
Heat (hot and cold exposure) 1,455
Meningitis 1,426
Maternal disorders 1,171
Hepatitis 1,081
Tuberculosis 939
Natural disasters 103
Terrorism 95
Conflict 9
Malaria 0
0
200,000
400,000
600,000
800,000
Source: IHME, Global Burden of Disease
OurWorldInData.org/causes-of-death • CC BY

What contributes to death?



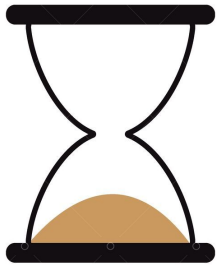
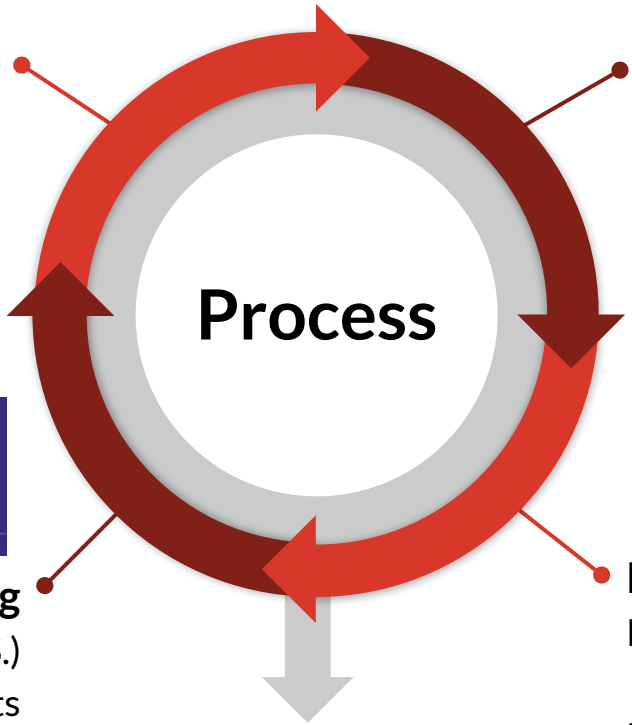


Image 1: Getty Images/iStockphoto / Creator/copyright:: colematt // Image 2:: Dreamstime.com / Copyright © Mike301 // Image 3: https://pixlr.com/stock/details/1001344267-hourglass/



Exploratory Data Analysis

Data set variations

Visualizations

Feature Engineering & Selection

40 features

Correlations

Feature analysis

Data Collection & Cleaning

Patient-level, national-level (U.S.)

