Eleven routine clinical features predict COVID-19 severity uncovered by machine learning of longitudinal measurements

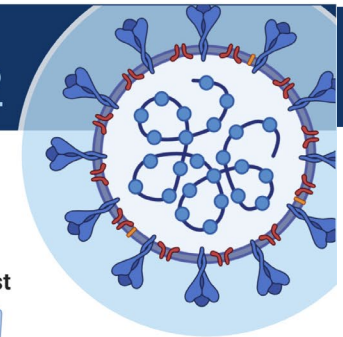
Dr. Yaoting Sun

Laboratory of Proteomics Big Data
www.guomics.com

西湖大學
WESTLAKE UNIVERSITY

Background

SARS-CoV-2



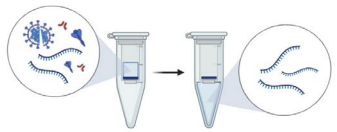
EPIDEMIOLOGY

Difficulty breathing or shortness of breath

- Non-severe (81%)
- Severe (14%)
- Critical (5%)

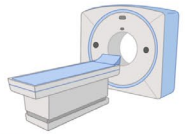
TRADITIONAL DIAGNOSIS

SARS-CoV-2 nucleic acid test



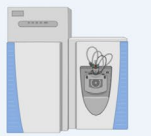
- This is a qualitative test showing whether the patient is infected or not.

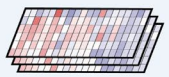
CT Scan-Chest



- About 20% of COVID-19 patients show no obvious imaging changes in the lung.

OUR NEW TECHNOLOGY: Proteome, Metabolome or Clinical Factors & Machine Learning





Proteome & Metabolome

Basic information

Days of symptom onset

Underlying diseases

Body temperature

Blood-gas analyzer

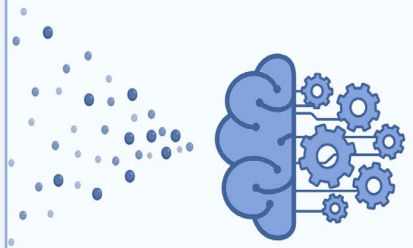
Symptoms

Blood tests

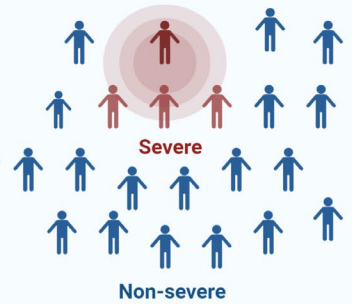
Blood pressure

Urine tests

Flow Cytometer



Machine Learning



Prediction

Background
Proteomic and metabolomic characterization of COVID-19 patient sera

Cell

CellPress

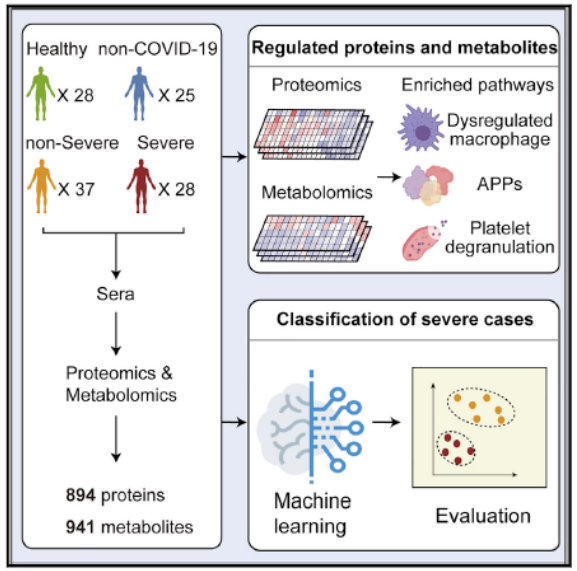
Article

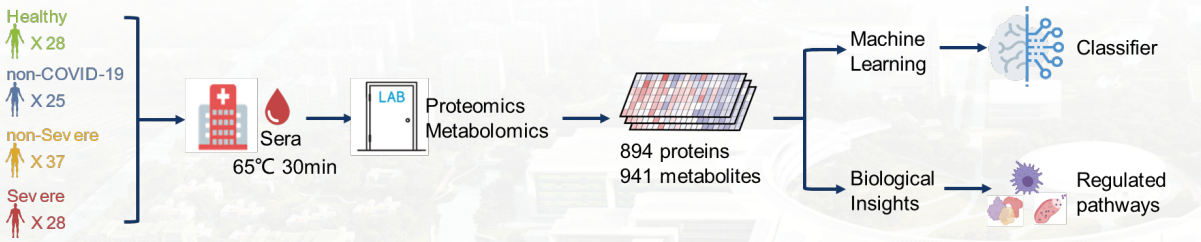
Proteomic and Metabolomic Characterization of COVID-19 Patient Sera

