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

Kai Zhou [a,1], Yaoting Sun [b,c,d,1], Lu Li [b,c,d,1], Zelin Zang [e,1], Jing Wang [a,1], Jun Li [a,1], Junbo Liang [a,1], Fangfei Zhang [b,c,d], Qiushi Zhang [f], Weigang Ge [f], Hao Chen [f], Xindong Sun [b,c,d], Liang Yue [b,c,d], Xiaomai Wu [a], Bo Shen [a], Jiaqin Xu [a], Hongguo Zhu [a], Shiyong Chen [a], Hai Yang [a], Shigao Huang [g], Minfei Peng [a], Dongqing Lv [a], Chao Zhang [a], Haihong Zhao [a], Luxiao Hong [a], Zhehan Zhou [a], Haixiao Chen [a], Xuejun Dong [h], Chunyu Tu [h], Minghui Li [h], Yi Zhu [b,c,d,*], Baofu Chen [a,*], Stan Z. Li [e,*], Tiannan Guo [b,c,d,*]

[a] Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai 317000, Zhejiang Province, China
[b] Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, China
[c] Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou 310024, Zhejiang Province, China
[d] Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou 310024, Zhejiang Province, China
[e] School of Engineering, Westlake University, Hangzhou 310024, Zhejiang Province, China
[f] Westlake Omics (Hangzhou) Biotechnology Co., Ltd., Hangzhou 310024, Zhejiang Province, China
[g] Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, 999078 Macau SAR, China
[h] Shaoxing People's Hospital, Shaoxing Hospital, Zhejiang University School of Medicine, Shaoxing 312000, Zhejiang Province, China

ARTICLE INFO

ABSTRACT

Severity prediction of COVID-19 remains one of the major clinical challenges for the ongoing pandemic. Here, we have recruited a 144 COVID-19 patient cohort, resulting in a data matrix containing 3,065 readings for 124 types of measurements over 52 days. A machine learning model was established to predict the disease progression based on the cohort consisting of training, validation, and internal test sets. A panel of eleven routine clinical factors constructed a classifier for COVID-19 severity prediction, achieving accuracy of over 98% in the discovery set. Validation of the model in an independent cohort containing 25 patients achieved accuracy of 80%. The overall sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were 0.70, 0.99, 0.93, and 0.93, respectively. Our model captured predictive dynamics of lactate dehydrogenase (LDH) and creatine kinase (CK) while their levels were in the normal range. This model is accessible at https://www.guomics.com/covidAI/ for research purpose.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The ongoing COVID-19 pandemic, caused by the SARS-CoV-2 virus, poses an unprecedented public health crisis to the entire world. To date, SARS-CoV-2 has spread to more than 200 countries and regions around the world. Most confirmed cases are classified as mild or moderate, while the other 14% and 5% are severe and critical cases, respectively [1]. These patients need to be hospital-

ized and monitored with intensive care to prevent deterioration of the disease that may lead to fatality without timely diagnosis and treatment.

Currently, the diagnosis of COVID-19 mainly depends on the virus RNA test of SARS-CoV-2 [2,3]. This is a qualitative test showing whether the patient is infected by the virus [4]. Computed tomography (CT) is a complementary strategy for COVID-19 diagnosis [5]. CT images have been used to facilitate severity determination of COVID-19 patients in an artificial intelligence (AI)-assisted model [5]. However, about 20% of COVID-19 patients showed no obvious imaging changes in the lung [6], bringing difficulties to physicians when making decisions for suitable clinical treatments. To better evaluate the disease conditions of COVID-19 patients, routine laboratory tests, including but not limited to complete blood cell count, blood biochemistry and immune tests are taken regularly by physicians. Physicians then make clinical decisions and prescribe treatments accordingly. However, this is laborious and sometimes biased, depending on the physician's own experience, especially when facing with such a heavy medical burden in the pandemic. Therefore, automatic integration and interpretation of routine laboratory indexes in an unbiased way will be beneficial to severity stratification and prognosis evaluation for COVID-19 patients.

Machine learning has been applied in medicine for clinical data such as images for disease diagnosis and classification. During the COVID-19 pandemic, machine learning has widely applied in the diagnosis, prognosis, and vaccine development of COVID-19 [7]. Studies focusing on building predictive models for supporting clinical decision-making have been reported, including prediction of outcome [8], evaluation of mortality risk [9], development of critical illness [10], and monitoring the severity [11,12], etc. Risk factors associated with the severity of the COVID-19 disease have been widely reported [7]. Several clinical features, such as age, gender, lactic dehydrogenase (LDH), C-reactive protein (CRP), and lymphocyte count, have been reported to be highly correlated with the severity of COVID-19 patients [11]. Recently, a Chinese team found that three key features (LDH, CRP, and lymphocyte) can be used to predict the mortality of COVID-19 with over 90% accuracy [13]. These models were built from a limited amount of data and might be biased [11]. In a paper published by Liang et al., a risk-scoring algorithm based on some key characteristics of COVID-19 patients at the time of admission to the hospital was developed, which may assist in predicting a patient's risk of developing into critical illness [14]. More recently, they also built a survival model to predict critical cases by comorbidity and several laboratory indexes [10]. These models are built on datasets containing static information at certain time points, for example, upon admission or discharge, mostly when the patients were in severe or critical status. However, the time from symptom onset to admission varied in a large range due to different local medical resources and personal medical treatment intention. The time of onset of symptoms as a starting point is more objective and reasonable. As the disease progresses, the physiological condition and respective laboratory indicators of the patient are constantly changing. However, few models make use of the longitudinal data to predict disease severity due to the lack of dynamic data. Therefore, it remains challenging to stratify the COVID-19 patients when they are in the transient stage from mild to severe.

In the present study, we built a machine learning model based on the longitudinal measurement of a panel of 124 clinical indicators in a retrospective patient cohort containing 144 COVID-19 patients to predict the disease severity. Eleven key clinical factors were prioritized to be highly associated with COVID-19 severity. The model achieved 93% accuracy in distinguishing severe patients among the infected cases in the whole datasets.

2. Materials and methods

2.1. Patients and samples

From January 17 to March 10 in 2020, 841 patients have been screened by SARS-CoV-2 nucleic acid test in Taizhou Hospital. From them, 144 patients were diagnosed as COVID-19 patients by reverse transcriptase-polymerase chain reaction (RT-PCR) and chest CT according to the Pneumonia Diagnosis and Treatment Scheme for New Coronary Virus Infections (Trial Edition 5, Revision). A total number of 124 indicators from 17 categories of laboratory tests have been regularly monitored over 52 days. Clinical data for these patients were curated from the hospital information system (HIS), included epidemiology, gender, age, BMI, underlying diseases, chest CT, presenting symptoms, length of stay (LOS). Laboratory data included complete blood count parameters, blood biochemistry and immune index, blood coagulation indicator, lymphocyte subsets, cytokines, and arterial blood gas (ABG). Another twenty-five independent test readings were collected from Shaoxing People's Hospital following the same criteria with the Taizhou test set.

This study was approved by the Medical Ethics Committee of Taizhou Hospital, Shaoxing People's Hospital, and Westlake University. The informed consent was obtained from each enrolled subject. Besides, the case of minors enrolled in the study was approved by parents and/or legal guardians.

