

Development and validation of a risk stratification model for screening suspected cases of COVID-19 in China

Jing Ma^{1,*}, Xiaowei Shi^{2,3,*}, Weiming Xu^{4,*}, Feifei Lv¹, Jian Wu^{2,3}, Qiaoling Pan^{2,3}, Jinfeng Yang^{2,3}, Jiong Yu^{2,3}, Hongcui Cao^{2,3}, Lanjuan Li^{2,3}

¹Department of Laboratory Medicine, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China

²State Key Laboratory for The Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China

³Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou 310003, China

⁴Taizhou Enze Medical Center (Group) Enze Hospital, Taizhou 318050, China

*Equal contribution

Correspondence to: Hongcui Cao; **email:** hccao@zju.edu.cn

Keywords: severe acute respiratory syndrome coronavirus 2, COVID-19, risk stratification, model, suspected cases

Received: April 28, 2020

Accepted: June 25, 2020

Published: July 29, 2020

Copyright: Ma et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

How to quickly identify high-risk populations is critical to epidemic control. We developed and validated a risk prediction model for screening SARS-CoV-2 infection in suspected cases with an epidemiological history. A total of 1019 patients, ≥ 13 years of age, who had an epidemiological history were enrolled from fever clinics between January 2020 and February 2020. Among 103 (10.11%) cases of COVID-19 were confirmed. Multivariable analysis summarized four features associated with increased risk of SARS-CoV-2 infection, summarized in the mnemonic COVID-19-REAL: radiological evidence of pneumonia (1 point), eosinophils $< 0.005 \times 10^9/L$ (1 point), age ≥ 32 years (2 points), and leukocytes $< 6.05 \times 10^9/L$ (1 point). The area under the ROC curve for the training group was 0.863 (95% CI, 0.813 - 0.912). A cut-off value of less than 3 points for COVID-19-REAL was assigned to define the low-risk population. Only 10 (2.70%) of 371 patients were proved to be SARS-CoV-2 positive, with a negative predictive value of 0.973. External validation was similar. This study provides a simple, practical, and robust screening model, COVID-19-REAL, able to identify populations at high risk for SARS-CoV-2 infection.

INTRODUCTION

At the end of December 2019, an outbreak of pneumonia caused by a novel coronavirus (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2) was reported in Wuhan, China [1]. Transmission takes place through respiratory droplets and other routes such as ocular surfaces [2–4]. This highly contagious virus spread rapidly to other cities of China, and gave rise to a

global outbreak. As of Mar 23, 2020, over 300,000 cases of COVID-19 have been confirmed worldwide, and more than 10,000 have died. The number of confirmed cases is still increasing. One study estimates the basic reproductive number (R0) to be 2.68, and the epidemic doubling time to be 6.4 days [5]. The control of COVID-19 must include detection and isolation of latent infection. A considerable proportion of COVID-19 cases are infected by those who only had mild

symptoms [6, 7]. COVID-19 patients have the highest viral load near symptom presentation [8]. Moreover, the rapid spread of COVID-19 has meant that large numbers of patients with suspicious symptoms are often crowded into fever clinics for diagnosis.

At present, cases are confirmed by a positive result with high-throughput sequencing or real-time reverse-transcriptase polymerase-chain-reaction (RT-PCR) assay of samples from nasal or pharyngeal swabs [9]. However, nucleic acid tests are not available to all suspected patients in pandemic areas due to the shortage of equipment and reagents [10, 11]. Testing for all cases with mild symptoms and/or an epidemiological history can lead to competition for resources. In addition, undiagnosed mild-type COVID-19 patients who were not properly isolated could become sources of infection as their viral load peaks near symptom presentation, which could explain the rapid spread of this epidemic [12]. A large proportion of infected cases continue to test negative for viral RNA, even after they develop clinical manifestations, and positive chest CT (computed tomography) results [13, 14]. This dilemma demands a fast and accurate model for early screening for SARS-CoV-2 infections to prioritize high-risk patients for clinical care, isolation, and contact tracking. Previous studies reported that a number of COVID-19 patients exhibit lymphopenia and thrombocytopenia

[15–17]. Blood counts and high-sensitivity C-reactive protein (hsCRP) are commonly used for early identification of fever [18], and CT is used to assess pneumonia. These tests are simple and fast, and nearly all patients with fever or respiratory symptoms can be tested. We first compared alterations of hematological parameters between cases with and without SARS-CoV-2 infection, then developed and validated a novel score-based prognostic model (COVID-19-REAL) for SARS-CoV-2 infection.

