PATTER, Volume *1*

Supplemental Information

A Learning-Based Model to Evaluate Hospitalization Priority in COVID-19 Pandemics

Yichao Zheng, Yinheng Zhu, Mengqi Ji, Rongpin Wang, Xinfeng Liu, Mudan Zhang, Jun Liu, Xiaochun Zhang, Choo Hui Qin, Lu Fang, and Shaohua Ma

Supplemental Information

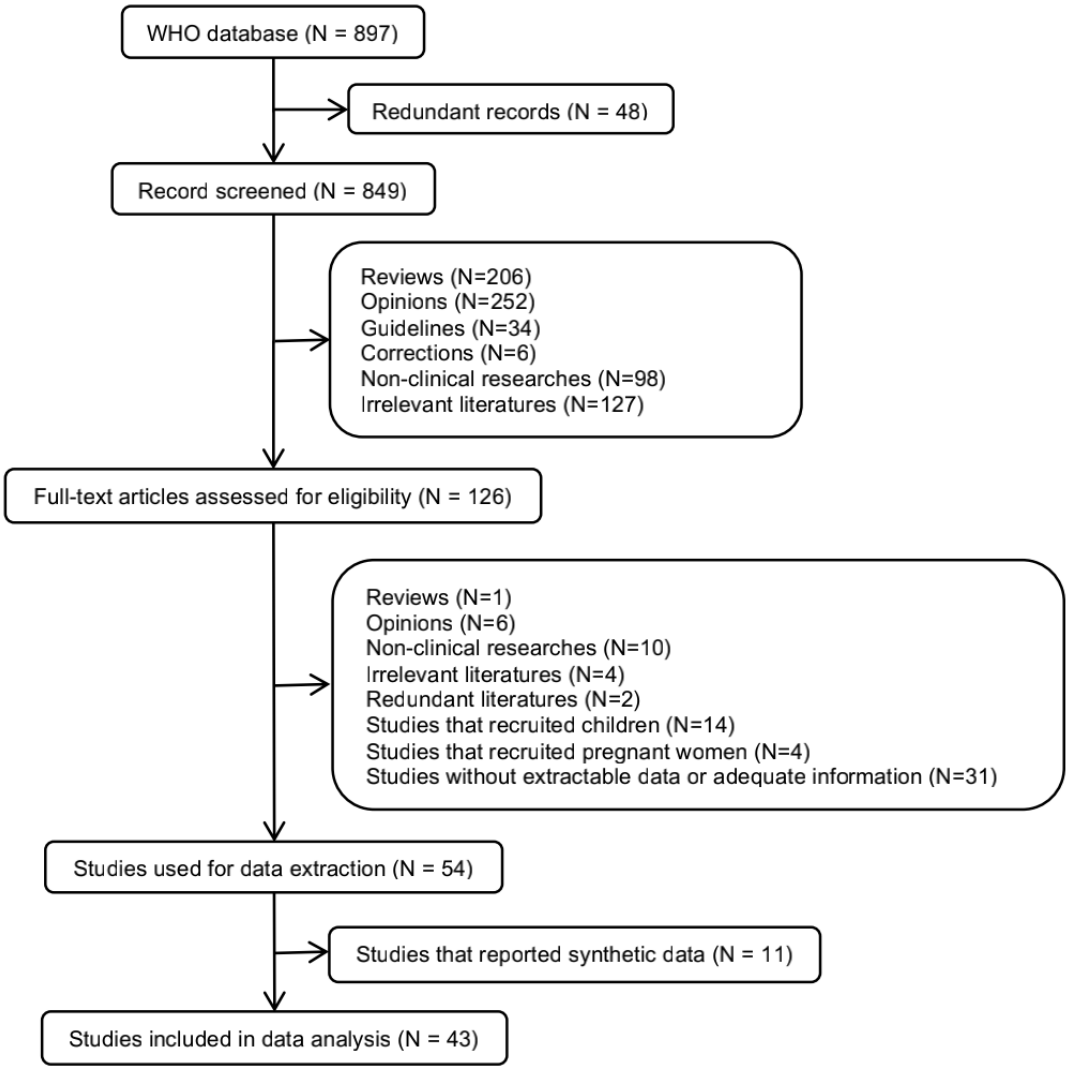


