Predict ICU Admission For Patients With Specific Diseases

Project Walkthrough

Objectives

- Timely decision of ICU (Intensive Care Unit) admission for patients. Physicians rely on guidelines or scoring systems to evaluate the need for ICU admission. However, disease specific guidelines for intensive care do not reflect real-world decision making. This project aims at using machine learning algorithms to predict the need for ICU admission for patients with specific diseases within 24 hours of admission.
- This walkthrough material presents results using randomly generated dataset, the python code in the directory demonstrates the methods behind a real-world project using electronic health records.

Disease Relevant Features

Dimension	Item	Data Type	Remark
Demographics	Age, Sex	Numerical, Categorical	*Use age at admission
Comorbid Conditions	Comorbid Condition 1, Comorbid Condition 2, Comorbid Condition 3, Comorbid Condition 4, Comorbid Condition 5	Categorical (Y, N)	*Comorbid conditions categorized based on ICD (International Classification of Diseases) Codes
Pathogens	Pathogen A	Categorical (Y, N)	Patients with no positive results within admission plus 1 minus 3 days are considered negative.
Lab Exams	LabItemA, LabItemB, LabItemC	Numerical	Data sampled within admission plus 1 minus 3 days. Missing value replaced with median.
Vital signs	Body Temperature (Maximum), Pulse (Maximum), Systolic Pressure (Minimum), Systolic Pressure (Minimum)	Numerical	Recorded within admission plus/ minus 1 day. Missing value replaced with median

* Feature selected by domain experts.

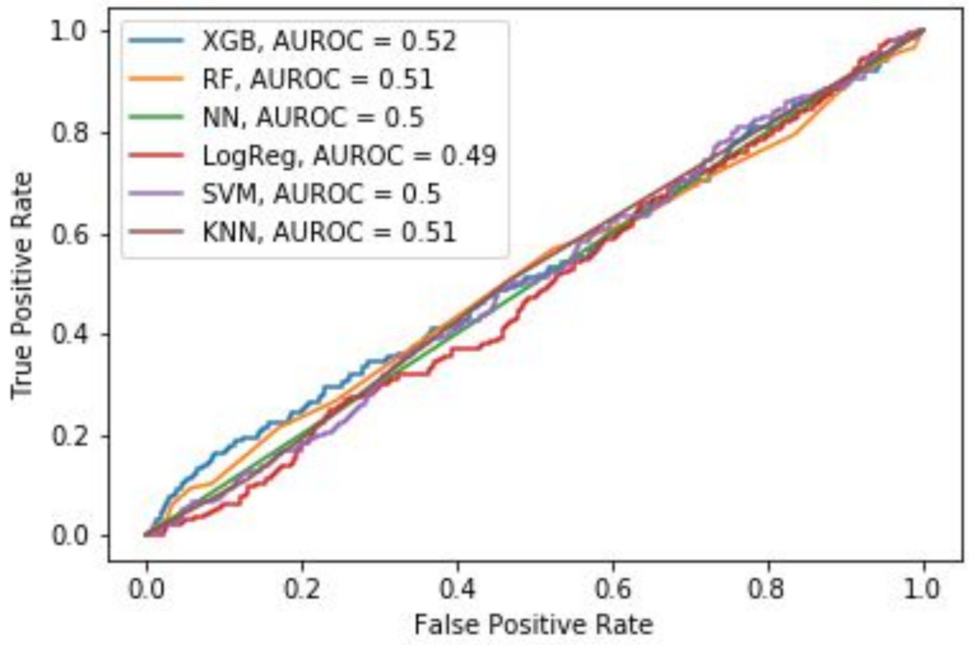
Predict ICU Admission .

Model	Accuracy	Precision	Recall	F1-Score	AUROC
Boosted Trees	.88	.20	.01	.02	.52
Random Forest	.89	.50	.01	.01	.51
Feedforward Neural Net	.89	.00	.00	.00	.50
Logistic Regression	.89	.00	.00	.00	.49
SVM	.89	.00	.00	.00	.50
KNN	.87	.11	.02	.03	.51

Results gained on test dataset, model tuned on training set with cross validation.

Parameters
XGBoost: max depth= 3, learning rate= 0.4.
Random Forest: # of estimators = 50, gini impurity index, minimum sample per leaf = 1.
Feedforward Neural Net: Layer = 3 (ReLU hidden layer, nodes = 3, 3, Sigmoid for output layer), Adam optimized, epoch = 100, learning rate = .01
Logistic Regression: threshold =.5, L2 regularized (lambda=1), SVM: squared hinge loss, soft margin, C = 1, L2 regularized. KNN: K =5, standard normalized.

The Receiver Operating Characteristic Curves



Exploratory Analysis

Demographics And Severity

	ICU Admission 24H	Not ICU
Admissions	1,183	8,321
Male (%)	579 (48.9)	4,138(59.7)
Age Median (IQR)	14.3 (10.0-19.1)	14.5 (10.0-19.2)
In Hospital Days Median (IQR)	14 (7-22)	15 (8-23)

BACK

N = 9,504
Data cut date: 2001.01-2010.12

Comorbid Conditions

	ICU Admission 24H (N = 1,183)	**Not ICU** (N = 8,321)
Comorbid Condition 1 # (%)	192 (16.2)	1,237 (14.9)
Comorbid Condition 2 # (%)	109 (9.2)	856 (10.3)
Comorbid Condition 3 # (%)	324 (27.4)	2,370 (28.5)
Comorbid Condition 4 # (%)	249 (21.0)	1,867 (22.4)
Comorbid Condition 5 # (%)	315 (26.6)	2,193 (26.4)

BACK

N = 9,504
Data cut date: 2001.01-2010.12

Exam And Pathogen Results

	ICU Admission 24H (N = 1,183)	**Not ICU** (N = 8,321)
Lab Item A median, IQR	1.0 (0.5-1.5)	1.0 (0.5-1.5)
Lab Item B median, IQR	10.3 (6.5-14.3)	10.6 (6.9-14.3)
Lab Item C median, IQR	139 (129-150)	140 (129-150)
Pathogen A # (%)	40 (3.4)	281 (3.4)

BACK

N = 9,504
Data cut date: 2001.01-2010.12

Vital Signs

	ICU Admission 24H (N = 1,183)	**Not ICU** (N = 8,321)
Body Temperature (Max), Median (IQR)	35.7(35.3-36.2)	35.7(35.3-36.2)
Pulse (Max), Median (IQR)	75(69-83)	74(69-83)
Systolic Pressure (Min), Median (IQR)	95(84-104)	95(84-104)
Diastolic Pressure (Min), Median (IQR)	60(52-66)	61 (53-66)

BACK

N = 9,504
Data cut date: 2001.01-2010.12