CIRPMC
An online model with simplified inflammatory signature to predict the occurrence of critical illness in patients with COVID-19
Gao, Yue; Chen, Lingxi; Zeng, Shaoqing; Feng, Xikang; Chi, JianHua; Wang, Ya; Li, Huayi; Jiang, Tengping; Yu, Yang; Jiao, XiaoFei; Liu, Dan; Feng, XinXia; Wang, SiYuan; Yu, RuiDi; Yuan, Yuan; Xu, Sen; Cai, Guangyao; Xiong, Xiaoming; Chen, Pingbo; Mo, Qingqing; Jin, Xin; Wu, Yuan; Ma, Ding; Li, Chunrui; Li, Shuai Cheng; Gao, Qinglei
Published in: Clinical and Translational Medicine
Published: 01/10/2020

Document Version:
Final Published version, also known as Publisher's PDF, Publisher’s Final version or Version of Record

License:
CC BY

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1002/ctm2.210

Publication details:
Gao, Y., Chen, L., Zeng, S., Feng, X., Chi, J., Wang, Y., Li, H., Jiang, T., Yu, Y., Jiao, X., Liu, D., Feng, X., Wang, S., Yu, R., Yuan, Y., Xu, S., Cai, G., Xiong, X., Chen, P., ... Gao, Q. (2020). CIRPMC: An online model with simplified inflammatory signature to predict the occurrence of critical illness in patients with COVID-19. Clinical and Translational Medicine, 10(6), [e210]. https://doi.org/10.1002/ctm2.210

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.
Dear editor,

The coronavirus disease 2019 (COVID-19) is characterized by heterogeneous clinical features and multiple organ damage. Many patients with mild symptoms can suddenly develop into critical illness and progress to a refractory state that has significantly increased mortality, indicating the necessity to promptly identify patients at high risk of physiologic deterioration before the occurrence of critical COVID-19. With both innate and adaptive immune compartments contribution, cytokine storm in covid-19 is widely concerned. Hyperinflammatory response induced by immune dysfunction is reported to underpin critical COVID-19.1 Uncontrolled release of cytokines results in tissue damage and further leads to multiple organ failure, which is the major cause of death in patients with COVID-19.2 As expected, the differences of multiple cytokines and immune features between critical ill and noncritical ill patients were observed in clinical practice. Besides, early seroconversion and high antibody titer were linked with less severe clinical symptoms. Using inflammatory/immune factors to predict the risk of developing critical COVID-19 under the assistance of machine learning (ML) is promising to aid management of the disease, but rarely reported.

Electronic health records (EHRs) harbor valuable resources generated from routine medical activities and have been widely used. However, medical data are often complex, multidimensional, nonlinear, heterogeneous, and required to be analyzed using more effective statistical methods than traditional logistic regression. ML is a subfield of artificial intelligence that encapsulates statistical and mathematical algorithms, which enables facts interrogation and complex decision making through a given set of data. The combination of EHRs and ML shows potential applications in predicting the risk of atherosclerotic cardiovascular disease and gestational diabetes.

In this multicenter study, we developed an online model with four inflammatory factors (C reactive protein [CRP], tumor necrosis factor α [TNF-α], interleukin 2 receptor [IL-2R], and interleukin 6 [IL-6]) that enabled accurate identification of COVID-19 patients prone to critical illness approximately 20 days in advance. The model was validated in an internal validation cohort (SFV cohort) and an external validation cohort (OV cohort). Study design is presented in Figure 2A. The detailed demographic and characteristics of patients are shown in Table S1.

A total of 15 raw inflammatory/immune features were collected from COVID-19 patients at admission. After feature filtering (Figure S1) and data imputation (Figure S2),3 eight features were fitted into Least Absolute Shrinkage and Selection Operator (LASSO)4 logistic regression for feature selection (Figure 1A). As illustrated in Figure 1B, we considered features whose coefficients equaled to zero as redundant and less predictive features. As a result, LASSO analysis identified four features (CRP, TNF-α, IL-2R, and IL-6) for the development of critical illness classifier.