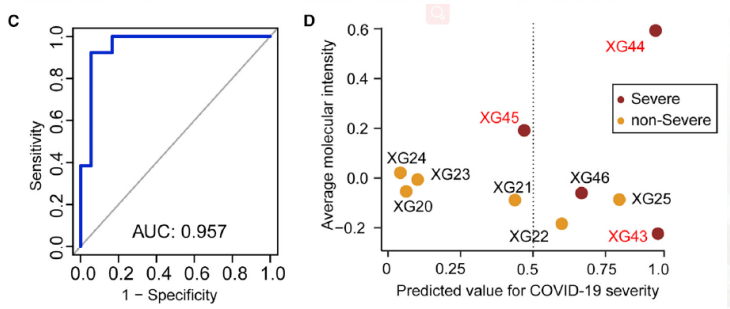
Bo Shen,[1,6] Xiao Yi,[2,3,6] Yaoting Sun,[2,3,6] Xiaojie Bi,[1,6] Juping Du,[1,6] Chao Zhang,[4,6] Sheng Quan,[4,6] Fangfei Zhang,[2,3] Rui Sun,[2,3] Liujia Qian,[2,3] Weigang Ge,[2,3] Wei Liu,[2,3] Shuang Liang,[2,3] Hao Chen,[2,3] Ying Zhang,[1] Jun Li,[1] Jiaqin Xu,[1] Zebao He,[1] Baofu Chen,[1] Jing Wang,[1] Haixi Yan,[1] Yufen Zheng,[1] Donglian Wang,[1] Jiansheng Zhu,[1] Ziqing Kong,[4] Zhouyang Kang,[4] Xiao Liang,[2,3] Xuan Ding,[2,3] Guan Ruan,[2,3] Nan Xiang,[2,3] Xue Cai,[2,3] Huanhuan Gao,[2,3] Lu Li,[2,3] Sainan Li,[2,3] Qi Xiao,[2,3] Tian Lu,[2,3] Yi Zhu,[2,3,5,*] Huafen Liu,[4,5,*] Haixiao Chen,[1,5,*] and Tiannan Guo[2,3,5,7,*]

[1]Taizhou Hospital, Wenzhou Medical University, 150 Ximen Street, Linhai 317000, Zhejiang Province, China
[2]Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China
[3]Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China
[4]Calibra Lab at DIAN Diagnostics, 329 Jinpeng Street, Hangzhou 310030, Zhejiang Province, China
[5]Senior author
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: zhuyi@westlake.edu.cn (Y.Z.), liuhf1@dazd.cn (H.L.), chenhx@enzemed.com (H.C.), guotiannan@westlake.edu.cn (T.G.)
https://doi.org/10.1016/j.cell.2020.05.032