3 data sets

671,435 data points

Model Testing

Different models →→ decision tree-based methods

Metric: **Recall**

Hyperparameter tuning

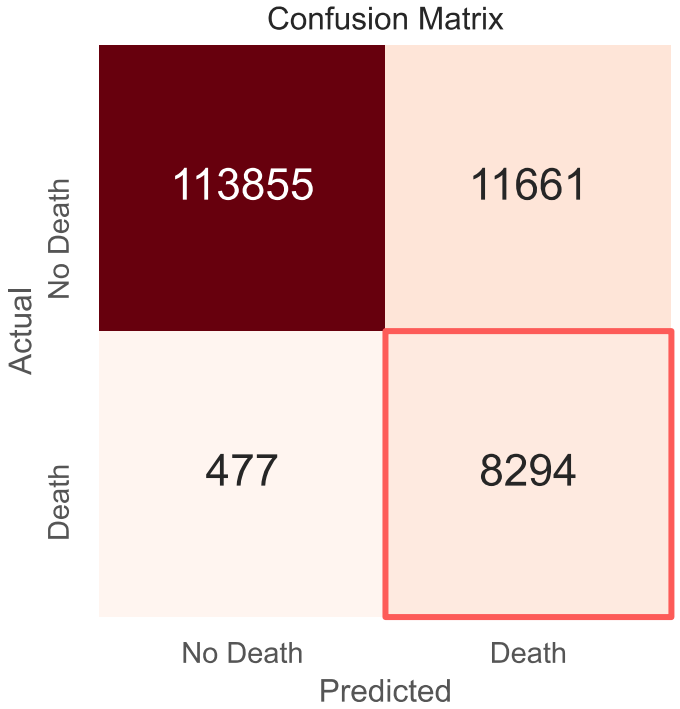
Final Model

XGBoost

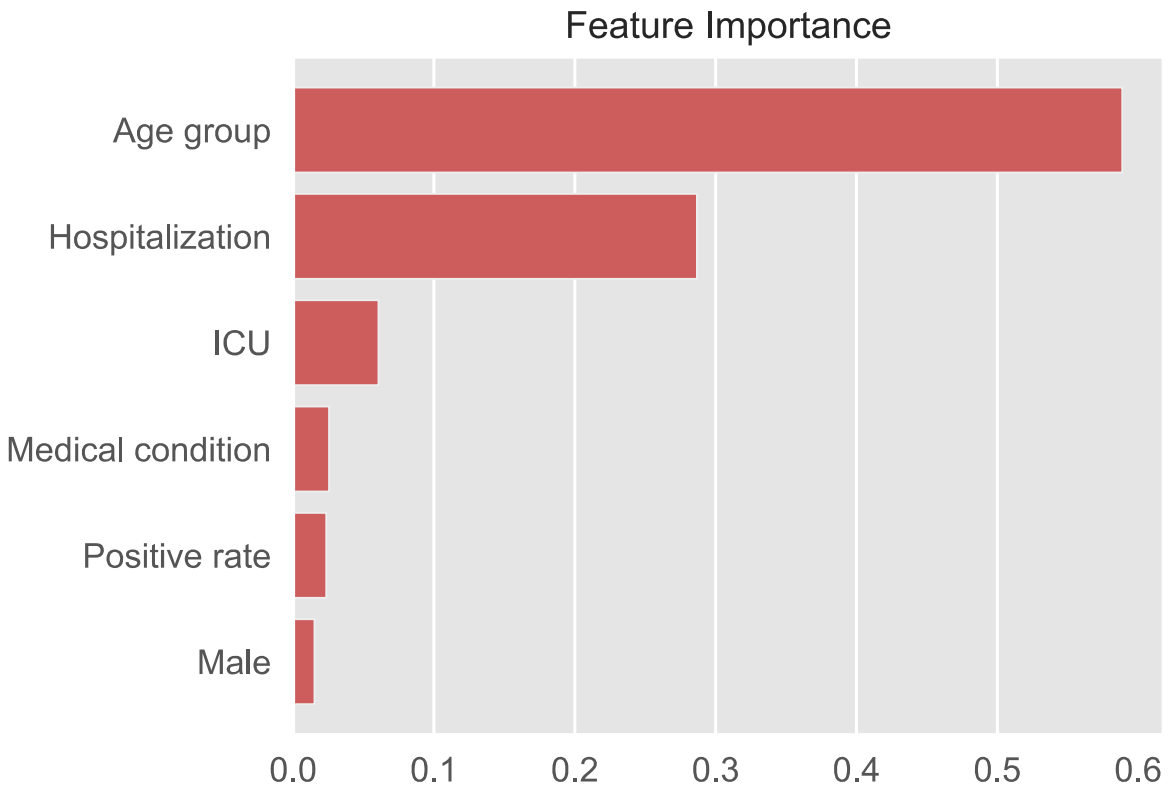
6 features

Final Model: XGBoost

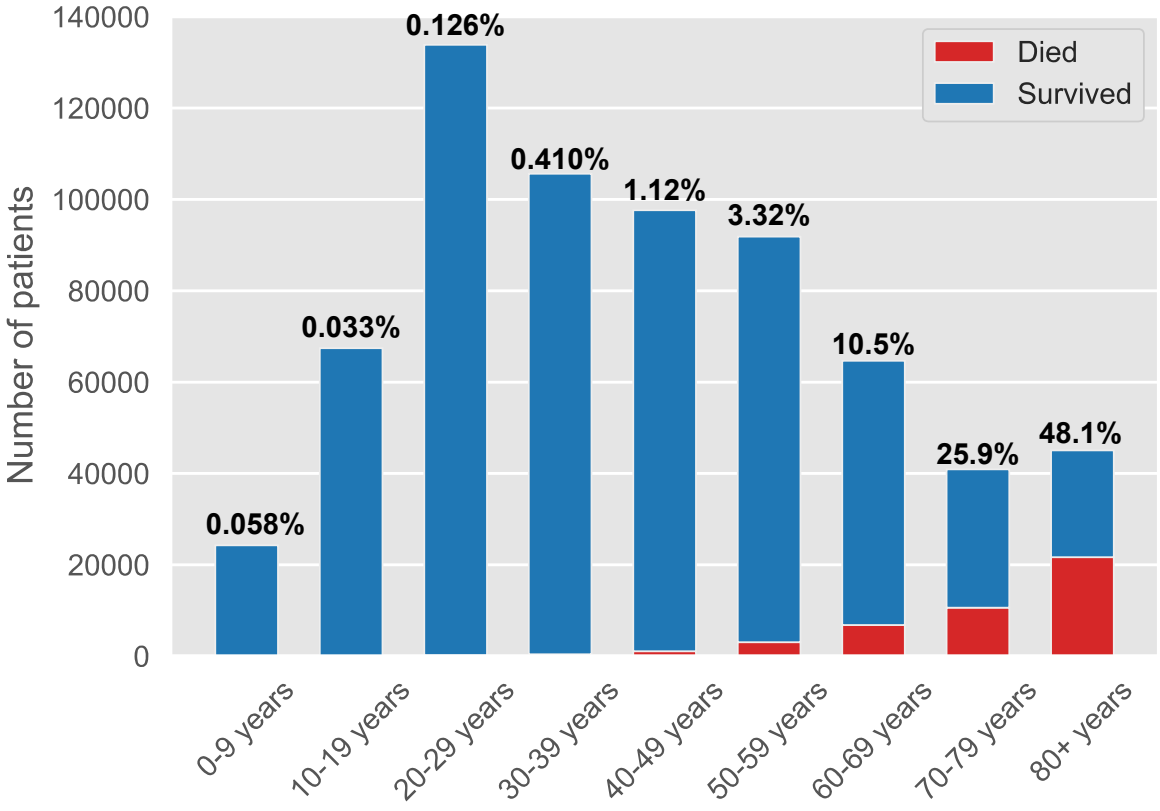
Metric	Test	Train
Recall	**0.939**	0.946
Precision	0.416	0.419
F-beta (beta=2)	**0.756**	**0.756**