Figure S1. Flowchart of the study inclusion and exclusion

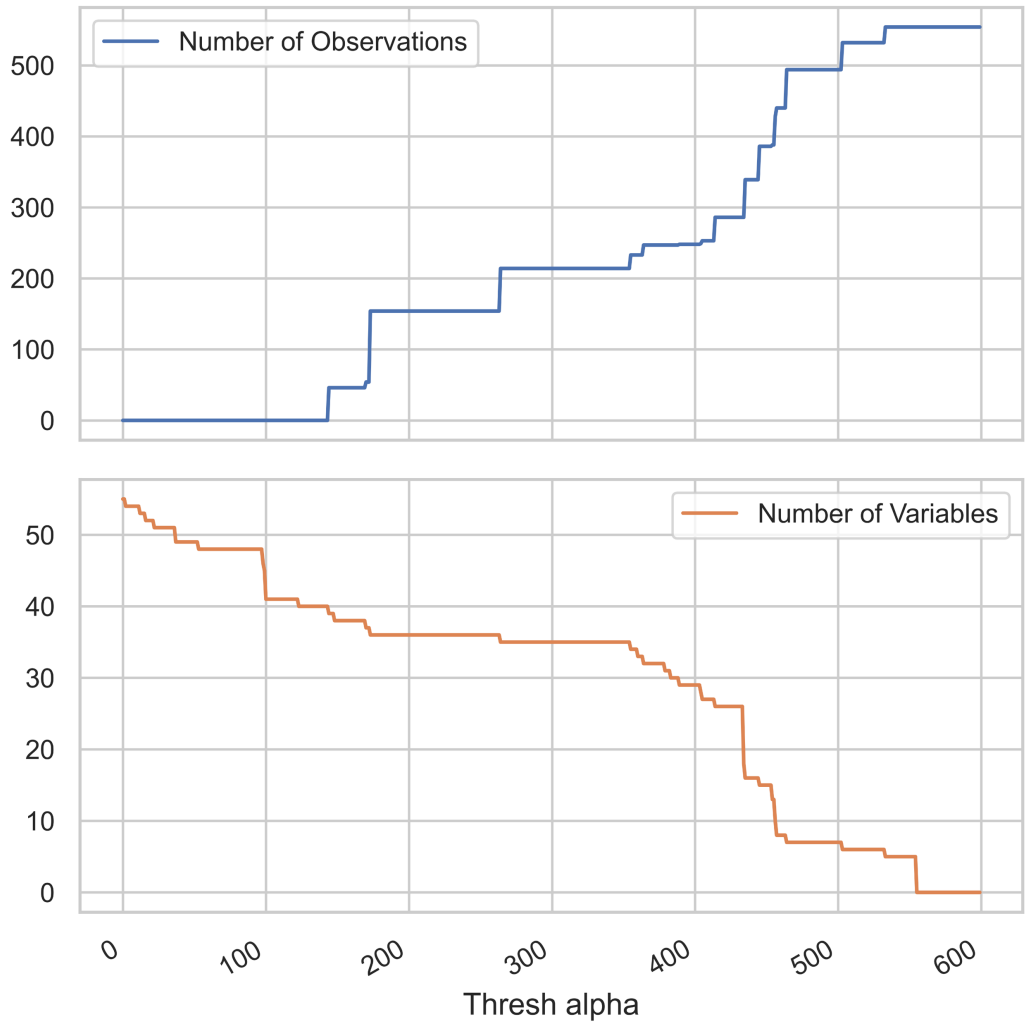


Figure S2. Strategy to remove the missing values. In this first step, the variables with missing data that exceeded the threshold alpha were removed (bottom figure). Next, observations (that were COVID-19 cases) with missing data for any of the resulting variables were removed.