Graphical Abstract

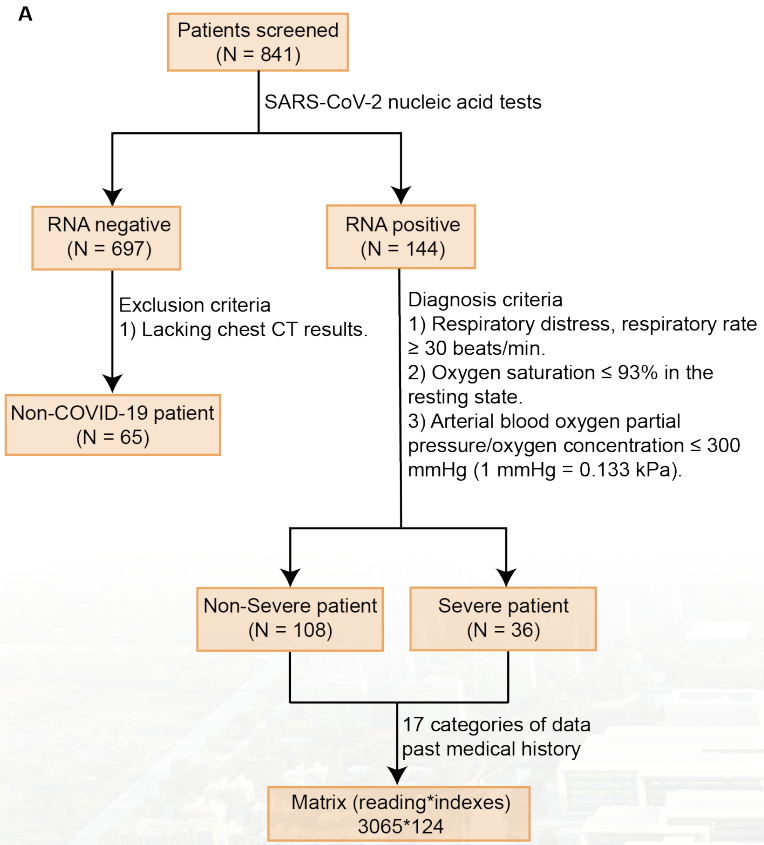




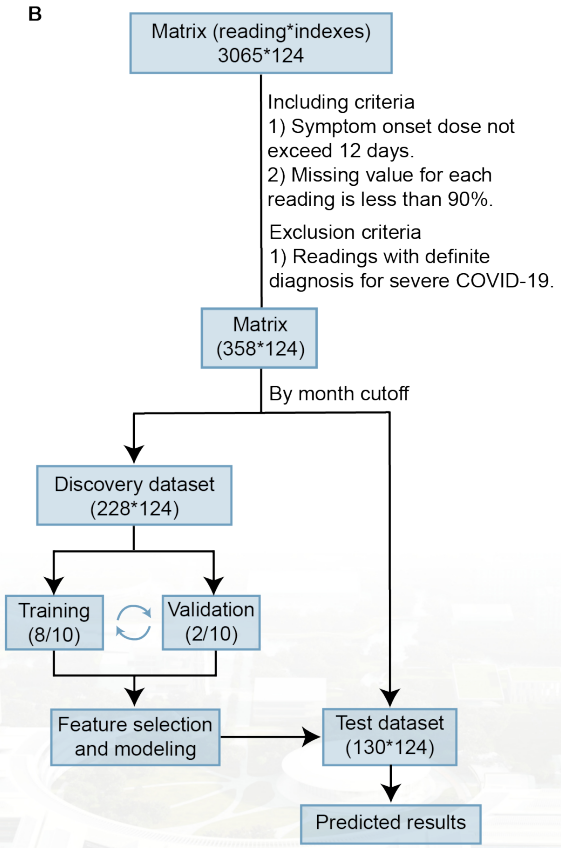


Proteomic and Metabolomic Characterization of COVID-19 Patient Sera
Cell. *2020, 182(1): 59-72 e15 10.1016/j.cell.2020.05.032*

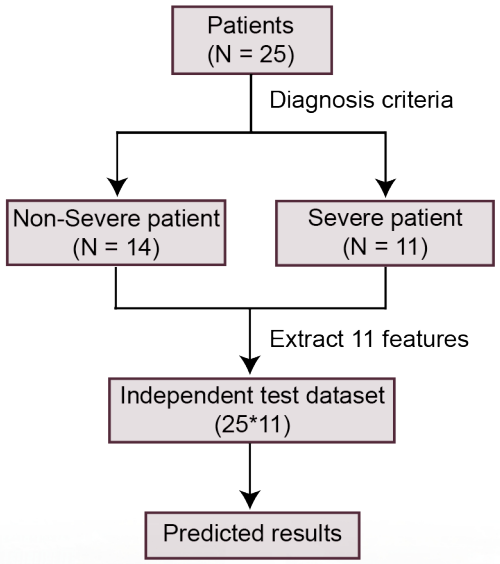
The study design and modeling workflow



The COVID-19 patient cohort



The workflow of the modeling process.

