LONG-TERM MORTALITY PREDICTION BASED ON MIMIC-III CLINICAL DATABASE

YUQIN SHEN, TONG FU, MAGI LIU, XIANG LI

INTRODUCTION

CONTINUITY OF CARE

PROJECT GOAL

Data visulization

Long-term mortality prediction

Compare our model with the baseline model

MIMIC-III DATABASE

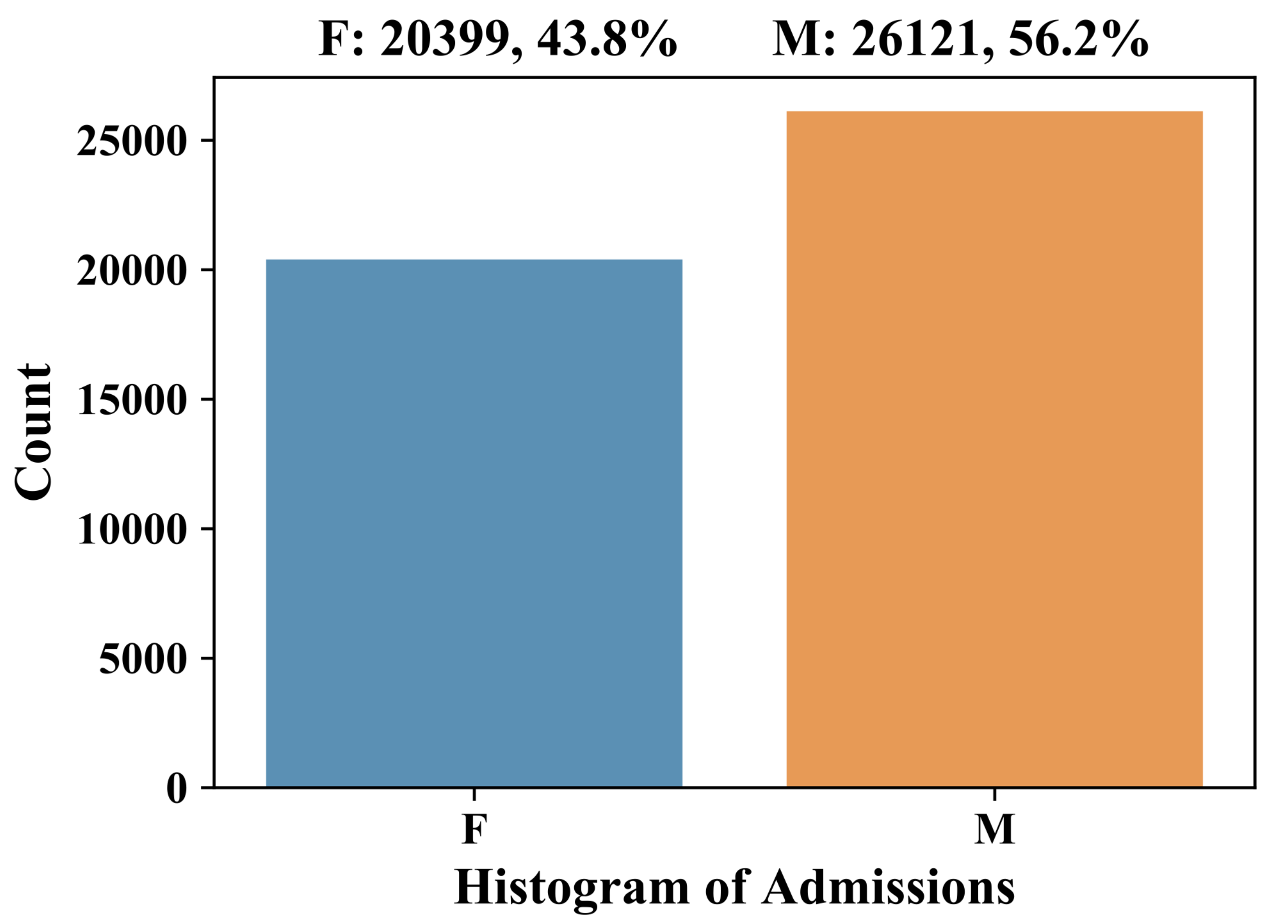
INPUTEVENTS	fluids administered to the patient
OUTPUTEVENTS	fluids excreted by / extracted from the patient
CHARTEVENTS	all charted data available for a patient
LABEVENTS	all laboratory measurements for a given patient

STEPS

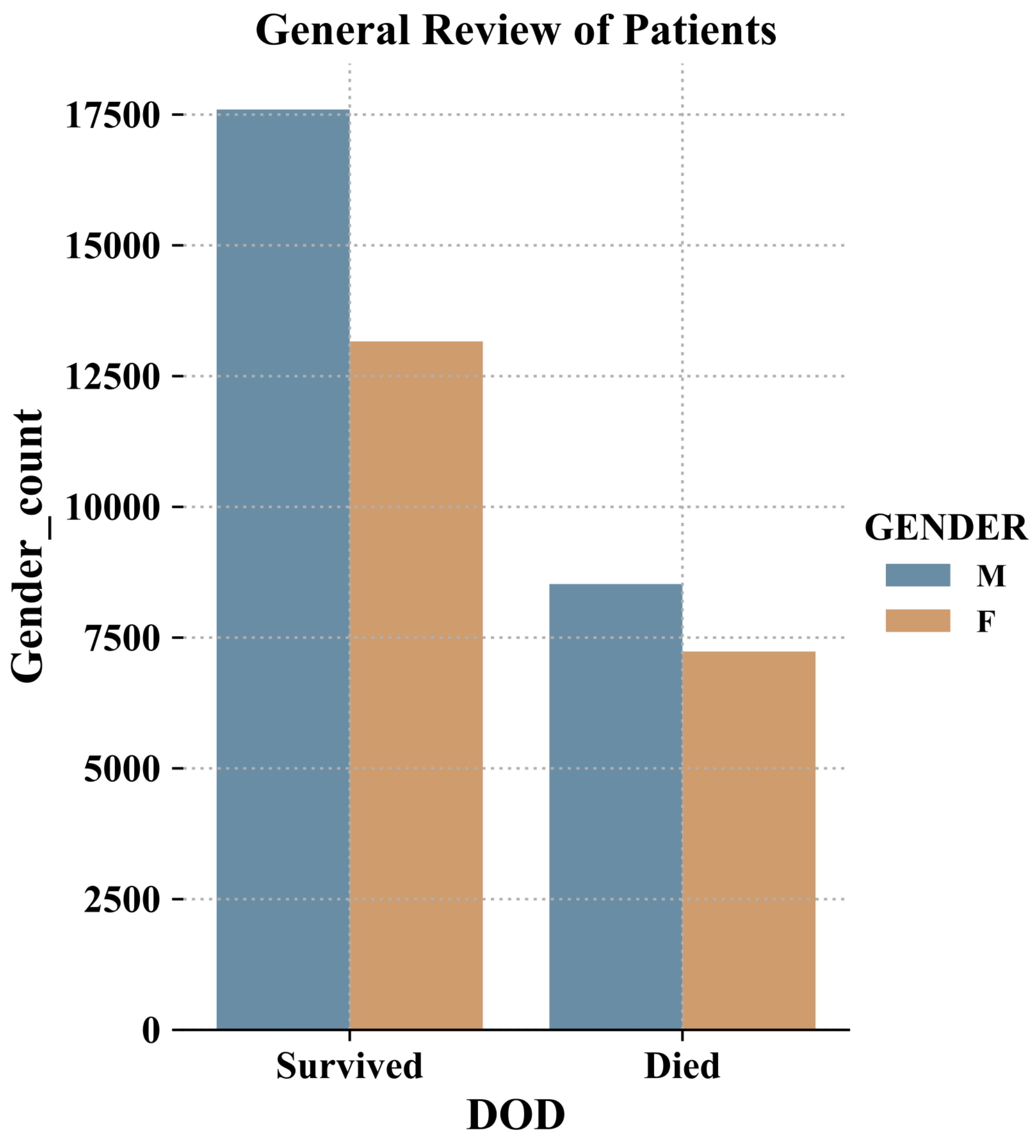
- Data Extraction & Cleaning
- Visualization
- Machine Learning Models
- Comparison

