COVID-19 ICU Predictions

Candace Warr
May , 2023

MOTIVATION

The COVID-19 pandemic has shown us the unpreparedness of our current healthcare system and services. We need to optimize the allocation of medical resources to maximize the utilization of resources. I have prepared this Machine Learning model based on the clinical data of patients admitted to the Hospital Sírio-Libanês, São Paulo and Brasilia. This will help us to predict the need of ICU for a patient in advance. With this information hospitals can plan the flow of operations and make critical decisions like shifting a patient to another hospital or determining the arrangement of resources within the time so that patients' lives can be saved.

ABOUT THE DATASET

The dataset used in this study is published by Hospital Sírio-Libanês, São Paulo and Brasilia. The data was anonymized following the best international practices. It includes clinical details of 384 COVID19 positive patients. Link: [COVID-19 — Clinical Data to assess diagnosis | Kaggle]()

METHODOLOGY

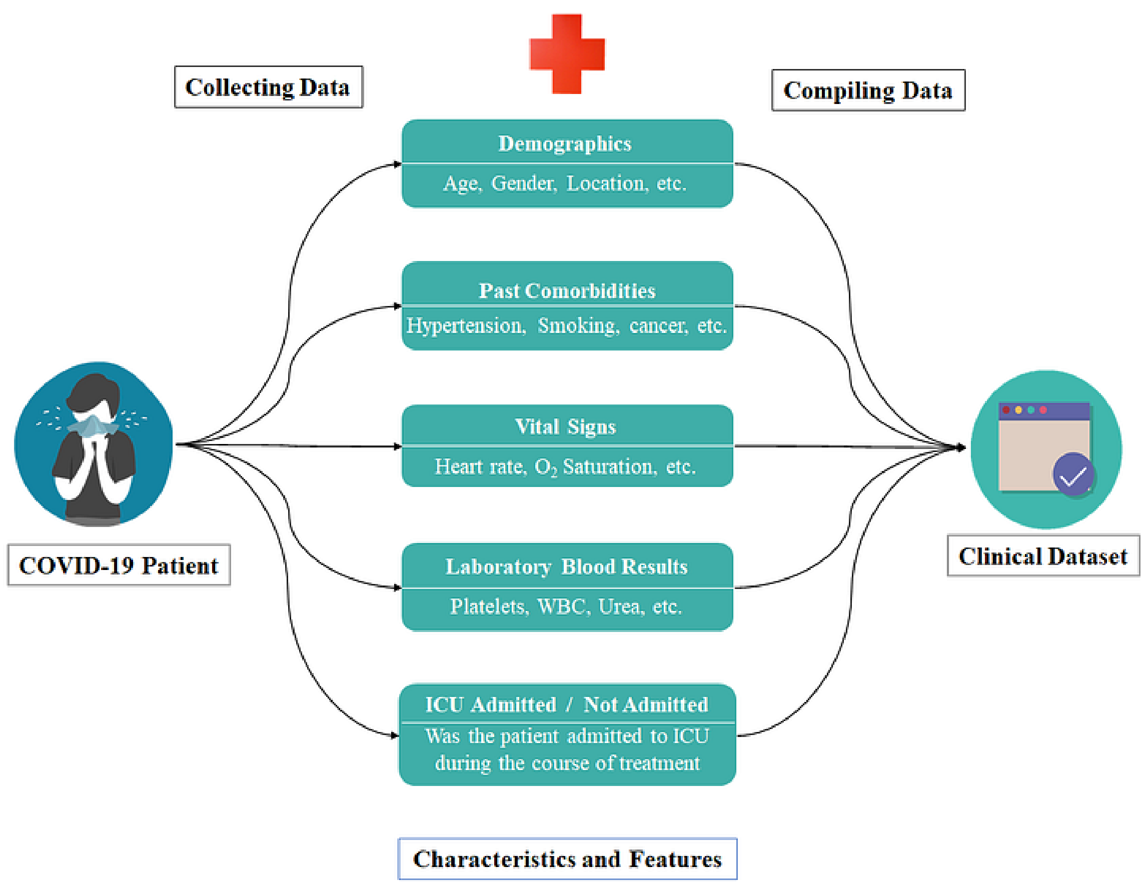
1. Data exploration: exploratory data analysis and visualization
2. Data cleaning and preprocessing
3. Training and testing ML models
4. Model result analysis

DATASET FEATURES

The data is given in a time window format i.e. for each patient data is provided for five different time windows: 0–2, 2–4, 4–6, 6–12 and 12+ hours from the time of admission to the hospital. Data is cleaned and scaled by column according to the Min Max Scaler to fit between -1 and +1. There are 54 features in the dataset that are further expanded by calculating mean, median, max, min, diff and relative diff. This results in the total number of columns equal to 231. The shape of the dataset is (1925, 231).