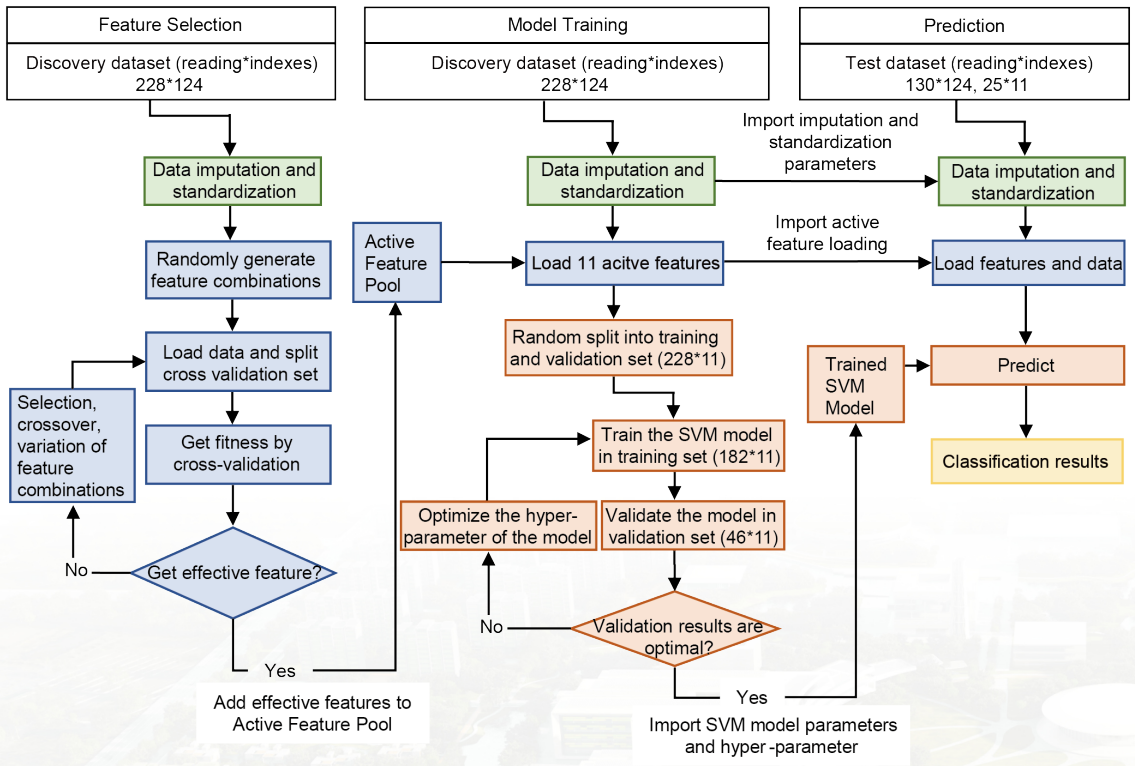


Independent test patient cohort.

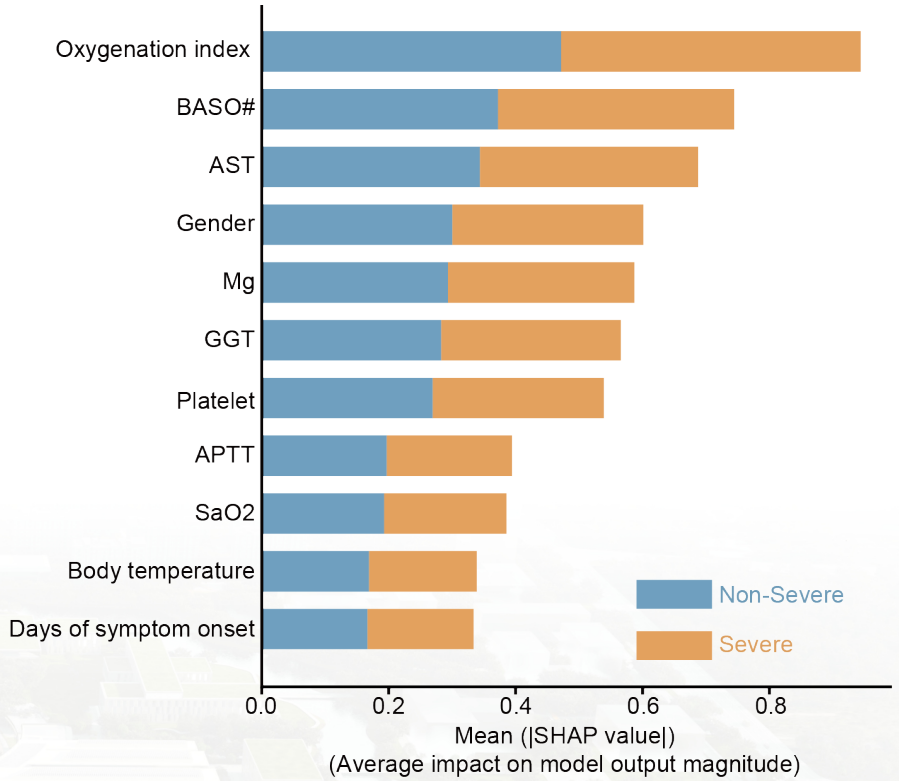
Eleven Routine Clinical Features Predict COVID-19 Severity
medRxiv 2020.07.28.20163022; **(Unpublished, not peer-reviewed)**

