Predicting the Need for Intubation among COVID-19 Patients Using Machine Learning Algorithms: A Single-Center Study

Raouf Nopour¹, Mostafa Shanbehzadeh², Hadi Kazemi-Arpanahi³,⁴*

Received: 7 Sep 2021 Published: 4 Apr 2022

Abstract

Background: Owing to the shortage of ventilators, there is a crucial demand for an objective and accurate prognosis for 2019 coronavirus disease (COVID-19) critical patients, which may necessitate a mechanical ventilator (MV). This study aimed to construct a predictive model using machine learning (ML) algorithms for frontline clinicians to better triage endangered patients and priorities who would need MV.

Methods: In this retrospective single-center study, the data of 482 COVID-19 patients from February 9, 2020, to December 20, 2020, were analyzed by several ML algorithms including, multi-layer perception (MLP), logistic regression (LR), J-48 decision tree, and Naïve Bayes (NB). First, the most important clinical variables were identified using the Chi-square test at P < 0.01. Then, by comparing the ML algorithms' performance using some evaluation criteria, including TP-Rate, FP-Rate, precision, recall, F-Score, MCC, and Kappa, the best performing one was identified.

Results: Predictive models were trained using 15 validated features, including cough, contusion, oxygen therapy, dyspnea, loss of taste, rhinorrhea, blood pressure, absolute lymphocyte count, pleural fluid, activated partial thromboplastin time, blood glucose, white cell count, cardiac diseases, length of hospitalization, and other underline diseases. The results indicated the J-48 with F-score = 0.868 and AUC = 0.892 yielded the best performance for predicting intubation requirement.

Conclusion: ML algorithms are potentials to improve traditional clinical criteria to forecast the necessity for intubation in COVID-19 in-hospital patients. Such ML-based prediction models may help physicians with optimizing the timing of intubation, better sharing of MV resources and personnel, and increase patient clinical status.

Keywords: COVID-19, Coronavirus, Machine Learning, Intubation, Prognosis, Mechanical Ventilator

Introduction

The Coronavirus Disease of 2019 (COVID-19) is a highly contagious viral infection that has to propagate speedily around the world as of the first advent in early December 2019 reported in Wuhan, Hubei province, Chi-

*What is “already known” in this topic:

Despite the effective and large-scale vaccination programs, tolls of COVID-19 new cases, caused by extensive dissemination of multiple variants, have increased. This pandemic overwhelmed the health care systems across the world with severe shortages in critical medical resources.

→What this article adds:

In this study, we applied several machine learning algorithms to predict the likelihood of the need for Mechanical ventilation in hospitalized COVID-19 patients based on routine clinical data collected at the first time of admission. Based on the research’s results, it found that machine learning algorithms enable a reasonable level of accuracy in predicting the risk of intubation among hospitalized COVID-19 patients.
Predicting COVID-19 Mechanical Intubation

na (1, 2). The COVID-19 is characterized by a varied and multi-dimensional clinical picture. The disease severity ranged from asymptomatic infection to mild symptoms, and baseline comorbidities appear after one week following infection onset, and even serious progressive complications in a small proportion of patients requiring the intensive care unit (ICU) admission (3, 4). Despite the effective and large-scale vaccination plans, tolls of COVID-19 new cases, caused by extensively contagious multiple variants, have plateaued (5-7). Old-age, male sex, pre-existing conditions, and hypoxemia demonstrated as significant factors leading to the critical stage (8-10). The critical or grave stage of COVID-19 is characterized by serious complications such as acute respiratory distress syndrome (ARDS), cytokine storm syndrome, and multi-system organ dysfunction (MOF) (10, 11).

The COVID-19 patients with acute respiratory insufficiency required a medical ventilator (MV) and supplemental oxygen (12). Therefore, to manage the MV scarcity, a clinical judgment is required to decide the need for early or postponing intubation and who doesn’t necessitate it (13). Furthermore, the COVID-19 course and outcome are unpredictable that complex this situation. There is a high degree of vagueness in the deterioration of the patient’s clinical status and in the speed at which cases develop respiratory distress demanding MV. Estimation of the number of patients that need MV has been considered in previous researches (8-11, 14, 15). To address these problems, in this study, we aimed to develop machine learning (ML)-based prediction models for frontline clinical workers and public health authorities to better triage endangered patients and priorities who would need MV.