RESULTS

Patient characteristics

A total of 1019 patients were enrolled in this study out of the 1076 patients who presented to fever clinics until 5 February 2020. Fifty-seven patients were excluded, including one with stroke, two with organ transplantation, one with HIV, 12 with cancer, one with active tuberculosis, 18 with age < 12 years, and 22 unconfirmed cases until 10 February 2020 (Figure 1). Of the 1019 patients, 485 (48%) were female, and the median age was 34 years (range 13 to 91 years). The characteristics of the patients are shown in Table 1. All received sequencing or nucleic acid testing using RT-PCR; 103 (10.11%) tested positive for SAR-CoV-2 (Supplementary Table 1).

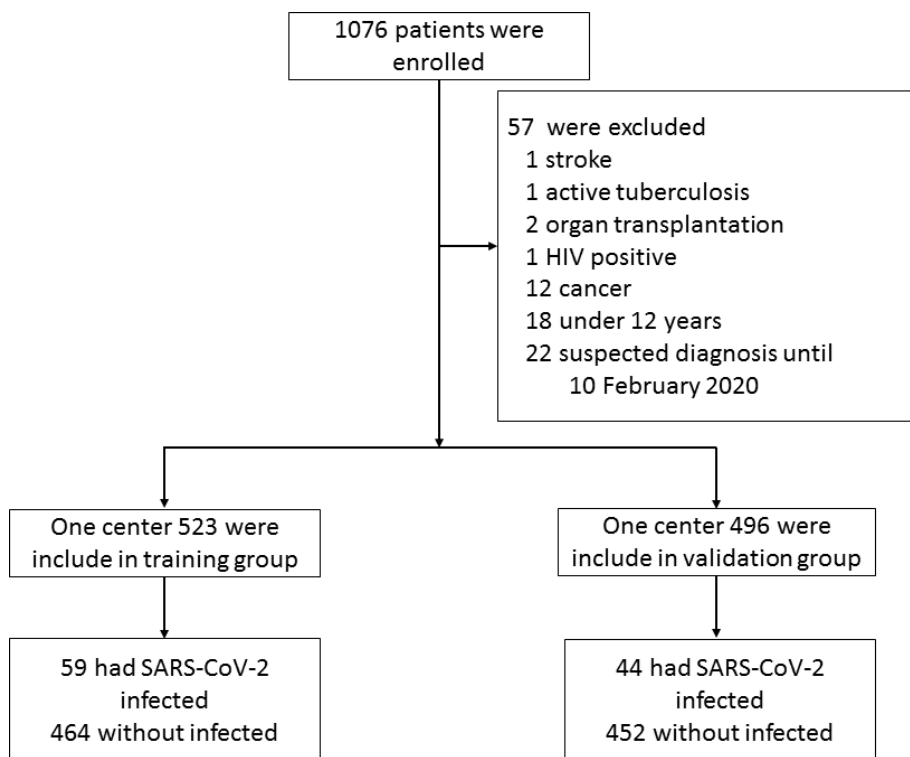


Figure 1. Flowchart of patient selection.

Table 1. Characteristics of patients in this study.