DATA VISUALIZATION

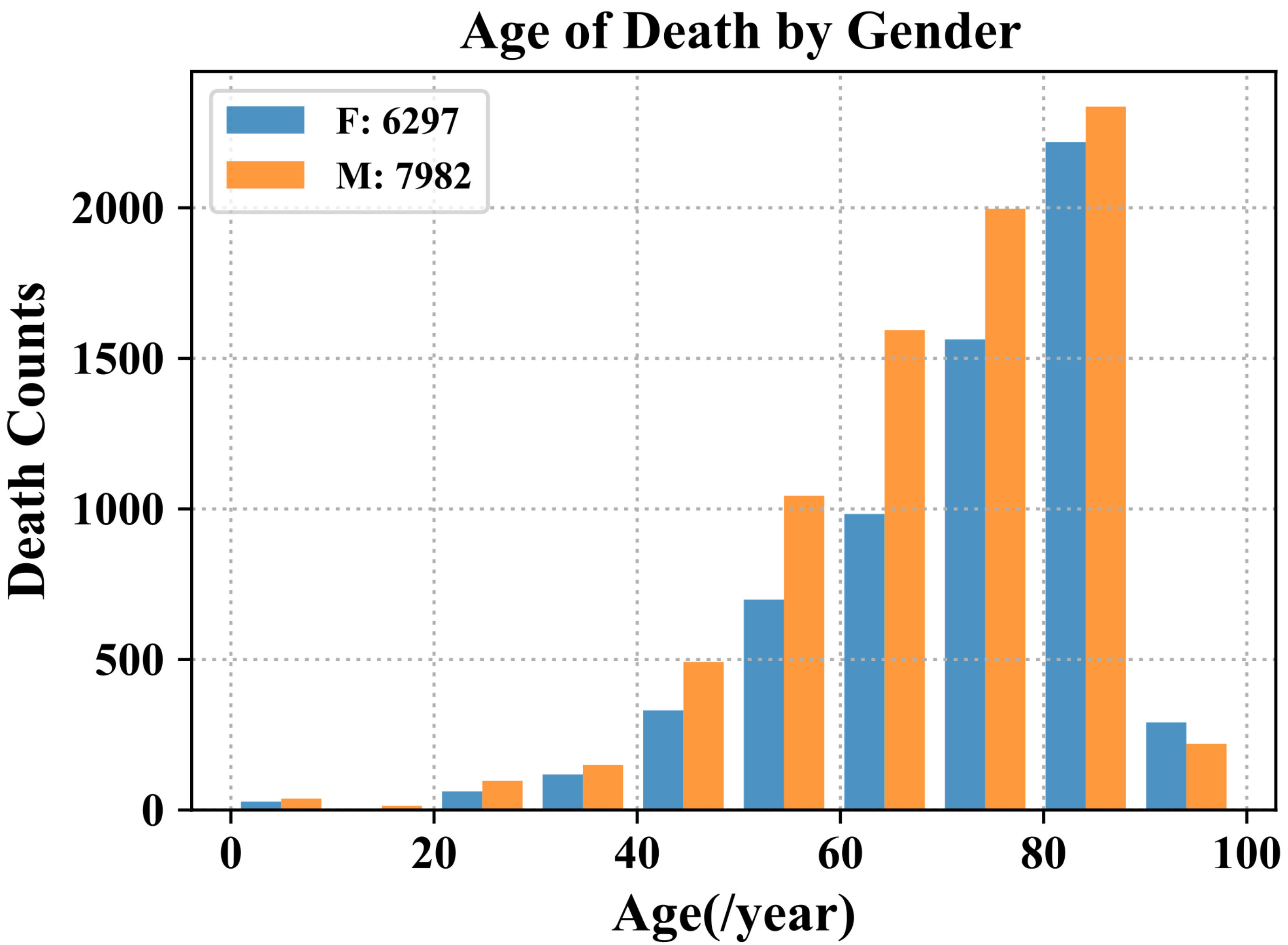
PATIENTS


F: 20399, 43.8%
M: 26121, 56.2%
Count
25000
20000
15000
10000
5000
0
F
M
Histogram of Admissions

PATIENTS


General Review of Patients
17500
15000
12500
10000
7500
5000
2500
0
Gender_count
Survived
Died
DOD
GENDER
M
F

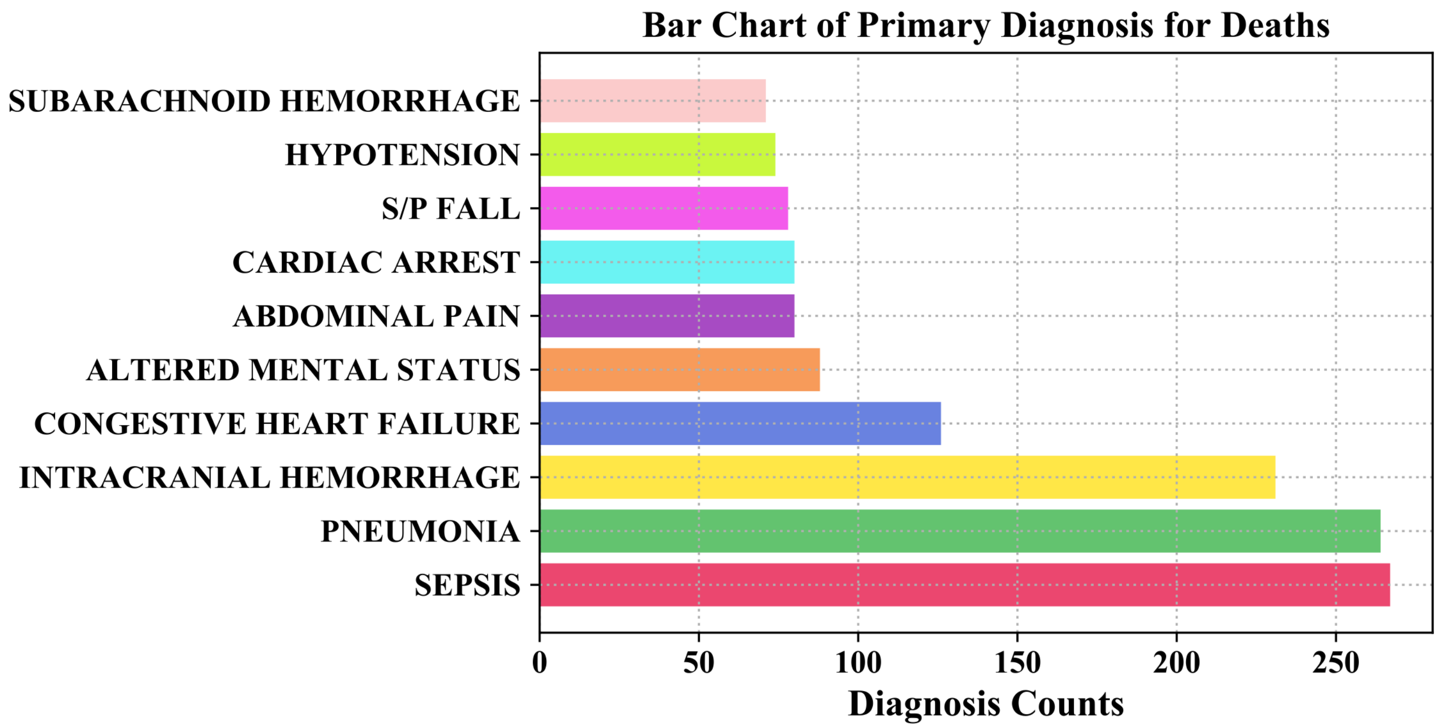
PATIENTS



- Fewer females died than males in each age range except for the 90-100 year old range.
- Most people died in between their 80s to 90s.

