



# Severity Detection for the Coronavirus Disease 2019 (COVID-19) Patients Using a Machine Learning Model Based on the Blood and Urine Tests

Haochen Yao<sup>1†</sup>, Nan Zhang<sup>2†</sup>, Ruochi Zhang<sup>3†</sup>, Meiyu Duan<sup>3</sup>, Tianqi Xie<sup>4</sup>, Jiahui Pan<sup>1</sup>, Ejun Peng<sup>5</sup>, Juanjuan Huang<sup>1</sup>, Yingli Zhang<sup>2</sup>, Xiaoming Xu<sup>2</sup>, Hong Xu<sup>2\*</sup>, Fengfeng Zhou<sup>3\*</sup> and Guoqing Wang<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Liang Cheng,  
Harbin Medical University, China

### Reviewed by:

Kun Xiong,  
Independent Researcher, Changsha,  
China  
Zhaocai Zhang,  
Zhejiang University, China

### \*Correspondence:

Hong Xu  
chxuhong@163.com  
Fengfeng Zhou  
ffzhou@jlu.edu.cn  
Guoqing Wang  
qing@jlu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Molecular Medicine,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

Received: 11 June 2020

Accepted: 06 July 2020

Published: 31 July 2020

### Citation:

Yao H, Zhang N, Zhang R,  
Duan M, Xie T, Pan J, Peng E,  
Huang J, Zhang Y, Xu X, Xu H, Zhou F  
and Wang G (2020) Severity  
Detection for the Coronavirus Disease  
2019 (COVID-19) Patients Using  
a Machine Learning Model Based on  
the Blood and Urine Tests.  
*Front. Cell Dev. Biol.* 8:683.  
doi: 10.3389/fcell.2020.00683

<sup>1</sup> Department of Pathogenobiology, The Key Laboratory of Zoonosis, Chinese Ministry of Education, College of Basic Medical Science, Jilin University, Changchun, China, <sup>2</sup> The First Hospital of Jilin University, Jilin University, Changchun, China, <sup>3</sup> BioKnow Health Informatics Lab, College of Software, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, China, <sup>4</sup> School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, United States, <sup>5</sup> Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

The recent outbreak of the coronavirus disease-2019 (COVID-19) caused serious challenges to the human society in China and across the world. COVID-19 induced pneumonia in human hosts and carried a highly inter-person contagiousness. The COVID-19 patients may carry severe symptoms, and some of them may even die of major organ failures. This study utilized the machine learning algorithms to build the COVID-19 severeness detection model. Support vector machine (SVM) demonstrated a promising detection accuracy after 32 features were detected to be significantly associated with the COVID-19 severeness. These 32 features were further screened for inter-feature redundancies. The final SVM model was trained using 28 features and achieved the overall accuracy 0.8148. This work may facilitate the risk estimation of whether the COVID-19 patients would develop the severe symptoms. The 28 COVID-19 severeness associated biomarkers may also be investigated for their underlining mechanisms how they were involved in the COVID-19 infections.

**Keywords:** severity detection, COVID-19, model, blood and urine tests, biomarkers

## INTRODUCTION

Multiple cases of pneumonia patients were linked to the coronavirus disease-2019 (COVID-19) occurred in December 2019 (Zhu et al., 2020). The virus 2019-nCoV demonstrated a substantial capability of inter-human transmissions (Chan et al., 2020) and has rapidly spread around the world, in particular South Korea and Japan (Li Q. et al., 2020). Patients infected with COVID-19 had significantly varied symptoms and their outcomes ranged from mild to death, and the mortality rate was approximately 4.3% (Wang et al., 2020). It is necessary to mention that 61.5% of the COVID-19 pneumonia patients with critical symptoms died within 28 days after admission (Yang X. et al., 2020). The discrimination of severely ill patients with COVID-19 from those with mild symptoms may help understand the individualized variations of the COVID-19 prognosis. The knowledge may also facilitate the establishing of early diagnosis of the COVID-19 severeness.

The diagnosis of COVID-19 heavily relies on the epidemiological features, clinical characteristics, imaging findings, and nucleic acid screening (Shi et al., 2020), etc. The delivery of the diagnosis result by these technologies was time consuming and error prone (Xie et al., 2020). Multiple types of clinical data were collected for a patient with COVID-19 infection and they were manually integrated by the clinicians to make the diagnosis decisions. The stochastic transmission model was also used to investigate how the COVID-19 transmitted locally and globally (Kucharski et al., 2020). Machine learning algorithms were widely used to integrate the heterogeneous biomedical data sources for the diagnosis decision (Thompson et al., 2018; Hu et al., 2019). So they may also be utilized to produce more delicate prediction models for the severeness diagnosis of the COVID-19 patients. The biomarkers used for an accurate diagnosis model of patients with COVID-19 may serve as the drug targets for this global infectious disease.

This study investigated the detection of severely ill patients with COVID-19 from those with mild symptoms using the clinical information and the blood/urine test data. The clinical information consisted of age, sex, body temperature, heart rate, respiratory rate, and blood pressure. The blood/urine tests may be carried out using the technically easy and cost-efficient procedures. An accurate severeness detection model of the patients with COVID-19 based on those features above may improve the prognosis of this disease in large scale clinical practices. The following sections will firstly describe the data collection and modeling methods, and then utilized the popular machine learning algorithms to build the best severeness detection model.

## MATERIALS AND METHODS

### Data Collection

This study recruited 137 clinically confirmed cases of COVID-19, which were collected from the Tongji Hospital Affiliated to Huazhong University of Science and Technology. Patients were hospitalized from January 18, 2020, to February 13, 2020. The cohort consisted of 17 mild cases, 45 moderate ones and 75 severely ill patients. 21 of the severe cases eventually died. This study investigated the binary classification problem between 75 severe/deceased cases and 62 mild/moderate ones. Each participant was regarded as a sample in this study. This study was approved by the Ethics Commission of the First Hospital of Jilin University (2020-236). With informed consent was waived for this emerging infectious disease.

Patient information including age, sex, body temperature, heart rate, respiratory rate, blood pressure and the blood/urine tests data. Each clinically obtained value was regarded as a feature in this study. In summary, each sample has 100 features, consisting of 8 clinical, 76 blood test, and 16 urine test values.

### Data Pre-processing

The missing entries were filled in the following procedure. We assumed a missing entry to be within the normal range

and filled this entry with the median of that normal range. If there is no normal range for a missing entry, we filled it with zero (0). The samples were randomly split into 80% as training and 20% as test datasets in a stratified fashion. Features in continuous values were normalized by the values in the training dataset. The categorical features were encoded by the one-hot strategy.

### Feature Selection

The principle of Occam's razor suggested that a model using fewer features was preferred over a complicated model with a similar prediction performance (Koller and Sahami, 1996). Feature selection algorithms may be utilized to remove those unrelated features (Ebrahimpour et al., 2017) and may usually increase the model prediction performances (Zhang et al., 2019; Yang S. et al., 2020).

The student *t*-test (abbreviated as *T*-test) is a filter algorithm and it evaluates the statistical association of each feature with the disease severeness of a sample. The features with the *T*-test calculated *P*-values below 0.05 were usually considered to be statistically significantly associated with the disease severeness (Govindan et al., 2019; Peng et al., 2019).

### Prediction Algorithms

This study evaluated several classification algorithms to build the prediction models of the severely ill patients with COVID-19. The predictive logistic regression (LR) model is a regression analysis for the dataset with the binary dependent variable, i.e., the class label (Kleinbaum et al., 2002). LR has been widely used to build clinical decision models (Luo et al., 2019; Heijnen et al., 2020). A probability is calculated by LR to describe whether the sample belongs to a class and a threshold for the probability is usually utilized to make the predictive decision. LR firstly calculates the  $\log - \text{odds} = \log_b \left[ \frac{p}{1-p} \right] = \beta_0 + \beta_1 x + \dots + \beta_n x$ , and the probability  $p = 1/[1 + \beta^{-(\beta_0 + \beta_1 x + \dots + \beta_n x)}]$ , where  $\beta_i$  is the model parameter.