Characteristic	Development group	Validation group	P-value
Number	523	496	
Female	253 (48.38%)	232 (46.77%)	0.609
Age (years)	33 (24-45)	32 (26-40)	0.895
Symptom			
Fever	412 (78.78%)	367 (73.99%)	0.072
Dry cough	209 (39.96%)	171 (34.48%)	0.070
Fatigue	45 (8.60%)	43 (8.669%)	0.970
Pharyngalgia	84 (16.06%)	89 (17.94%)	0.424
Diarrhea	12 (2.29%)	13 (2.62%)	0.736
Coexisting comorbidity			
Hypertension	29 (5.54%)	34 (6.85%)	0.386
Cardiovascular diseases	6 (1.15%)	5 (1.01%)	0.83
Diabetes	11 (2.10%)	7 (1.41%)	0.48
Chronic lung disease	0 (0.00%)	3 (0.60%)	0.115
Chronic liver disease	11 (2.10%)	19 (3.83%)	0.103
Chronic renal disease	1 (0.19%)	2 (0.40%)	0.615
Blood parameters			
Leucocyte (109/L)	6.9 (5.30-8.80)	7.0 (5.20-9.03)	0.74
hsCRP (mg/L)	5.07 (0.90-15.95)	9.10 (2.75-22.56)	<0.001
Monocyte (109/L)	0.50 (0.40-0.70)	0.55 (0.41-0.76)	0.477
RBC (1012/L)	4.78 (4.44-5.22)	4.74 (4.37-5.14)	0.031
Hematocrit (%)	0.42 (0.40-0.46)	0.42 (0.39-0.46)	0.538
Lymphocyte (109/L)	1.30 (0.90-1.80)	1.25 (0.86-1.69)	0.592
MCH (pg)	30.30 (29.30-31.00)	30.30 (29.48-31.20)	0.074
MCHC (g/L)	339.00 (333.00-345.00)	339.00 (332.00-345.00)	0.251
MPV	10.00 (9.60-10.60)	10.00 (9.40-10.60)	0.04
Basophilic granulocyte (109/L)	0.02 (0.01-0.02)	0.02 (0.01-0.03)	<0.001
Eosinophil (109/L)	0.04 (0.01-0.08)	0.03 (0.01-0.09)	0.612
Hemoglobin (g/L)	143 (133-157)	144.00 (132-156)	0.318
PDW (%)	11.70 (10.80-12.85)	11.20 (10.10-12.60)	0.003
Platelet (109/L)	216 (181-256)	212 (173-256)	0.874
Platelet hematocrit (%)	0.22 (0.18-0.25)	0.21 (0.18-0.25)	0.37
Neutrophil (109/L)	4.70 (3.40-6.60)	4.75 (3.30-7.10)	0.7
Radiological evidence of pneumonia	92 (17.59%)	63 (12.70%)	0.03
Confirmed with COVID-19	59 (11.28%)	44 (8.87%)	0.202

Abbreviations: SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; HsCRP: high-sensitivity C-reactive proteins; RBC: red blood cell; MCH: mean corpuscular hemoglobin; MPV: mean platelet volume; MCHC: mean corpuscular hemoglobin concentration; PDW: platelet distribution width; CT: chest computed tomography scan.

Association factors for SARS-CoV-2 infection

The association between age and infection rate is presented in Figure 2A. The rate of SARS-CoV-2 infection increased with age. After stratifying patients by age quartile, the positive rate of SARS-CoV-2 infection from first to fourth quartile was 2.90%, 3.06%, 12.14%, and 23.81% in the training group, and 2.97%, 3.45%, 6.72%, and 23.28% in the validation group (Figure 2B,

C). The risk of infection in last two quartiles was relatively higher than the first two quartiles. The infection rate was lower (less than 5%) for patients with age < 32 years. Subgroup analyses were performed for patients with age ≥ 32 years to stratify those as high-risk population.