The clinical characteristics are divided into the following categories:

1. **Demographics**: Gender, Age Above 65, and Age Percentile.
2. **Past Comorbidities:** Anonymized Disease Grouping 1–6, Hypertension, Immunocompromised, and Other.
3. **Blood Test Results (36):** Albumin, BE Arterial, BE Venous, BIC Arterial, BIC Venous, Bilirubin, Blast, Calcium, Creatinin, FFA, GGT, Glucose, Hematocrite, Hemoglobin, INR, Lactate, Leukocytes, Linfocitos, Neutrophiles, P02 Arterial, P02 Venous, PC02 Arterial, PC02 Venous, PCR, PH Arterial, PH Venous, Platelets, Potassium, SAT02 Arterial, Sat02 Venous, Sodium, TGO, TGP, TTPA, Urea ,and Dimer.
4. **Vital Signs (6):** Blood Pressure (Diastolic and Systolic), Heart rate, Respiratory rate, Temperature, and Oxygen saturation.

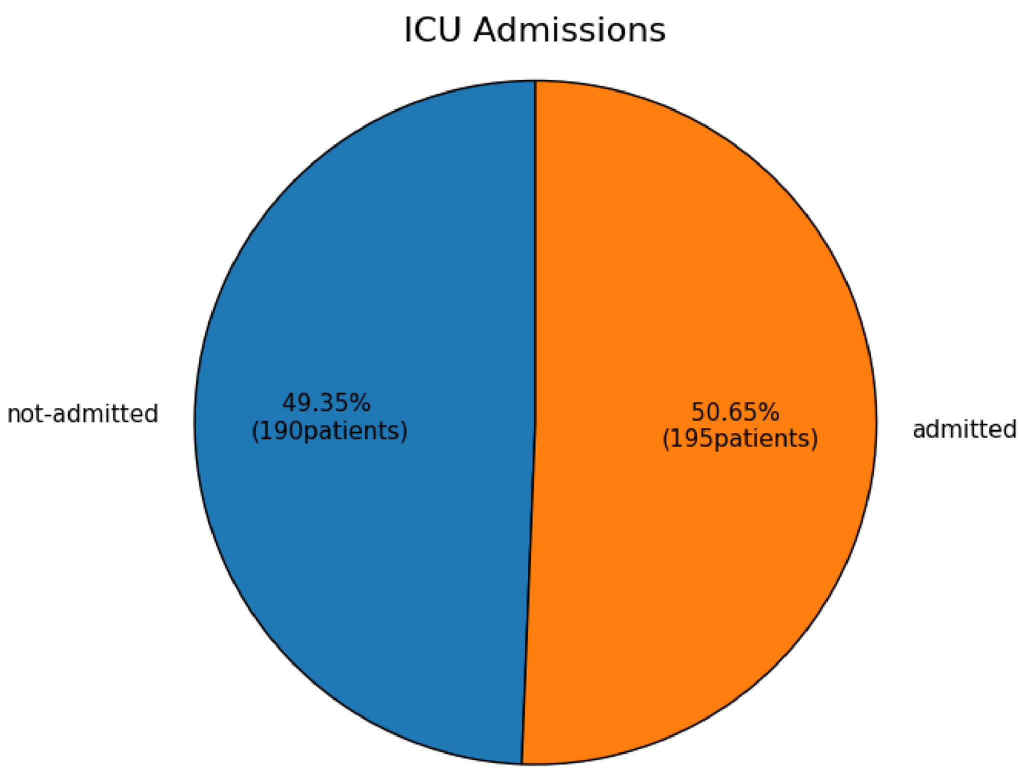


DATA ANALYSIS AND VISUALIZATION

Before preprocessing the data, I performed an in-depth analysis on the dataset and used various types of graphs and plots to understand the given data. The analysis was done on the clinical characteristics demographics and past comorbidities (disease groupings).