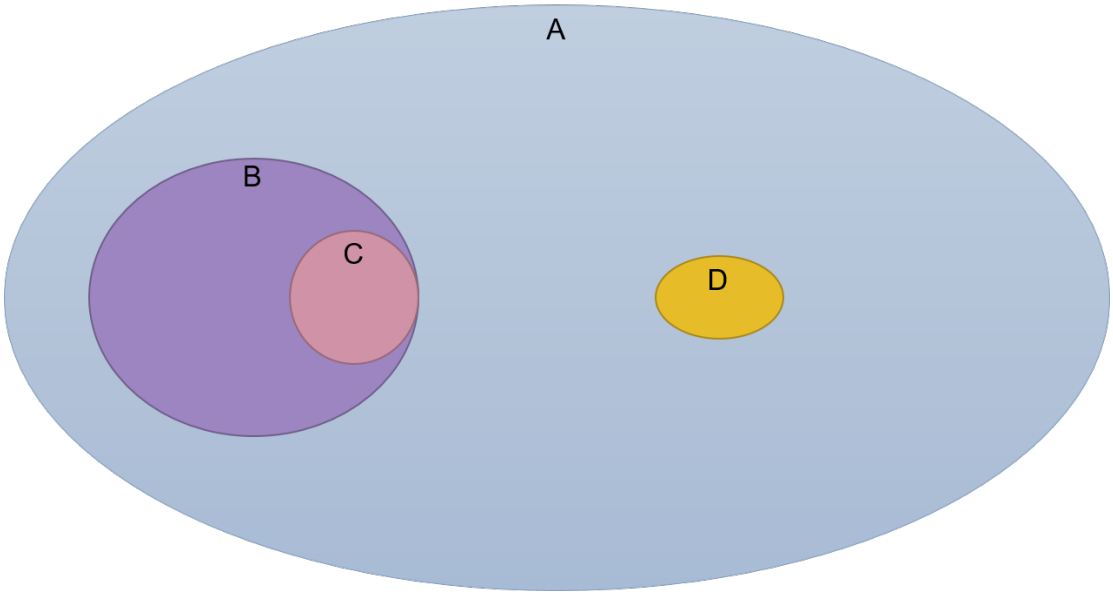


Figure S3. Venn Diagram.
The Venn Diagram was used to represent the relationship of different case sets. Set A contained all the 601 cases; Set B contained the 214 cases after removing the missing values (Figure S2). Set C was a subset of the set B and contained 65 cases that were used for the holdout testing. Set D contained the 39 progressed severe cases, which had no missing values in term of the selected 4 variables and had no intersection with the Set B. The 4 selected variables were lymphocyte, lactic dehydrogenase, C-reactive protein and neutrophil.

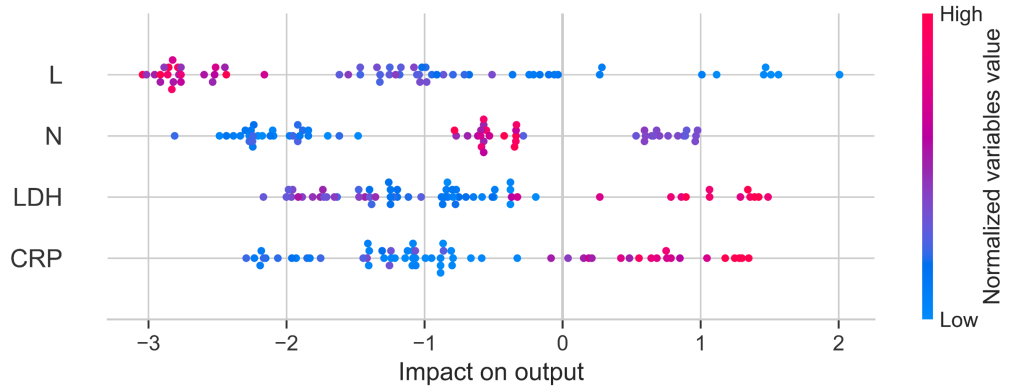


Figure S4. SHapley Additive exPlanations (SHAP) Plot of the key variables identified to construct the model. The SHAP values were plotted to visualize how the value of each variable would impact the prediction of disease severity. Each dot corresponded to an individual patient in the study. And the colors encoded the values of key variables, while the SHAP values represented the impact of key variables on the prediction.
Abbreviations: L, lymphocyte; LDH, lactic dehydrogenase; CRP, C-reactive protein; N, neutrophil.