The factors associated with a positive result of SARS-CoV-2 infection in univariate analysis are shown in

Table 2. Compared to non-COVID-19 patients, COVID-19 patients had a lower count of leukocytes ($5.10 \times 10^9/L$ vs $7.15 \times 10^9/L$, $p < 0.001$), monocytes ($0.40 \times 10^9/L$ vs $0.55 \times 10^9/L$, $p < 0.001$), lymphocytes ($1.10 \times 10^9/L$ vs $1.30 \times 10^9/L$, $p = 0.02$), eosinophils ($0.01 \times 10^9/L$ vs $0.04 \times 10^9/L$, $p < 0.001$), neutrophils ($3.40 \times 10^9/L$ vs $5.00 \times 10^9/L$, $p < 0.001$), and platelets ($192 \times 10^9/L$ vs $220 \times 10^9/L$, $p < 0.001$). They had a higher age (47 years vs 32 years, $p < 0.001$) in the training group, and similar characteristics were found in validation group (Supplementary Tables 2). After multivariate analysis, age, leukocytes, and eosinophils remained as significant factors; lymphocytes, leukocytes, monocytes, platelets, and neutrophils were not significant indicators (Table 2).

A COVID-19 prediction model based on age, leukocyte, and eosinophil and radiological evidence of pneumonia

The AUROC value for the prediction of leukocytes and eosinophils in the training group for COVID-19 diagnosis were 0.747 and 0.729, respectively. This was comparable to the validation group, where the AUROC value for leukocytes and eosinophils were 0.763 and 0.772 (Supplementary Figure 1). Using Youden's index, the optimal cut-off value for leukocytes and eosinophils were $6.05 \times 10^9/L$ and $0.005 \times 10^9/L$.

Significantly higher infection rate was observed in those with leukocytes $< 6.05 \times 10^9/L$ (23.66% vs 4.45% in leukocytes $\geq 6.05 \times 10^9/L$), and eosinophils $< 0.005 \times 10^9/L$ (33.72% vs 6.68% in eosinophils $\geq 0.005 \times 10^9/L$) in the training group. The trend was

consistent in the validation group, where the infection rate was 18.13% vs 3.5% for leukocyte subgroups, and 28.13% vs 4.25% for eosinophil subgroups (Figure 3).

Based on multivariate logistic regression analysis, the major criterion was age ≥ 32 years (2 point). Minor criteria included leukocytes $< 6.05 \times 10^9/L$ (1 point), eosinophils $< 0.005 \times 10^9/L$ (1 point), and radiological evidence of pneumonia (1 point) (Table 3). The model showed good discrimination (AUROC = 0.863, 95% CI, 0.81 - 0.91) and calibration. Internal verification shows AUROC = 0.863 (95% CI, 0.81 - 0.91) and external verification showed good discrimination (AUROC = 0.871, 95% CI, 0.82-0.93) (Table 4, Supplementary Figure 2)

The following four risk groups were developed: very low risk (0 point), with a risk of infection of 0.84%; low risk (1 - 2 points), with a risk of 3.57%; moderate risk (3 points), with a risk of 19.05%; and high risk (4 - 5 points), with a risk of 61.70%. For the validation group, the infection risk was 0% (0 point); 3.49% (1 - 2 points); 10.87% (3 points); and 55.32% (4 - 5 points) (Figure 4). A cut-off value of less than 3 points for COVID-19-REAL was used to stratify 371 out of 523 (70.94%) cases as low risk, of whom only 10 (2.70%) were infected with SARS-CoV-2 in the training group. The remaining 152 patients were classified as higher risk of infection; about 49 (32.24%) were infected with SARS-CoV-2. According to the cut-off value of 3 points, the sensitivity, specificity, positive predictive value, and negative predictive value was 0.778, 0.831, 0.322, and 0.973 respectively (Table 4).