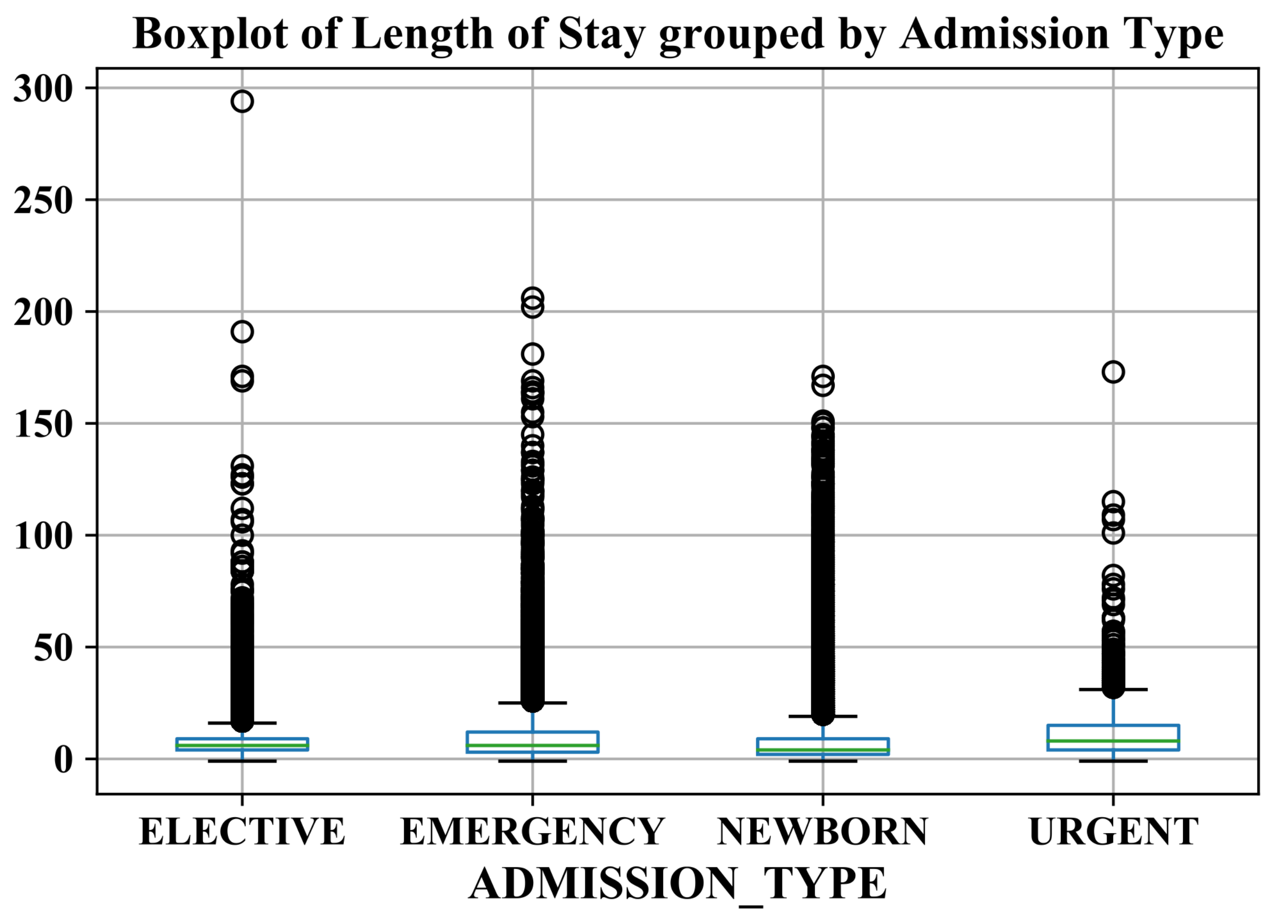
ADMISSIONS

- Top 10 leading diagnosis of death by their counts.
- Sepsis is the #1 cause of death.



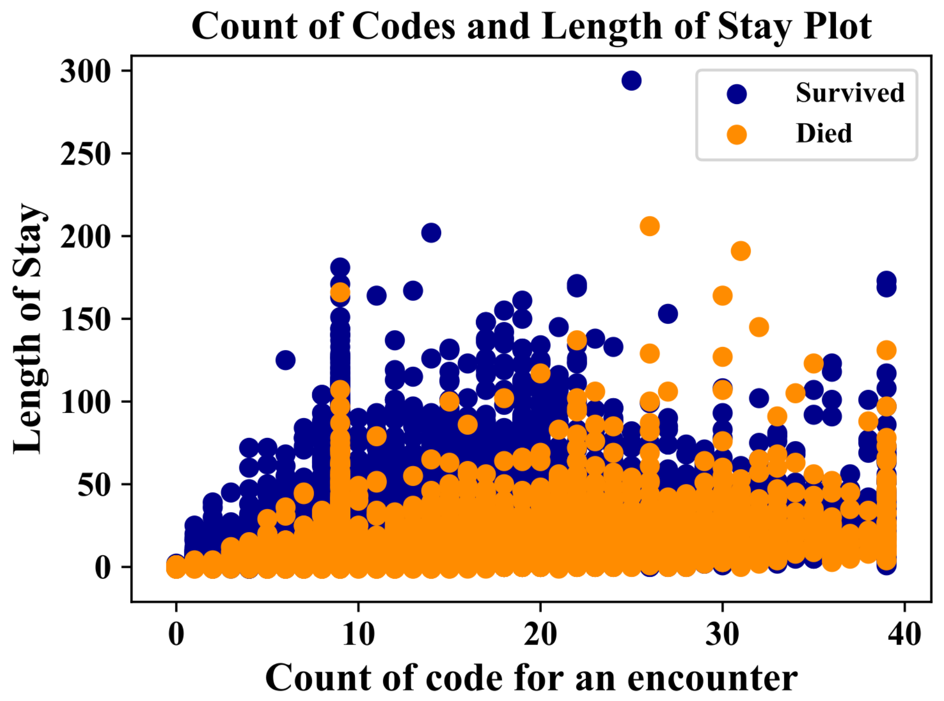
ADMISSIONS

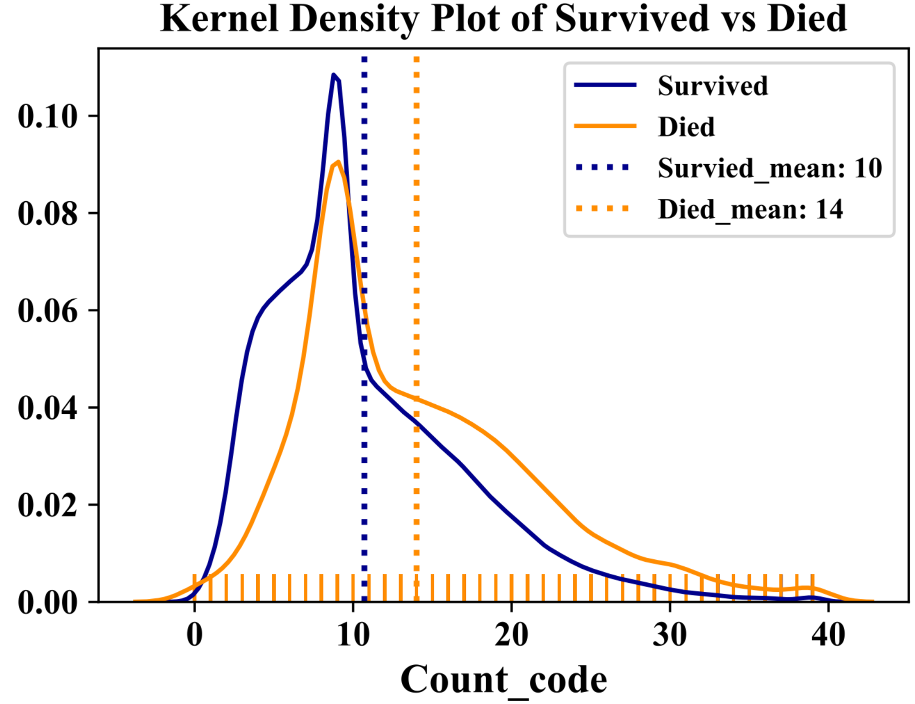
- Most patients were admitted to the hospital in the emergency category.
- "Urgent" has the least but the most sparse average length of stay.
- “Emergency” has the longest length of stay.