ML as a sub-category of artificial intelligence (AI) is increasingly employed for COVID-19 screening, diagnosis, prediction, and prognosis outcomes (16, 17). It can rapidly synthesize and analyze large dimensional data. ML algorithms are employed to generate the prognostic models that can be used to support and improve clinical decision-making for a wide diversity of outcomes (18, 19). In the prior studies, a large number of ML-based models were developed for estimating the risk of COVID-19’ severity and patient illness deteriorating (16, 20), ICU admission (20-24), and deaths (21, 22, 25-30). Thus, this study aimed to construct and compare several ML-based prediction models for predicting the COVID-19 patients’ severity requiring MV.

Methods

This retrospective single-center study aimed at predict of the need for MV among COVID-19 hospitalized patients using four popular ML algorithms.

Dataset definition

In this study, a COVID-19 hospital-based registry from Ayatollah Taleghani hospital, Abadan city, Southwest of Khuzestan, Iran, was retrospectively reviewed from February 9, 2020, to December 20, 2020. During this period, a total of 6854 suspected cases with COVID-19 had been referred to this center, of whom 1853 cases were introduced as positive COVID-19, 2472 as negative, and 2529 as unspecified.

The inclusion criteria for patient selection were: 1- hospitalized patients with confirmed COVID-19, 2- patients who were greater than 18 years of age, 3- those with qualitative and comprehensive medical documentation (missing less than 70%), and 4-. On the other hand, the exclusion criteria for patient selection were: 1- non-COVID-19 cases or non-hospitalized COVID-19 or patients with unknown disposition, 2- patients who were less than 18 years of age, The under 18 age patients should be included in the scope of pediatric exploration, 3- incomplete case records (missing more than 70%), and 4- discharged / death from the emergency department or unknown patient disposition. The data on 1853 positive RT-PCR patients were extracted from the Ayatollah Taleghani hospital registry database. Based on the Table 1, the number of 53 clinical features in five classes including patient’s demographic data (five features), clinical features (14 features), history of personal diseases (five features), epidemiological (two features), laboratory results (26 features), remedies (one feature) and an output variable (0: non-intubation and 1: intubation) are extracted from the dataset. Table 1 demonstrates all different determinant factors associated with the prediction of intubation.

Dataset normalization and preprocessing

In this study, first, all included cases were investigated by two health information managers (R: N and H: KA)

Table 1. All extracted clinical features from the dataset

<table>
<thead>
<tr>
<th>Mode</th>
<th>Feature classes</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs Basic</td>
<td>Age, Sex, height, weight, and blood group</td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>Cough, nausea, headache, gastrointestinal (GI) manifestation, chill, loss of taste and smell, rhinorhea, sore throat, confusion, fever, muscular pain, vomiting, dyspnea</td>
<td></td>
</tr>
<tr>
<td>History of diseases Laboratory</td>
<td>Cardiac disease, pneumonia, hypertension, diabetes, and other underline diseases</td>
<td></td>
</tr>
<tr>
<td>red-cell count, hematocrit, hemoglobin, absolute lymphocyte count, blood calcium, blood potassium, absolute neutrophil count, alanine aminotransferase (ALT), magnesium, activated partial, prothrombin time, alkaline phosphatase, platelet count, hypersensitive troponin creatinine, white cell count, aspartate aminotransferase (ASP), blood glucose, total bilirubin, erythrocyte sedimentation rate (ESR), c-reactive protein, albumin, activated partial thromboplastin time, lactate dehydrogenase (LDH), blood phosphorus, blood sodium, and blood urea nitrogen (BUN)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epidemiological</td>
<td>Smoking, alcohol addiction</td>
<td></td>
</tr>
<tr>
<td>Remedy</td>
<td>Oxygen therapy</td>
<td></td>
</tr>
<tr>
<td>Output Outcome</td>
<td>Endotracheal intubation (Yes, No)</td>
<td></td>
</tr>
</tbody>
</table>
with consulting two infectious diseases and virology specialists. After reviewing all patients’ records, those with more than 70% missing values were omitted from the analysis. For other missing fields, the average of the existing available values and the K-Nearest Neighborhood (KNN) Euclidean distance for the quantitative and qualitative variables were used, respectively, in the Rapid Miner Studio V-7.1.001 environment.