2.2. Clinical laboratory examinations

Blood samples were collected at each time point since admission. 2 mL EDTA-K2 anticoagulant peripheral blood samples were measured for complete blood count using a Sysmex 2100D routine hematology analyzer (Kobe, Japan). Erythrocyte sedimentation rate (ESR) was calculated using the Alifax Test 1 automatic ESR analyzer (UDINE, Italy). Cytokines and lymphocyte subsets were measured using BD FACSCantoTM II within 6 h. Sodium citrate plasma samples were centrifuged at 1500 g for 15 min. Coagulation parameters were determined using a Sysmex CS 5100i automatic hemagglutination analyzer (Kobe, Japan). ABG analysis was performed using GEM Premier 3500 (Instrumentation Laboratory, US).

Serum samples were centrifuged at 1500 g for 10 min for measuring biochemical data including electrolyte, liver and kidney functional proteins, immunoglobulin serial index, blood lipid and glucose, myocardial enzymes (except myoglobin, creatine kinase MB and troponin-I), and infection index (except procalcitonin) using Beckman automatic biochemical analyzer (AU5821). Myoglobin, creatine kinase MB and troponin-I were detected by Beckman automatic immunological analyzer (UniCel DXI-800). Procalcitonin (PCT) levels were determined using the Roche Cobas e411 electrochemiluminescence analyzer (Basel, Switzerland).

Throat swab and sputum specimens collected during hospitalization were sent to the PCR laboratory (BIOSafety Laboratory II) in a biosafety transportation box. Total nucleic acid extraction from the samples was performed using Nucleic Acid Extraction Kit (Shanghai Zhijiang) and RT-PCR was performed using a commercial kit specific for 2019-nCoV detection (triple fluorescence PCR, Shanghai Zhijiang, China, NO. P20200105) approved by the China Food and Drug Administration (CFDA).

The patients (except pregnant) had a chest CT examination at the time of admission. For each patient, the chest CT was performed according the needs of the disease changes during the period of her/his hospitalization. A CT examination was also operated at the time of her/his discharge. The points of CT score are defined by the following rules. Based on the lesion involvement and lesion properties, each infected lobe adds one point. Presence of ground-

glass opacity adds one point. Two points were added in the presence of consolidation lesions, while three points were added in case of fibrosis lesions. The score was reduced by 0.5 point if the CT is improved compared to the previous CT scan. Otherwise, the score was increased by 0.5 point.

The abbreviations, symbols and markings that we used throughout the text are provided in Supplementary Table 1.

2.3. Machine learning

The data matrix from Taizhou Hospital was divided into a discovery dataset and a test dataset. This method defined the samples collected before February 2, 2020 as the discovery dataset (228*124) and the samples collected after February 2, 2020 as the test dataset (130*124). Feature selection and model training were performed in the discovery dataset. The model was then tested in the test dataset. We also included an independent test data matrix (25*11) from Shaoxing People's Hospital for further testing.

The data analysis included four steps: data preprocessing, feature selection, model training and testing. In the data preprocessing step, all missing values were first filled with a median value of all patients. Then as normalization, the mean value of all features in the discovery set was converted to 0 and the standard deviation was converted to 1. The same normalization parameters were applied to the test set. In the feature selection step, the standard genetic algorithm (GA) method in the Python deap library was used. We set the gene locus in the method as the index of the feature, and set the length of the gene chain to 20 (this indicates that the method can select up to 20 different features). Then we set the crossover probability to 0.3, the mutation probability to 0.5, the number of genes in the population to 500, and the number of iterations to 30. The accuracy rate of the 10-fold cross-validation on the verification set was taken as the fitness of the gene chain. Finally, we selected eleven non-repeated and effective features. In the model training step, the discovery dataset was divided into training set and the validation set. The training set was used to train the model, while the validation set was used to optimize the model parameters.

Thereafter, the test dataset was used to test the model for prediction accuracy. The support vector machine (SVM) model with 'rbf' kernal in Python's scikit-learn library was employed. Hyperparameters including regularization parameter C and threshold T. Kernel coefficient were optimized. Regularization parameter C was selected from [1.0, 1.2, 1.4, 1.6, 1.8], and Kernel coefficient was selected from ['rbf', 'linear', 'poly', 'sigmoid']. Finally, we chose Kernel='rbf' and C = 1.2 with the highest area under the curve (AUC) (Supplementary Table 2). The above parameters were optimized and evaluated according to AUC.

The detection result of patient i on a certain feature f is $D_{i,f}$, then the data composed of the selected features of the patient is D_i. Then we have $D_i = \{D_{i,f_1}, D_{i,f_2}, D_{i,f_3}, \cdots, D_{i,f_N}\}$ as the number of selected features. Enter D_i into the trained model $M(\cdot)$, and the model will output a predicted score S_i.

$$S_i = M(D_i)$$

We made the diagnosis based on predicted score S_i.

$$Y_i = \begin{cases} 1 & \text{if } S_i < T \\ 0 & \text{if } S_i \geq T \end{cases}$$

When Y_i is 1, patient i is diagnosed as severe. When Y_i is 0, patient i is diagnosed as non-severe. T is the threshold for diagnosis.

Due to the heterogeneity of the positive and negative samples, it is not possible to directly use $T = 0.5$ for diagnosis. The threshold T is therefore determined by maximizing the correct rate of diagnosis in the validation set, and the same T was applied to the evaluation test set.

2.4. Statistical analysis

Statistical analysis of clinical characteristics was performed by SPSS (version 19.0). Continuous variables were represented by median and range, Kruskal-Wallis H test was used between multiple groups. Categorical variables were expressed as numbers (percentages), and a comparison between groups was made by chi-square. Further investigations were performed by R software (version 3.6.3). The comparison of continuous variables between two groups was performed using Student's t-test for normally distributed variables and the Mann-Whitney U statistics for non-normally distributed variables. P-value in boxplot was calculated by the unpaired two-tailed Student's t-test, and in violin plot p-value was adjusted by Benjamini & Hochberg method. The smooth plot was fitted by locally estimated scatterplot smoothing (LOESS) using the geom smooth function in ggplot. PPV and NPV were adjusted by the ratio of severe cases followed by the published formula [15].

3. Results and discussion

3.1. Clinical characteristics of the study cohort

A total of 841 patients have been screened with the SARS-CoV-2 nucleic acid test from January 17 to March 10, 2020 in Taizhou Hospital, with 144 patients with positive virus RNA (COVID-19) and 697 non-COVID-19 individuals (Fig. 1A). From the non-COVID-19 group, outpatients and patients lacking chest CT results were excluded. 65 patients were recruited as the control group, with their epidemiological information collected. Meanwhile, 144 COVID-19 patients were stratified into severe (N = 36) and non-severe (N = 108) patients based on the clinical diagnosis guideline [6] (Fig. 1A). 124 types of measurements from 17 categories over 52 days were collected and manually curated for these 144 COVID-19 patients, resulting in a data matrix containing 3,065 readings for 124 types of measurements (3065*124). The 17 categories of data included basic information, clinical symptoms and signs, chest CT results, and laboratory tests, as detailed in Supplementary Table 3 that were regularly recorded during their hospitalization.

This cohort has 55.6% male in the severe COVID-19 group and 52.8% male in the non-severe COVID-19 group. The median age was 55.0 years for the severe group and 44.5 years for the non-severe group. The median BMI was 25.5 kg/m^2 in the severe group and 23.8 kg/m^2 in the non-severe group (Table 1). These parameters are consistent with previous observations [16].