ADMISSIONS

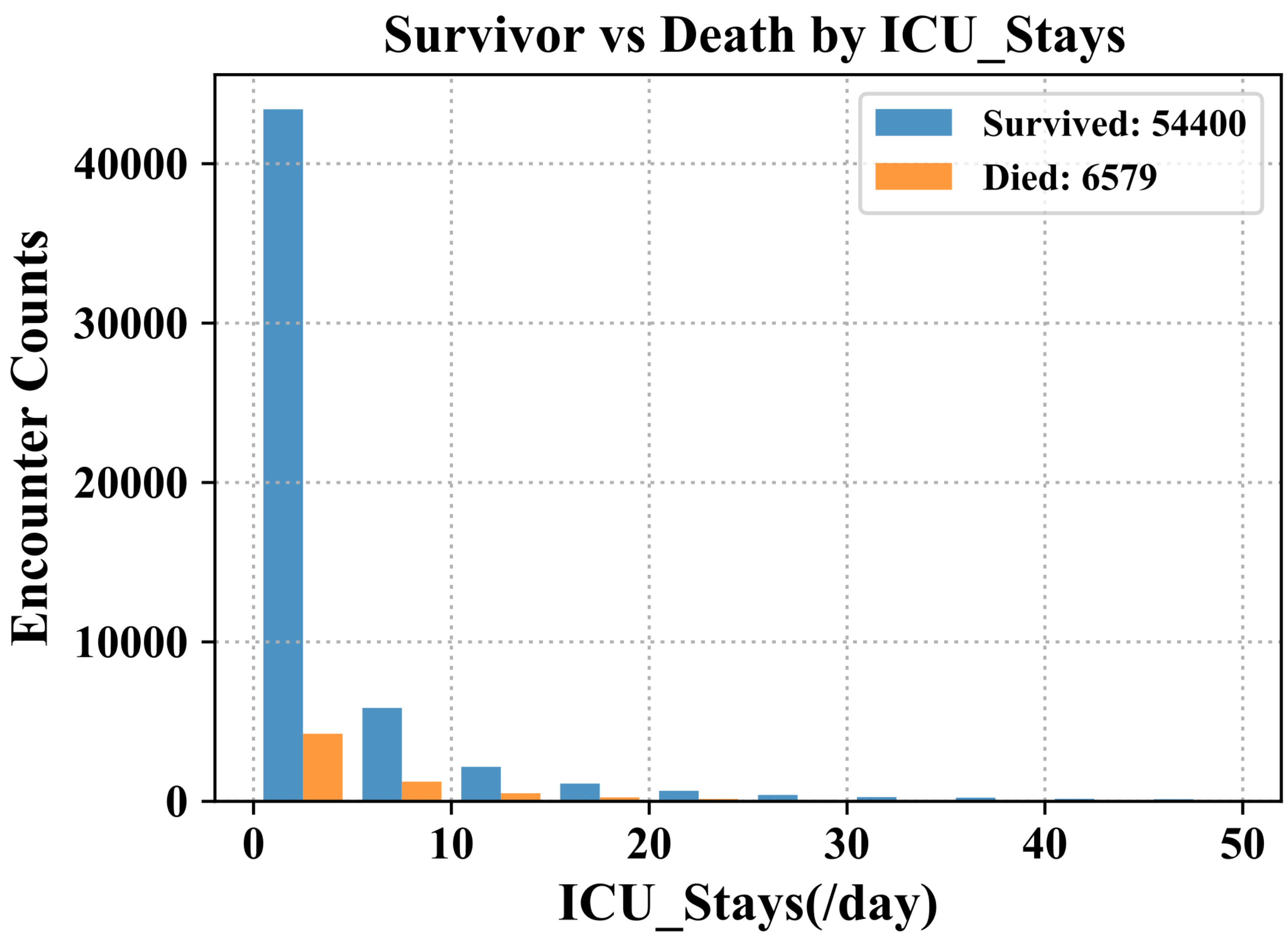
- If an encounter has fewer code, the patient has a higher probability to survive.
- In average, people who survived has a longer length of stay than people who died (more serious)





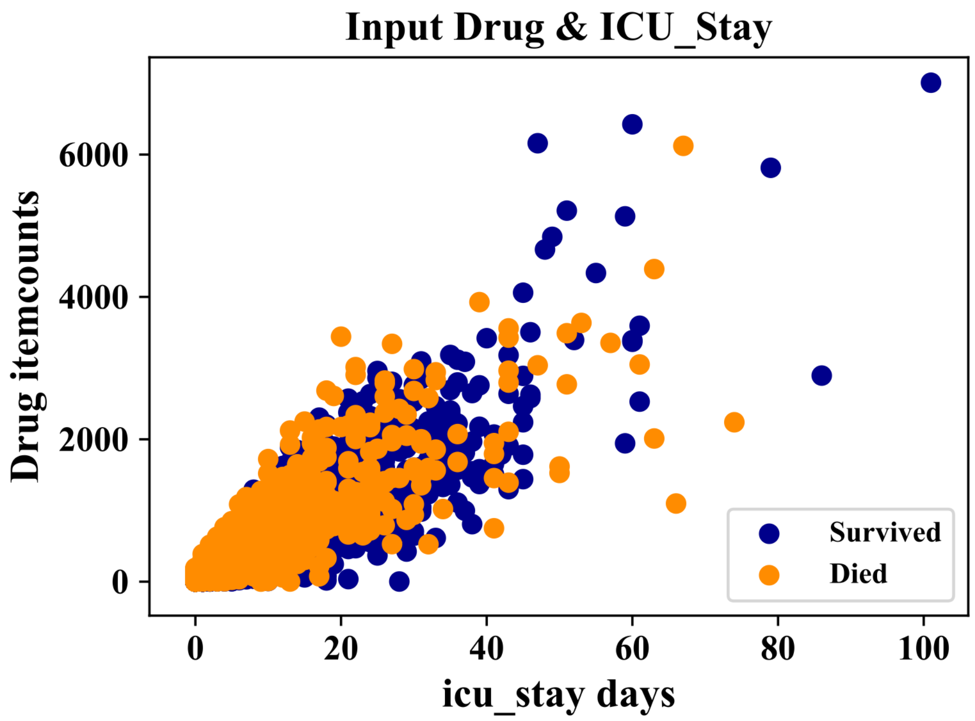
DRUG & ICU STAYS

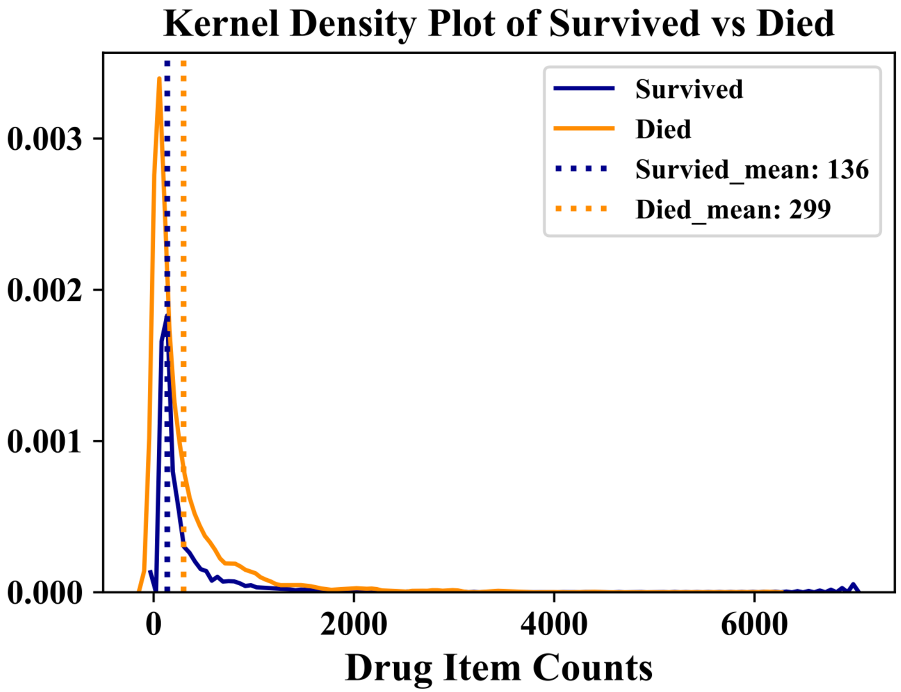
- More than one ICU encounters / admission
- Most patients, regardless of their mortality status, were in the range of 0-5 days of ICU stay.
- There were also very few people who died had more than 30 days of ICU stay.



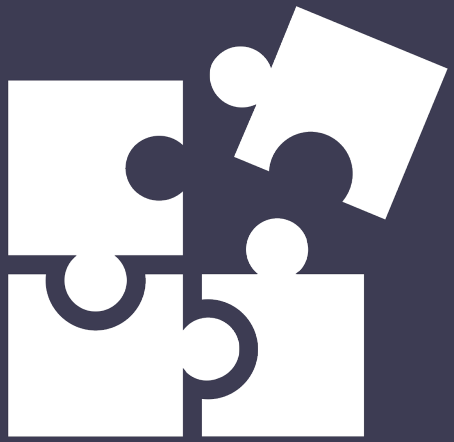
DRUG & ICU STAYS

- positive correlation - longer ICU stays renders more drug dispensed.
- Highly dense at the lower left corner, meaning that most people, regardless of whether they survived, have short ICU stays and few drug given.