Support vector machine (SVM) is a supervised machine learning algorithm that may accomplish both classification and regression tasks (Suykens and Vandewalle, 1999). SVM tries to find a hyperplane to separate data by the highest margin. The learning strategy of SVM is spacing maximization, which can be formalized as a problem of solving convex quadratic programming (Shilton et al., 2005). This algorithm has been widely used to build the prediction models using the data of blood test (Li et al., 2014, 2018) and urine test (Osredkar et al., 2019; Zhou et al., 2019).

Random forest (RF) is an ensemble algorithm that summarizes the prediction results of multiple tree-based classifiers (Pal, 2005). RF may improve the model performances and avoid over-fitting by averaging the results of models trained over various sub-samples of the dataset. Its model complexity renders itself computation-intensive and RF runs slower than many prediction algorithms. RF is another popular algorithm for

building the prediction models using the clinical data (Zhang et al., 2018; Wu et al., 2019).

K nearest neighbor (KNN) is an instance-based learning algorithm and summarizes the prediction based on the class labels of the query sample's  $k$  nearest neighbors (Dudani, 1976). KNN simply assigns the query sample with the class label of its majority nearest neighbors. And its prediction performance heavily relies on the definition of the inter-sample distances. Nicholas Schaub et al. (2009) demonstrated that selecting the best biomarkers may be essential to improve the KNN models.

The boosting-based algorithm AdaBoost iteratively trained weak learners and summarized these weak learners' results into a weighted sum (Ratsch et al., 2001). Multiple variants of Adaboost were proposed for recognizing human actions (Lv and Nevatia, 2006), diagnosing the dog hypoadrenocorticism (Reagan et al., 2019), and predicting protein binding sites (Qiao and Xie, 2019), etc.

The above algorithms are implemented using Python programming language (version 3.6) and Scikit-learn package (version 0.22).

## Prediction Performance Evaluation Metrics

The binary classification model was evaluated using four classification performance metrics, as defined in the followings. The severely ill patients were regarded as positive samples and the other patients constituted the negative dataset. The number of correctly predicted positive samples was defined as true positive (TP), and the number of the other positive samples was false negative (FN). The true negative (TN) and the false positive (FP) were defined as the numbers of correctly and incorrectly predicted negative samples, respectively. So the overall accuracy Acc was defined as  $Acc = (TP + TN)/(TP + FN + TN + FP)$ . The model's sensitivity ( $S_n$ ) and specificity ( $S_p$ ) were defined as  $S_n = TP/(TP + FN)$  and  $S_p = TN/(TN + FP)$ . The three metrics Acc,  $S_n$  and  $S_p$  measured the percentages of correctly predicted all, positive and negative samples, respectively. The Matthew's Correlation Coefficient (MCC) described the overall correlation of the predicted and the real class labels, and MCC was defined as  $MCC = (TP \times TN - TP \times FN)/\sqrt{[(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)]}$ , where  $\sqrt{\phantom{x}}$  was the square root function (Khurana et al., 2018; Cogan et al., 2019).

Each model was randomly trained for twenty runs with different random seeds and the metric averaged accuracy  $aAcc = [Acc(1) + Acc(2) + \dots + Acc(20)]/20$ , where  $Acc(i)$  was the accuracy of the  $i$ th model. The metric  $aAcc$  was used to find the best prediction model. The metrics  $aS_n$ ,  $aS_p$  and  $aMCC$  were the averaged  $S_n$ , averaged  $S_p$ , and averaged MCC over the twenty random runs.

## Ethics Statement

This study was approved by the Ethics Commission of the First Hospital of Jilin University (2020-236). With informed consent was waived for this emerging infectious disease.

## RESULTS

### Baseline Characteristics of the 2019-nCoV Pneumonia Participants

This study recruited 137 COVID-19 patients to build the detection model of severely ill (positive) samples against the patients with mild symptoms. All the 100 features were screened for their association with the class label, i.e., Positive or Negative. There were 8 clinical values, 76 blood test values and 16 urine test values, respectively. Thirty-two features achieved the  $T$ -test  $P$ -value  $< 0.05$ , and were kept for further analysis in the following sections, as summarized in the **Supplementary Table S1**.

The feature of the patient's age at diagnosis (Age) demonstrated a significant difference ( $P$ -value =  $1.75e-6$ ) between the two groups of samples, and the severely ill patients were on average 13.5695 years older than the patients with mild symptoms. This supported the observation that patients aged around 65 years old tended to have more severe symptoms than those aged around 51 years old (Chen et al., 2020). The sex also demonstrated severe-specific  $P$ -value =  $7.71e-5$ , suggesting that male patients were at higher risks of developing severe symptoms (Huang et al., 2020), as shown in **Figure 1A**.

We also summarized three blood test values and three urine test values with the most significant differences between the two groups of samples, as shown in **Figure 1A**. Overall, the blood test values demonstrated much more significant inter-group differences than the urine test values. The summary data suggested that the percentage of neutrophil cells was significantly enriched in the blood of the severely ill patients, with  $P$ -values  $4.14e-11$ . In addition, the serum calcium level and the monocyte percentage were also significantly lower in the severely ill patients than those mild ones.

Three urine test values demonstrated weak inter-group differential significances. The two values "Urine | Urine protein" and "Urine | Red blood cell (occult)" demonstrated the elevated levels in the severely ill patients with  $P$ -values  $1.44e-2$  and  $2.83e-2$ , respectively. But their variations were very larger, which rendered neither of them as good disease severeness biomarkers. A minor decrease ( $0.1028$ ) in the urine pH value [feature "Urine | PH(Urine)"] in the severely ill patients achieved the inter-group differential significance  $P$ -value  $4.25e-2$ .

In the following sections. The detailed summary may be found in the **Supplementary Table S1**.

### Evaluation of Feature Correlations With the Group Labels

We firstly evaluated the correlation between the 32 features and the class label using Pearson Correlation Coefficient (PCC), as shown in **Figure 1B**. The PCC value ranges between  $-1$  and  $1$ . This study focused on the whether a feature was correlated with the class label. So the absolute value of PCC was calculated in **Figure 1B**.