The most common symptom at disease onset was fever (64.8% in non-severe and 94.4% in the severe group), followed by cough and pharyngalgia (Table 1). Before admission, 40.2% of patients had underlying diseases, and the ratio was substantially higher in the severe group (50.0%) than that in the non-severe group (37.0%). Hypertension (15.3%) and diabetes (9.7%) were common comorbidities.

Upon admission, 142 (98.6%) COVID-19 patients had image changes in chest CT. The pulmonary plaque was the top 1 abnormal pattern (52.8%), with ratios of abnormality of 66.7% for severe patients and 48.2% for non-severe patients, respectively. The second abnormal pattern was ground-glass opacity (severe vs. non-severe was 41.7% vs. 31.5%). Ratios of pulmonary fibrosis and consolidation did not show a statistical difference between severe patients and non-severe ones.
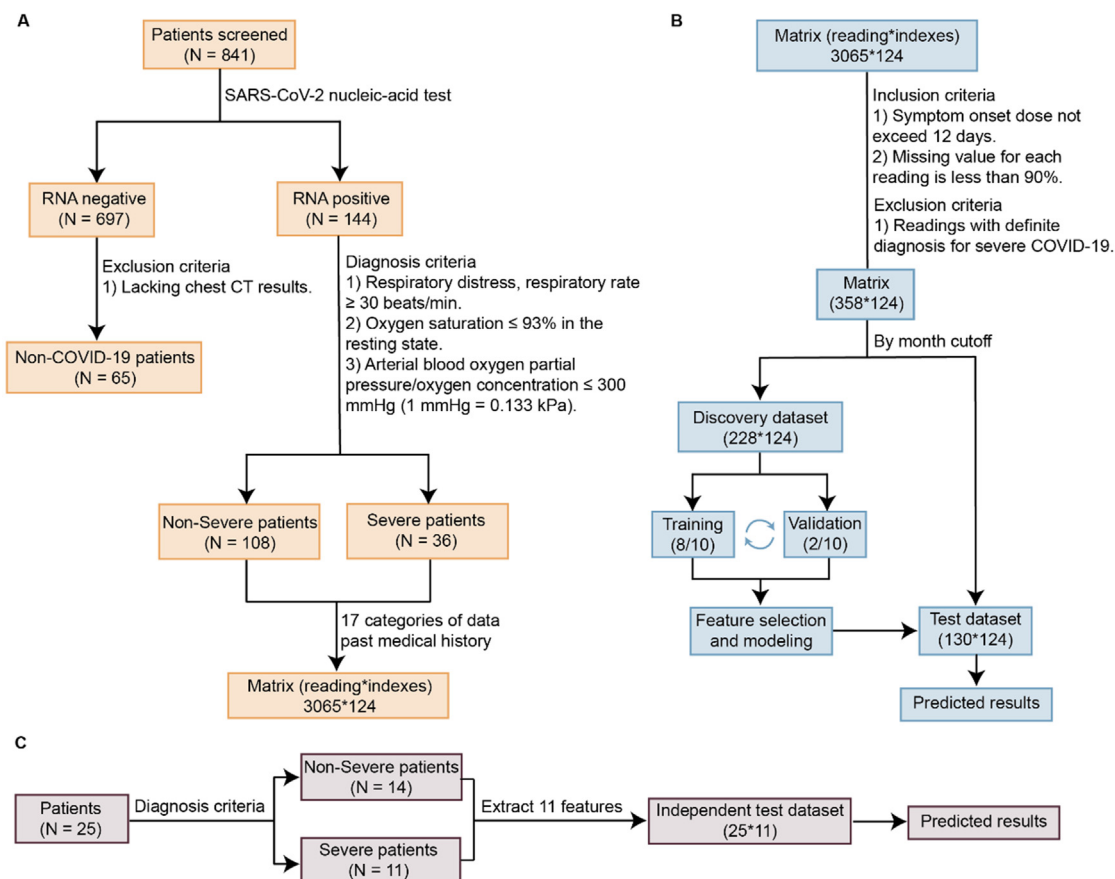
Fig. 1. The study design and modeling workflow. (A) The COVID-19 patient cohort. (B) The workflow of the modeling process. (C) Independent test patient cohort.

In severe patients, intervals from onset and admission to the diagnosis of severe cases were 9 (median) days and 2 (median) days. All 144 patients were followed up during their entire course in hospital. In this study, only one patient was admitted to the intensive care unit (ICU) and underwent invasive mechanical ventilation. They were all cured and discharged eventually. The median LOS for the patients was 20 days for non-severe group and 23 days for severe group, respectively. In summary, this is a well-annotated and curated COVID-19 cohort with comprehensively and systematically recorded information from the disease onset till convalesce and discharge, which provided the potential for subsequent model construction.

3.2. Feature selection and machine learning

To establish a model for severity prediction, we filtered the matrix including 3065*124 readings for all patients over 52 days in total (Fig. 1B). The readings with a definite diagnosis for severe COVID-19 were excluded. Then the readings recorded from symptom onset to the 12th day for all patients were included for severity prediction by machine learning, in which the ratio of severe verse non-severe was close to 20% prevalence of severe cases according to previous studies [1]. At this time point (12th day), no patient had been clinically diagnosed as severe COVID-19 cases. After removing readings with more than 90% missing values, a much smaller data matrix containing 358*124 readings remained. The discovery dataset contains 228*124 readings collected from patients admitted before 1 February 2020, while the test dataset includes 130*124 readings from patients admitted after 1 February 2020. Based on the discovery dataset, a machine learning model was built up by cross training and validation (Fig. 1B). To further

test our model, we also collected an independent test dataset from another clinical center, employing the same criteria as the previous one (Fig. 1C).

The modeling contains three parts, including feature selection, model training, and prediction (Fig. 2A). Firstly, the missing value for each test item in the discovery dataset was filled with a relevant median value for each gender (female or male). We randomly generated feature combinations and loaded the data with selected features. Then feature selection was performed by using \GA\ [17], one of the most advanced and the widely used algorithm for feature selection, assisted with 10-fold cross-validation in the discovery dataset, including both training (9/10) and validation (1/10) sets. The method of GA which selected a panel of features could avoid being trapped into a local optimal solution. As a result, a panel of eleven key clinical factors associated with COVID-19 severity was evolved from 124 characteristics, and included in the 'Active Feature Pool'.

Then we applied the selected features and established the classifier using the SVM [18]. In this step, we set up ten random seeds, and the SVM model was trained in the training set (8/10) and then validated in the validation set (2/10). The hyper-parameters of the model were evaluated by AUC and the threshold was optimized to maximize accuracy. The threshold was set at 0.45, which was determined by maximizing the correct diagnosis rate in the validation set. Via this way, a trained SVM model including the eleven active features and classifier was generated for predicting severe cases in the test dataset (130*124) and an independent test dataset (25*11) using the same threshold of 0.45 (Fig. 2A). We further tested six other classifiers using LightGMB, Catboost, Random Forest, AdaBoost, Nearest Neighbors and Decision Tree. Four of six achieved AUC around 90%; the other achieved more than 70% (Sup-

Table 1
Demographic and clinical characteristics of the patients.