DATA CLEANING



PATIENT SELECTION PROCESS

MIMIC III
N = 58976

Patient data < 1 year
Patient age > 0 year
N = 58600

Survived
N = 44073

Dead
N = 12657

DATASETS

I

39 raw features extracted from SAPSII

II

150 features selected

DATASET 1 PREPROCESSING

Select ITEMID, VALUE and Categorical data based on 17 features used in SAPS-II scoring system

Filter out data that were collected 1 year after the admission

Reshape data for analysis (39 features were selected in total)

DATASET 2 PREPROCESSING

SELECT SUBJECT_ID, HADM_ID, ITEMID, AMOUNT or VALUE from CHARTEVENT, OUTPUTEVENT, INPUTEVENT

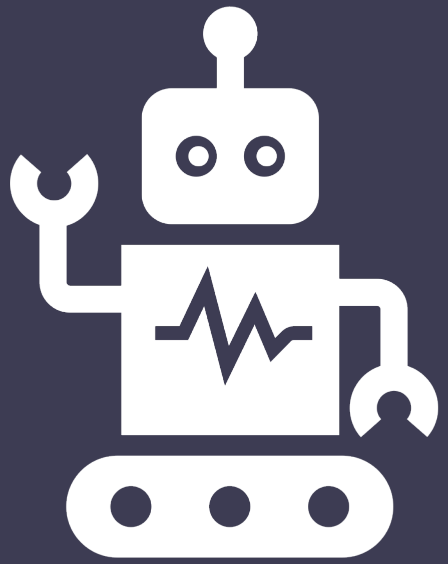
Filter out data that were collected 1 year after admission

Compute mean for time series data

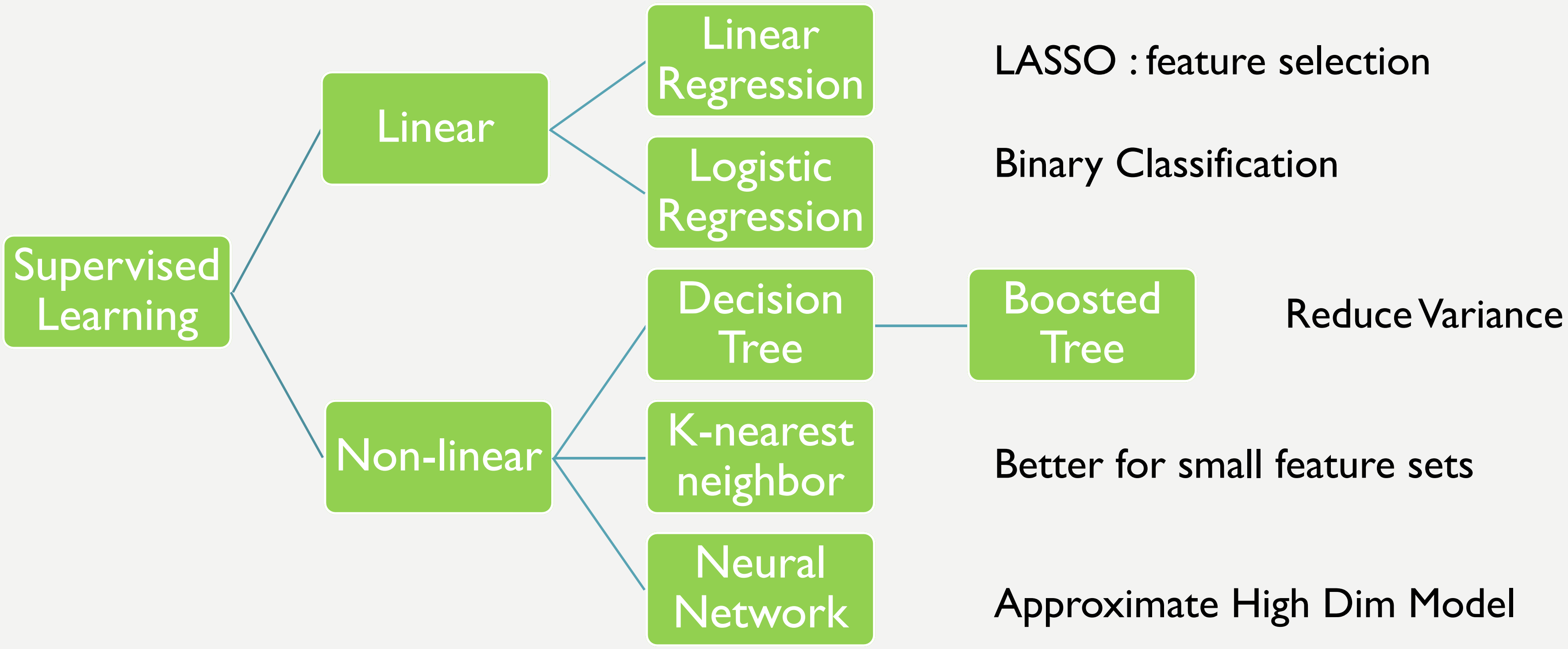
Select features with low missing rate at a threshold of 5%

Reshape data for analysis (150 features were selected in total)