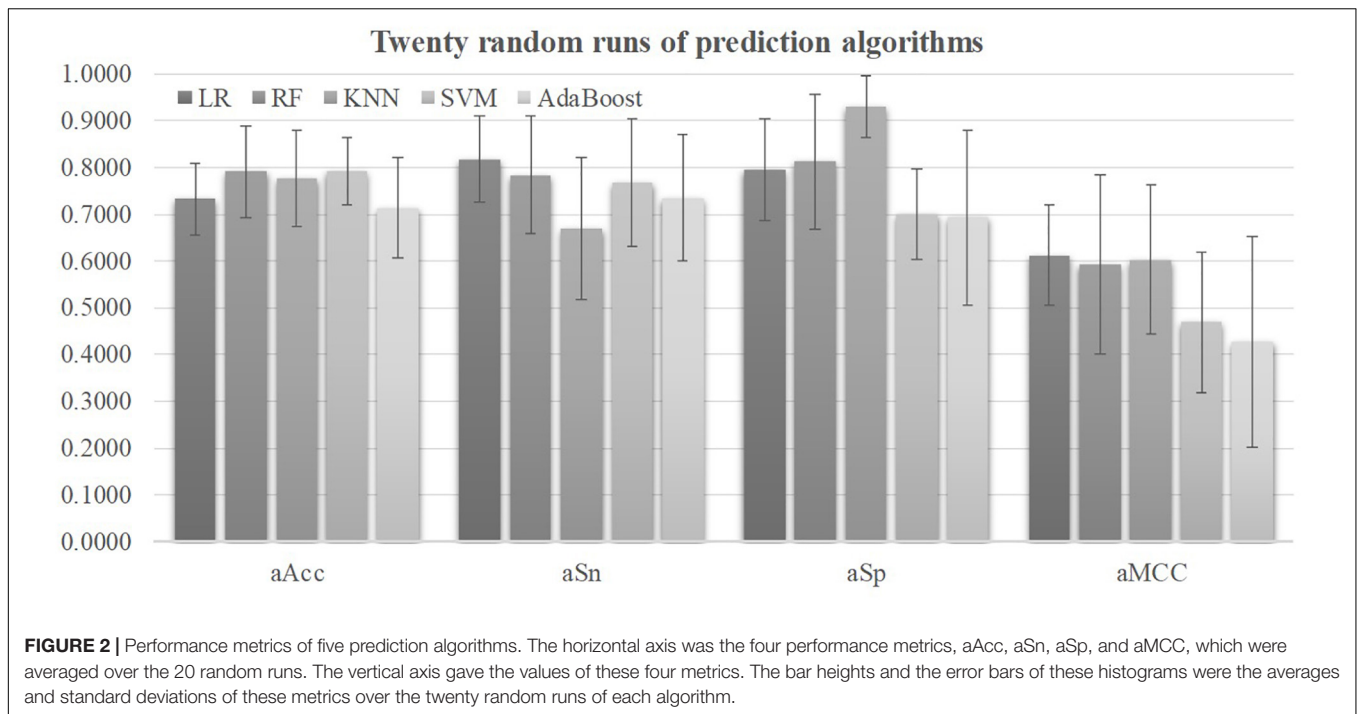
Some features showed strong correlations with the 2019-nCoV pneumonia severeness, which was the class label. The feature "Blood | Neutrophil percentage" demonstrated the largest PCC =  $0.53$  with the disease severeness (class label).

**A**

Features	Positive samples (75)	Positive Std	Negative samples (62)	Negative Std	Pvalue
Age	65.0533	14.5462	51.4839	17.2196	1.75E-06
Sex (Male/Female)	59/16	0.4124	29/33	0.5030	7.71E-05
Blood   Neutrophil percentage	82.5440	14.5598	58.7823	23.7638	4.14E-11
Blood   Calcium	1.0819	0.1884	2.2119	0.1149	9.55E-10
Blood   Monocyte percentage	8.6567	9.7006	17.3052	12.9063	1.62E-05
Urine   Urine protein	0.4467	0.7692	0.1613	0.3708	8.41E-03
Urine   Red blood cell(occult)	0.4400	0.8052	0.1774	0.5364	2.98E-02
Urine   PH(Urine)	6.1633	0.3275	6.2661	0.2432	4.25E-02