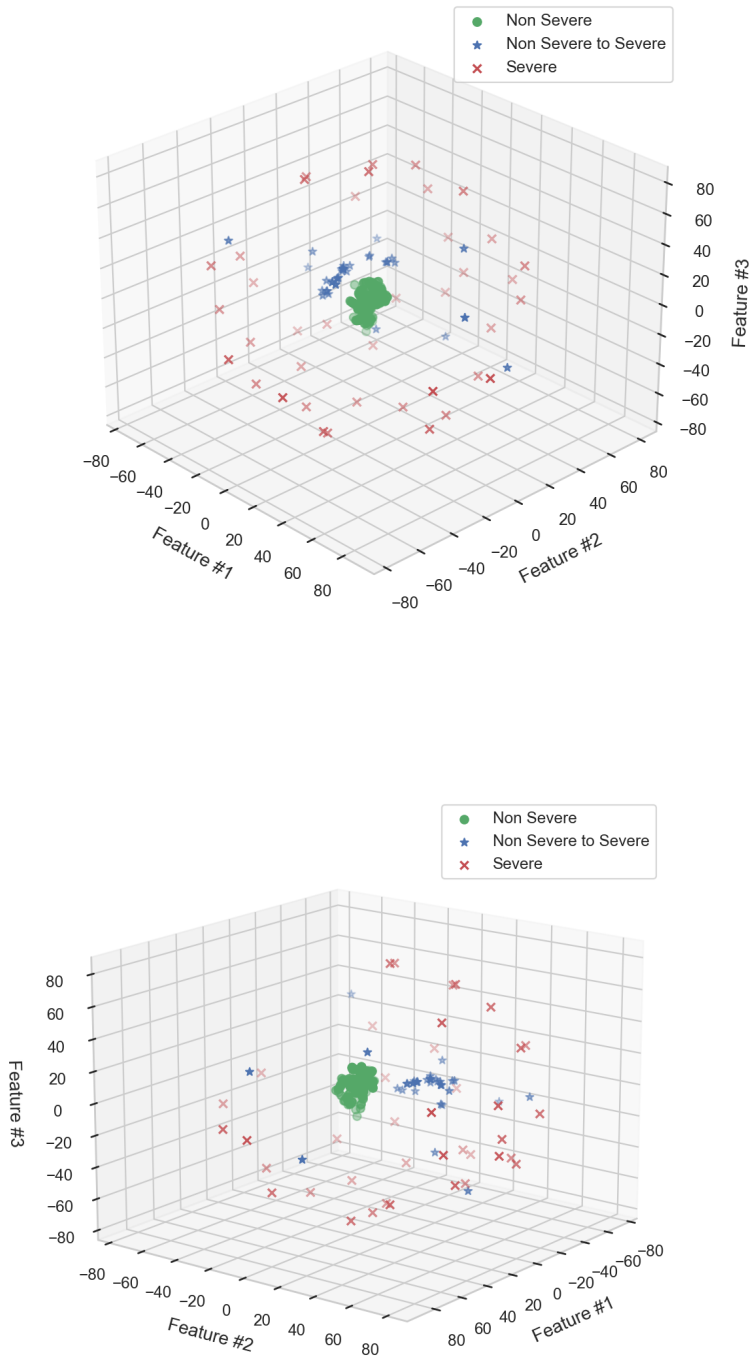


Figure S5. t-SNE visualization
The 3D scatter visualization was generated though the t-SNE algorithm, which projected 4D (lymphocyte, lactic dehydrogenase, C-reactive protein, neutrophil) samples into a 3D feature space. The space was defined by the first, second and third component (Features #1, #2, and #3). The green dots, blue stars and red crosses denoted the non-severe, progressed severe (non-severe to severe progression) and severe cases respectively. Green dots clustered in the center while red crosses scattered in the outer layer; the blue stars located in-between.