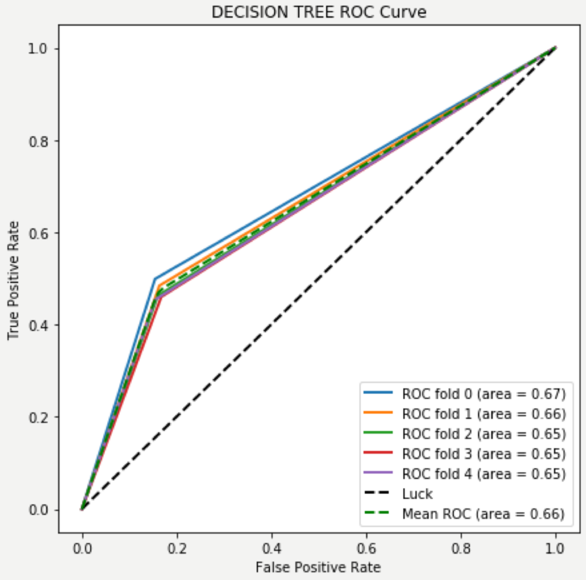
MACHINE LEARNING MODELS

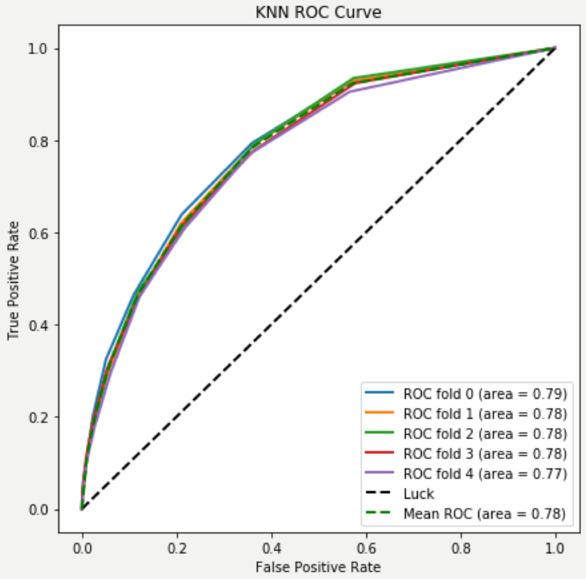


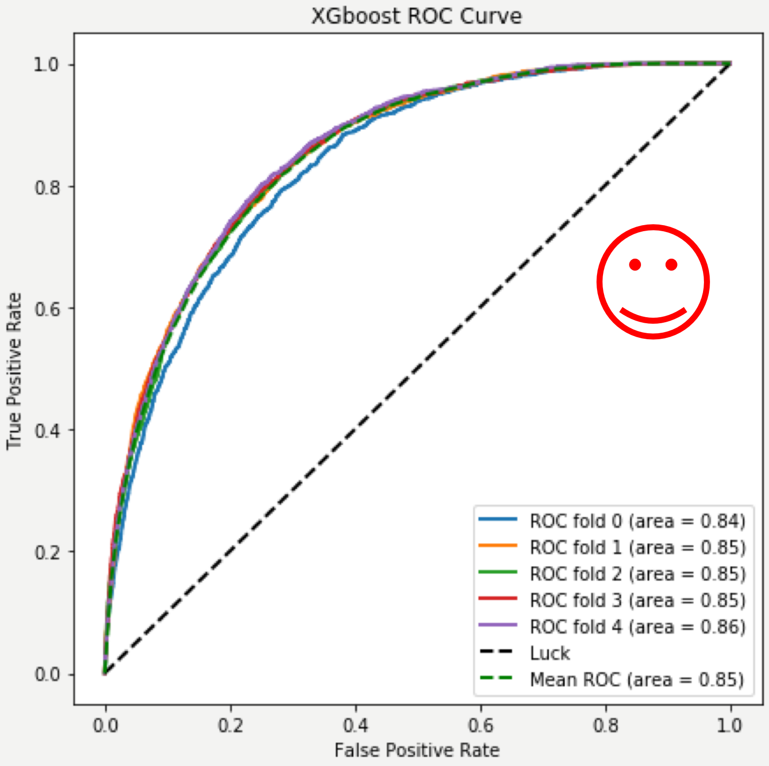
MODEL SELECTION

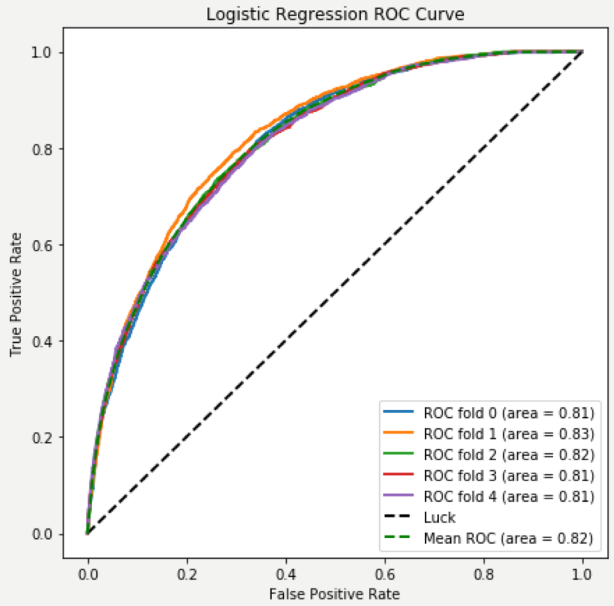


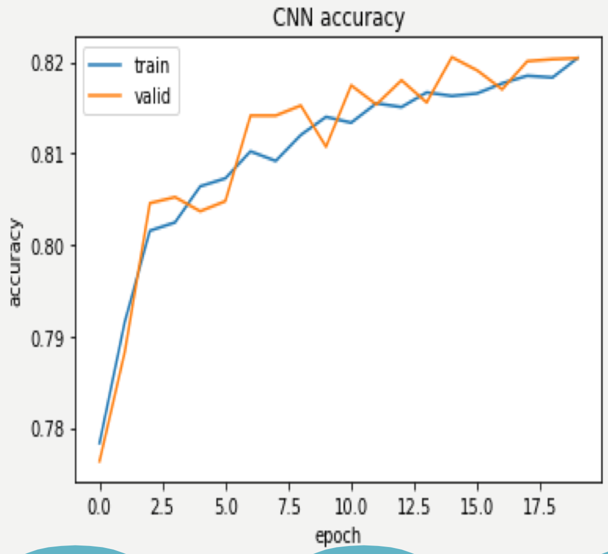
ROC CURVES


DECISION TREE ROC Curve
True Positive Rate
False Positive Rate
ROC fold 0 (area = 0.67)
ROC fold 1 (area = 0.66)
ROC fold 2 (area = 0.65)
ROC fold 3 (area = 0.65)
ROC fold 4 (area = 0.65)
Luck
Mean ROC (area = 0.66)

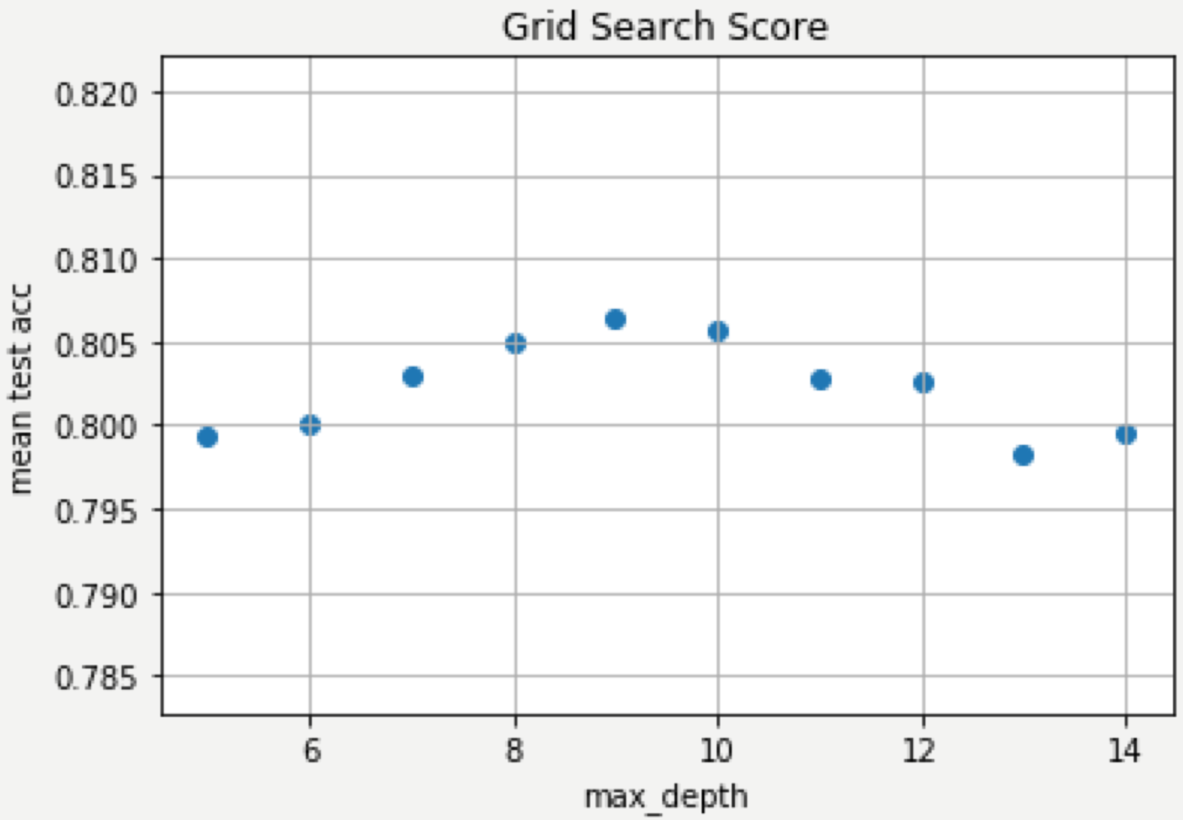

KNN ROC Curve
True Positive Rate
False Positive Rate
ROC fold 0 (area = 0.79)
ROC fold 1 (area = 0.78)
ROC fold 2 (area = 0.78)
ROC fold 3 (area = 0.78)
ROC fold 4 (area = 0.77)
Luck
Mean ROC (area = 0.78)


XGboost ROC Curve
True Positive Rate
False Positive Rate
ROC fold 0 (area = 0.84)
ROC fold 1 (area = 0.85)
ROC fold 2 (area = 0.85)
ROC fold 3 (area = 0.85)
ROC fold 4 (area = 0.86)
Luck
Mean ROC (area = 0.85)


Logistic Regression ROC Curve
True Positive Rate
False Positive Rate
ROC fold 0 (area = 0.81)
ROC fold 1 (area = 0.83)
ROC fold 2 (area = 0.82)
ROC fold 3 (area = 0.81)
ROC fold 4 (area = 0.81)
Luck
Mean ROC (area = 0.82)


CNN accuracy
train
valid
accuracy
epoch

HYPERPARAMETER SELECTION



DecitionTreeClassifier*(criterion='gini', splitter='best', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, class_weight=None, presort=False)*