The detailed process of building up the machine learning model

A



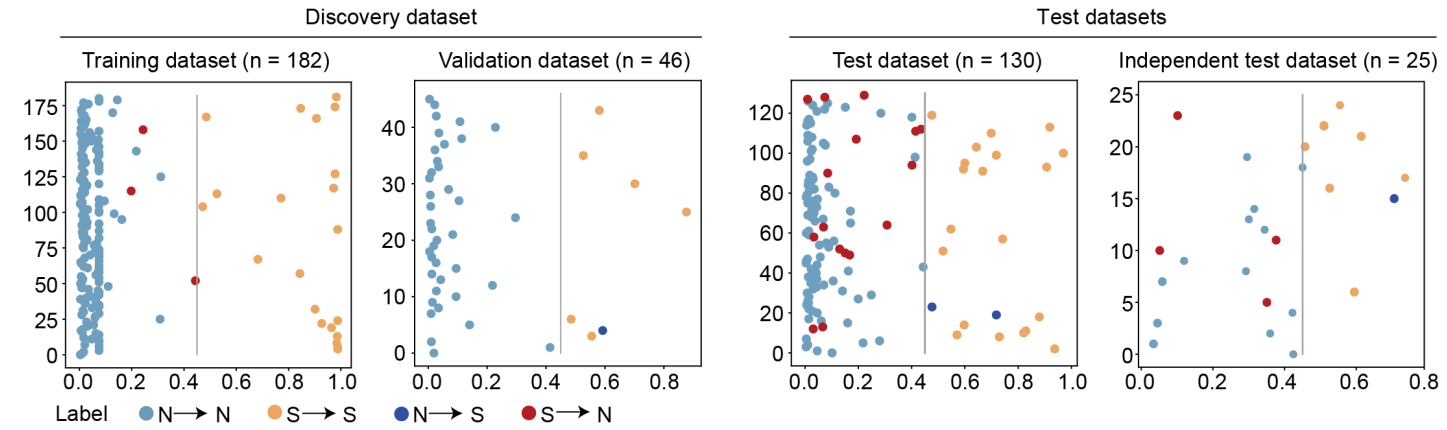
The detailed workflow contains four major steps: i) data preprocessing (green); ii) feature selection (blue); iii) building up the machine learning model (pink); iv) prediction results (yellow).