Characteristic	non-COVID-19 (N = 65)	COVID-19			P value
		Non-severe (N = 108)	Severe (N = 36)	All (N = 144)	
Gender - no. [a] (%)					
Male	41 (63.1)	57 (52.8)	20 (55.6)	77 (53.5)	0.7723
Female	24 (36.9)	51 (47.2)	16 (44.4)	67 (46.5)	
Age - yr. [b]					
Median	36.0	44.5	55.0	47.0	0.0001
Range	(3.0–79.0)	(4.0–86.0)	(33.0–79.0)	(4.0–86.0)	
BMI, kg/m^2					
Median	22.0	23.8	25.5	24.4	0.0050
Range	(14.0–32.9)	(16.0–31.2)	(21.6–31.3)	(16.0–31.3)	
Onset to severe, days					
Median	NA	NA	9	NA	NA
Range	NA	NA	(2–26)	NA	
Admission to severe, days					
Median	NA	NA	2	NA	NA
Range	NA	NA	(0–7)	NA	
Length of stay, days					
Median	4	20	23	22	0.0154
Range	(1–31)	(6–44)	(9–40)	(6–44)	
Smoking - no. (%)	5 (7.7)	11 (10.2)	2 (5.6)	13 (9.0)	0.5182
Familial aggregation - no. (%)	7 (10.8)	43 (39.8)	11 (30.6)	54 (37.5)	0.3203
Symptoms and signs - no. (%)					
Fever	35 (53.9)	70 (64.8)	34 (94.4)	104 (72.2)	0.0006
Pharyngalgia	14 (21.5)	15 (13.9)	2 (5.7)	17 (11.8)	0.2405
Cough	38 (58.5)	47 (43.5)	18 (50.0)	65 (45.1)	0.4985
Expectoration	25 (38.5)	19 (17.6)	7 (19.4)	26 (18.1)	0.8025
Muscle soreness	1 (1.5)	4 (3.7)	2 (5.6)	6 (4.2)	0.6399
Headache	1 (1.5)	9 (8.3)	7 (19.4)	16 (11.1)	0.1208
Diarrhea	24 (36.9)	3 (2.8)	3 (8.3)	6 (4.2)	0.1652
Chest tightness	4 (6.2)	7 (6.5)	4 (11.1)	11 (7.6)	0.4676
Heart rate					
Median	96	82	84	82	0.5106
Range	(72–126)	(57–147)	(57–106)	(57–147)	
Respiratory rate					
Median	19	19	19	19	0.6265
Range	(18–28)	(12–26)	(12–22)	(12–26)	
Comorbidities - no. (%)					
Hypertension	7 (10.8)	14 (13.0)	8 (22.2)	22 (15.3)	0.1811
Diabetes	6 (9.2)	9 (8.3)	5 (13.9)	14 (9.7)	0.3397
Cardiovascular disease	2 (3.1)	1 (0.9)	2 (5.6)	3 (2.1)	0.1543
Cerebrovascular disease	4 (6.2)	2 (1.9)	1 (2.8)	3 (2.1)	1.000
Chronic bronchitis	3 (4.6)	4 (3.7)	0 (0.0)	4 (2.8)	0.5721
Tuberculosis	0 (0.0)	4 (3.7)	0 (0.0)	4 (2. 8)	0.5721
Malignant tumor	2 (3.1)	1 (0.9)	1 (2.8)	2 (1.4)	0.4388
Thyroid disease	2 (3.1)	2 (1.9)	3 (8.3)	5 (3.5)	0.0998
Hepatitis	1 (1.5)	5 (4.6)	2 (5.6)	7 (4.9)	1.0000
Chronic kidney disease	1 (1.5)	1 (0.9)	1 (2.8)	2 (1.4)	0.4388
Digestive system disease	3 (4.6)	6 (5.6)	1 (2.8)	7 (4.9)	0.6803
Chest CT [e] - no. (%)					
Involvement of chest radiographs	47 (72.3)	106 (98.1)	36 (100.0)	142 (98.6)	1.0000
Unilateral lung	14 (21.5)	28 (25.9)	3 (8.3)	31 (21.5)	0.0261
Bilateral lung	33 (50.8)	65 (60.2)	31 (86.1)	96 (66.7)	0.0043
Pulmonary plaques	13 (20.0)	52 (48.2)	24 (66.7)	76 (52.8)	0.0539
Ground-glass opacity (GGO)	2 (3.1)	34 (31.5)	15 (41.7)	49 (34.0)	0.2640
Pulmonary fibrosis	3 (4.6)	7 (6.5)	1 (2.8)	8 (5.6)	0.6796
Pulmonary consolidation	11 (16.9)	23 (21.3)	9 (25.0)	32 (22.2)	0.6434

[a] no. (%): number.
[b] yr.: year.

plementary Table 4). These results indicate that our selected features are robust.

SHapley Additive exPlanations (SHAP) algorithm [19,20] was performed to interpret the eleven features in this model, namely oxygenation index, basophil counts (BASO#), aspartate aminotransferase (AST), gender, magnesium (Mg), gamma-glutamyl transpeptidase (GGT), platelet counts, activated partial thromboplastin time (APTT), oxygen saturation (SaO2), body temperature and days of symptom onset. Their importance to the model was evaluated by SHAP values (Fig. 2B). The oxygenation index was the most important one, with a SHAP value of 0.94. In clinical prac-

tice, the oxygenation index is also a critical factor to evaluate the state of disease severity.