We conducted the Spearman correlation analysis between the four features and critical illness status. Figure S3A indicates that the positive correlation at varying degrees existed across five features. The top weighted features, IL-6 ($R = 0.49$), CRP ($R = 0.47$), IL-2R ($R = 0.43$), and TNF-α ($R = 0.37$), were consisted with previously reported risk factors that were highly correlated with poor outcome of COVID-19. Standard box plots presented significant differences ($P < 2.2e-16$) of the four features between critically ill and noncritically ill COVID-19 patients (Figure S3B). The median (IQR) expression of TNF-α (17.0, 10.5-29.3, pg/mL), CRP (182.7, 103.2-258.6, mg/L), IL-2R (1447.0, 993.0-2327.5, U/mL), and IL-6 (169.4, 58.1-640.9, pg/mL) was significantly higher in critically ill patients compared with TNF-α (8.1, 6.1-10.5, pg/mL), CRP (16.1, 2.5-57.8, mg/L), IL-2R (520.0, 299.0-770.5, U/mL), and IL-6 (5.0, 2.1-18.4, pg/mL) in noncritically ill patients.
During model development stage, five models (support vector machine [SVM], logistic regression [LR], gradient boosted decision tree [GBDT], k-nearest neighbor [KNN], and neural network [NN]) were trained for risk prediction. In general, all five models showed varying but promising critical illness risk prediction performance in the internal and external validation cohorts. CIRPMC (critical illness risk prediction model for COVID-19) derived from SVM achieved the highest predictive performance. Relative feature importance rank of SVM is shown in Figure S4. As binary classifier, CIRPMC outputted critical illness risk probability \( P \) ranged from 0 to 1 for each patient, and stratified patients with \( P < .5 \) as low risk, otherwise high risk. For SFV cohort, CIRPMC achieved an AUC (area under the receiver operating characteristics curve) of 0.946 (95% CI 0.923-0.969) to identify patients having high risk of developing critical illness with an accuracy of 92.7% (95% CI 90.4%-94.6%). For OV cohort, CIRPMC
demonstrated an AUC of 0.969 (95% CI 0.945-0.992) and an accuracy of 96.6% (95% CI 95.1-97.7%) (Figure 1C, D). The calibration curve of CIRPMC in two validation cohorts is depicted in Figure S5. Intriguingly, CIRPMC also displayed the minimal Brier score of 0.057 for SFV cohort and 0.028 for OV cohort. All other metrics and the performance of other models are listed in Table 1.

With critical illness as status and time from admission to critical illness or discharge as the endpoint, Kaplan-Meier analysis further confirmed the risk stratification ability of the model. CIRPMC robustly stratified high-risk patients and low-risk patients with $P < .0001$ in both internal and external validation cohorts. The univariate Cox analysis also demonstrated the positive correlation between CIRPMC predicted critical illness subgroup and the ground truth critical illness survival for internal (HR: 22.52, 95% CI 14.69-34.53) and external (HR: 54.30, 95% CI 32.21-91.52) validation cohorts, respectively (Figure 1E, F).

Additionally, we opened up an online calculator based on CIRPMC to input the values of features needed for risk prediction of COVID-19 patients (https://cirpmc.deepomics.org/). After the clinicians fill in the online form with corresponding features, CIRPMC returns a personalized probability and risk group of critical illness. Illustration of an example of the online prediction system is presented in Figure 2B. In this study, CIRPMC was developed to identify COVID-19 patients with high risk of developing critical illness and achieved high predictive performance with an AUC range from 0.946 to 0.969 across the internal and external validation cohorts. The accurate and rapid risk stratification is critical to ensure health systems agile and hopefully will optimize patient outcomes where “time is life.”

Certain interpretability is a strength of CIRPMC. In accord with previous reports, we found that the expression levels of four contributive inflammatory cytokines (CRP, TNF-α, IL-2R, and IL-6) were significantly higher in critically ill patients than those in noncritically ill patients.5 Another strengths of CIRPMC are its stability and generalizability. Four features used for prediction are readily accessible and frequently monitored in routine clinical practice. Besides, they are relatively objective, solid, and less susceptible to human memory bias, suggesting that CIRPMC is not susceptible to human interference and has strong generalization to be extended to other medical institutions.

During the pandemic, there has emerged many studies on prognosis prediction of COVID-19.6–8 However, the sample size of most studies is small, thus harboring risks of overfitting.6,7 Moreover, most studies lack independent external validation or the number of patients within external validation is limited,9,10 which can impair the reproducibility and credibility of models. Our study is with larger sample size, independent external validation, detailed patient description, and relatively long observation time (18-20 days).