Feature selection
In this study, for reducing the dataset dimension, we used the Chi-square ($\chi^2$) test for determining the relationship between each independent variable (53 variables) with the dependent (intuition: Yes or No) as the output class in SPSS software V25. The P<0.01 has been considered as a statistically significant level in this respect.

ML algorithms
The four ML algorithms have been utilized in this study for building the prediction models for intuition risk assessment among COVID-19 hospitalized patients in Weka V3.9, because of their high rate usage of these algorithms in recent articles, and also, their higher performance in terms of data classification process than other data mining algorithms.

Multi-layer Perception (MLP): MLP is one of the most popular Artificial Neural Networks (ANNs) utilized for knowledge modeling in different scientific domains. An MLP consists of at least three layers of nodes: input, hidden, and output layers. Each node has its weight for communication with other nodes. The input layers have consisted of variables affecting the study output(s). The number of nodes in this layer is equal to the number of independent variables that existed in the study. The hidden or processing layer is included different nodes with a specific number of layers that can perform different calculations using math function in the logistic activation method for giving the suitable output values depending on different amounts of inputs. The number of the output layer is equal to the output variable and this layer gives the results of calculations in ANNs using the linear activation method that existed between multiple nodes (31-34). In this research, the back-propagation neural network (BPNN) along with tanSigmoid activation methods have been used to train the prediction model.

Logistic Regression (LR): LR has various applications, especially in health domains for example estimating the outcomes from different influencing factors and making a beneficial model for prognostic models (35, 36). In reality, this is a probabilistic and statistical model that can predict the dependent variable(s) in two situations: 1- the dependent variables are qualitative with two or more values also known as binominal and poly nominal variables, respectively and 2- the independent variables are highly correlated concerning the output class. Hence in this situation, we can evaluate the effects of variables on each other in predicting the probability of the output class. The formula of the LR has been represented in Equation 1. In this equation, y is equal to the anticipated output, (a) demonstrates the intercept term or bias, and (b) is the coefficient for the sole input value of (x) (37-39).

\[ y = \frac{e^{a+b\times x}}{1+e^{a+b\times x}} \]

J-48: The C4.5 decision tree algorithm known as J-48 in the Weka data mining environment is a more advanced algorithm developed from the ID3 decision tree algorithm. Extracting the rule sets from this algorithm causes that this algorithm has more widely applicable than other algorithms. In this decision tree type similar to others, the classes or dependent variables lie in the leaf of the tree, and the input variables lie in the paths from the root nodes, which lies the independent variable with the highest Information gain (IG) to the leaf nodes. These paths are called the branches (nodes from roots to leaf), and the rules can be extracted from them. The IG is a classification method in splitting the nodes and building the decision tree by finding the differences between weighted entropies of each tree branch and main entropies. Equation 2 demonstrates a simple calculation formula of the IG. The C is the dataset class and the P, represents the probability of selecting an element of the class (i) randomly.

\[ IG = -\sum_i \frac{p_i}{N} \times \log_2 \frac{p_i}{N} \]

Generally, some beneficial features existed in this algorithm, such as pruning the decision tree by setting the confidence factor, abilities in the classification of the continuous and numerical variables, considering the missing values in sample classification, and rule derivation which caused that this algorithm has become better than other algorithms, especially decision trees (40-42).