Furthermore, nine features with continuous records (12 days) were selected and compared in severe and non-severe groups. Their median values (discovery dataset) were calculated and shown in boxplots (Fig. 2C). Six of nine features were significantly dysregulated (p-value less than 0.05). Among them, oxygen index and SaO2 were decreased in severe cases. This was directly associated with pulmonary functions. Two selected features, AST and GGT, were up-regulated, indicating that COVID-19 may induce slight hepatic injury in the early stage. Although within the normal
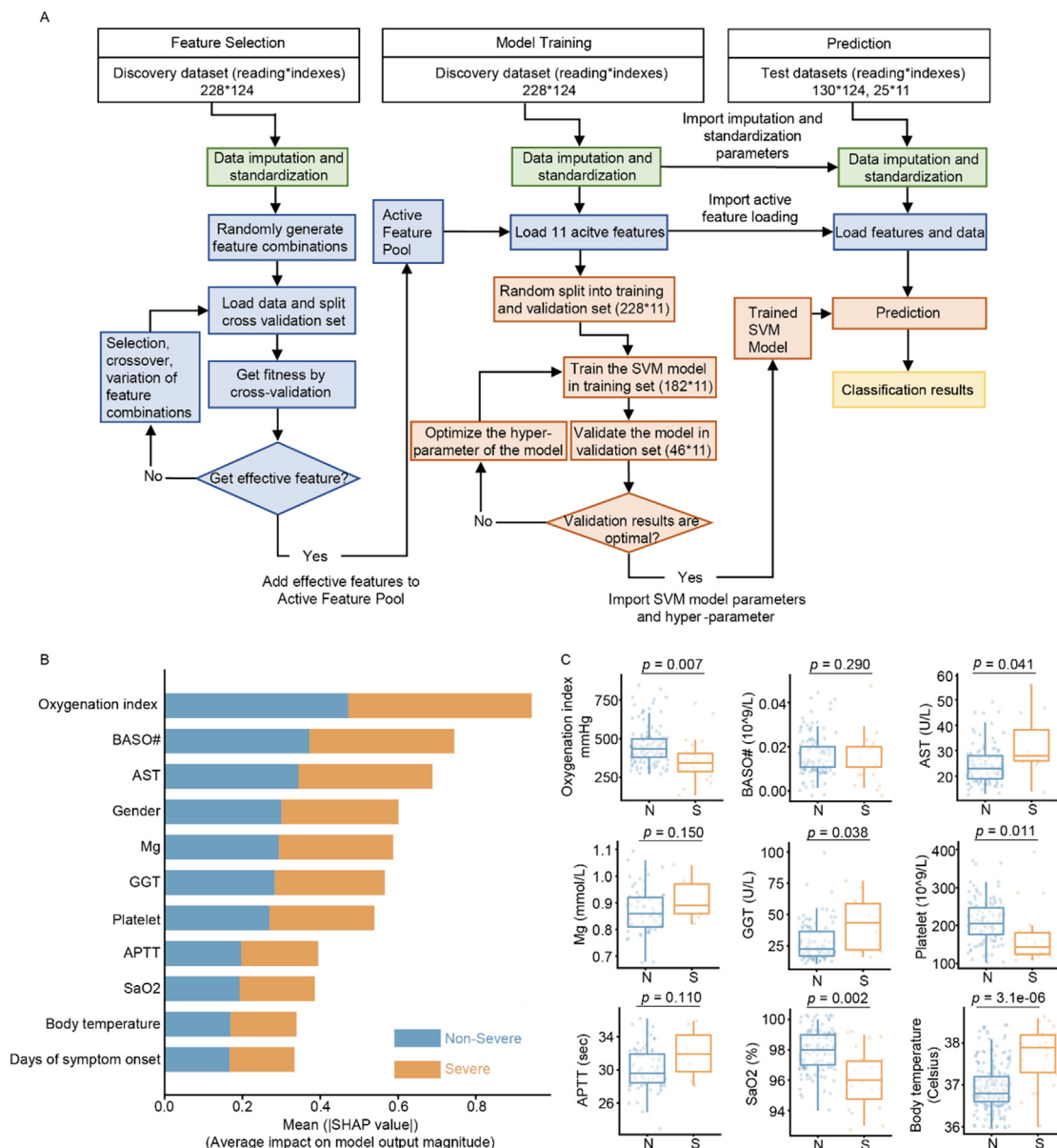
Fig. 2. The machine learning model structure and selected features. (A) The detailed workflow contains four major steps: i) data preprocessing (green); ii) feature selection (blue); iii) building up the machine learning model (pink); iv) prediction results (yellow). (B) The eleven selected key clinical features displayed in two groups according to the severity of the disease (blue: non-severe; orange: severe) from the discovery cohort. (C) The comparison of each key feature between non-severe and severe COVID-19 patients in the discovery dataset. *P*-value was calculated by the unpaired two-tailed Student's *t*-test. N, non-severe; S, severe. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

range, the count of platelets was reduced significantly in severe cases. Three characteristics, namely APTT, Mg, and BASO#, showed no significant change between severe and non-severe COVID-19 patients, and a comprehensive comparison was further performed. As shown in Supplementary Fig. 1, readings from four groups of individuals, severe COVID-19, non-severe COVID-19, non-COVID-19 patients with flu-like symptoms, and healthy people with physical examination were included for systematic comparison. The data showed that APTT was substantially up-regulated in the severe COVID-19 group in the entire dataset, including the discovery and the test datasets, and the count of basophils was decreased in all COVID-19 patients compared with the healthy group. However, Mg showed no difference across the four groups. To validate the importance of Mg in the model, we used the rest ten features except for Mg and investigated the performance of the model in

ROC (Supplementary Fig. 2). The AUC values were the same in the training and validation dataset and only decreased by 0.01 in the test dataset without the feature of Mg, which indicated the little contribution of Mg to the model.

The remaining two indicators not shown in the box diagram are days of symptom onset and gender. This is not surprising because when the patients were evolving from non-severe to severe status, the symptoms grew worse. Gender was selected as another important feature. In our dataset, male patients were more likely to be infected SARS-CoV-2 than female patients in both non-severe and severe groups, consistent with the literature reporting that male COVID-19 patients had a worse outcome and that 70% of patients who died of COVID-19 were male in an Italy cohort [21]. The vulnerability of males has also been found in the SARS-CoV epidemic in 2003 [22]. It has been recently found that the plasma concentra-

tion of ACE2, a functional receptor for SARC-CoV-2 infection, is higher in men than that in women, as detected in two independent cohorts [23]. This may explain the association between gender and the fatality rate of COVID-19. Meanwhile, females develop enhanced innate and adaptive immune responses than males did, thus they are less susceptible to kinds of infections of bacterial, viral, parasitic, and fungal origin and malignancies [24]. In fact, the clinical manifestations of infectious or autoimmune diseases and malignant tumors differ between men and women.

3.3. Severity prediction by machine learning

The model assigned a score (from 0 to 1) to indicate the likelihood of disease severity, and a higher score indicated greater severity. The model is described in detail in Methods, while the data are shown in the scatter plot (Fig. 3A). Samples with a score greater than 0.45 were identified as a severe state. The receiver operating characteristics (ROC) plot achieved an AUC value of 1.00 and 0.98 for training and validation datasets, respectively (Fig. 3B). 224 out of 228 samples were correctly identified with

an accuracy of 0.98 for the discovery dataset. The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were 0.90, 0.99, 0.96 and 0.99, respectively (Fig. 3C).