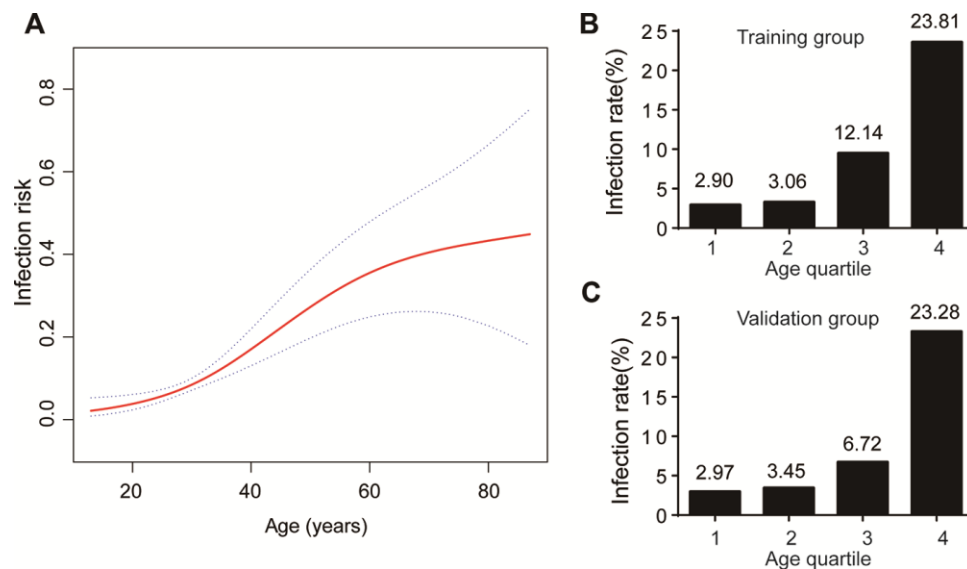


Figure 2. Age and COVID-19 infection. (A) The infection risk increased with increasing age; (B) Infection rate at age quartile in training group; (C) Infection rate at age quartile in validation group.

Table 2. Univariate and multivariate analyses of indicators for SARS-CoV-2 infection in training group.

Variable	non-COVID-19 N = 464	COVID-19 N = 59	Univariate		Multivariate	
			OR (95% CI)	P-value	OR (95% CI)	P-value
Age (years)	32 (23-42)	47 (38-56)	1.05 (1.04- 1.07)	<0.001	1.06 (1.04- 1.08)	<0.001
Leucocyte (109/L)	7.15 (5.70-9.03)	5.10 (4.05-6.05)	0.72 (0.63- 0.83)	<0.001	0.74 (0.64- 0.85)	<0.001
Monocyte (109/L)	0.55 (0.40-0.70)	0.40 (0.30-0.50)	0.06 (0.01- 0.24)	<0.001		
RBC (1012/L)	4.80 (4.45-5.24)	4.70 (4.25-5.01)	0.46 (0.27- 0.78)	0.004		
Lymphocyte (109/L)	1.30 (0.90-1.90)	1.10 (0.85-1.50)	0.57 (0.35- 0.91)	0.019		
Basophilic granulocyte (109/L)	0.02 (0.01-0.03)	0.01 (0.01-0.02)	0.00 (0.00- 45.46)	0.098		
Eosinophil (107/L)	4.00 (1.00-9.00)	1.00 (0.00-3.00)	0.88 (0.82- 0.95)	0.001	0.91 (0.85- 0.98)	0.009
Platelet (109/L)	220.00 (184.00-259.00)	192.00 (144.50-234.00)	0.99 (0.99- 1.00)	<0.001		
Neutrophil (109/L)	5.00 (3.60-6.80)	3.40 (0.80-22.20)	0.75 (0.65- 0.87)	<0.001		
Radiological evidence of pneumonia	68 (14.66%)	24 (40.68%)	3.99 (2.24- 7.13)	<0.001	4.00 (2.04- 7.86)	<0.001

Abbreviations: SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; RBC: red blood cell; HsCRP: high-sensitivity C-reactive proteins; CT: chest computed tomography scan; CI: confidence interval; OR: odds ratio.