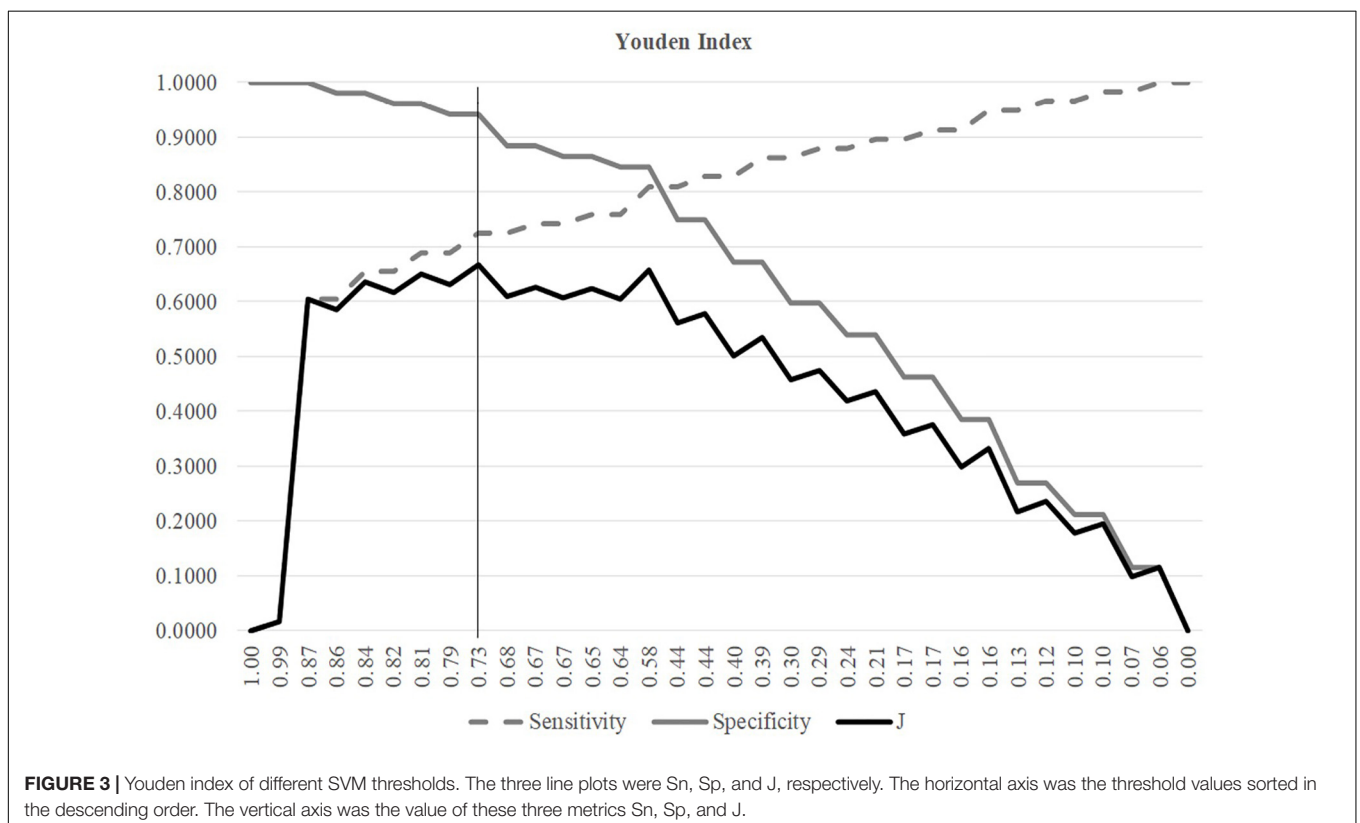
**B**

Age	1.00	0.14	0.36	0.25	0.29	0.32	0.14	0.32	0.19	0.14	0.22	0.21	0.23	0.06	0.30	0.12	0.27	0.16	0.18	0.08	0.19	0.18	0.07	0.14	0.03	0.12	0.02	0.17	0.17	0.14	0.08	0.28	0.40
Sex	0.14	1.00	0.17	0.22	0.21	0.16	0.21	0.02	0.11	0.16	0.08	0.20	0.21	0.09	0.20	0.15	0.03	0.17	0.08	0.08	0.16	0.07	0.07	0.18	0.02	0.21	0.04	0.12	0.12	0.09	0.09	0.17	0.33
Blood   Neutrophil percentage	0.36	0.17	1.00	0.48	0.43	0.41	0.33	0.34	0.37	0.06	0.15	0.28	0.22	0.58	0.21	0.19	0.30	0.25	0.23	0.11	0.14	0.20	0.14	0.14	0.08	0.15	0.12	0.14	0.04	0.22	0.25	0.11	0.53
Blood   Calcium	0.25	0.22	0.48	1.00	0.28	0.26	0.30	0.16	0.24	0.06	0.10	0.15	0.15	0.35	0.06	0.20	0.18	0.21	0.36	0.16	0.09	0.05	0.02	0.05	0.15	0.17	0.12	0.12	0.03	0.13	0.02	0.05	0.49
Blood   Monocyte percentage	0.29	0.21	0.43	0.28	1.00	0.38	0.21	0.46	0.24	0.09	0.68	0.28	0.09	0.21	0.10	0.17	0.42	0.12	0.64	0.12	0.21	0.29	0.15	0.07	0.09	0.09	0.11	0.13	0.02	0.12	0.15	0.11	0.36
Blood   Thrombin time	0.32	0.16	0.41	0.26	0.38	1.00	0.02	0.56	0.31	0.27	0.45	0.47	0.15	0.26	0.29	0.36	0.65	0.34	0.31	0.17	0.09	0.48	0.12	0.44	0.41	0.15	0.49	0.39	0.07	0.32	0.20	0.35	
Blood   hs-CRP	0.14	0.21	0.33	0.30	0.21	0.02	1.00	0.08	0.04	0.10	0.05	0.17	0.19	0.26	0.05	0.10	0.05	0.27	0.11	0.05	0.15	0.06	0.14	0.01	0.01	0.12	0.02	0.01	0.09	0.11	0.07	0.13	0.33
Blood   Neutrophil Number	0.32	0.02	0.34	0.16	0.46	0.56	0.08	1.00	0.14	0.31	0.55	0.37	0.08	0.18	0.20	0.01	0.73	0.04	0.39	0.17	0.03	0.67	0.14	0.02	0.18	0.13	0.29	0.32	0.14	0.28	0.22	0.08	0.32
Blood   Glucose	0.19	0.11	0.37	0.24	0.24	0.31	0.04	0.14	1.00	0.09	0.07	0.23	0.29	0.20	0.09	0.37	0.06	0.18	0.14	0.00	0.18	0.16	0.12	0.28	0.02	0.17	0.01	0.23	0.03	0.11	0.07	0.02	0.31
Blood   Leucocytes	0.14	0.16	0.06	0.06	0.09	0.27	0.10	0.31	0.09	1.00	0.19	0.14	0.00	0.11	0.38	0.09	0.21	0.06	0.12	0.12	0.12	0.13	0.09	0.03	0.10	0.02	0.15	0.12	0.20	0.12	0.11	0.07	0.31
Blood   Basophil percentage	0.22	0.08	0.15	0.10	0.68	0.45	0.05	0.55	0.07	0.19	1.00	0.28	0.00	0.14	0.14	0.13	0.69	0.06	0.55	0.23	0.11	0.45	0.04	0.04	0.16	0.12	0.20	0.11	0.08	0.05	0.09	0.12	0.29
Blood   Total bilirubin	0.21	0.20	0.28	0.15	0.28	0.47	0.17	0.37	0.23	0.14	0.25	0.90	0.14	0.14	0.20	0.20	0.31	0.51	0.23	0.11	0.04	0.46	0.02	0.54	0.39	0.11	0.42	0.07	0.03	0.17	0.21	0.33	0.28
Blood   Interleukin-2R	0.23	0.21	0.22	0.15	0.09	0.15	0.19	0.08	0.29	0.00	0.00	0.14	1.00	0.12	0.03	0.23	0.00	0.41	0.03	0.12	0.12	0.15	0.13	0.44	0.09	0.42	0.06	0.04	0.22	0.04	0.01	0.12	0.28
Blood   Packed cell volume	0.06	0.09	0.58	0.35	0.21	0.26	0.26	0.18	0.20	0.11	0.14	0.14	0.12	1.00	0.06	0.00	0.26	0.11	0.25	0.41	0.02	0.10	0.03	0.06	0.03	0.04	0.11	0.01	0.30	0.15	0.11	0.01	0.28
Blood   RBC distribution width SD	0.30	0.20	0.21	0.06	0.10	0.29	0.05	0.20	0.09	0.38	0.14	0.20	0.03	0.06	1.00	0.20	0.18	0.07	0.07	0.20	0.04	0.21	0.09	0.22	0.30	0.03	0.24	0.09	0.54	0.13	0.21	0.16	0.27
Blood   Interleukin-10	0.12	0.15	0.19	0.20	0.17	0.36	0.10	0.01	0.37	0.09	0.13	0.20	0.23	0.00	0.20	1.00	0.00	0.39	0.08	0.08	0.09	0.05	0.07	0.54	0.21	0.22	0.20	0.08	0.14	0.19	0.18	0.18	0.26
Blood   Prothrombin time	0.27	0.03	0.30	0.18	0.42	0.65	0.05	0.73	0.06	0.21	0.69	0.31	0.00	0.26	0.18	0.00	1.00	0.10	0.36	0.17	0.02	0.60	0.10	0.05	0.13	0.21	0.23	0.27	0.09	0.12	0.09	0.04	0.26
Blood   Serum ferritin	0.16	0.17	0.25	0.21	0.12	0.34	0.27	0.04	0.18	0.06	0.06	0.51	0.41	0.11	0.07	0.39	0.10	1.00	0.02	0.17	0.04	0.02	0.01	0.51	0.26	0.29	0.30	0.08	0.12	0.09	0.13	0.31	0.25
Blood   Eosinophil percentage	0.18	0.08	0.23	0.36	0.64	0.31	0.11	0.39	0.14	0.12	0.55	0.23	0.03	0.25	0.07	0.08	0.36	0.02	1.00	0.11	0.03	0.25	0.03	0.04	0.08	0.05	0.12	0.13	0.00	0.01	0.11	0.05	0.24
Blood   Platelet distributon width	0.08	0.08	0.11	0.16	0.12	0.17	0.05	0.17	0.00	0.12	0.23	0.11	0.12	0.41	0.20	0.08	0.17	0.17	0.11	1.00	0.08	0.08	0.09	0.15	0.17	0.01	0.20	0.07	0.01	0.17	0.07	0.01	0.24
Blood   Oxygen partial pressure	0.19	0.16	0.14	0.09	0.21	0.09	0.15	0.03	0.18	0.12	0.11	0.04	0.12	0.02	0.04	0.09	0.02	0.04	0.03	0.08	1.00	0.06	0.47	0.06	0.09	0.13	0.07	0.10	0.03	0.07	0.03	0.04	0.23
Blood   Alkaline phosphatase	0.18	0.07	0.20	0.05	0.29	0.48	0.06	0.67	0.16	0.13	0.45	0.46	0.15	0.10	0.21	0.05	0.60	0.02	0.25	0.08	0.06	1.00	0.08	0.02	0.08	0.11	0.14	0.38	0.19	0.14	0.12	0.05	0.23
Blood   Total carbon dioxide	0.07	0.07	0.14	0.02	0.15	0.12	0.14	0.14	0.12	0.09	0.04	0.02	0.13	0.03	0.09	0.07	0.10	0.01	0.03	0.09	0.47	0.08	0.98	0.09	0.08	0.02	0.03	0.23	0.08	0.14	0.04	0.08	0.22
Blood   Tumor necrosis factor-α	0.14	0.18	0.14	0.05	0.07	0.44	0.01	0.02	0.38	0.03	0.04	0.54	0.44	0.06	0.22	0.54	0.05	0.51	0.04	0.15	0.06	0.02	0.09	0.08	0.41	0.15	0.39	0.02	0.09	0.19	0.17	0.36	0.20
Blood   Chlorine	0.03	0.02	0.08	0.15	0.09	0.41	0.01	0.18	0.02	0.10	0.16	0.39	0.09	0.03	0.30	0.21	0.13	0.26	0.08	0.17	0.09	0.08	0.08	0.41	1.00	0.09	0.93	0.05	0.18	0.23	0.30	0.20	0.19
Blood   Alanine aminotransferase	0.12	0.21	0.15	0.17	0.09	0.15	0.12	0.13	0.17	0.02	0.12	0.11	0.42	0.04	0.03	0.22	0.21	0.29	0.05	0.01	0.13	0.11	0.02	0.15	0.09	1.00	0.04	0.31	0.24	0.06	0.02	0.05	0.19
Blood   Sodium	0.02	0.04	0.12	0.12	0.11	0.49	0.02	0.29	0.01	0.15	0.20	0.42	0.06	0.11	0.24	0.20	0.23	0.30	0.12	0.20	0.07	0.14	0.03	0.39	0.93	0.04	1.00	0.11	0.10	0.23	0.29	0.19	0.19
Blood   Aspartate aminotransferase	0.17	0.12	0.14	0.12	0.13	0.39	0.01	0.32	0.23	0.12	0.11	0.07	0.04	0.01	0.09	0.08	0.27	0.08	0.13	0.07	0.10	0.38	0.23	0.02	0.05	0.31	0.11	1.00	0.02	0.04	0.02	0.03	0.18
Blood   RBC distributon width CV	0.17	0.12	0.04	0.03	0.02	0.07	0.09	0.14	0.03	0.20	0.08	0.03	0.22	0.30	0.54	0.14	0.09	0.12	0.00	0.01	0.03	0.19	0.08	0.09	0.18	0.24	0.10	0.02	1.00	0.03	0.06	0.08	0.17
Urine   Urine protein	0.14	0.09	0.22	0.13	0.12	0.32	0.11	0.28	0.11	0.12	0.05	0.17	0.04	0.15	0.13	0.19	0.12	0.09	0.12	0.07	0.07	0.14	0.14	0.19	0.23	0.06	0.23	0.04	0.03	1.00	0.62	0.31	0.22
Urine   Red blood cell(occult)	0.08	0.09	0.25	0.02	0.15	0.32	0.07	0.22	0.07	0.11	0.09	0.21	0.01	0.11	0.21	0.18	0.09	0.13	0.11	0.07	0.03	0.12	0.04	0.17	0.30	0.02	0.29	0.02	0.06	0.62	1.00	0.16	0.19
Urine   PH(Urine)	0.28	0.17	0.11	0.05	0.11	0.20	0.13	0.08	0.02	0.07	0.12	0.33	0.12	0.01	0.16	0.18	0.04	0.31	0.05	0.01	0.04	0.05	0.08	0.36	0.20	0.05	0.19	0.03	0.08	0.31	0.16	1.00	0.17
Class	0.40	0.33	0.53	0.49	0.36	0.35	0.33	0.32	0.31	0.31	0.29	0.28	0.28	0.28	0.27	0.26	0.26	0.25	0.24	0.24	0.23	0.23	0.22	0.20	0.19								