DEMOGRAPHICS

Distribution of ICU Admissions



The number of patients admitted into ICU is about the same as the number of hospital patients not admitted into ICU.

Window time frame that a patient was admitted into ICU.

	PATIENT_VISIT_IDENTIFIER	ICU	WINDOW
0	0	[0, 0, 0, 0, 1]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
1	1	[1, 1, 1, 1, 1]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
2	2	[0, 0, 0, 0, 1]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
3	3	[0, 0, 0, 0, 0]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
4	4	[0, 0, 0, 0, 0]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
5	5	[0, 0, 0, 0, 0]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
6	6	[0, 0, 0, 0, 0]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
7	7	[0, 0, 0, 0, 0]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
8	8	[0, 0, 0, 0, 0]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
9	9	[0, 0, 0, 0, 0]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
10	10	[0, 0, 0, 0, 0]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
11	11	[0, 0, 0, 1, 1]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
12	12	[0, 0, 0, 0, 0]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
13	13	[0, 0, 0, 0, 1]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]
14	14	[0, 0, 1, 1, 1]	[0-2, 2-4, 4-6, 6-12, ABOVE_12]

Patient 0 was admitted into ICU after 12+ hours of being admitted into the hospital.
Patient 1 was admitted into ICU between 0-2 hours after being admitted into the hospital.
Patient 3 was not admitted into ICU.
Patient 11 was admitted into ICU between 6-12 hours after being admitted into the hospital.
Patient 14 was admitted into ICU between 4-6 hours after being admitted into the hospital.

Number of patients admitted (1) and not admitted (0) into ICU per window time frame.

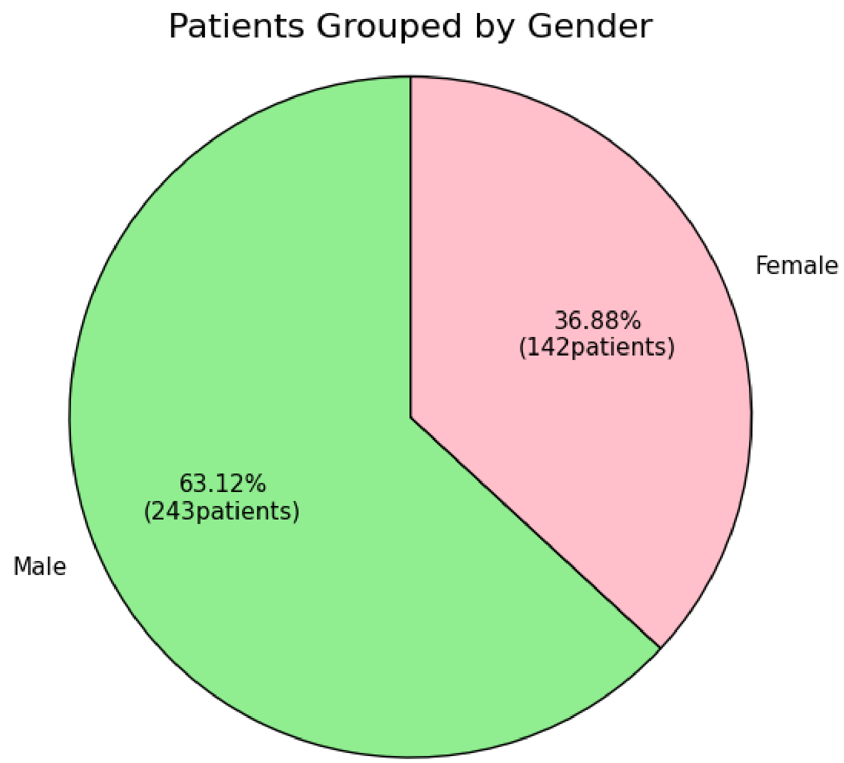
ICU WINDOW	0	1	Total
0-2	353	32	385
2-4	326	59	385
4-6	286	99	385
6-12	255	130	385
ABOVE_12	190	195	385

Once a patient is admitted into ICU, they will continue to be in ICU for the subsequent window time frames. In order to obtain the number of *new* patients admitted, subtract the previous window's number of ICU admissions from the desired window's ICU admission count.

