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

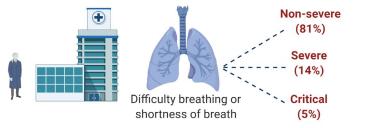




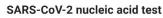
Background

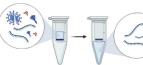
SARS-CoV-2

EPIDEMIOLOGY



TRADITIONAL DIAGNOSIS





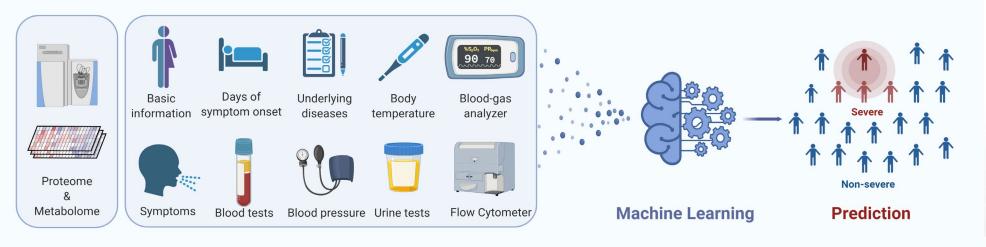
• This is a qualitative test showing whether the patient is infected or not.

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



1



WESTLAKE UNIVERSITY



Background Proteomic and metabolomic characterization of COVID-19 patient sera

Cell

CellPress

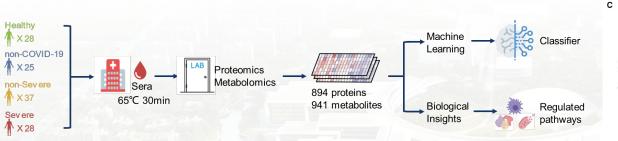
Article

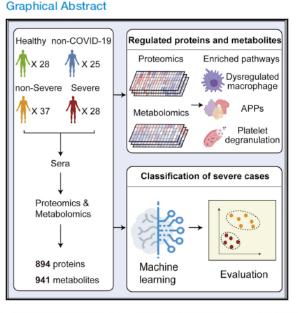
西湖大學

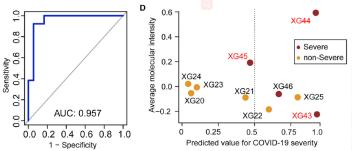
Proteomic and Metabolomic Characterization of COVID-19 Patient Sera

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,^{2,4} 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} Xua 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^{2,3,6,7,*} ¹Taizhou Hospital, Wenzhou Medical University, 150 Ximen Street, Linhai 317000, Zhejiang Province, China ²Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China ³Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China ⁴Calibra Lab at DIAN Diagnostics, 329 Jinpeng Street, Hangzhou 310030, Zhejiang Province, China ⁵Senior author ⁶These authors contributed equally ⁷Lead 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/i.cell.2020.05.032





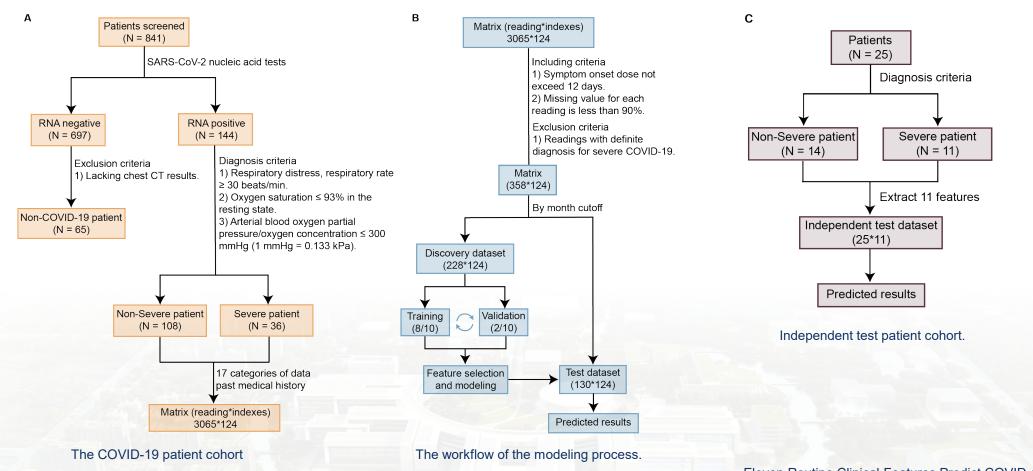


Proteomic and Metabolomic Characterization of COVID-19 Patient Sera **Cell**. 2020, 182(1): 59-72 e15 10.1016/j.cell.2020.05.032