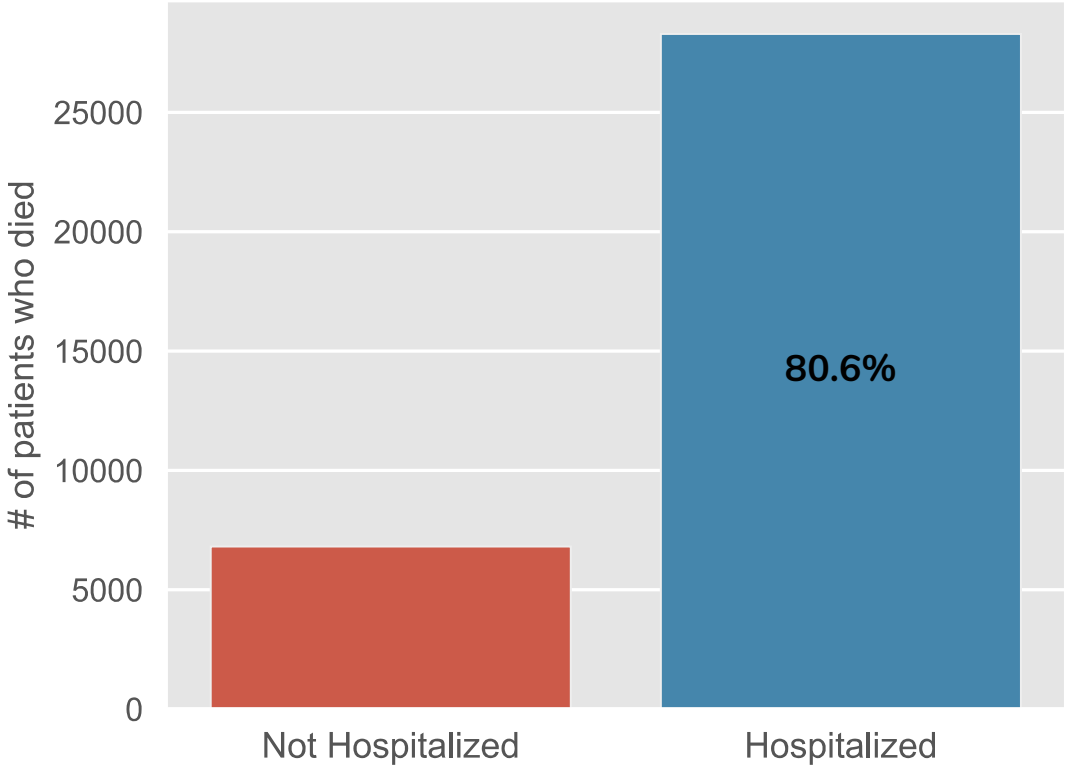
Conclusions



Outcome by Age Group



Hospitalizations Among Deaths



Application

Navigation

Explore?

- Home
- Predictor
- Data

Can we predict the outcome of a case of COVID-19?