Number of patients *admitted* into ICU per window time frame:
32 ICU admissions in window 0-2 hours.
27 new ICU admissions in window 2-4 hours.
40 new ICU admissions in window 4-6 hours.
31 new ICU admissions in window 6-12 hours.
65 new ICU admissions in window Above 12 hours.

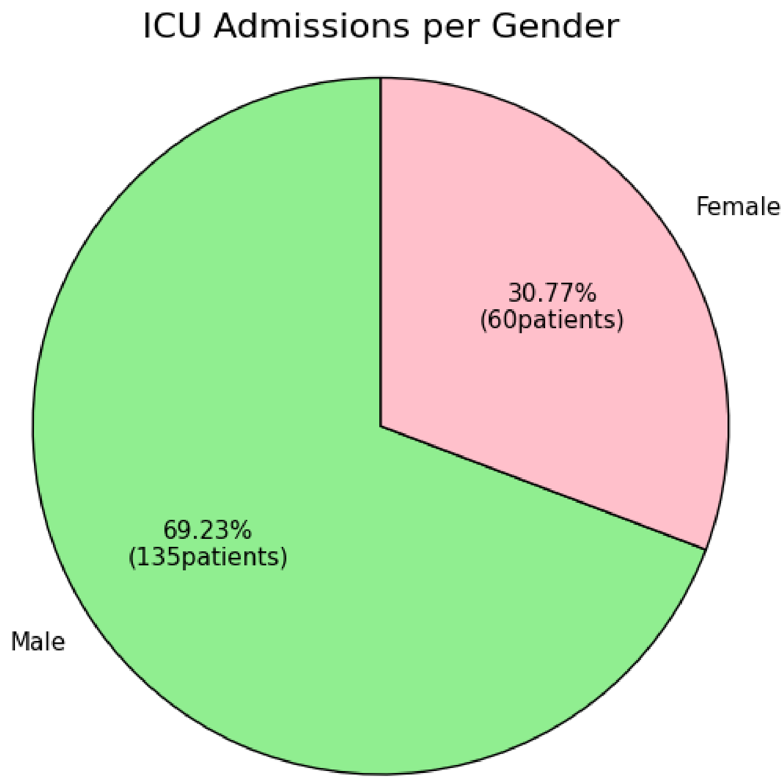
There are nearly twice as many ICU admissions in the Above 12 hour window time frame than any of the other windows.

Distribution of Patients by Gender



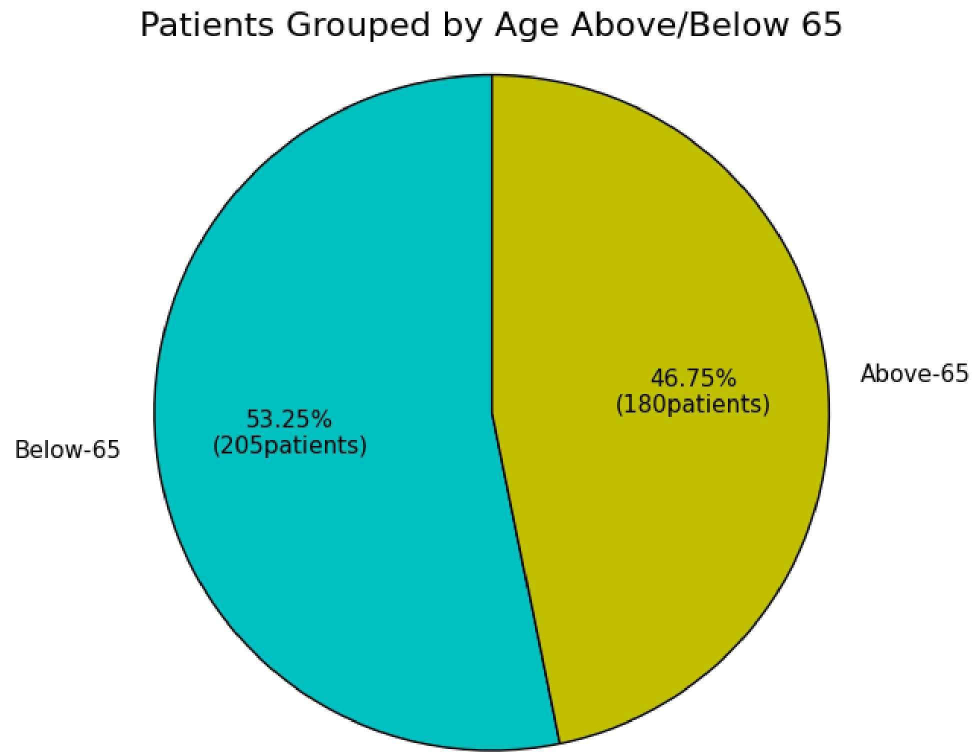
There are nearly twice as many male patients admitted to the hospital than female patients.