Naïve Bayes (NB): The NB is a simple algorithm that is based on the Bayes theory. In this theory, all features values that existed in databases are for predicting the output class are independently considered in contrast with most other algorithms such as LR as one of them with the hybrid correlation between input variables to predict output class, and all features are equal in determining the output. It can be used for mining datasets with high dimensions. Some outstanding features that existed in this algorithm are 1- linear training time associated with features in model classification, 2- Low variance: although there are highly biased in this algorithm’s samples classification, because of not utilizing the searching method, it is a low-variance algorithm), 3-Insensitive in associated with the missing values: in this algorithm, all the features existing in the database will be used in predicting the output class, and although there might be a missing value associated with one feature, the other features can be used for predicting with simultaneously, a slight diminishing in algorithm performance. Generally, because of using all features that existed in the database and the nature of probabilistic, this algorithm is less sensitive to the noise and missing values. The probability of predicting the output class using the NB can be calculated in Equation 3. In this equation, the probability of occurrence of Y provided the occurrence of the X is the probability of occurring the feature of X in the condition that the output class (Y) oc-

http://mjiri.iums.ac.ir
Med J Islam Repub Iran. 2022 (4 Apr); 36.30.
Predicting COVID-19 Mechanical Intubation

Table 2: Confusion matrix

<table>
<thead>
<tr>
<th>Results</th>
<th>Predicted cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>TP FP</td>
</tr>
<tr>
<td>-</td>
<td>FN TN</td>
</tr>
</tbody>
</table>

curs and the probability of the output class occurrence (P(Y)). This equation demonstrates the importance and independence of each input class in determining the occurrence of the output class distinctly (42-45).

Equation 3: \[ P(Y/X) = P(Y) * \frac{P(X/Y)}{P(X)} \]

Performance evaluation of selected ML algorithms

In this study, the confusion matrix (Table 2) has been used for measuring the capabilities of each data mining algorithm in classification. In this table, the True Positive (TP) represents the hospitalized COVID-19 patients who have performed the intubation and are truly classified by the data mining algorithms; True Negative (TN) has belonged to hospitalized COVID-19 patients without any intubation and is classified truly by the model. The False Negative (FN) and False Positive (FP) have belonged to hospitalized COVID-19 patients who had and had not done the intubation, respectively, and were falsely classified by the model. Based on the confusion matrix, the TP-Rate, FP-Rate, Precision, Recall, F-Score, Matthews Correlation Coefficient (MCC), Kappa statistics, and AUC (Area Under the ROC (Receiver Operator Curve)) of each algorithm have been measured, and then the capability of each data mining algorithm has been assessed using these evaluation criteria. 10% fold cross-validation has been considered in this regard. Finally, the best data mining algorithm has been explained in more detail.

Results

After applying exclusion criteria, ultimately, the 482 case records were selected for the study (191 and 291 cases were associated with intubated and non-intubated hospitalized COVID-19 patients, respectively.) (Fig. 1).

The results of using the Chi-square test for determining the association between each factor and intubation outcome demonstrated that the variables such as age (\( \chi^2=3.222 \) at \( P=0.124 \)), sex (\( \chi^2=6.222 \) at \( P=0.126 \)), height (\( \chi^2=2.256 \) at \( P=0.068 \)), weight (\( \chi^2=16.226 \) at \( P=0.285 \)), and blood group (\( \chi^2=4.446 \) at \( P=0.123 \)) as basic classes, and headache (\( \chi^2=1.114 \) at \( P=0.049 \)), GI manifestation (\( \chi^2=2.774 \) at \( P=0.171 \)), chill (\( \chi^2=21.552 \) at \( P=0.243 \)), loss of smell (\( \chi^2=4.771 \) at \( P=0.110 \)), sore throat (\( \chi^2=5.54 \) at \( P=0.086 \)), fever (\( \chi^2=13.446 \) at \( P=0.121 \)), muscular pain (\( \chi^2=21.256 \) at \( P=0.056 \)), and vomiting (\( \chi^2=14.954 \) at \( P=0.151 \)) as clinical manifestations, and red-cell count (\( \chi^2=3.223 \) at \( P=0.068 \)), hematocrit (\( \chi^2=6.532 \) at \( P=0.113 \)), hemoglobin (\( \chi^2=1.32 \) at \( P=0.081 \)), blood calcium (\( \chi^2=4.412 \) at \( P=0.095 \)), blood potassium (\( \chi^2=3.12 \) at \( P=0.072 \)), absolute neutrophil count (\( \chi^2=14.889 \) at \( P=0.171 \)), ALT (\( \chi^2=2.226 \) at \( P=0.144 \)), blood magnesium (\( \chi^2=1.112 \) at \( P=0.085 \)), alkaline phosphatase (\( \chi^2=5.847 \) at \( P=0.062 \)), platelet count (\( \chi^2=1.776 \) at \( P=0.041 \)), hypersensitive troponin (\( \chi^2=4.112 \) at \( P=0.075 \)), creatinine (\( \chi^2=7.412 \) at \( P=0.041 \)), ASP (\( \chi^2=2.745 \) at \( P=0.093 \)), total bilirubin (\( \chi^2=18.745 \) at \( P=0.166 \)), ESR (\( \chi^2=14.256 \) at \( P=0.083 \)), C-reactive protein (\( \chi^2=5.445 \) at \( P=0.143 \)), albumin (\( \chi^2=12.332 \) at \( P=0.121 \)), activated partial thromboplastin time (\( \chi^2=13.227 \) at \( P=0.165 \)), LDH (\( \chi^2=4.556 \) at \( P=0.064 \)), blood phosphorus (\( \chi^2=1.226 \) at \( P=0.082 \)), blood sodium (\( \chi^2=7.747 \) at \( P=0.188 \)), and BUN