WESTLAKE UNIVERSITY



The study design and modeling workflow



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

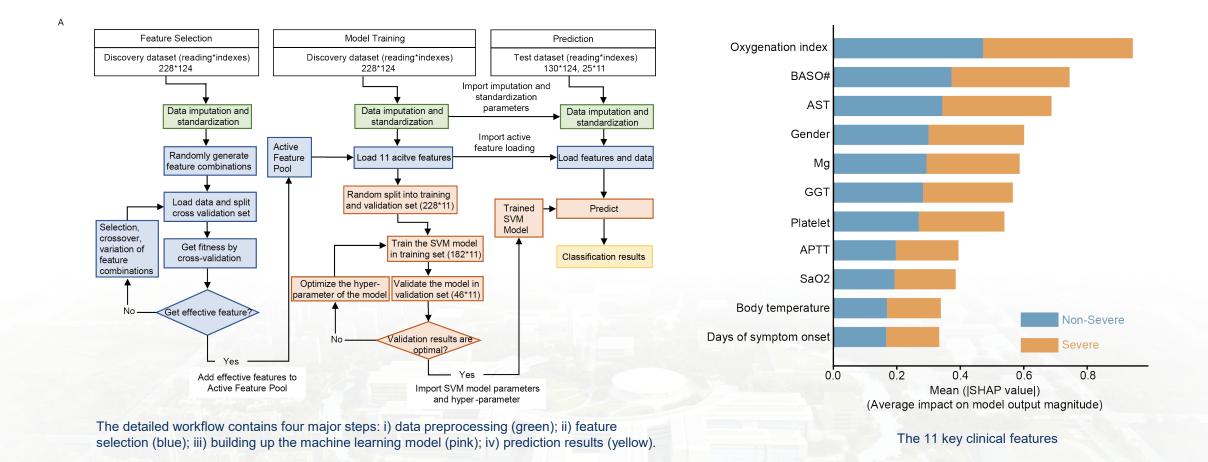




毌

湖大學

The detailed process of building up the machine learning model



WESTLAKE UNIVERSITY

8

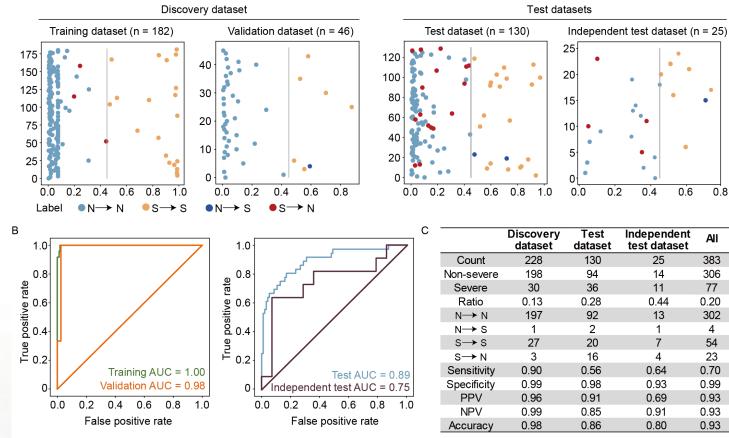


Evaluation of the performance of the model

А

Ŧ

湖大



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

Summary of the performance metrics.

9

0.6

0.8

All

383

306

77

0.20

302

4

54

23

0.70

0.99

0.93

0.93

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. Yaxis denotes the indexes of samples. $N \rightarrow S$ indicates a non-severe case which was predicted as a severe case.