Distribution of ICU Admissions per Gender



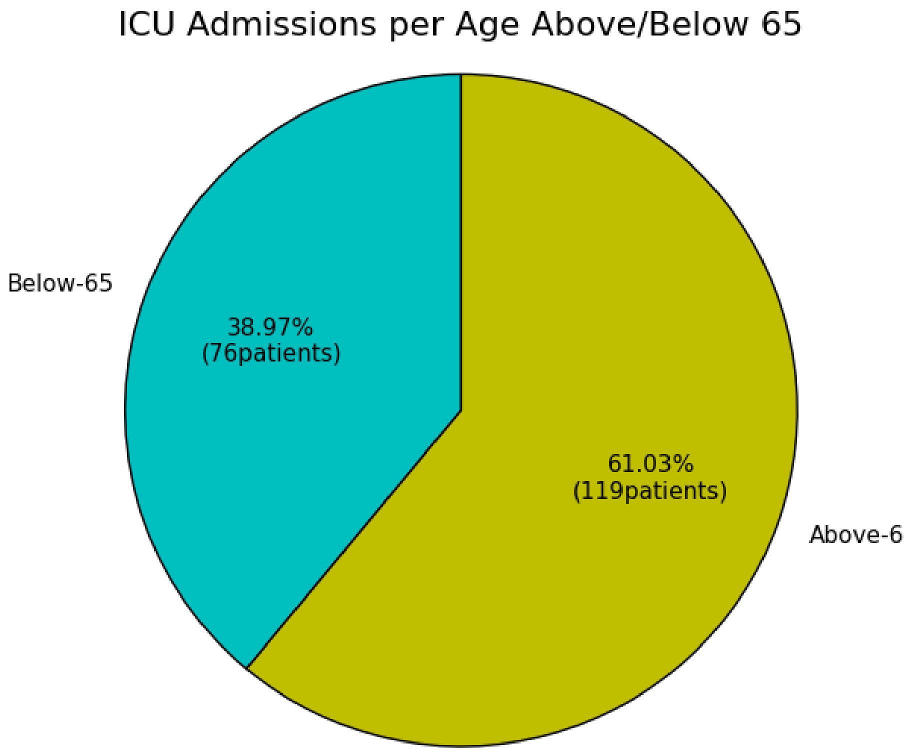
The number of male patients admitted into ICU is more than twice the number of female patients admitted into ICU. This is proportional to the number of male and female patients admitted to the hospital.

Distribution of Patients Above/Below Age 65



There are 25 more patients admitted into the hospital that are below the age of 65 than above the age of 65.

Distribution of ICU Admissions per Age Above/Below 65



There are nearly twice as many patients above the age of 65 admitted into ICU despite there being slightly less patients above the age of 65 than below the age of 65 admitted to the hospital.

Distribution of Patients per Age Percentile

AGE_PERCENTIL	
10th	41
20th	43
30th	41
40th	40
50th	38
60th	37
70th	39
80th	38
90th	31
Above 90th	37

The number of patients per age percentile (age bracket) is evenly distributed.

Distribution of ICU Admissions per Age Percentile

	AGE_PERCENTIL	ICU
0	10th	10
1	20th	12
2	30th	18
3	40th	15
4	50th	20
5	60th	20
6	70th	22
7	80th	26
8	90th	23
9	Above 90th	29

There are more ICU admissions as the age percentile (age bracket) increases despite the age percentile of patients admitted to the hospital being evenly distributed.

Of the varying demographic information provided, being male (gender) and of an older age (above age 65) gives indication that patients with COVID-19 are more likely to be admitted into ICU.