Fig. 1. Flow chart describing patient selection
mc, and Kappa statistics have been shown in Figure 2.

The results of classifying the samples using the AUC have been demonstrated in Figure 3. (The vertical and horizontal vertices show the TP-Rate and FP-Rate, respectively.)

Based on the information given in Figures 1 and 2, it has resulted that the J-48 decision tree algorithm with TP-Rate=0.669, FP-Rate=0.155, Precision=0.869, Recall=0.869, F-Score=0.868, MCC=0.725, Kappa=0.723, and AUC=0.892 had the best capability for early predicting the risk of intubation in COVID-19 hospitalized patients. On the other hand, the MLP with TP-Rate=0.614, FP-Rate=0.446, Precision=0.605, Recall=0.614, F-Score=0.607, MCC=0.175, Kappa=0.173 and AUC=0.639 gained the worst predictive performance. Therefore, the J-48 decision tree algorithm with confidence factors of 0.15 has been depicted in Figure 4.

Based on the J-48 decision tree algorithm, some clinical rules have been extracted, we have brought the two most important of them with the highest samples classified.

**Rule 1:** IF (Activated partial thromboplastin time <=31) THEN the Intubation=True. This rule can be interpreted as overall among the 64 research samples who had more than 31 of activated partial thromboplastin time, the 47 samples had the intubation process, and the variable as the root node in the J-48 decision tree was considered as the most important factor for determining the endotracheal intubation risk among hospitalized COVID-19 patients.

**Rule 2:** IF (Activated partial thromboplastin time >31 & Pleural fluid=Yes & White cell count <9200 & activated partial thromboplastin time <=43) THEN the endotracheal intubation risk is negative. In this study, 221 samples had this rule template, and among them, 187 samples have been classified correctly through this rule template as negative or low risk of endotracheal intu-