model. Youden’s (1950) index assigns equal weights for sensitivity and specificity and tries to maximize the index value  $J = (Sn + Sp - 1)$ . **Figure 3** illustrated the changing curves of Sn and Sp

with different thresholds for the prediction scores of the samples. The maximal value of  $J$  was achieved at the threshold 0.7318, and averaged accuracy of SVM was improved to 0.8148.



The Youden’s index was used to find the best threshold of the SVM models with different parameters and features in the following sections.

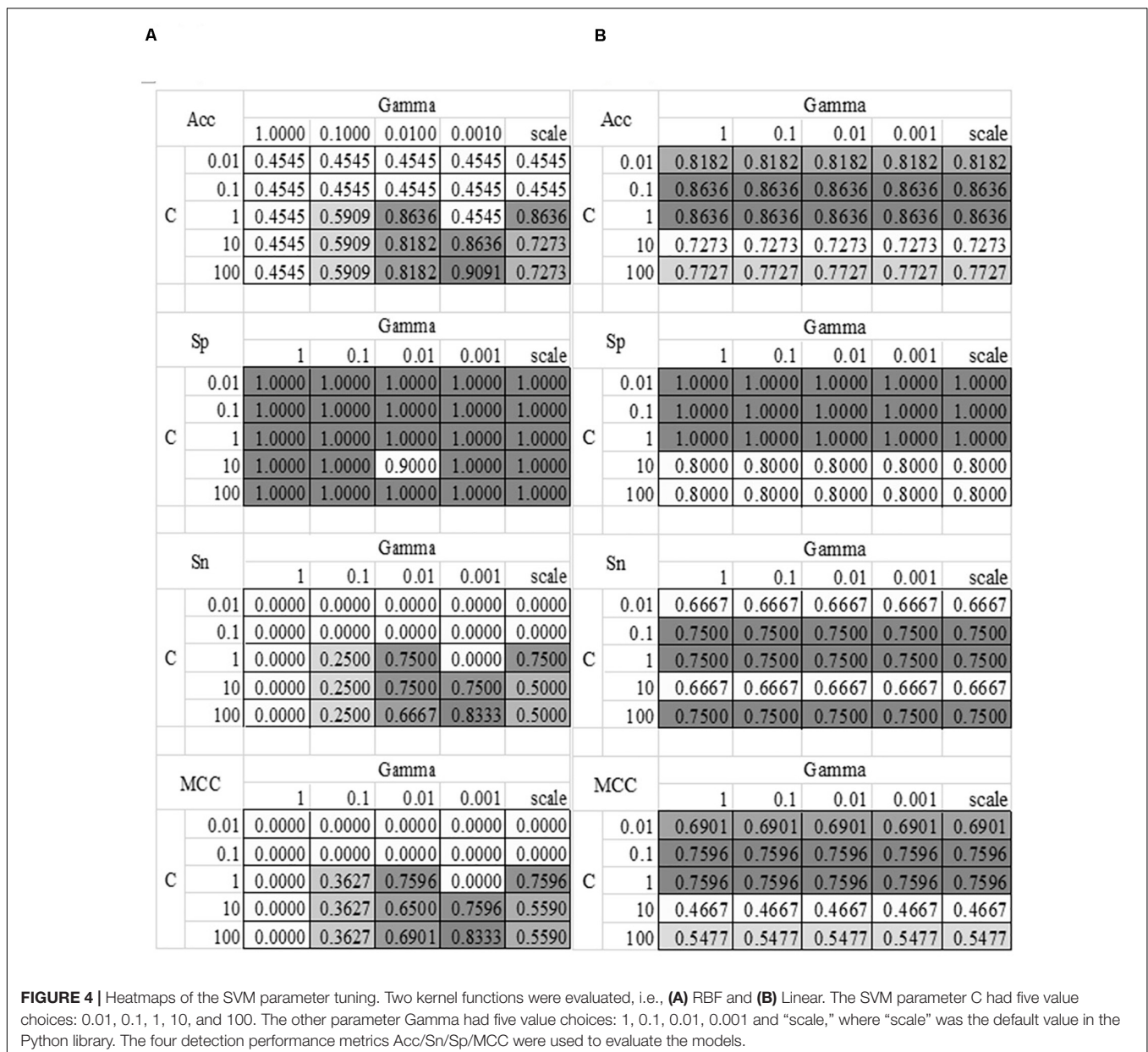
### Tuning the Parameters of the SVM Model

The grid search strategy was carried out to evaluate how different parameter values affected the disease severeness detection model, as shown in **Figure 4**. Parameter tuning was a time-consuming step. So this section randomly split the training dataset into 80% sub-training dataset and 20% validation dataset. Each model was trained using the sub-training dataset and the performance was calculated on the validation dataset. The model detection performance didn’t change with the linear kernel and different choices of the parameter Gamma, as shown in **Figure 4B**.