А

Ŧ

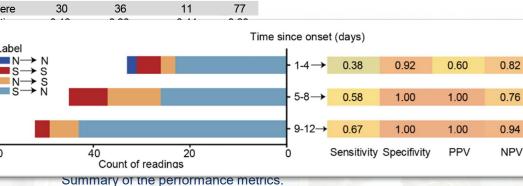
湖大

Evaluation of the performance of the model

Discovery dataset Test datasets Training dataset (n = 182) Test dataset (n = 130) Independent test dataset (n = 25) Validation dataset (n = 46) 25 175 120 40 150 • • 20 100 125 30 80 15 100 60 20 75 10 40 50 10 20 25 0 0.4 0.2 0.2 0.4 0.6 0.8 0.0 0.6 0.8 0.0 0.2 0.4 0.6 0.8 0.0 1.0 0.0 0.2 1.0 Label $N \rightarrow N$ $\bullet s \rightarrow s$ $\bullet N \rightarrow$ S • S⁻ → N В С Discovery Test Independent 1.0 1.0 dataset dataset 130 Count 228 198 94 0.8 0.8 Non-severe rate rate 30 36 Severe True positive ra 9.0 positive ר Label N→ N 9 L 0.2 S→S ■ N→ S S→ N Training AUC = 1.00 Test AUC = 0.89 Validation AUC = 0.98 VIndependent test AUC = 0.75 0 0.2 0.4 0.6 0.8 1.0 0.2 0.4 0.6 0.8 1.0 0.0 0.0 False positive rate False positive rate 60 40

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

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. Yaxis denotes the indexes of samples. $N \rightarrow S$ indicates a non-severe case which was predicted as a severe case.



0.4

test dataset

25

14

0.6

0.8

All

383

306

0.79

0.82

0.91

Accuracy

1.00

0.50

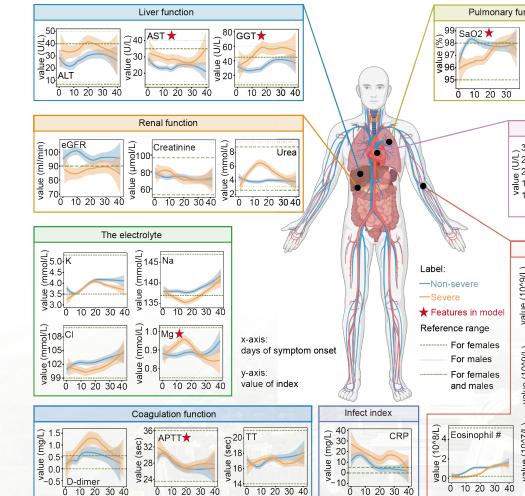
0.00

9



西湖大學

Dynamic changes of key clinical features over 7 weeks



ry function (Arterial blood gas analysis)	Covid-AI
Understanding (10 20 30) (10 20	All the data must be th Basic Informa Gender Female Age (Y)
Cardiac function	Blood Gas As
300 250 200 9200 100 100 0 10 20 30 40 100 0 10 20 30 40 0 10 20 30 40	Oxygen Saturation (SaO2) Uptake Oxygen: Yes No Electrolyte Magnesium (Mg) (0.75 - 1.
Blood	
$ \begin{array}{c} 1\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\$	Blood Routine Activated Partial Thrombo 35 second)
13^{-1}_{0} Lymphocyte # 10^{-1}_{0} Lymphocyte # 10^{-1}_{0} 2 10^{-1	Data share Please indicate your hospi share the data
TLZv01 Basophil #★ TLV01 and Display the second secon	Suggested Tr 1. Strengthen disease 2. Evaluate whether p 3. Consider using low
40 0 10 20 30 40 0 10 20 30 40	https://g

10

Jovia-Al		
All the data must be the data detected on the sa	me day. Enter the age by full year.	
Basic Information	Body Temperature (C)	
Female •	body reinperature (c)	
remaie		
Age (Y)	Days of Symptom Onset (D)	
Blood Gas Assay		
Oxygen Saturation (SaO2) (95 - 98 %)	Oxygenation Index (400 - 500 mmHg)	
Uptake Oxygen:		
Yes		
O No		
Electrolyte		
Magnesium (Mg) (0.75 - 1.02 mmol/L)		
Liver Free time Test		
Liver Function Test		
Glutamic Oxalacetic Transaminase (AST) (15 - 40 U/L)	Gamma Glutamyl Transpeptidase (GGT) (10 - 60 U/L)	
Blood Routine Examination		
Activated Partial Thromboplastin Time (APTT) (23.5 - 36 second)	Basophil Counts (BASO#) (0 - 0.6 10^9/L)	
	Platelet Counts (PLT#) (125 - 350 10^9/L)	
Data share		
Please indicate your hospital name if you want to		
share the data		
Suggested Treatment for Severe patient:		
1. Strengthen disease monitoring and respiratory support for patients		
	immunomodulatory drugs such as glucocorticoids, IVIG	
3. Consider using low molecular weight heparin	anticoagulants	
Submit		

https://guomics.shinyapps.io/covidAl/

WESTLAKE UNIVERSITY

ACKNOWLEDGEMENTS



THANK YOU