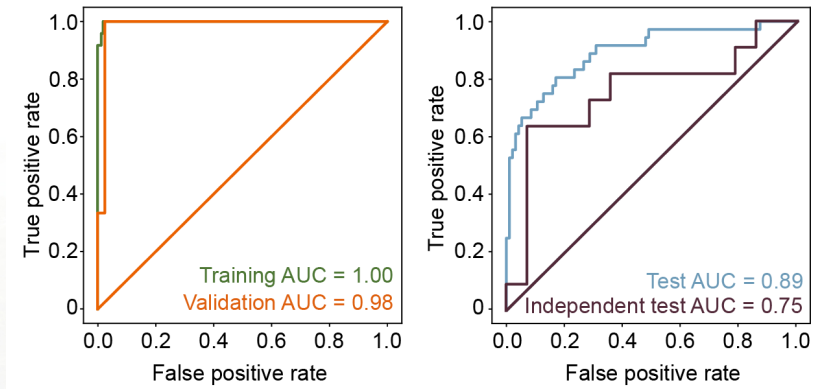
The 11 key clinical features

Evaluation of the performance of the model

A



B



C

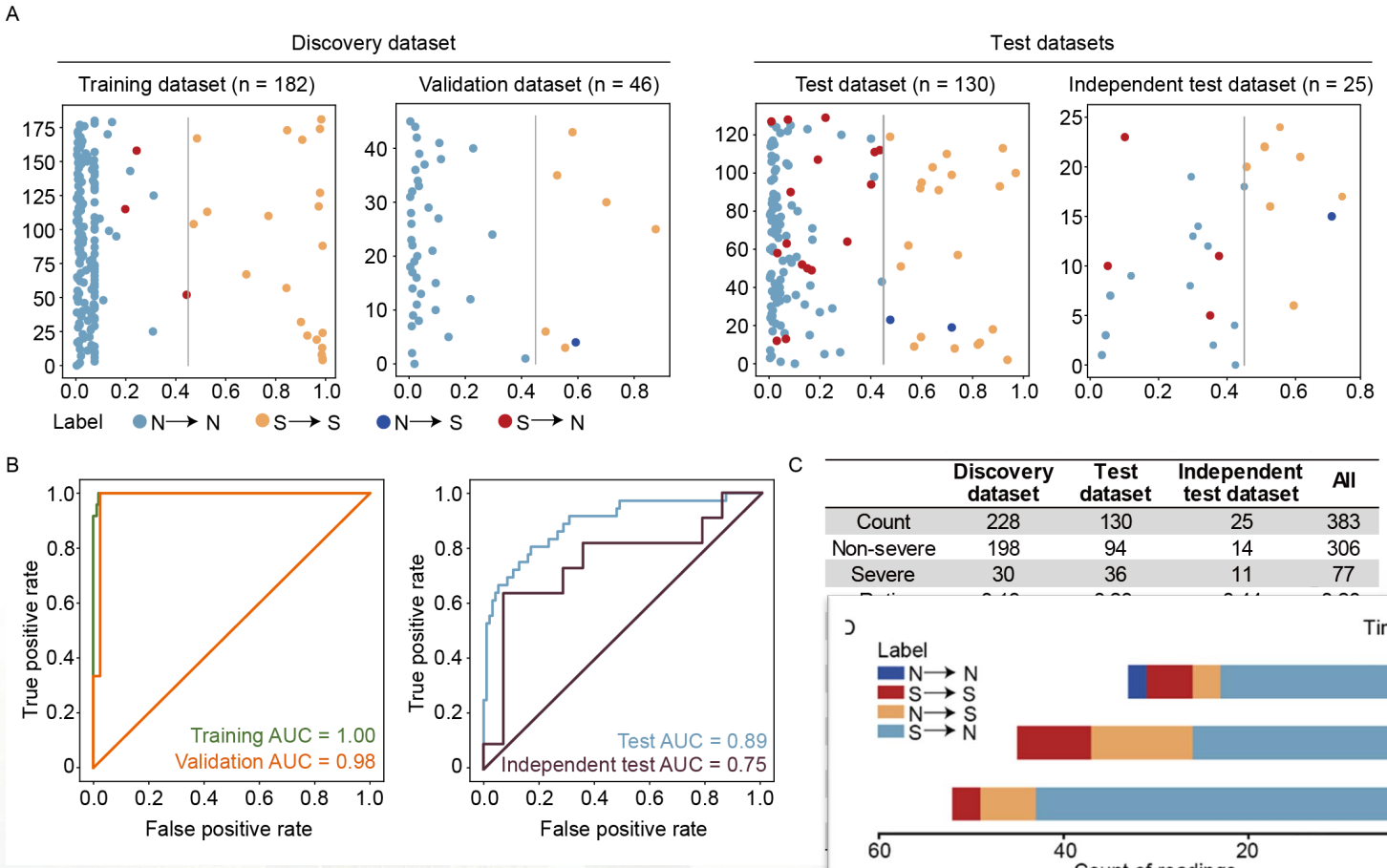
	Discovery dataset	Test dataset	Independent test dataset	All
Count	228	130	25	383
Non-severe	198	94	14	306
Severe	30	36	11	77
Ratio	0.13	0.28	0.44	0.20
N→ N	197	92	13	302
N→ S	1	2	1	4
S→ S	27	20	7	54
S→ N	3	16	4	23
Sensitivity	0.90	0.56	0.64	0.70
Specificity	0.99	0.98	0.93	0.99
PPV	0.96	0.91	0.69	0.93
NPV	0.99	0.85	0.91	0.93
Accuracy	0.98	0.86	0.80	0.93

Severe and Non-severe cases are shown as scatter plots in different colors (red: severe; blue: non-severe). The cutoff of the predicted score was 0.45. X-axis indicates the predicted scores, representing the probability of disease severity for each time point. Y-axis denotes the indexes of samples. N→S indicates a non-severe case which was predicted as a severe case.

ROC plots of the performance of support vector machine (SVM) for severity prediction.