And the best accuracy = 0.8636 of the linear kernel SVM was achieved when C = 0.1 or 1. The best model with the RBF kernel achieved Acc = 0.9091 for the validation dataset, where C = 100 and Gamma = 0.0010. The other three metrics Sp = 1.0000, Sn = 0.8333, and MCC = 0.8333 were also the best values in **Figure 4**. The SVM model with the above-mentioned parameters achieved Acc = 0.8148 on the independent test dataset. So the following sections used these two choices of the parameters C and Gamma.

### Remove Redundant Features to Improve the Model

The existence of strong inter-feature correlations in **Figure 1B** suggested that some features may be removed to further



improve the model. This section carried out a conservative recursive feature elimination (cRFE) strategy to eliminate the redundant features while ensuring the model performance was not decreased. The model performance was evaluated for its threshold-independent metric AUC value (Deepak and Ameer, 2019; Kourou et al., 2020). Firstly, all the 32 features were ranked by the ascending order of their  $T$ -test  $P$ -values. Then, the detection model was evaluated by eliminating each feature. A feature was eliminated if the model's AUC was improved with its removal. Otherwise that feature was kept. The final feature set was returned after all the features were evaluated.

The feature selection procedure should avoid using the test samples, so this section calculated the performance metrics on the validation dataset using the model trained over the sub-training dataset, as shown in **Figure 5**. The heuristic cRFE strategy ensured by its nature that the model performance would not be decreased, and the rising line segment indicated the removal of the feature on the horizontal axis. Four features were removed, i.e., "Age," "Blood | Interleukin-10," "Blood | Prothrombin time," and "Blood | Oxygen partial pressure." **Figure 1B** illustrated that all these four features were strongly correlated with some other features, with the PCC values at least 0.51.

The remaining 28 features achieved  $\text{Acc} = 0.9917$  on the validation dataset, and  $\text{Acc} = 0.8148$  on the independent test dataset. Although the COVID-19 severeness detection performance was not improved, the model complexity was reduced and the clinical screening cost was reduced with fewer features.

A web site was established to help the clinicians to try this COVID-19 infection severity estimation model, and the users may access: <http://dVirusSeverity.HealthInformaticsLab.org/>.

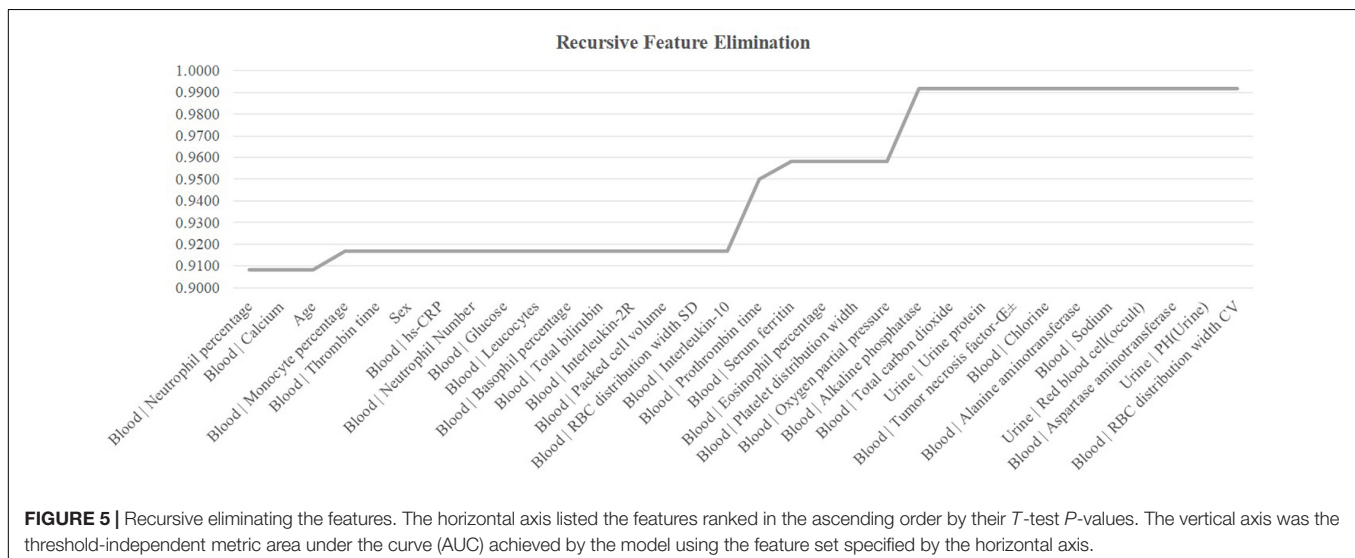
## DISCUSSION

The emergence of SARS-CoV-2 marked the third of highly pathogenic coronavirus in humans in the twenty-first century,

after severe acute respiratory syndrome (SARS) in 2003, and Middle East respiratory syndrome (MERS) in 2012 (Drosten et al., 2003; Zaki et al., 2012). SARS-CoV-2 belongs to the coronavirus family,  $\beta$ -coronavirus genera and belongs to the cluster of betacoronaviruses (Chen Y. et al., 2020). Based on Sequence analysis, the amino acid sequences of SARS-CoV-2 showed 94.4% identity with SARS-CoV (Zhou et al., 2020). It is suggested that SARS-CoV-2 was more closely related to SARS-like bat CoV. In comparison, SARS-CoV-2 was more distant from the MERS-CoV (Lu et al., 2020; Wu et al., 2020). The mortality of critically ill patients with COVID-19 is considerable. The survival time of the dead patients may be within 1–2 weeks after ICU admission (Yang X. et al., 2020).

The present diagnosis of COVID-19 didn't achieve a satisfying accuracy. Both false positives and false negatives need to be decreased (Li D. et al., 2020; Li Z. et al., 2020; Yan et al., 2020). The clinical decisions of COVID-19 infections are usually confirmed by epidemiological features, clinical manifestations, imaging factors, and nucleic acid screenings, etc. Some of the COVID-19 patients may develop severe symptoms and these patients are at a much higher mortality rate than the other patients. This challenge raised the scientific question of finding the COVID-19 severeness specific biomarkers, which may help reduce the overall mortality.

This study investigated the binary classification problem between 75 severely ill COVID-19 infected patients and the other 62 patients with mild symptoms. A comprehensive optimization procedure led to the best SVM-based COVID-19 severeness detection model using only 28 features. The experimental data suggested that the severely ill patients had a higher serum level of neutrophil percentage and lower serum levels of monocyte percentage and calcium compared with those mild ones. Urine test contributed three weak group-specific biomarkers, i.e., urine pH value, urine protein and urine red blood cell. Compared with the urine pH value, the variations of urine protein and urine red blood cell were very large and these two urine features may not serve well as COVID-19 infection



severeness biomarkers. The blood test features demonstrated much more significant inter-group differences than the urine test features. The summary data suggested these three blood test features as candidate severeness biomarkers, i.e., serum ferritin, hs-CRP, interleukin-2R, and tumor necrosis factor- $\alpha$ .