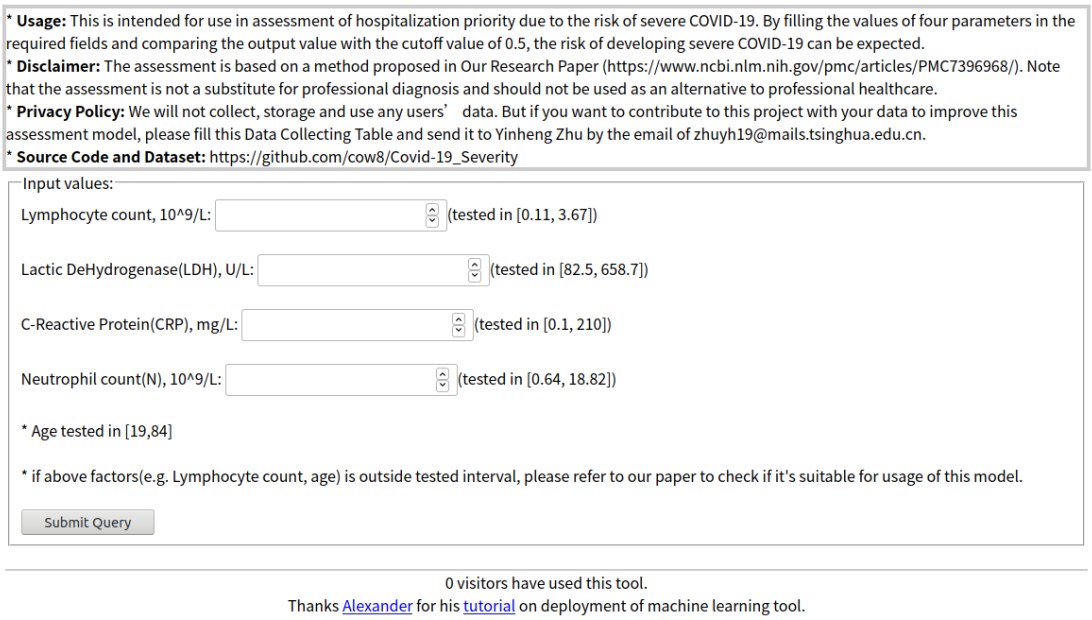


Figure S6. Interface of the online COVID-19 assessment program. (urls: covid-19.zyh.science:8888)

Table S1 List of eligible literature in WHO COVID-19 database

No.	Title
1	COVID-19: Serial CT Findings
2	COVID-19 in Taiwan
3	2019-nCoV severe adult respiratory distress syndrome in two cases in Italy
4	A familial cluster of pneumonia associated with the 2019-nCoV indicating person-to-person transmission
5	A Locally Transmitted Case of SARS-CoV-2 Infection in Taiwan
6	Analysis of the first cluster of cases in a family of novel coronavirus pneumonia in Gansu Province
7	Asymptomatic cases in a family cluster with SARS-CoV-2 infection
8	Asymptomatic novel coronavirus pneumonia patient outside Wuhan
9	Case of the Index Patient Who Caused Tertiary Transmission of Coronavirus Disease 2019 in Korea
10	Changes of CT findings in a 2019-nCoV pneumonia patient
11	Chest CT for Typical 2019-nCoV Pneumonia: Relationship to Negative RT-PCR Testing
12	Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury
13	Clinical characteristics and therapeutic procedure for four cases with COVID-19 receiving combined Chinese and Western medicine treatment
14	Clinical Features and Treatment of 2019-nCov Pneumonia Patients in Wuhan: Report of A Couple Cases
15	Clinical Features of Atypical COVID-19 with an initially Negative RT-PCR Assay
16	CT features of 2019-novel coronovirus pneumonia
17	CT Imaging of the COVID-19
18	CT Manifestations of Two Cases of 2019-nCoV Pneumonia
19	Early transmission patterns of COVID-19 in travellers from Wuhan to Thailand
20	Evolution of CT Manifestations in a Patient Recovered from COVID-19 in Wuhan
21	Evaluation of coronavirus in tears and conjunctival secretions of patients with SARS-CoV-2 infection
22	First Atypical case of 2019 novel coronavirus in Yan'an
23	First Case of 2019 Novel Coronavirus in the US
24	First case of COVID-19 in Taiwan
25	First cases of COVID-19 in France
26	First imported case of 2019 novel coronavirus in Canada
27	Identification of a novel coronavirus causing severe pneumonia in human
28	Importation and Human-to-Human Transmission
29	Journey of a Thai Taxi Driver and Novel Coronavirus
30	Novel Coronavirus Pneumonia Outbreak in 2019
31	Pathological findings of COVID-19 associated with ARDS
32	Pre- and Posttreatment Chest CT Findings
33	Pulmonary pathology of early phase COVID-19 in two patients with lung cancer
34	RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak
35	SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients
36	The COVID-19 with onset of oculomotornerve palsy
37	The course of clinical diagnosis and treatment of a case infected with COVID-19
38	The first 2019 novel coronavirus case in Nepal
39	The First Case of COVID-19 Imported into Korea from Wuhan
40	The first Vietnamese case of COVID-19 acquired from China
41	The report of two cases infection with 2019-ncov after kidney transplantation
42	Two clinical cases of NCP in renal transplant recipients
43	Viral Load Kinetics of SARS-CoV-2 Infection in First Two Patient