---

**Table 3. Important features related to the prediction of the need for MV**

<table>
<thead>
<tr>
<th>No</th>
<th>Variable</th>
<th>Variable’s type</th>
<th>Frequency or Mean±/ SD</th>
<th>χ²</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cough</td>
<td>Nominal</td>
<td>Yes (401) No (81)</td>
<td>5.949</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>Contusion</td>
<td>Nominal</td>
<td>Yes (180) No (302)</td>
<td>4.997</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3</td>
<td>Oxygen therapy</td>
<td>Nominal</td>
<td>Yes (437) No (45)</td>
<td>7.01</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>Dyspnea</td>
<td>Nominal</td>
<td>Yes (442) No (40)</td>
<td>15.023</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5</td>
<td>Loss of taste</td>
<td>Nominal</td>
<td>Yes (124) No (358)</td>
<td>7.722</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>6</td>
<td>Rhinorrhea</td>
<td>Nominal</td>
<td>Yes (202) No (280)</td>
<td>10.239</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>7</td>
<td>Blood pressure</td>
<td>Nominal</td>
<td>Yes (189) No (293)</td>
<td>7.281</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>8</td>
<td>Absolute lymphocyte count</td>
<td>Numeric</td>
<td>21.702±12.01</td>
<td>23.46</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>9</td>
<td>Pleural fluid</td>
<td>Nominal</td>
<td>Yes (275) No (78)</td>
<td>19.583</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>10</td>
<td>Activated partial thromboplastin time</td>
<td>Numeric</td>
<td>35.453±9.25</td>
<td>17.458</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>11</td>
<td>Blood glucose</td>
<td>Numeric</td>
<td>148.4±96.946</td>
<td>12.884</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>12</td>
<td>White cell count</td>
<td>Numeric</td>
<td>968±1241</td>
<td>14.424</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>13</td>
<td>Cardiac diseases</td>
<td>Nominal</td>
<td>Yes (157) No (325)</td>
<td>12.491</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>14</td>
<td>Length of hospitalization</td>
<td>Numeric</td>
<td>5.03±2.188</td>
<td>2.713</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>15</td>
<td>Other underline diseases</td>
<td>Nominal</td>
<td>Yes (339) No (143)</td>
<td>13.277</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

(χ²=2.266 at P=0.121) as laboratory findings and alcohol consumption (χ²=16.227 at P=0.075) and smoking (χ²=8.887 at P=0.111) as epidemiological and pneumonia (χ²=4.536 at P=0.162) and diabetes (χ²=1.1447 at P=0.061) as history of diseases, gained the P>0.01, and therefore, were not considered as the important factor predicting the intubation among hospitalized COVID-19 patients and were excluded from the analysis process.

For 15 variables, there was a meaningful relationship with output class (intubation prediction) at P<0.01, and so has been shown in Table 3.

Based on the information given in Table 3, the 15 variables obtained the meaningful association at P<0.01. Of these, five variables including the history of cardiac diseases (χ²=12.491, P<0.001), pleural fluid (χ²=19.583, P<0.001), absolute lymphocyte count (χ²=23.46, P<0.001), cough (χ²=5.949, P<0.001), and dyspnea (χ²=15.023, P<0.001) yielded the highest association at P<0.001 to predict the need for MV among hospitalized COVID-19 patients. The results of classifying the samples using the confusion matrix have been shown in Table 4.

Based on the information provided in Table 4, the J-48 decision tree algorithm with TP=266 and TN=153 yielded the highest performance in the prediction of the need for MV. Also, this algorithm with FP=24 and FN=39 had the lowest incorrectly classified samples than others. The results of the performance of selected ML algorithms based on the TP-Rate, FP-Rate, Precision, Recall, F-Score, MCC, and Kappa statistics have been shown in Figure 2.

### Table 4. The data mining algorithm’s confusion matrix

<table>
<thead>
<tr>
<th>No</th>
<th>Algorithm</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MLP</td>
<td>212</td>
<td>78</td>
<td>108</td>
<td>84</td>
</tr>
<tr>
<td>2</td>
<td>LR</td>
<td>241</td>
<td>49</td>
<td>95</td>
<td>97</td>
</tr>
<tr>
<td>3</td>
<td>J-48</td>
<td>266</td>
<td>24</td>
<td>39</td>
<td>153</td>
</tr>
<tr>
<td>4</td>
<td>NB</td>
<td>195</td>
<td>95</td>
<td>56</td>
<td>136</td>
</tr>
</tbody>
</table>

---

**Table 4. The data mining algorithm’s confusion matrix**

<table>
<thead>
<tr>
<th>No</th>
<th>Algorithm</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MLP</td>
<td>212</td>
<td>78</td>
<td>108</td>
<td>84</td>
</tr>
<tr>
<td>2</td>
<td>LR</td>
<td>241</td>
<td>49</td>
<td>95</td>
<td>97</td>
</tr>
<tr>
<td>3</td>
<td>J-48</td>
<td>266</td>
<td>24</td>
<td>39</td>
<td>153</td>
</tr>
<tr>
<td>4</td>
<td>NB</td>
<td>195</td>
<td>95</td>
<td>56</td>
<td>136</td>
</tr>
</tbody>
</table>
Predicting COVID-19 Mechanical Intubation