Decision Tree for Dataset1

Best 'max_depth' : 7

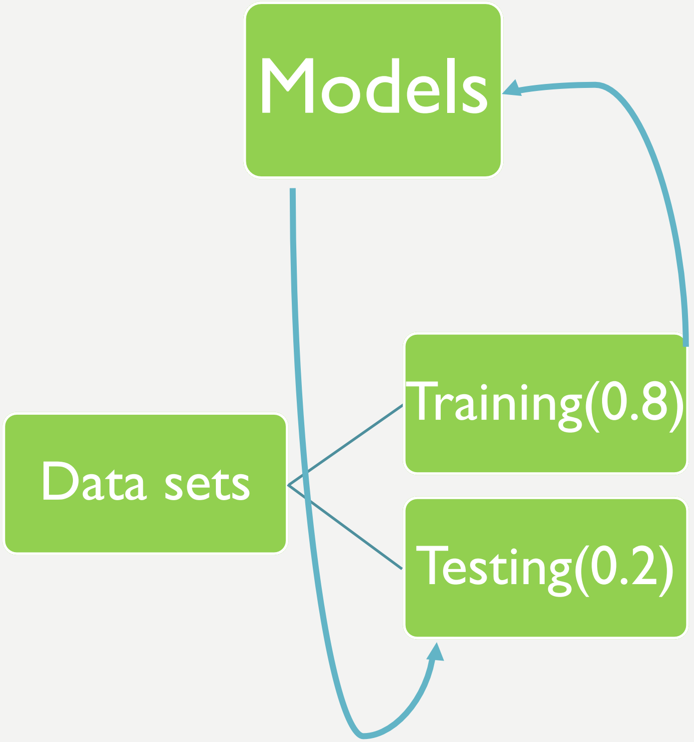
Grid Search + Cross Validation

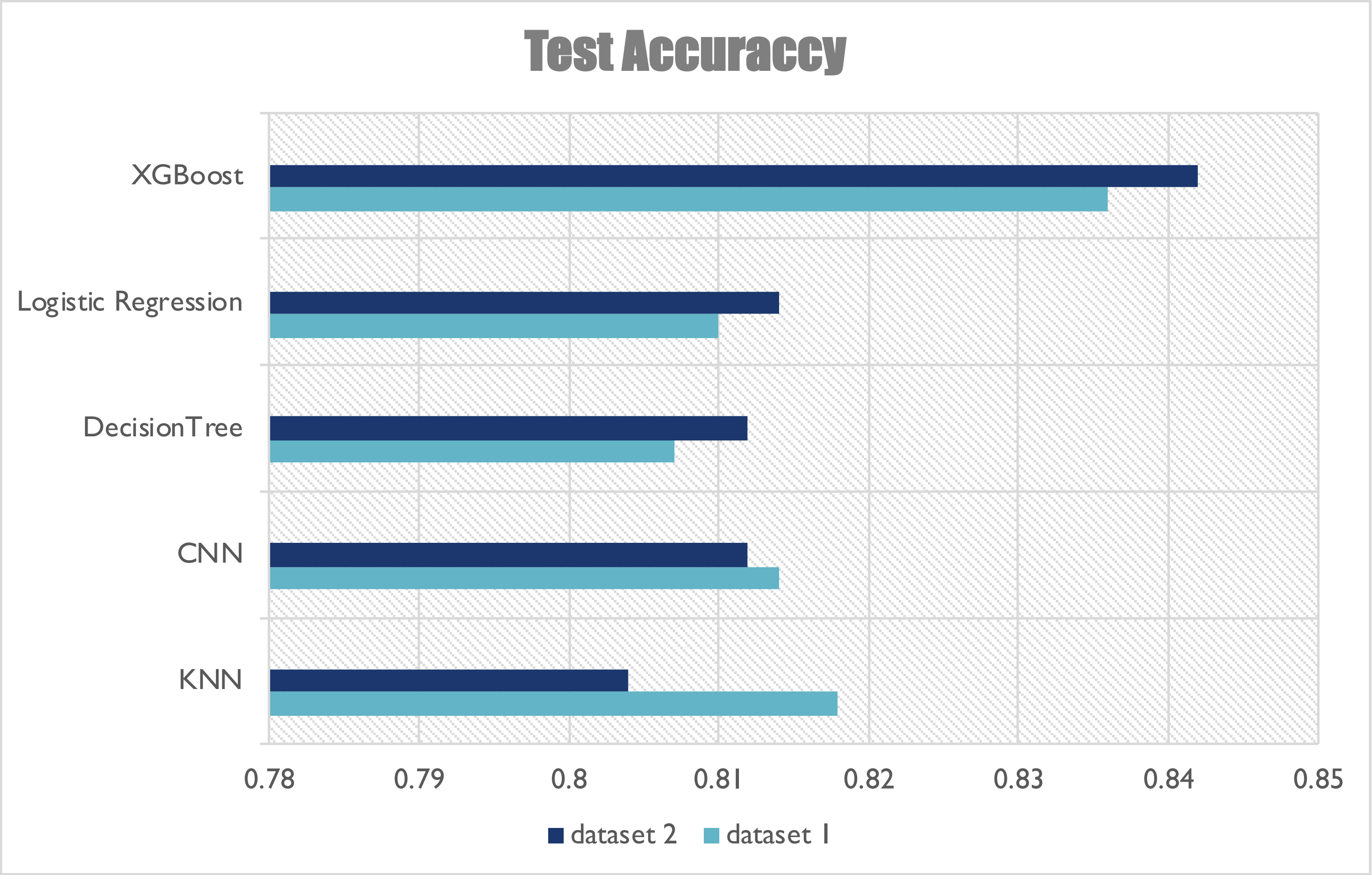
MODEL COMPARISON

- **Decision tree**
  - Min samples split: 2
  - Max depth: 9
- **XGBoost**
  - Max depth: 7
  - Subsample: 0.5
  - Learning rate: 0.1
- **CNN**
  - Epoch: 20
  - Units: 32
  - Optimizer: Adam
- **Logistic Regression**
  - C: 0.1
- **KNN**
  - K: 4

Hyperparameters for tuning:
Prevent Overfitting
Reduce Variance
Reduce Bias
Improve Training Speed

MODEL EVALUATION





SUMMARY & IMPROVEMENT

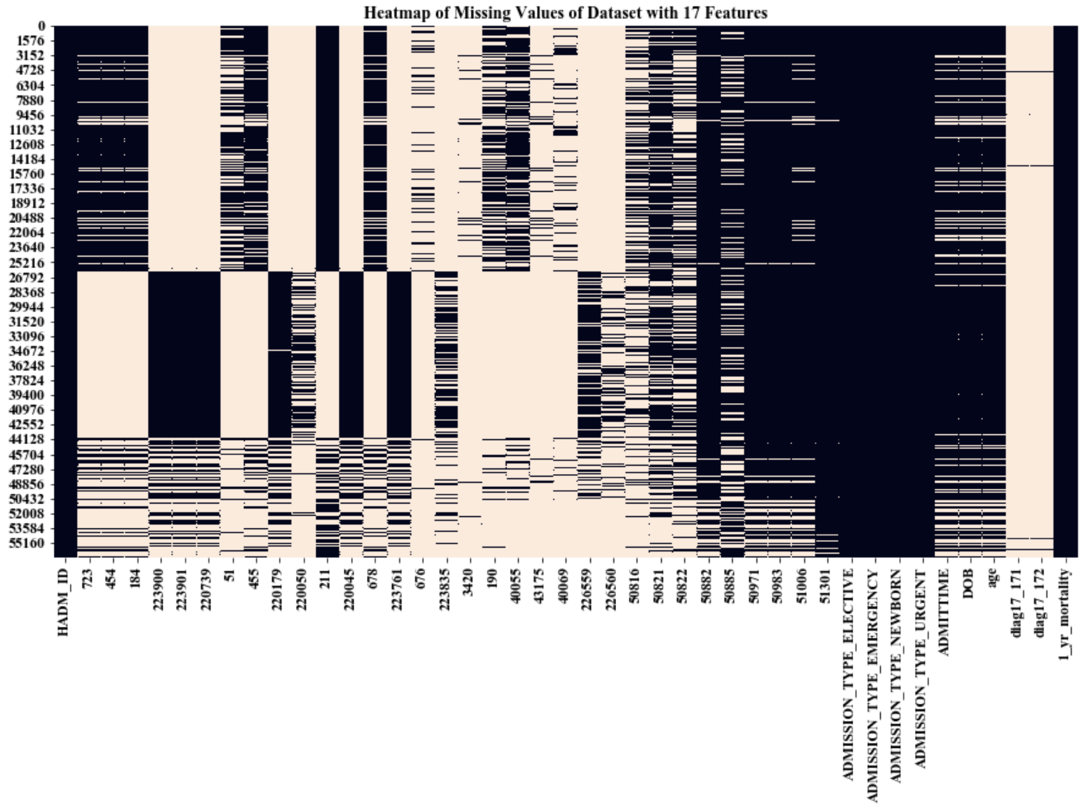
Important Features (GAIN>0.016)	Patient Die(1586) [mean,dev]	Patient Survive(9760) [mean,dev]	Nan (%)	Gain
Morphine Sulfate	12.3,35.8	6.3,11.0	0.979	0.022
Morphine Sulfate	8.7,14.1	2.9,9.4	0.918	0.035
Chest Tubes CTICU CT 1	35.8,26.3	31.9,17.2	0.919	0.021
Chest Tube 1	116.9,208.5	45.1,73.6	0.942	0.019
Verbal Response	3.1,1.6	4.0,1.2	0.521	0.026
50983?	138.7,4.465	138.8,3.3	0.091	0.053
ADMISSION_TYPE_ELECTIVE	0.05,0.23	0.1,0.3	0	0.153
ADMISSION_TYPE_EMERGENCY	0.91,0.28	0.7,0.5	0	0.061

Improvement: Feature Selection

Dimension Reduction:
PCA/LASSO/Variable Importance

Data Imputer
Data Correlation
Data Density

Doctor Opinion



Missing Value Heat Map
(17 features)

Missing Value Heat Map
(150 features)

SUMMARY & IMPROVEMENT

- ❑ Recommend using XGboost model to predict 1- year mortality with 39 feature sets
- ❑ Optimize patient selection process
- ❑ Better handling on time series data

