CT-based Patient Triage of COVID-19: Radiomics Prediction of ICU Admission, Mechanical Ventilation, and Death of Patients

Xianghao Zhan[1], Yiheng Li[2]

[1] Department of Bioengineering
[2] Department of Biomedical Data Science

CS229 Spring 2020, Stanford University

Introduction

In order to achieve rapid stratification and timely intensive care of COVID-19 patients, as well as the optimization of medical resource allocation under this unprecedented public health emergency, in this study, prediction models, which take radiomics based features, are built to predict high-risk inpatients at the time of admissions.

Data and Study Design

1. Flowchart of the study

   ![Flowchart](image)

   - Inpatients with laboratory-confirmed SARS-CoV-2 infection from 39 designated hospitals
   - Patients were classified into two groups based on their clinical status:
     - Cohort 1: n = 1662
     - Cohort 2: n = 700

2. Data Types and Feature Engineering

   - Data Types
     1. Radiomics features (Radiom)
     2. Laboratory results (Lab)
     3. Clinical features (Clin)
     4. Radiologist findings.

   - Feature Engineering Methods
     1) SMOTEENN [1]
     2) PCA [2-4]
     3) LASSO feature selection [2]
     4) GUS (generic univariate selection)
     5) FPR (false positive rate test).

Prediction Models

- Logistic Regression (LR)
- Random Forest (RF)
- Support Vector Machine (SVM)
- Multilayer Perceptron (MLP)
- LightGBM

Results

- We first tested the models on Cohort 1

<table>
<thead>
<tr>
<th></th>
<th>ICU</th>
<th>MV</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>AUROC</td>
<td>ACC</td>
<td>AUROC</td>
</tr>
<tr>
<td>Radio</td>
<td>0.752</td>
<td>0.760</td>
<td>0.521</td>
</tr>
<tr>
<td>Clin</td>
<td>0.836</td>
<td>0.826</td>
<td>0.630</td>
</tr>
<tr>
<td>Lab</td>
<td>0.837</td>
<td>0.824</td>
<td>0.590</td>
</tr>
<tr>
<td>RadioClin</td>
<td>0.836</td>
<td>0.826</td>
<td>0.630</td>
</tr>
</tbody>
</table>

- Then optimal models are chosen and tested on Cohort 2

<table>
<thead>
<tr>
<th></th>
<th>ICU</th>
<th>MV</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>AUROC</td>
<td>ACC</td>
<td>AUROC</td>
</tr>
<tr>
<td>Radio</td>
<td>0.836</td>
<td>0.826</td>
<td>0.630</td>
</tr>
<tr>
<td>Clin</td>
<td>0.837</td>
<td>0.824</td>
<td>0.590</td>
</tr>
<tr>
<td>Lab</td>
<td>0.837</td>
<td>0.824</td>
<td>0.590</td>
</tr>
<tr>
<td>RadioClin</td>
<td>0.836</td>
<td>0.826</td>
<td>0.630</td>
</tr>
</tbody>
</table>

Conclusion

- The models were trained on a cohort of 1662 COVID-19 patients and the performances were validated on an unseen cohort of 700 COVID-19 patients with high AUROC and AUPRC.
- The model with radiomics, demographics, clinical symptoms, comorbidity and lab tests showed the best performances on three prediction tasks while various models based on different types of data could also enable the users to flexibly choose based on the data available.
- This work enables the risk evaluation of COVID-19 patients and facilitates the resource allocation to those with high risks.