Image source: Getty Images, via yalemedicine.org

Future Work

- Improve precision
  - More data
  - More features
- Geography
  - State/local level
- Application

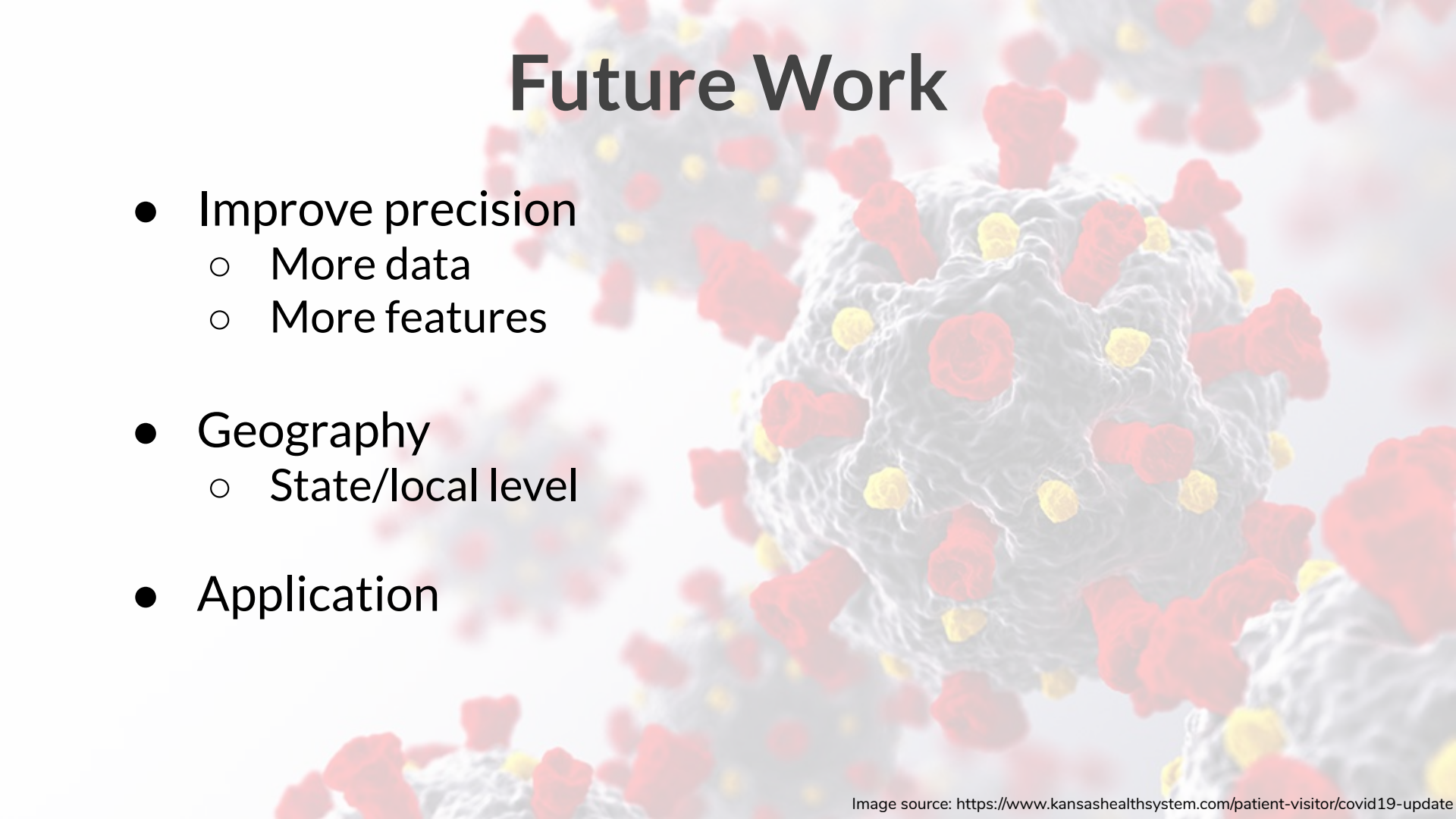


Image source: https://www.kansashealthsystem.com/patient-visitor/covid19-update

Questions?

Thank you!