![MLP, LR, J-48, NB comparison](image)

**Fig. 2.** Visual comparison of ML algorithm capabilities for prediction of the need for MV

<table>
<thead>
<tr>
<th></th>
<th>MLP</th>
<th>LR</th>
<th>J-48</th>
<th>NB</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP-Rate</td>
<td>0.614</td>
<td>0.701</td>
<td>0.869</td>
<td>0.687</td>
</tr>
<tr>
<td>FP-Rate</td>
<td>0.446</td>
<td>0.365</td>
<td>0.155</td>
<td>0.306</td>
</tr>
<tr>
<td>Precision</td>
<td>0.605</td>
<td>0.696</td>
<td>0.869</td>
<td>0.702</td>
</tr>
<tr>
<td>Recall</td>
<td>0.614</td>
<td>0.701</td>
<td>0.869</td>
<td>0.687</td>
</tr>
<tr>
<td>F-Score</td>
<td>0.607</td>
<td>0.692</td>
<td>0.868</td>
<td>0.69</td>
</tr>
<tr>
<td>MCC</td>
<td>0.175</td>
<td>0.358</td>
<td>0.725</td>
<td>0.373</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.173</td>
<td>0.3504</td>
<td>0.7237</td>
<td>0.3681</td>
</tr>
</tbody>
</table>

**Fig. 3.** The ROC diagrams of selected ML algorithms

**Fig. 4.** The pruned J-48 decision tree algorithm

bation. Generally, this rule with the most classified samples has been recognized as the most important decision rule in this research.

**Discussion**

Given the high spectrum of COVID-19 clinical manifestations, it is important to construct models for estimating the likelihood of intubation by using ML techniques.
Thus, we trained four ML-based models according to the top related parameters affecting the risk of intubation that derived from a statistical analysis. The ML methods employed herein included ANN, LR, J-48, and NB techniques which were trained using the most important forecasters from 482 hospitalized laboratory-confirmed COVID-19 patients at the time of admission. Finally, based on our analysis, we found that the J-48 classifier with an F-score of 0.868 and AUC of 0.892 has better performance than other selected ML algorithms.

During the COVID-19 pandemic, the requirement for informed decision-making is most imperative, specifically, where the healthcare system runs into an increasing surge of patients and scarcities in intensive care resources such as ICU beds and ventilators (46, 47).Clinicians have stated trouble in forecasting the disease progression of COVID-19 in-hospital patients, along with problems in the detection of patients who are susceptible to fast decompensation (48). In response to this life-saving infection, the design and implementation of clinical decision support systems (CDSS), will be critical to hold the optimal use of limited hospital resources and supporting clinical decisions (16, 49). CDSSs equipped with ML can help clinical decisions by informing caregivers and recommending interventions based on objective and generalizable experimental data (50). Our study proves that ML algorithms, particularly the J-48 algorithm, augment the analytic precision and the discriminative efficacy of these variables, increasing their usage for estimating the need for MV among COVID-19 hospitalized patients.