Summary of the performance metrics.

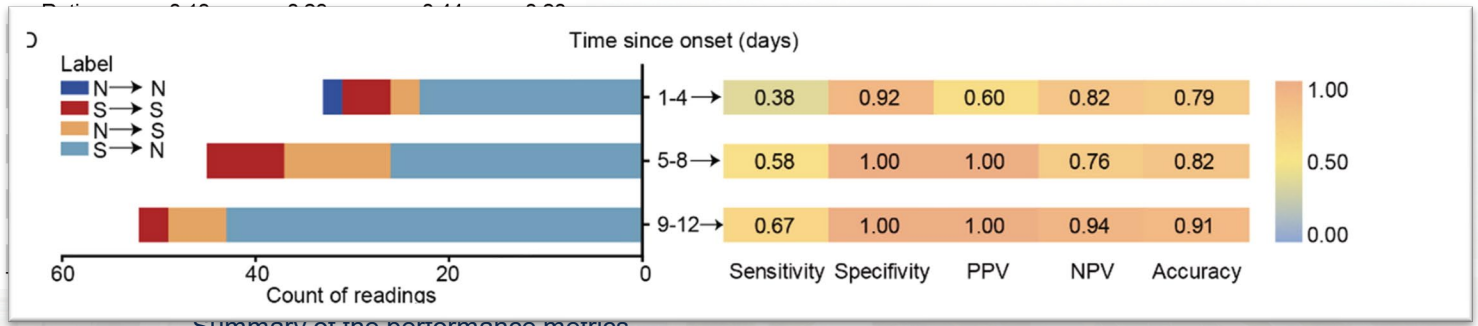
Evaluation of the performance of the model



ROC plots of the performance of support vector machine (SVM) for severity prediction.

C

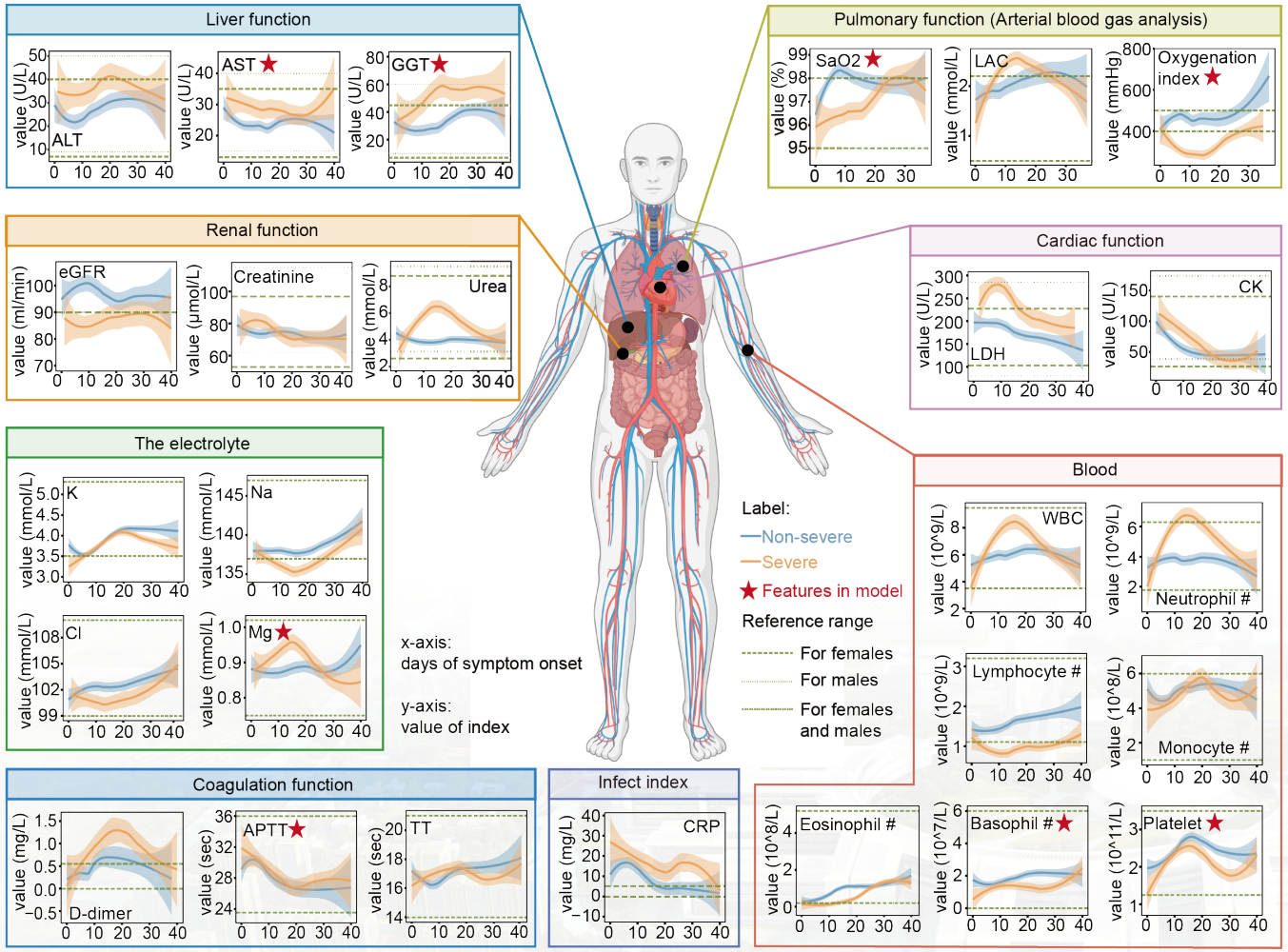
	Discovery dataset	Test dataset	Independent test dataset	All
Count	228	130	25	383
Non-severe	198	94	14	306
Severe	30	36	11	77