Table S2. Univariate analysis of difference in clinical variables between severe and non-severe COVID-19 cases excluding missing values

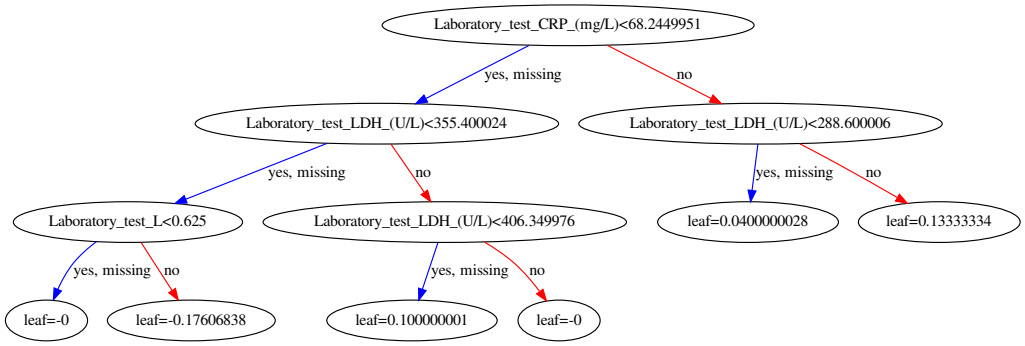
Clinical features	P values
Age	< 0.001
Gender	0.708
Underlying comorbidities	
Hypertension	0.994
Endocrine diseases	0.998
Cardiovascular diseases	0.955
Chronic lung diseases	0.391
Gastrointestinal or hepatobiliary diseases	0.326
Cerebrovascular diseases or neurological disorder	1.000
Immune disorder	1.000
Others	0.814
Signs or symptoms	
Fever	0.003
Cough	0.072
Dyspnea	< 0.001
Fatigue	0.073
Nausea or emesis	0.864
Myalgia	1.000
Dizziness or headache	0.955
Laboratory findings	
White blood cell	0.100
Lymphocyte	< 0.001
Neutrophil	< 0.001
C-reactive protein	< 0.001
Lactate dehydrogenase	< 0.001
D-dimer	0.001
Alanine aminotransferase	0.003
Aspartate aminotransferase	< 0.001
Total bilirubin	0.160
Albumin	< 0.001
Serum creatinine	0.395
Creatine kinase	< 0.001

Table S3. Definition of the evaluation metrics.

Metrics	Definition
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Sensitivity	$\frac{TP}{TP + FP}$
Specifity	$\frac{TP}{TN + FP}$
F1 weighted	$\frac{1}{TP + FP} * \frac{2TP}{TP + FN + 1} + \frac{1}{TN + FN} * \frac{2TN}{TN + FP + 1}$

Abbreviations: TP, true positive; TN, true negative; FP, false positive; FN, false negative

Table S4. Previously validated Single-tree XGBoost model for identification of severe COVID-19 risk. A decision rule using three key features and their thresholds in absolute values was shown in the figure and the performance of the single-tree XGBoost model in identification of severe COVID-19 cases on admission was shown in the table.



AUC	F1 weighted	Accuracy	Sensitivity	Specificity
0.680	0.769	0.800	0.231	0.942

Abbreviations: L, lymphocyte; LDH, lactic dehydrogenase; CRP, C-reactive protein; AUC: area under the curve.