COVID-19 severeness detection model achieved the overall accuracy 0.8148 on the independent test dataset with only 28 clinical biomarkers. Twenty-one out of these 28 biomarkers were investigated in the coronavirus. Two serum values “Blood | Tumor necrosis factor- $\alpha$ ” (56 papers) and “Blood | Sodium” (57 papers) were known to be associated with the coronavirus infections. The tumor necrosis factor-alpha (TNF-alpha) was observed to have elevated expression levels in the serum of the coronavirus-infected mice (Zalinger et al., 2015). The serum sodium level was slightly increased by 2.01% in the severely ill patients in the cohort used in this study. Hoffman et al. (2018) proposed that the pulmonary complication were more frequently observed in the hypernatremia patients. So it would be interesting to investigate the underlining mechanism of how the serum sodium may induce the COVID-19 severeness. The feature “Urine | PH(Urine)” is the pH level in the urine, and quite a few investigations observed the aberrant pH levels in the body fluid or fecal matter of the coronavirus-infected animals (Raabis et al., 2018; Yuan et al., 2018). Although the urine pH level was not investigated in the coronavirus-infected animals, this may be worth of an investigation. The sex bias was also observed that coronavirus tended to infect males (Habib et al., 2019; Petrarca et al., 2019). Our data suggested that males were at a higher risk to be infected by COVID-19 and to develop more severe symptoms.

An accurate severeness detection model of the patients with COVID-19 based on those features may improve the prognosis of this disease in large scale clinical practices, and reduce the incidence of COVID-19 severeness and mortality. The biomarkers used for an accurate diagnosis model of patients with COVID-19 may serve as the drug targets for this global infectious disease.

There are some limitations that should be noted. First, the number of patients with COVID-19 is relatively small, which may limit the accuracy of severeness detection model. Second, since all subjects in our study were Chinese patients with COVID-19, the results may not be applied to other ethnicities. Third, the data of this study is only the preliminary establishment of COVID-19 severeness detection model. Further studies are still needed.

This study utilized the machine learning algorithms to detect the COVID-19 severely ill patients from those with only mild symptoms. Our experimental data demonstrated strong correlations with the COVID-19 severeness. And the final COVID-19 severeness detection model achieved the accuracy 0.8148 on the independent test dataset using only 28 clinical biomarkers. The detection model itself is in urgent need for the current epidemic situation that the severely ill patients are at a very high mortality rate. The 28 biomarkers may also be investigated for their

underlining mechanisms of their roles in the COVID-19 severely ill patients.

## DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Commission of the First Hospital of Jilin University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

NZ collected the data. HY and RZ analyzed the data and wrote the manuscript. MD, TX, JP, EP, JH, YZ, and XX did literature search. HX, FZ, and GW conceived and designed this study. The corresponding author had full access to all the study data and had final responsibility for the decision to submit for publication. All authors read and approved the final manuscript.

## FUNDING

This work was supported by grants from the epidemiology, early warning and response techniques of major infectious diseases in the Belt and Road Initiative (#2018ZX10101002), National Natural Science Foundation of China (#81871699), Jilin Provincial Key Laboratory of Big Data Intelligent Computing (20180622002JC), the Education Department of Jilin Province (JJKH20180145KJ), Foundation of Jilin Province Science and Technology Department (#172408GH010234983), and the startup grant of the Jilin University. This work was also partially supported by the Bioknow MedAI Institute (BM CPP-2018-001), the High Performance Computing Center of Jilin University, and the Fundamental Research Funds for the Central Universities, JLU.

## ACKNOWLEDGMENTS

We thank all patients.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2020.00683/full#supplementary-material>



## REFERENCES

- Chan, J. F., Yuan, S., Kok, K. H., To, K. K., Chu, H., Yang, J., et al. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 395, 514–523. doi: 10.1016/S0140-6736(20)30154-9
- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 395, 507–513. doi: 10.1016/S0140-6736(20)30211-7
- Chen, Y., Liu, Q., and Guo, D. (2020). Emerging coronaviruses: genome structure, replication, and pathogenesis. *J. Med. Virol.* 92, 418–423. doi: 10.1002/jmv.25681
- Cogan, T., Cogan, M., and Tamil, L. (2019). MAPGI: accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning. *Comput. Biol. Med.* 111:103351. doi: 10.1016/j.compbimed.2019.103351
- Deepak, S., and Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Comput. Biol. Med.* 111:103345. doi: 10.1016/j.compbimed.2019.103345
- Drosten, C., Günther, S., Preiser, W., Werf, S., Brodt, H., Becker, H., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1967–1976. doi: 10.1056/NEJMoa030747
- Dudani, S. A. (1976). “The distance-weighted k-nearest-neighbor rule,” in *Proceedings of the IEEE Transactions on Systems, Man, and Cybernetics*, Piscataway, NJ.
- Ebrahimpour, M. K., Zare, M., Eftekhari, M., and Aghamolaei, G. (2017). Occam’s razor in dimension reduction: using reduced row Echelon form for finding linear independent features in high dimensional microarray datasets. *Eng. Appl. Artif. Intellig.* 62, 214–221. doi: 10.1016/j.engappai.2017.04.006
- Govindan, R. B., Massaro, A., Vezina, G., Chang, T., and du Plessis, A. (2019). Identifying an optimal epoch length for spectral analysis of heart rate of critically-ill infants. *Comput. Biol. Med.* 113:103391. doi: 10.1016/j.compbimed.2019.103391
- Habib, A. M. G., Ali, M. A. E., Zouaoui, B. R., Taha, M. A. H., Mohammed, B. S., and Saquib, N. (2019). Clinical outcomes among hospital patients with Middle East respiratory syndrome coronavirus (MERS-CoV) infection. *BMC Infect. Dis.* 19:870. doi: 10.1186/s12879-019-4555-5
- Heijnen, B. J., Bohringer, S., and Speyer, R. (2020). Prediction of aspiration in dysphagia using logistic regression: oral intake and self-evaluation. *Eur. Arch. Otorhinolaryngol.* 277, 197–205. doi: 10.1007/s00405-019-05687-z
- Hoffman, H., Verhave, B., and Chin, L. S. (2018). Hypernatremia is associated with poorer outcomes following aneurysmal subarachnoid hemorrhage: a nationwide inpatient sample analysis. *J. Neurosurg. Sci.* doi: 10.23736/S0390-5616.18.04611-8
- Hu, F., Zeng, W., and Liu, X. A. (2019). Gene signature of survival prediction for kidney renal cell carcinoma by multi-omic data analysis. *Int. J. Mol. Sci.* 20:720. doi: 10.3390/ijms20225720
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi: 10.1016/0010-7824(81)90078-0
- Khurana, S., Rawi, R., Kunji, K., Chuang, G. Y., Bensmail, H., and Mall, R. (2018). DeepSol: a deep learning framework for sequence-based protein solubility prediction. *Bioinformatics* 34, 2605–2613. doi: 10.1093/bioinformatics/bty166
- Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., and Klein, M. (2002). *Logistic Regression*. Berlin: Springer.
- Koller, D., and Sahami, M. (1996). *Toward Optimal Feature Selection*. Stanford, CA: Stanford InfoLab.
- Kourou, K., Rigas, G., Papaloukas, C., Mitsis, M., and Fotiadis, D. I. (2020). Cancer classification from time series microarray data through regulatory dynamic bayesian networks. *Comput. Biol. Med.* 116, 103577. doi: 10.1016/j.compbimed.2019.103577
- Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., et al. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* 20, 553–558. doi: 10.1016/S1473-3099(20)30144-4
- Li, C., Hou, L., Sharma, B. Y., Li, H., Chen, C., Li, Y., et al. (2018). Developing a new intelligent system for the diagnosis of tuberculous pleural effusion. *Comput. Methods Programs Biomed.* 153, 211–225. doi: 10.1016/j.cmpb.2017.10.022
- Li, D., Feng, S., Huang, H., Chen, W., Shi, H., Liu, N., et al. (2014). Label-free detection of blood plasma using silver nanoparticle based surface-enhanced Raman spectroscopy for esophageal cancer screening. *J. Biomed. Nanotechnol.* 10, 478–484. doi: 10.1166/jbn.2014.1750
- Li, D., Wang, D., Dong, J., Wang, N., Huang, H., Xu, H., et al. (2020). False-negative results of real-time reverse-transcriptase polymerase chain reaction for severe acute respiratory syndrome coronavirus 2: role of deep-learning-based ct diagnosis and insights from two cases. *Korea. J. Radiol.* 21, 505–508. doi: 10.3348/kjr.2020.0146
- Li, Q., Guan, X., Wu, P., Li, Q., Guan, X., Wu, P., et al. (2020). Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* 382, 1199–1207. doi: 10.1056/NEJMoa2001316
- Li, Z., Yi, Y., Luo, X., Xiong, N., Liu, Y., Li, S., et al. (2020). Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J. Med. Virol.* 10.1002/jmv.25727. doi: 10.1002/jmv.25727
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8
- Luo, C. L., Rong, Y., Chen, H., Zhang, W., Wu, L., Wei, D., et al. (2019). A logistic regression model for noninvasive prediction of AFP-negative hepatocellular carcinoma. *Technol. Cancer Res. Treat.* 18:1533033819846632. doi: 10.1177/1533033819846632
- Lv, F., and Nevatia, R. (2006). “Recognition and segmentation of 3-d human action using hmm and multi-class adaboost,” in *Computer Vision – ECCV 2006. ECCV 2006. Lecture Notes in Computer Science*, Vol. 3954, eds A. Leonardis, H. Bischof, and A. Pinz (Berlin: Springer).
- Osredkar, J., Gosar, D., Macek, J., Kumer, K., Fabjan, T., Finderle, P., et al. (2019). Urinary markers of oxidative stress in children with autism spectrum disorder (ASD). *Antioxidants* 8:187. doi: 10.3390/antiox8060187
- Pal, M. (2005). Random forest classifier for remote sensing classification. *Intern. J. Remote Sens.* 26, 217–222. doi: 10.1080/01431160412331269698
- Peng, T., Trewh, M. L., and Malik, A. (2019). Predictive modeling of drug effects on electrocardiograms. *Comput. Biol. Med.* 108, 332–344. doi: 10.1016/j.compbimed.2019.03.027
- Petrarca, L., Nenna, R., Frassanito, A., Pierangeli, A., Di Mattia, G., Scagnolari, C., et al. (2019). Human bocavirus in children hospitalized for acute respiratory tract infection in Rome. *World J. Pediatr.* 16, 293–298. doi: 10.1007/s12519-019-00324-5
- Qiao, L., and Xie, D. (2019). MlonSite: ligand-specific prediction of metal ion-binding sites via enhanced AdaBoost algorithm with protein sequence information. *Anal. Biochem.* 566, 75–88. doi: 10.1016/j.ab.2018.11.009
- Raabis, S. M., Ollivett, T. L., Cook, M. E., Sand, J. M., and McGuirk, S. M. (2018). Health benefits of orally administered anti-IL-10 antibody in milk-fed dairy calves. *J. Dairy Sci.* 101, 7375–7382. doi: 10.3168/jds.2017-14270
- Ratsch, G., Onoda, T., and Muller, K.-R. (2001). Soft margins for AdaBoost. *Mach. Learn.* 42, 287–320.
- Reagan, K. L., Reagan, B. A., and Gilor, C. (2019). Machine learning algorithm as a diagnostic tool for hypoadrenocorticism in dogs. *Domest. Anim. Endocrinol.* 72:106396. doi: 10.1016/j.domaniend.2019.106396
- Schaub, N. P., Jones, K. J., Nyalwidhe, J. O., Cazares, L. H., Karbassi, I. D., Semmes, O. J., et al. (2009). Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients. *J. Am. Coll. Surg.* 208, 970–978. doi: 10.1016/j.jamcollsurg.2008.12.024
- Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O., Gu, J., et al. (2020). Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect. Dis.* 20, 425–434. doi: 10.1016/S1473-3099(20)30086-4
- Shilton, A., Palaniswami, M., Ralph, D., and Tsoi, A. C. (2005). Incremental training of support vector machines. *IEEE Trans. Neural Netw.* 16, 114–131. doi: 10.1109/TNN.2004.836201
- Suykens, J. A., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Proc. Lett.* 9, 293–300. doi: 10.1023/A:1018628609742
- Thompson, J. A., Christensen, B. C., and Marsit, C. J. (2018). Methylation-to-expression feature models of breast cancer accurately predict overall survival,

