Comparison of Logistic Regression model with Machine Learning models to predict Acute Liver Injury in patients hospitalized with COVID-19

Anjali Byale, MD1; Shivabal Kanthavarayan Ramu, MD2; Deeksha Acharya, MD1; Sivanesan, MD1; Sneha Gupta, MD1; Kanika Mahajan, MD1; Achintya Singh, MD1

1UPMC McKeesport; 2Mayo Clinic, Rochester; 3PGIMER, India; 4MetroHealth Medical Center, Cleveland

Background
- The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has an an average mortality rate approaching 6.5%.
- The impact of liver injury on adverse clinical outcomes in coronavirus disease 2019 (COVID-19) patients remains unclear.
- ALT/AST elevation was common and independently associated with adverse clinical outcomes in COVID-19 patients.
- There is a need to develop a normogram risk prediction tool in patients who have been admitted with COVID19 to appropriately manage this sector of patients.

Methods
- We reviewed and analyzed data from 1844 COVID-19 patients from Stony Brook University. A total of 10 variables were used to build multivariable logistic regression model. Missing values were imputed using random forest approach in R.
- Subsequently, the data was randomly split into training (80%) and testing (20%). We developed a new nomogram for quick and easy prediction of COVID-19 patients with and without hepatic injury.

Results
- Logistic regression analysis showed an accuracy of 95.7% on the validation dataset.
- Amongst the machine learning models, Random Forest, Gradient Boosting, and Support Vector Machine each showed accuracy of 100%. Decision tree displayed the lowest accuracy of 50%.
- In the development group, best predictor for the new nomogram as selected by the multivariate logistic regression model was triglyceride/neutrophil ratio, followed by age, platelet count, albumin, AST/ALT, neutrophil lymphocyte ratio, total cholesterol, AST/ALT ratio, procalcitonin, lactate dehydrogenase, alanine transaminase, aspartate transaminase, and creatinine. The nomogram was then simplified by removing variables with poor performance.

Objectives
- We aimed to develop a novel mortality scoring system for patients with COVID-19 based on demographic factors, laboratory findings, and outcomes.
- We proposed a new mortality scoring system for COVID-19 patients with the lowest accuracy of 50%.

Discussion
- We developed a novel regression model with novel biomarkers that can predict the outcome of COVID-19 patients with COVID-19 (COVID-19) with 95% accuracy.
- The AUCROC of the LR model was 95%. The machine learning models, SAGA, GBR and SGB outperformed the LR model.
- The overall survival was older in patients who did not have hepatic injury compared to those with hepatic injury (median, 55 days, and 6 days, respectively; hazard ratio, 0.38; 95% CI, 0.25 to 0.56) (P<0.001 for interaction). Whereas median duration of mechanical ventilation in patients with hepatic injury was higher than those without hepatic injury (median, 40 days, and 10 days, respectively; hazard ratio, 0.33; 95% CI, 0.2 to 0.49) (P<0.001 for interaction) (Figures 3a & 3b).

Table 1. Comparison of performance metrics of the models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Area Under Curve</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.957</td>
<td>0.971</td>
<td>0.964</td>
<td>0.967</td>
</tr>
<tr>
<td>Gradient Boosting</td>
<td>0.500</td>
<td>0.964</td>
<td>0.946</td>
<td>0.959</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>1.000</td>
<td>0.978</td>
<td>0.918</td>
<td>0.979</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.100</td>
<td>0.992</td>
<td>0.901</td>
<td>0.992</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.826</td>
<td>0.820</td>
<td>0.875</td>
<td>0.863</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.650</td>
<td>0.974</td>
<td>0.967</td>
<td>0.973</td>
</tr>
</tbody>
</table>

Figure 1. Work flow of the study showing the models used.

Figure 2. Nomogram to predict the probability of acute liver injury in COVID-19 patients.

Figure 3a. Kaplan-Meier survival curves with different LOS in patients with and without hepatic injury.

Figure 3b. Kaplan-Meier survival curves with different duration of mechanical ventilation in patients with and without hepatic injury.

Figure 4. Distribution of patients and Kaplan-Meier survival curves: Grey curve: Tg/Ni<28 and Blue curve: Tg/Ni>28.