Image source: https://www.kansashealthsystem.com/patient-visitor/covid19-update

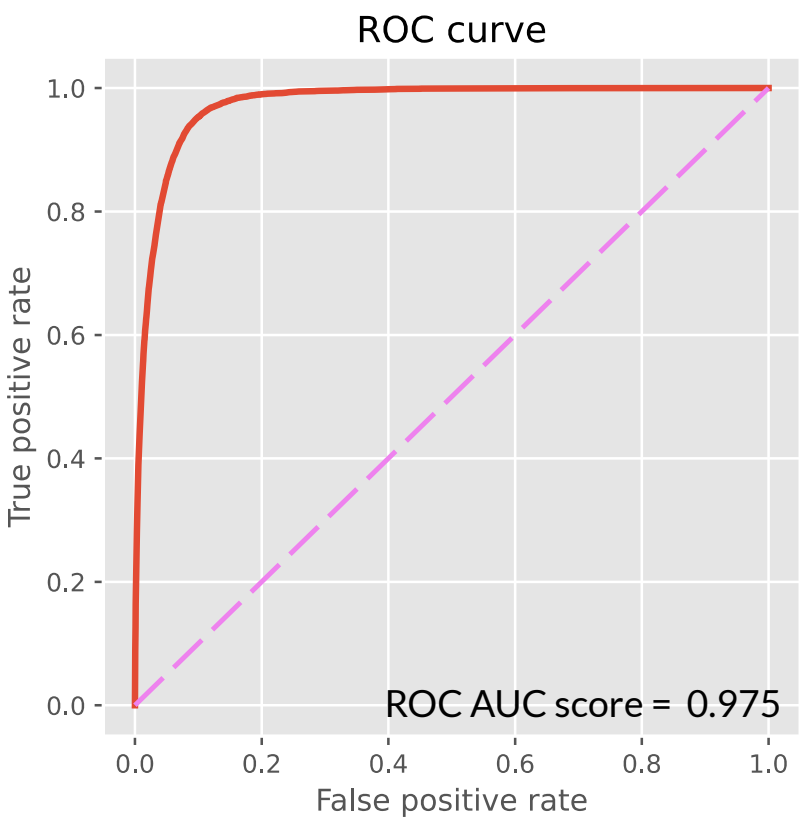
Appendix

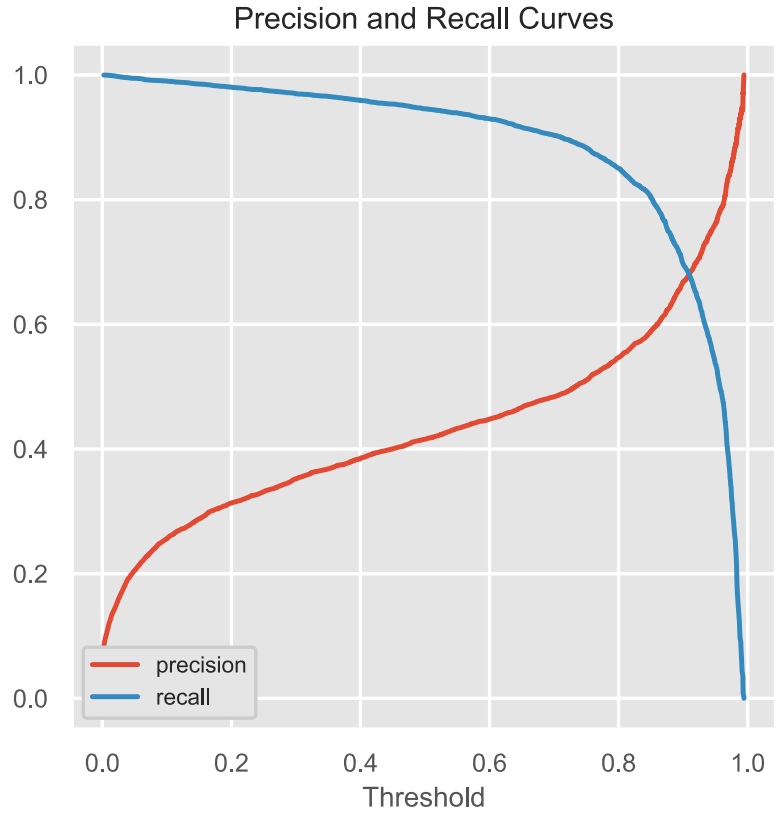
Metric	Test	Train
Accuracy	0.910	0.910
Recall	**0.939**	0.946
Precision	0.416	0.419
F1	0.580	0.576
F-beta (beta=2)	**0.756**	0.756

Classification report

	precision	recall	f1-score	support
0	1.00	0.91	0.95	125516
1	0.42	0.95	0.58	8771
accuracy			0.91	134287
macro avg	0.71	0.93	0.76	134287
weighted avg	0.96	0.91	0.93	134287

Appendix


ROC curve
True positive rate
False positive rate
ROC AUC score = 0.975
0.0
0.2
0.4
0.6
0.8
1.0


Precision and Recall Curves
precision
recall
Threshold
0.0
0.2
0.4
0.6
0.8
1.0