However, the study has some limitations. First, patients included are primarily locals in Wuhan. Data from multiple provinces or countries could further improve the applicability and robustness of models. Besides, the prognostic implication of CIRPMC has not been evaluated in prospective cohorts due to the retrospective nature of this study.
In conclusion, this retrospective, multicenter study showed CIRPMC with readily available features holds great potential in accurately and timely (approximately 20 days in advance) identifying COVID-19 patients prone to develop into critical illness. The model held strong stability, generalizability, universality, and wide prediction horizon to be easily extended to areas with limited medical resources. The proposed model potentially assists clinicians to locate the patients with a higher priority to be early intervened and intensively monitored, and eliminate delays to maximize the number of survivors during the rapidly developing global emergency. Equipped with high predictive performance, the online calculator CIRPMC deserves to be proceeded with. However, these findings warrant further validations in prospective clinical trials.

| SFV cohort | SVM | LR | GBDT | KNN | NN | OV cohort | SVM | LR | GBDT | KNN | NN |
|------------|-----|----|------|-----|----|-----------|-----|----|------|-----|----|
| **AUC (95% CI)** | 0.946 (0.923-0.969) | 0.944 (0.921-0.966) | 0.928 (0.912-0.946) | 0.984 (0.983-0.994) | 0.939 (0.935-0.966) |
| **Accuracy (95% CI)** | 0.633 (0.546-0.747) | 0.728 (0.659-0.840) | 0.663 (0.559-0.759) | 0.918 (0.885-0.943) | 0.949 (0.894-0.981) |
| **SN (95% CI)** | 0.976 (0.959-0.987) | 0.916 (0.889-0.936) | 0.942 (0.919-0.969) | 0.972 (0.954-0.984) | 0.972 (0.954-0.984) |
| **SP (95% CI)** | 0.939 (0.915-0.963) | 0.990 (0.974-0.992) | 0.932 (0.916-0.949) | 0.991 (0.988-0.994) | 0.990 (0.981-0.997) |
| **PPV (95% CI)** | 0.076 (0.000-0.100) | 0.067 (0.000-0.100) | 0.092 (0.000-0.100) | 0.100 (0.000-0.100) | 0.100 (0.000-0.100) |
| **NPV (95% CI)** | 0.000 (0.000-0.000) | 0.000 (0.000-0.000) | 0.000 (0.000-0.000) | 0.000 (0.000-0.000) | 0.000 (0.000-0.000) |
| **Brier** | 0.017 | 0.016 | 0.019 | 0.017 | 0.016 |

**Abbreviations:** AUC, area under the curve; SVM, supported vector machine; GBDT, gradient boosted decision tree; KNN, k-nearest neighbour; LR, logistic regression; NN, neural network; PPV, positive predictive value; NPV, negative predictive value; SN, sensitivity; SP, specificity.

**ACKNOWLEDGMENTS**

We are grateful to all health-care workers and people nationwide and worldwide, who are involved in the fighting against COVID-19. The opinions expressed reflect the collective views of the coauthors. The study was supported by the National Science and Technology Major Sub-Project (2018ZX10301402-002), the Technical Innovation Special Project of Hubei Province (2018ACA138), the National Natural Science Foundation of China (81572570, 81974405, 81772878, 81873452, 81702572, 81702574, and 82072889), Hubei Natural Science Foundation (2019CFB453), and the Fundamental Research Funds for the Central Universities (2017JYXJJ025, 2018JYXJJ001, and 2019kyXMZ024).

**CONFLICTS OF INTEREST**

The authors have no conflicts of interest to declare.

**ETHICS APPROVAL AND CONSENT TO PARTICIPATE**

This study was approved by the Research Ethics Commission of Tongji Hospital of Huazhong University of Science and Technology (TJ-IRB202000406) in view of the retrospective nature of the study and all the procedures performed were part of the routine care. The trial has been registered in the Chinese Clinical Trial Registry (ChiCTR2000032161). The informed consents were waived by the Ethics Commission of Tongji Hospital of Huazhong University of Science and Technology.