So far, several studies have been evaluating the application of ML techniques in predicting the COVID-19 poor outcomes. Saha et al. (2021) designed an intelligent system based on some ML algorithms using a dataset of 1023 patients’ data to predict future intubation among hospitalized patients with COVID-19. Finally, the best performance was yielded by the DT algorithm with an AUC of 0.84 (51). Alotaibet et al. (2021) in their study, assessed the performance of three ML algorithms for early prediction of disease severity using patient history and laboratory findings of patients with COVID-19, and the best performance in all the applied techniques is yielded by the Random Forest (RF) (AUC = 0.897) (52). In one study performed by Cobre (2021) the data of 5,643 COVID-19 negative and positive samples were analyzed to predict the individual severity by selected ML models. The results showed DT algorithm has a good discriminative ability with an accuracy of 86% (53). Accordingly, Yadaw et al. (2020) assessed the performance of four ML algorithms using a dataset including 3841 COVID-19 records for the prediction of COVID-19 deterioration and severity. Finally, the DT model with an AUC of 0.92 was introduced as the most appropriate algorithm (29). Pan and their colleagues (2020) assessed the performance of four ML algorithms to anticipate patient condition deterioration with COVID-19 and the best performance was reported from RF the model (AUC = 0.92) (24). Gao and their colleagues (2020) retrospectively studied the 2520 COVID-19 hospitalized patients’ medical records with 13 physical features to construct an intelligent predictive model through selected ML algorithms for physiological deterioration and endotracheal intubation prediction. Finally, the DT model with an AUC of 0.9760% gained the best performance (25). Similarly, in the current study, the results showed that the J-48 decision tree with an F-score of 0.868 and AUC of 0.892 has the best capability for early prediction of the risk of intubation in COVID-19 hospitalized patients.

The high predictive measures attained by the developed J-48 model in our study reveal that it has the capability of correct judgment amid COVID-19 patients at high risk against low risk of demanding MV. The innovation of the current study lies in the fact that contrast to the prior studies, we predict the intubation possibility based on the most pertinent predictors derived from the performing feature selection. Furthermore, to precisely detect predictors for intubation in infected patients with COVID-19, we evaluated the patient’s features at the first time of admission and not at the progressive or severe stage of the disease. For this reason, some important laboratory features such as increased ALT/ASP, high BUN, elevated C-reactive protein, and increased lymphocyte or neutrophil are not identified as intubation predictors in our study because these factors may only develop in the advanced stage that was omitted from analysis in our study.

Zhou(11), Choron (54), Allenbach (23), Lei (55), and Yadaw (29) stated that some predictors such as age (elderly), BMI (high), gender (male sex), ALT/ASP (raised), C-reactive protein (elevated) and oxygen saturation (decreased); had been related to COVID-19 poor outcomes and patient deterioration condition. However, these factors are likewise very predominant in COVID-19 moderate or asymptomatic presentation. But our analysis in this study does not demonstrate the association between these variables with intubation as a critical outcome of COVID-19. This hole may originate from the analysis of the only selected admitted patients in the hospital instead of population-based investigation. Henceforth, if validated, these predictors could be used for estimating the risk for patients’ intubation and may support the effective patients’ triaging.

This work has some limitations that need to be addressed. First, as analysis of a single-center and retrospective dataset with limited sample size and the outcome of intubation for model prediction in our study is rare, the study design might be affected by several hypothesis testing biases. Thus, external validation is essential to be conducted in further studies. Second, the dynamic variations of some significant variables must be followed up to better and timely recognize patients at higher risks of poor outcomes. Finally, the selected dataset lacks some important clinical variables, such as radiological indicators. In the future, the performance accuracy of our model and its generalizability will be enhanced if we test more ML techniques at the larger, multicenter, and prospective dataset which is equipped with more qualitative and validated data.

**Conclusion**

In this article, we analyzed the data from a hospital reg-
Predicting COVID-19 Mechanical Intubation

ists to develop and test models capable of predicting the need for MV in hospitalized COVID-19 patients according to 15 baseline clinical features. The results disclosed a satisfactory performance and tuning of the J-48 decision tree model, which indicates that adopting the models is acceptable. Given the considerable challenges concerning hospital resources, including MV, during the COVID-19 pandemic, an exact estimation of patients to be expected to require intubation may aid to provide vital guidance regarding priority patients toward assigning the restricted resources to patients whom emergency required. Further, timely detection of such people may allow for planned intubation measures and decrease some known risks related to urgent intubation. These developed prediction models may therefore be an advantage in better care delivery, lessen clinician workload, lessen illness and death in the COVID-19 pandemic.

Acknowledgment
We thank the research deputy of the Abadan University of Medical Sciences for financially supporting this project. (IR.ABADANUMS.REC.1400.071)

Conflict of Interests
The authors declare that they have no competing interests.

References
31. Bayram S, Ocal ME, Laptai Oral E, Atis CD. Comparison of multi