To further evaluate the performance of the eleven-clinical feature classifier, we analyzed 130*124 readings in the test dataset and 25*11 readings in the independent test dataset from a different hospital. The classifier achieved AUC value of 0.89 (Fig. 3B) and correctly classified 112 of 130 readings with accuracy, sensitivity, specificity, PPV, and NPV of 0.86, 0.56, 0.98, 0.91, and 0.85, respectively. Sixteen of 18 incorrectly classified readings were from the severe COVID-19 group, especially in the early stage of COVID-19, which showed similar clinical signs with the non-severe group. Most of the incorrect predictions (10 of 18 readings) were from four COVID-19 patients, indicating an individual effect on the model. As to the independent test dataset of much smaller size, the classifier achieved an AUC value of 0.75 and accuracy, sensitivity, specificity, PPV, and NPV of 0.80, 0.64, 0.93, 0.69, and 0.91, respectively (Fig. 3C). All incorrectly identified cases in this dataset belonged to the early stage of COVID-19 with days of symptom onset less than eight days.
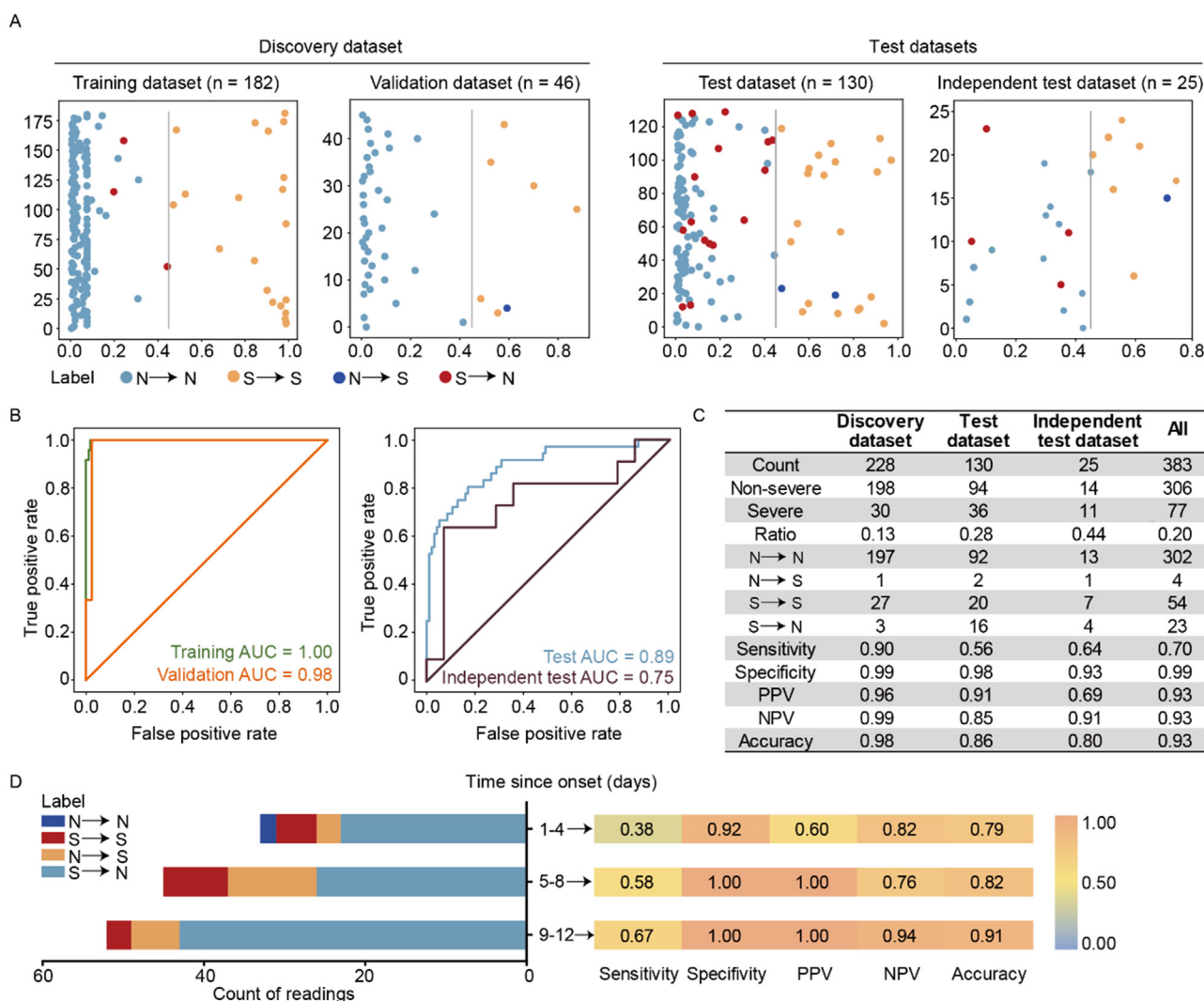

Fig. 3. The performance of the predictive model. (A) Severe and Non-severe cases are shown as scatter plots in different colors (red: severe; blue: non-severe). X-axis indicates the predicted scores, representing the probability of disease severity for each timepoint. The cutoff of the predicted score was 0.45. Y-axis denotes the indexes of samples. N → S indicates a non-severe case which was predicted as a severe case. (B) ROC plots of the performance of support vector machine (SVM) for severity prediction. (C) Summary of the performance metrics. (D) The test dataset was further divided into three longitudinal parts at a four-day interval. The X-axis represents the count of the readings for four types of predicted outcomes in each part (N → N, S → S, N → S, and S → N). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We also examined the performance of the model from the longitudinal perspective. The test dataset was divided into three parts according to the length of symptom lasting period since disease onset. In general, the longer the disease progressed, the better the model predicted. As the onset time increased, the accuracy of prediction elevated from 0.79 (1–4 days), 0.82 (5–8 days) to 0.91 (9–12 days) in the test set (Fig. 3D). There were two incorrectly identified readings in non-severe patients. They were both from the initial three days since onset, probably due to the low oxygen index (less than 380 mmHg), which was out of the normal range (400–500 mmHg). As to the part of 9–12 days, three readings of the severe group were classified into the non-severe group, while two of them were scored with 0.42 and 0.44, which were very close to the threshold.

3.4. Dynamic changes of key clinical features over seven weeks

The unique curated longitudinal data permitted detailed temporal investigation of specific parameters. We investigated the temporal changes of eight out of eleven critical clinical features in both severe and non-severe COVID-19 patients during the first 40 days since disease onset. In addition, we also analyzed 18 more clinical features (Fig. 4) that were found to be related to the progression of COVID-19 disease [11,25,26].

Based on longitudinal changes of blood cell counts (such as platelet), our data showed that the severity of the disease intensified in the second week since symptom onset (Fig. 4). The counts of platelets, lymphocytes, eosinophils, and basophils were reduced in the severe group, while neutrophils and WBC were dramatically elevated (Fig. 4).

Decreased number of lymphocytes has been reported in association with the COVID-19 severity [27], probably due to apoptosis and necrosis of lymphocytes [28]. Dysregulation of neutrophils was another predominant alteration in the complete blood count. Neutrophils are recruited early to sites of infection where they kill pathogens (bacteria, fungi, and viruses) by oxidative burst and phagocytosis [29].

A significant rise in neutrophil counts features the severity of COVID-19. An increase of neutrophils may induce the formation of neutrophil extracellular traps (NETs) which could trigger a cascade of inflammatory reactions [30]. The latter induces damages to surround tissues, facilitates micro thrombosis, and results in permanent organ injuries to the pulmonary, cardiovascular, and renal systems [29], which are three commonly affected organ systems in severe COVID-19 [13].