COMORBIDITIES

The labels for the past comorbidities were anonymized by the dataset author to respect the privacy of the patients. These past comorbidities generally include previous health issues and lifestyle habits affecting health like smoking, drinking, etc.

```
Disease Group 1
----------------------------------
ICU                     0    1  Total
DISEASE GROUPING 1
1.0                   136   72    208
```

```
Disease Group 2
---------------------------------
ICU                     0    1  Total
DISEASE GROUPING 2
1.0                    28   26     54
```

```
Disease Group 3
----------------------------------
ICU                     0    1  Total
DISEASE GROUPING 3
1.0                   118   70    188
```

```
Disease Group 4
---------------------------------
ICU                     0    1  Total
DISEASE GROUPING 4
1.0                    21   17     38
```

```
Disease Group 5
----------------------------------
ICU                     0    1  Total
DISEASE GROUPING 5
1.0                   155   91    246
```

```
Disease Group 6
---------------------------------
ICU                     0    1  Total
DISEASE GROUPING 6
1.0                    67   23     90
```

```
HTN Group
---------------------
ICU     0     1   Total
HTN
1.0   240   169     409
```

```
Immunocompromised Group
----------------------------------
ICU                       0     1   Total
IMMUNOCOMPROMISED
1.0                     218    86     304
```

```
Disease Group Other
-----------------------
ICU        0     1   Total
OTHER
1.0     1154   401    1555
```

Past comorbidities have influence on the need for ICU treatment. Patients with hypertension and comorbidities in disease groupings 2 and 4 have the highest percentages of ICU admissions among the different disease groupings (comorbidities).

While the characteristics of a majority of patients admitted into ICU fall under being male (gender), above age 65, and belonging to disease groupings 2, 4, and HTN (past comorbidities), there overall is no feature that is ***highly*** correlated with whether or not a patient is going to the intensive care unit. It appears to be a "synergistic" combination from multiple variables that results in the need for a patient to be admitted (or not) into ICU. In complex analyses like this, machine learning algorithms can help.

DATA CLEANING AND PREPROCESSING

1. **Object Type Handling:** Time Window and Age Percentile were of type String, so I Label Encoded them to numerical data for effective analysis.
2. **Handling Missing Data**: There are some missing values in the dataset. Those values are filled based on the neighboring windows of that particular patient, because it is safe to assume that there is little to no variation in the vital signs and blood test results within the neighboring windows. (Forward fill and backward fill).
3. **Handling time-window format:** Each patient's data is divided in five different time windows i.e. 0-2, 2–4, 4–6, 6–12 and 12+ hours after admission to the hospital. Once a patient is admitted into ICU, their data cannot be used for modeling as we need to predict the need for future ICU admission rather than at the present moment. So we give the machine learning algorithm data from the windows before ICU admission. For the model to make more clinically relevant predictions, it will receive data only from the first window time frame of 0-2 hours and then from all windows except the last, above 12 hours. The earlier the model can predict the need for ICU treatment of the patient, the more it becomes clinically relevant.

TRAINING AND TESTING ML MODELS

Sklearn library is used to train and test various machine learning models with 30% of the database as the test sample, and the stratify parameter is used to further balance the division of the train and test data. Variable y is set as the target variable (patients admitted into ICU).

Models included in this study:

1. Logistic Regression: to predict the output of a categorical dependent variable (y) - the probability of each patient going into ICU
2. Decision Tree: to perform both classification and regression tasks to predict the need of ICU well in advance so hospitals have more time to make critical, life saving decisions
3. K-Neighbors: to predict ICU admissions based on the classification of neighboring data points

MODEL RESULT ANALYSIS

After analyzing results of all three classifiers, the best performance was achieved by Logistic Regression and Decision Tree algorithms. Both models consistently show correct predictions and efficiency for ICU admissions at around 85-90% with limited false positive and false negative results. In conclusion, logistic regression and decision tree algorithms can be used to predict the need for confirmed COVID-19 patients to be admitted into ICU given clinical data. Our healthcare systems must incorporate innovative use of technologies like machine learning to improve the overall efficiency of hospitals and use of the intensive care unit. Machine learning models can provide solutions and assist doctors in critical decisions for saving lives.