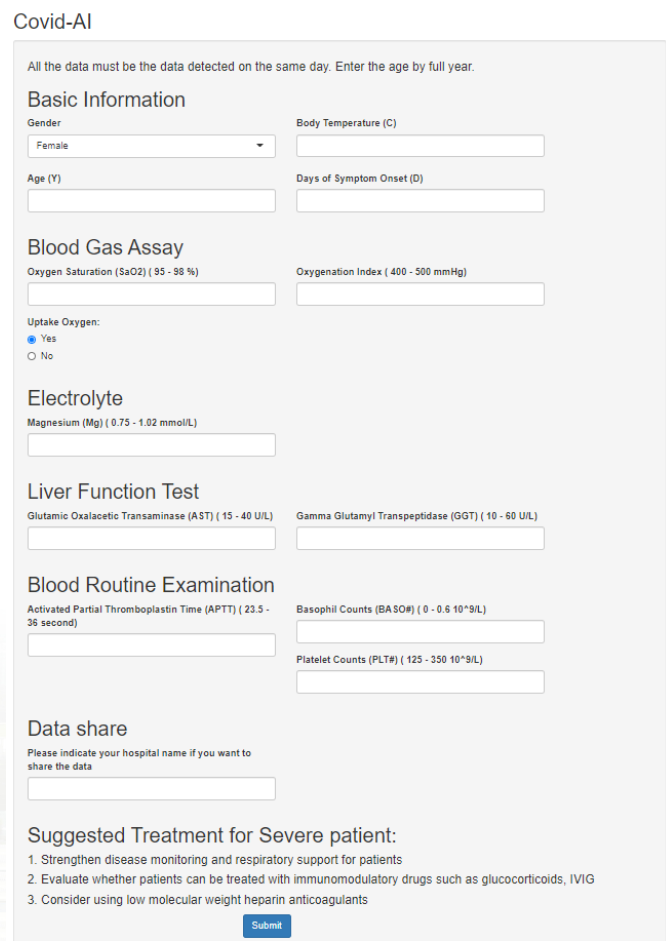


Summary of the performance metrics.

Severe and Non-severe cases are shown as scatter plots in different colors (red: severe; blue: non-severe). The cutoff of the predicted score was 0.45. X-axis indicates the predicted scores, representing the probability of disease severity for each time point. Y-axis denotes the indexes of samples. N→S indicates a non-severe case which was predicted as a severe case.

Dynamic changes of key clinical features over 7 weeks





https://guomics.shinyapps.io/covidAI/

ACKNOWLEDGEMENTS

UOMICS -since 2017

西湖大學 WESTLAKE UNIVERSITY

Stan.Z. Li Lab

台州恩泽医疗中心(集团) TAIZHOU ENZE MEDICAL CENTER

Bo Shen, Jing Wang, Kai Zhou *et al*

ThermoFisher SCIENTIFIC

THANK YOU

西湖大學
WESTLAKE UNIVERSITY