- distant-recurrence free survival, and pathologic complete response in multiple cohorts. *Sci. Rep.* 8:5190. doi: 10.1038/s41598-018-23494-0
- Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., et al. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 323, 1061–1069. doi: 10.1001/jama.2020.1585
- Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., et al. (2020). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microb.* 27, 325–328. doi: 10.1016/j.chom.2020.02.001
- Wu, J., Bai, J., Wang, W., Xi, L., Zhang, P., Lan, J., et al. (2019). ATBdiscrimination: an in silico tool for identification of active tuberculosis disease based on routine blood test and T-SPOT.TB detection results. *J. Chem. Inf. Model.* 59, 4561–4568. doi: 10.1021/acs.jcim.9b00678
- Xie, X., Zhong, Z., Zhao, W., Zheng, C., Wang, F., and Liu, J. (2020). Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. *Radiology* 200343. doi: 10.1148/radiol.2020200343
- Yan, G., Lee, C. K., Lam, L. T. M., Yan, B., Chua, Y. X., Lim, A. Y. N., et al. (2020). Covert COVID-19 and false-positive dengue serology in Singapore. *Lancet Infect. Dis.* 20:536. doi: 10.1016/S1473-3099(20)30158-4
- Yang, S., Li, B., Zhang, Y., Duan, M., Liu, S., Zhang, Y., et al. (2020). Selection of features for patient-independent detection of seizure events using scalp EEG signals. *Comput. Biol. Med.* 119:103671. doi: 10.1016/j.compbiomed.2020.103671
- Yang, X., Yu, Y., Xu, J., Shu, H., Xia, J., Liu, H., et al. (2020). Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir. Med.* 8, 475–481. doi: 10.1016/S2213-2600(20)30079-5
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer* 3, 32–35. doi: 10.1002/1097-0142(1950)3:1<32::aid-cnrcr2820030106>3.0.co;2-3
- Yuan, P., Yang, Z., Song, H., Wang, K., Yang, Y., Xie, L., et al. (2018). Three main inducers of alphacoronavirus infection of enterocytes: sialic acid, proteases, and Low pH. *Intervirology* 61, 53–63. doi: 10.1159/000492424
- Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367, 1814–1820. doi: 10.1056/NEJMoa1211721
- Zalinger, Z. B., Elliott, R., Rose, K. M., and Weiss, S. R. (2015). mda5 is critical to host defense during infection with murine coronavirus. *J. Virol.* 89, 12330–12340. doi: 10.1128/JVI.01470-15
- Zhang, C., Leng, W., Sun, C., Lu, T., Chen, Z., Men, X., et al. (2018). Urine proteome profiling predicts lung cancer from control cases and other tumors. *Ebiomedicine* 30, 120–128. doi: 10.1016/j.ebiom.2018.03.009
- Zhang, Y., Chen, C., Duan, M., Liu, S., Huang, L., and Zhou, F. (2019). BioDog, biomarker detection for improving identification power of breast cancer histologic grade in methylomics. *Epigenomics* 11, 1717–1732. doi: 10.2217/epi-2019-0230
- Zhou, H., Li, L., Zhao, H., Wang, Y., Du, J., Zhang, P., et al. (2019). A Large-scale, multi-center urine biomarkers identification of coronary heart disease in TCM syndrome differentiation. *J. Proteome Res.* 18, 1994–2003. doi: 10.1021/acs.jproteome.8b00799
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. doi: 10.1056/NEJMoa2001017

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yao, Zhang, Zhang, Duan, Xie, Pan, Peng, Huang, Zhang, Xu, Zhou and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.