Eosinophils were drastically reduced after SARS-CoV-2 infection and were merely detectable in the acute stage (Fig. 4). The increase of eosinophils in non-severe patients happened earlier than that in
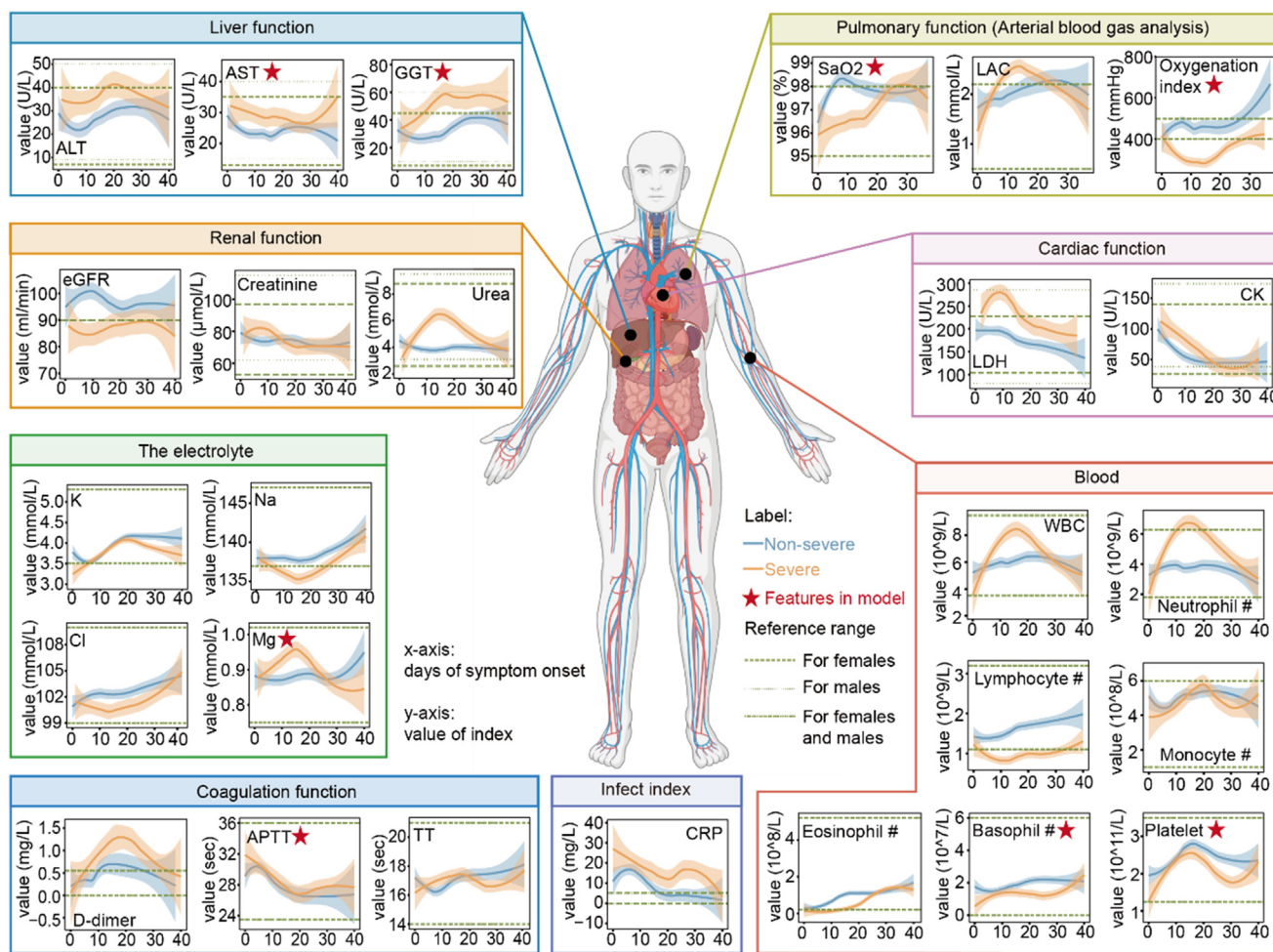

Fig. 4. Temporal dynamics of selected clinical indexes in COVID-19 patients. The x-axis represents the time of disease course since symptom onset, and y-axis shows the measured values for each feature over time in severe (orange line) and non-severe (blue line) patients. The solid line was fitted by locally estimated scatterplot smoothing (LOESS). Abbreviations: ALT, alanine transaminase; LAC, lactic acid; eGFR, estimated glomerular filtration rate; LDH, lactate dehydrogenase; CK, creatine kinase; K, potassium; Na, sodium; Cl, chlorine; WBC, white blood cell; TT, thrombin time; CRP, C-reactive protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

severe cases, suggesting that the increase of eosinophils can be the signs for COVID-19 recovery. The predictive value of eosinophil count observed in our study is supported by a recent independent study [10]. The count of basophils is another predictive feature of our model. Basophils are the least abundant granulocytes, representing less than 1% of peripheral blood leukocytes. Recent studies have shown that the count of basophils was reduced in COVID-19 patients [31], agreeing with our data. Moreover, an independent study that proposed a multivariate Cox regression model exploring risk factors for lethal COVID-19 patients also nominated progressive increase of basophils [32].

Coagulation dysfunction has been found in multiple epidemiologic studies of COVID-19 [13,33]. Our data showed that platelet counts in blood decreased in severe patients significantly (Fig. 4), consistent with a previous meta-analysis reporting low platelet count as a risk factor for COVID-19 severity and mortality [34]. Furthermore, coagulation-related parameters such as D-dimer elevated too, especially in death cases [33]. Thrombin time (TT) and APTT are often used to monitor the coagulation function of patients, which are mostly increased in the presence of heparin or heparin-like substances.

CRP is an acute-phase protein, the expression level of which in the blood is associated with the severity of inflammation [35]. As an indicator of infection, in our dataset, CRP was significantly increased in both mild and severe cases, far beyond the normal range. CRP decreased as the disease progressed and recovered.

We also evaluated the functions of organs since it has been reported that SARS-CoV-2 infection systematically induced multi-organ dysfunctions, such as lung, liver, kidney, and heart, particularly in severe and critical cases [36,37]. The commonly used indicators in clinic include pulmonary function indicators including SaO2, blood-gas lactate (LAC) and oxygen index, liver function indicators (GGT, AST, and ALT), renal function indicators (eGFR and creatinine), and cardiac function indicator (LDH), were monitored. ABG analysis is a key approach to evaluate the function of the lungs by measuring acidity and the level of oxygen and carbon dioxide, which provides important indexes and direct evidence for indicating pulmonary function and the severity of COVID-19. Our data showed that oxygen-related indexes, namely SaO2 and oxygen index, decreased in severe COVID-19 patients (Fig. 4), suggesting a high degree of sensitivity for severity prediction and thus selected as pivotal features in our model. The level of blood-gas LAC reached a peak in the first ten days in the severe group, agreeing with the literature [25].

Regarding liver functions, the dynamic change of GGT, AST and ALT showed that hepatic functions got worse around the third week (Fig. 4). Moreover, levels of the three indexes were higher in severe patients than in non-severe patients from the beginning of the disease. This may be partly due to the side effect of multi-drug administration according to ALT and GGT levels. AST was a predominant feature in our predictive model, which was elevated in the severe group. A multicenter retrospective study including 5771 adult patients reported that AST increased first, followed by ALT in severe cases with liver injury. This alternation was associated with dysregulation of lymphocyte and neutrophil counts [38]; the latter was selected by our model as well.

The estimated glomerular filtration rate (eGFR) is a direct indicator of renal function, and its level in severe patients was significantly lower than that in non-severe patients. On the opposite, the level of urea was elevated in the severe group in the early stage (first ten days) while remained stable in the non-severe group. Electrolytes were also closely related to renal functions, and most of them were within the reference interval except for sodium and potassium in the severe group (Fig. 4C).

LDH is widely expressed in various organs, and it has been reported as a key cardiac marker closely associated with COVID-19 severity [39]. In our study, LDH peaked around the 10th day in severe COVID-19 patients comparing with non-severe cases, although the peak value did not exceed the upper limit of the normal range of LDH for males. Another cardiac indicator, CK, was also in the normal range; however, its dynamics differed between severe and non-severe patients.

The major limitation of this study is the relatively small size of the patient cohort, although some interesting clues have been uncovered. Therefore, the model still requires further validation in sizeable cohorts from multiple clinical centers.

4. Conclusions

In summary, we built a customized machine learning model for COVID-19 severity prediction based on the longitudinal measurement of the feature clinical factors over time. The model composed of eleven routine clinical features which are widely available might be further developed into a practical tool for COVID-19 management.

5. Key points

Question: 1) To predict whether a COVID-19 patient would develop into severe cases based on routine clinical information using machine learning. 2) To monitor how critical clinical parameters change over the disease course of COVID-19.

Findings: We present a novel support vector machine model for evaluating the severity of COVID-19 patients based on routine clinical parameters. Days after symptom onset, basophil counts, magnesium, and gamma-glutamyl transpeptidase have not been reported previously. Our model is freely available at webserver https://www.guomics.com/covidAI/. In addition, we depicted the dynamics of key factors over the disease course.