**AUTHOR CONTRIBUTIONS**

QG had full access to all data in the study, took responsibility for the integrity of data, and the accuracy of the data analysis. YG designed the study. LC did the analysis. YC, LC, and HL interpreted the data and wrote the paper. SZ, XF, YW, TJ, YY, JC, XJ, DL, XF, SW, R, YY, SX, XX, PC, QM, XJ, and YW provided patients’ samples and clinical data, entered the data into database, and double-checked...
the data. QG, SL, CL, and DM advised on the conception and design of the study. All authors vouched for the respective data and analysis, approved the final version, and agreed to publish the manuscript.

DATA AVAILABILITY STATEMENT
The data contain information that could compromise research participant privacy, and so are not publicly available. Data supporting the findings of this study are available from the corresponding author upon reasonable request.

Yue Gao\textsuperscript{1,2,3}  
Lingxi Chen\textsuperscript{4}  
Shaoqing Zeng\textsuperscript{1,2,3}  
Xikang Feng\textsuperscript{5}  
JianHua Chi\textsuperscript{1,2,3}  
Ya Wang\textsuperscript{1,2,3}
Huayi Li1,2,3
Tengping Jiang5
Yang Yu1,2,3
XiaoFei Jiao1,2,3
Dan Liu1,2,3
XinXia Feng7
SiYuan Wang1,2,3
RuiDi Yu1,2,3
Yuan Yuan1,2,3
Sen Xu1,2,3
Guangyao Cai1,2,3
Xiaoming Xiong1,2,3
Pingbo Chen1,2,3
Qingqing Mo1,2,3
Xin Jin1,2,3
Yuan Wu1,2,3
Ding Ma1,2,3
Chunrui Li8
Shuai Cheng Li4,9
Qinglei Gao1,2,3

1 National Medical Center for Major Public Health Events, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, People's Republic of China
2 Cancer Biology Research Center (Key Laboratory of Chinese Ministry of Education), Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, People's Republic of China
3 Department of Gynecology and Obstetrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, People's Republic of China
4 Department of Computer Science, City University of Hong Kong, Kowloon Tong, Hong Kong
5 School of Software, Northwestern Polytechnical University, Xi'an, People's Republic of China
6 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China
7 Department of Gastroenterology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, People's Republic of China
8 Department of Hematology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
9 Department of Biomedical Engineering, City University of Hong Kong, Kowloon Tong, Hong Kong

Correspondence
Prof. Qinglei Gao, Cancer Biology Research Center, Key Laboratory of the Ministry of Education, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, 1095 Jiefang Ave, Wuhan 430000, China.
Email: qingleigao@hotmail.com
Or Prof. Shuai Cheng Li, Department of Computer Science & Biomedical Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong 999077, Hong Kong.
Email: shuaicli@cityu.edu.hk

Yue Gao, Lingxi Chen, Shaqing Zeng, Xikang Feng, and JianHua Chi contributed equally.

ORCID
Chunrui Li https://orcid.org/0000-0001-5134-7133
Qinglei Gao https://orcid.org/0000-0002-9448-3423

REFERENCES
1. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol. 2020;20(6):355-362.
2. Chen R, Sang L, Jiang M, et al. Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. J Allergy Clin Immunol. 2020;146(1):89-100.
3. Stekhoven DJ, Bühlmann P. MissForest–non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28(1):112-118.
4. Fu H, Zhu Y, Wang Y, et al. Identification and validation of stromal immunotype predict survival and benefit from adjuvant chemotherapy in patients with muscle-invasive bladder cancer. Clin Cancer Res. 2018;24(13):3069-3078.
5. Tian J, Yuan X, Xiao J, et al. Clinical characteristics and risk factors associated with COVID-19 disease severity in patients with cancer in Wuhan, China: a multicentre, retrospective, cohort study. Lancet Oncol. 2020;21(7):903-903.
6. Bai X, Fang C, Zhou Y, et al. Predicting COVID-19 malignant progression with AI techniques. 2020. https://doi.org/10.1101/2020.03.20.20037325
7. Gong J, Ou J, Qiu X, et al. A tool for early prediction of severe coronavirus disease 2019 (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. Clin Infect Dis. 2020;71(15):333-840.
8. Yue H, Yu Q, Liu C, et al. Machine learning-based CT radiomics method for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. Ann Transl Med. 2020;8(14):859.
9. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care. 2020;24(1):108.
10. Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. PLoS One. 2020;15(3):e0230548.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article.