APPENDIX

Value

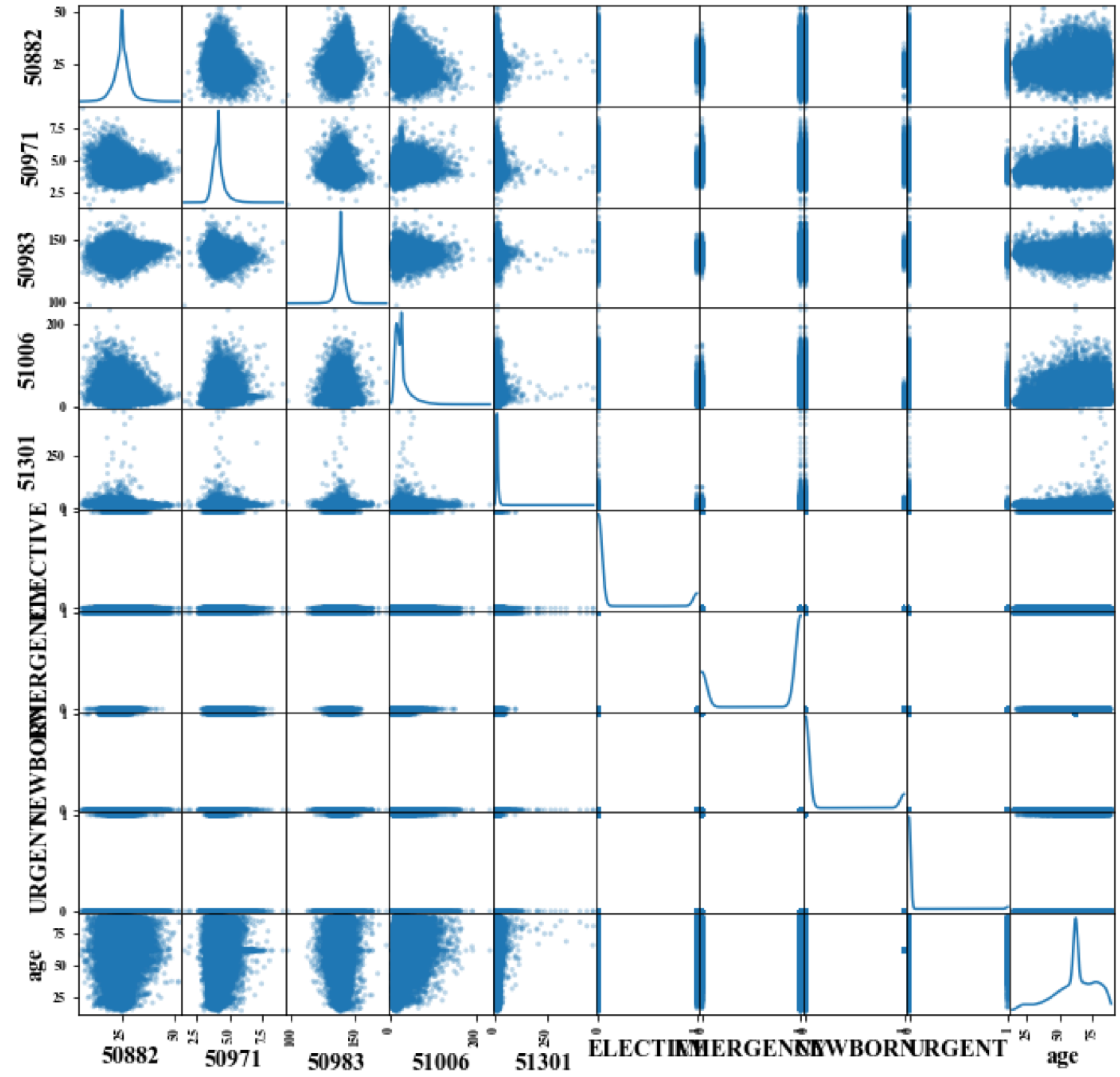
Feature item ID

1 year mortality

subject

51	184	190	211	442	454	455	676	678	723	3420	3422	6701	220045	220050	220179	220739	223761	223762	223835	223900	223901	1_yr_mortality
117.79279	3.3448276	0.4112903	92.376471		5.5	115.01538	36.951042	97.262501	2.3793103													1
	4	0.725	90.354839		6	123		98.066667	5													0
153.60526	3.9259259		86.549451		6	153.84483		97.8	5													0
			137.92857							31.285714	28.166667											0
149.71649	1	0.535	85.994819		1.9078947			97.637838	1													1
			152.24623																			0
	4		84.948718		5.8823529	106.71053		98.277779	4.9411765													1
152.66495	2.9565217	0.4957447	87.669767		5.057971	146.19048		99.028302	1.942029													1
133.18182	3.7826087	0.56	98.017241		5.6086957	107.25	37.511842	98.824999	4.3043478													0
106.86207	3.5263158	0.6666667	88.5		5.4736842	110.75	37.33	98.145455	3.9473684													0
	3.7142857		107.66667		5.7142857	142.4		99.385715	4.4285714													0
	4		98.416667		6	146.70833		98.371428	4.2													0
129.5	3	0.625	79.121951		4.3333333	118.33333	36.561765	96.800003	3													0
118.14737	3.4266667	0.7	76.756098		5.8157895	101.75		97.69011	4.5921053													1
	3.3333333		101.875		6	141.53846		98.5	2.3333333													0
104.5	3.1428571	0.6	90.139535		4.5714286	92.916667	37.21	98.25	3.8571429				76.694444	129.29412	107.0625	4	97.342857			5	6	0
139.33333	4		67.923077		6	119.15385		97.566666	5													0
116.5	4		76.321839		6	104.63636		98.470589	5													0
	4		70.58		6	98.234043		98.889999	5													0
123	3.4285714	0.64	89.03125		5.2857143	99.5	37.033333		3.8571429													0
	4		78.090909		6	108.8		97.433333	5													0
102.55191	1.0173913	0.3673913	54.814208		1.4173913	94		98.86087	1													1
114.45263	3.8695652	0.8222222	88.301471		5.9347826	120		98.671429	4.6521739													0
	4		64.324324		6	115		98.427272	5													0
	4	1	80.851852		5.9	130.81481		98.5	4.9				55.864865		128.05556	3.5555556	97.375		60	4.5555556	5.8888889	0
	4	1	80.851852		5.9	130.81481		98.5	4.9				55.864865		128.05556	3.5555556	97.375		60	4.5555556	5.8888889	1
114.46626	3.6969697	0.5727273	74.04321		5.8181818	121	37.321818	98.563636	3.9090909													0
110.6092	2.5833333	0.5217021	98.849421		3.9375	111.42609	37.076563	98.738236	2.75				85.586538		127.95	3.5	99.076		57.586207	2.0909091	5.9545455	0
	3.8571429		81.09375		6	123.16667		99.366667	4.4285714													0
106.73034	3.3733333	0.4187135	88.505956		5.4933333	94.774359		99.335443	1.32													0
			131.91753							83.523364	30.85											0
141.62838	3.8852459	0.4638889	74.012903		5.9672131	135.23077		97.845454	4.2131148													1
	4		72.958333		6	117.19565		97.14875	5													0

SCATTER MATRIX



REFERENCE

https://www.mcall.com/health/mc-inspire-health-how-important-is-continuity-of-care-story.html

Sudhakar-Krishnan, V., & Rudolf, M. C. (2007). How important is continuity of care?. Archives of disease in childhood, 92(5), 381–383. doi:10.1136/adc.2006.099853

Jeffers, H., & Baker, M. (2016). Continuity of care: still important in modern-day general practice. The British journal of general practice : the journal of the Royal College of General Practitioners, 66(649), 396–397. doi:10.3399/bjgp16X686185

Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/content/101/23/e215.full]; 2000 (June 13).

MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available from: http://www.nature.com/articles/sdata201635

https://arxiv.org/pdf/1710.08531.pdf

https://www.researchgate.net/publication/308007023_Mortality_prediction_in_the_ICU_based_on_MIMIC-II_results_from_the_Super_ICU_Learner_Algorithm_SICULA_project

Pirracchio, Romain. "Mortality Prediction in the ICU Based on MIMIC-II Results from the Super ICU Learner Algorithm (SICULA) Project." Secondary Analysis of Electronic Health Records, 2016, pp. 295–313., doi:10.1007/978-3-319-43742-2_20.

Doell, Linda, and Linda Doell. "How Important Is Continuity of Care?" Mcall.com, The Morning Call, 28 Mar. 2019, www.mcall.com/health/mc-inspire-health-how-important-is-continuity-of-care-story.html.

QUESTIONS