Meaning: The model composed of eleven routine clinical features which are widely available might be further developed into a practical tool for COVID-19 management.

Declaration of Competing Interest

T.G. and Y.Z. are shareholders of Westlake Omics Inc. W.G., Q.Z., and H.C. are employees of Westlake Omics Inc. The other authors declare no competing interests in this paper.

Acknowledgments

Author contributions

T.G., B.C., J.Liang., S.L., B.S., J.Li., H.C. and Y.Z. supervised the project. T.G. and Y.S. designed the study. K.Z., J.W., M.P., X.Z., Z.Zhou., H.Zhao., J.X., H.Zhu., J.Li., S.C., H.Y., D.L., C.Z., L.H., and X.W. collected clinical data and provided clinical supervision. S.H. organized the data, Y.S., L.L., J.W., K.Z., F.F., X.S., Q.Z., W.G., L.Y., X.S., and H.C. conducted data analysis and visualization. Z.Zang. established machine learning model. X.D., C.T., and M.L. collected the validation data. Data were interpreted and presented by all co-authors.

Y.S., L.L., Y.Z., and T.G. wrote the manuscript with inputs from all co-authors.

Ethics approval and consent to participate

This study was approved by the Medical Ethics Committee of Taizhou Hospital, Shaoxing People's Hospital and Westlake University, Zhejiang province of China, and informed consent was obtained from each enrolled subject. Besides, the case of minors enrolled in the study was approved by parents and/or legal guardians.

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.06.022.

References

[1] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72314 cases from the Chinese Center for disease control and prevention. JAMA 2020;323(13):1239. https://doi.org/10.1001/jama.2020.2648.

[2] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Yi, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395 (10223):497–506.

[3] Asselah T, Durantel D, Pasmant E, Lau G, Schinazi RF. COVID-19: Discovery, diagnostics and drug development. J Hepatol 2021;74(1):168–84.

[4] Broughton JP, Deng X, Yu G, Fasching CL, Servellita V, Singh J, et al. CRISPR-Cas12-based detection of SARS-CoV-2. Nat Biotechnol 2020;38(7):870–4.

[5] Zhang K, Liu X, Shen J, Li Z, Sang Ye, Wu X, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed Tomography. Cell 2020;182(5):1360. https://doi.org/10.1016/j.cell.2020.08.029.

[6] Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020.

[7] Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos Solitons Fractals 2020;139:110059. https://doi.org/10.1016/j.chaos.2020.110059.

[8] Ning W, Lei S, Yang J, Cao Y, Jiang P, Yang Q, et al. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. Nat Biomed Eng 2020;4(12):1197–207.

[9] Barda N, Riesel D, Akriv A, Levy J, Finkel U, Yona G, et al. Developing a COVID-19 mortality risk prediction model when individual-level data are not available. Nat Commun 2020;11(1). https://doi.org/10.1038/s41467-020-18297-9.

[10] Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, et al. Early triage of critically ill COVID-19 patients using deep learning. Nat Commun 2020;11(1). https://doi.org/10.1038/s41467-020-17280-8.

[11] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328.

[12] Gallo Marin B, Aghagoli G, Lavine K, Yang L, Siff EJ, Chiang SS, et al. Predictors of COVID-19 severity: A literature review. Rev Med Virol 2021;31:1–10.

[13] Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nature. Machine Intelligence 2020.

[14] Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. JAMA Intern Med 2020;180(8):1081. https://doi.org/10.1001/jamainternmed.2020.2033.

[15] Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ (Clinical research ed) 1994;309:102-..

[16] Rolf JD. Clinical characteristics of Covid-19 in China. N Engl J Med 2020;382:1860.

[17] Mitchell M. Elements of Generic Algorithms. An Introduction to Generic Algorithms 1998:158..

[18] Huang C-L, Wang C-J. A GA-based feature selection and parameters optimizationfor support vector machines. Expert Syst Appl 2006;31 (2):231–40.

[19] Roth AE, editor. The Shapley value: essays in honor of Lloyd S. Shapley: Cambridge University Press; 1988.

[20] Shapley LS. Notes on the n-Person Game – II: The Value of an n-Person Game. Santa Monica, Calif: RAND Corporation; 1951.

[21] Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA 2020.

[22] Karlberg J, Chong DS, Lai WY. Do men have a higher case fatality rate of severe acute respiratory syndrome than women do?. Am J Epidemiol 2004;159:229–31.

[23] Sama IE, Ravera A, Santema BT, van Goor H, Ter Maaten JM, Cleland JGF, et al. Circulating plasma concentrations of angiotensin-converting enzyme 2 in men and women with heart failure and effects of renin-angiotensin-aldosterone inhibitors. Eur Heart J 2020;41:1810–7.

[24] Jaillon S, Berthenet K, Garlanda C. Sexual dimorphism in innate immunity. Clin Rev Allergy Immunol 2019;56(3):308–21.

[25] Cecconi M, Piovani D, Brunetta E, Aghemo A, Greco M, Ciccarelli M, et al. Early predictors of clinical deterioration in a cohort of 239 patients hospitalized for Covid-19 infection in Lombardy, Italy. J Clin Med 2020;9(5):1548. https://doi.org/10.3390/jcm9051548.

[26] The Lancet Global H. Decolonising COVID-19. Lancet Glob Health 2020;8: e612..

[27] Henry BM. COVID-19, ECMO, and lymphopenia: a word of caution. Lancet Respir Med 2020;8(4):e24. https://doi.org/10.1016/S2213-2600(20)30119-3.

[28] Tan Li, Wang Qi, Zhang D, Ding J, Huang Q, Tang Y-Q, et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. Signal Transduct Target Ther 2020;5(1). https://doi.org/10.1038/s41392-020-0148-4.

[29] Barnes BJ, Adrover JM, Baxter-Stoltzfus A, Borczuk A, Cools-Lartigue J, Crawford JM, et al. Targeting potential drivers of COVID-19: Neutrophil extracellular traps. J Exp Med 2020;217..

[30] Jorch SK, Kubes P. An emerging role for neutrophil extracellular traps in noninfectious disease. Nat Med 2017;23(3):279–87.

[31] Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, et al. Dysregulation of immune response in patients with COVID-19 in Wuhan, China. Clin Infect Dis 2020..

[32] Chen R, Sang L, Jiang M, Yang Z, Jia N, Fu W, 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.

[33] Wang D, Hu Bo, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020;323(11):1061. https://doi.org/10.1001/jama.2020.1585.

[34] Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. Clin Chim Acta 2020;506:145–8.

[35] Sproston NR, Ashworth JJ. Role of C-reactive protein at sites of inflammation and infection. Front Immunol 2018;9:754.

[36] Feng Y, Ling Y, Bai T, Xie Y, Huang J, Li J, et al. COVID-19 with different severities: A multicenter study of clinical features. Am J Respir Crit Care Med 2020;201(11):1380–8.

[37] Liu PP, Blet A, Smyth D, Li H. The science underlying COVID-19: Implications for the cardiovascular system. Circulation 2020;142(1):68–78.

[38] Lei F, Liu Y-M, Zhou F, Qin J-J, Zhang P, Zhu L, et al. Longitudinal association between markers of liver injury and mortality in COVID-19 in China. Hepatology 2020;72(2):389–98.

[39] Mo P, Xing Y, Xiao Y, Deng L, Zhao Q, Wang H, et al. Clinical characteristics of refractory COVID-19 pneumonia in Wuhan, China. Clin Infect Dis 2020.