DISCUSSION

Beginning in mid-January 2020, a large number of people living in Wuhan left the area via public transportation due to Chinese New Year, leading to a dramatic increase in confirmed or suspected cases nationwide. The management of these suspected cases is of major concern. Nucleic acid testing is currently the main diagnostic method, but the sensitivity and specificity of nucleic acid tests are yet to be verified, and the overall detection rate is constrained by virus concentration and sampling method. Another problem is that some patients with positive chest CT images test negative for COVID-19 by RT-PCR [14]. With such issues in mind, we proposed a robust, high-throughput screening model to help prioritize high-risk patients. We used the data of routine blood tests and CT images to develop a score system (COVID-19-REAL) that can stratify patients into risk groups. Suspected cases with 0 - < 3 points had a predicted probability of 99.16% in training and 97.3% in validation groups for not being infected by SARS-CoV-2. This risk classification can be employed by clinicians and medical institutions, especially those with inadequate detection reagents or equipment, to make rational allocation of resources.

Previous investigations have revealed valuable information about demographics for COVID-19. Most patients with COVID-19 are older [16]. We first stratified patients according to age. Two earlier studies stated the median age of the patients was 56 and 59 years [15, 19]. In our study, the median age was 47 years. We found the risk of infection significantly increased with age, from less than 3% to over 23% from the first to last quartile.

The level of leukocytes, monocytes, lymphocytes, eosinophils, neutrophils, and platelets was dramatically lower in COVID-19 patients. Our results are consistent with previous research that patients exhibited leukopenia, lymphopenia, and thrombocytopenia after SARS-CoV-2 infection [15, 20]. Some researchers suggested a decreased level of white blood cells could serve as an auxiliary diagnosis [20]. Similar patterns emerged in SARS-CoV, with cases of lymphopenia and neutropenia [21, 22], and decreased levels of leukocytes and platelets [23]. A SARS-CoV model showed that neutrophils, lymphocytes, and leukocytes were significantly reduced the day after infection [24]. In a SARS-CoV MA15 infection model, the decrease of peripheral blood cells was explained by inflammatory cell infiltration to the lungs [25]. The N protein of

SARS-CoV enhances eosinophilic infiltration into the lungs and aggravates lung inflammation [26]. Lung lesions were the most important feature of SARS-CoV-2 infections [20], and eosinophilopenia may indicate a poor prognosis of COVID-19 [27]. These results shed light on the neglected role that eosinophils might play in the progression of respiratory disease.

To better stratify SARS-CoV-2 infection risk for the suspected cases, four criteria including leukocytes $< 6.05 \times 10^9 / L$ (1 point), eosinophils $< 0.005 \times 10^9 / L$ (1 point), radiological evidence of pneumonia (1 point), and age ≥ 32 years (2 point) were used to determine the likelihood of SARS-CoV-2 infection. We defined four risk groups: very low risk (0 point), low risk (1 - 2 points), moderate risk (3 points), and high risk (4 - 5 points). According to the cut-off value that was assigned as less than 3 points of COVID-19-REAL score, the number of suspected cases who required priority examination and hospitalization decreased by 70.94% and 71.98%, while maintaining a false negative rate of 2.70% and 2.24% in training and validation group, respectively.

Clinical decision models have been explored to predict infection of SARS-CoV-2. Sun et al. [28] studied 788 cases in Singapore to identify populations at high risk for COVID-19. From their large population-based

study, a model that combined laboratory blood tests, clinical findings, and radiology was proposed, and the AUROC was 0.88 (95% CI: 0.83- 0.93). Similar to our cohort, those authors found that eosinophils and CT imaged pneumonia were strong predictors. However, their conclusions were limited by a lack of external verification, clinical inapplicability caused by redundant parameters, and missing data in laboratory blood tests.

The advantage of present study is that a simple and applicable prediction model, COVID-19-REAL, which combines age, radiological image, and two functionally related hematological indicators (i.e., leukocytes and eosinophils) has been developed to stratify and distinguish between high- and low-risk populations suspected of SARS-CoV-2 infection. This evaluation of suspected cases based on age, radiological image, and two dichotomous criteria could be easily implemented in routine clinical practice. In clinical settings where resources and testing kits are limited, patients with advanced respiratory symptoms are usually tested first. However, those undiagnosed mild-type COVID-19 patients who were not properly isolated would become sources of infection as the viral load peaked near symptom presentation. This score system will be of great help for early infection screening and offer more information for physicians to help prioritize high-risk patients.

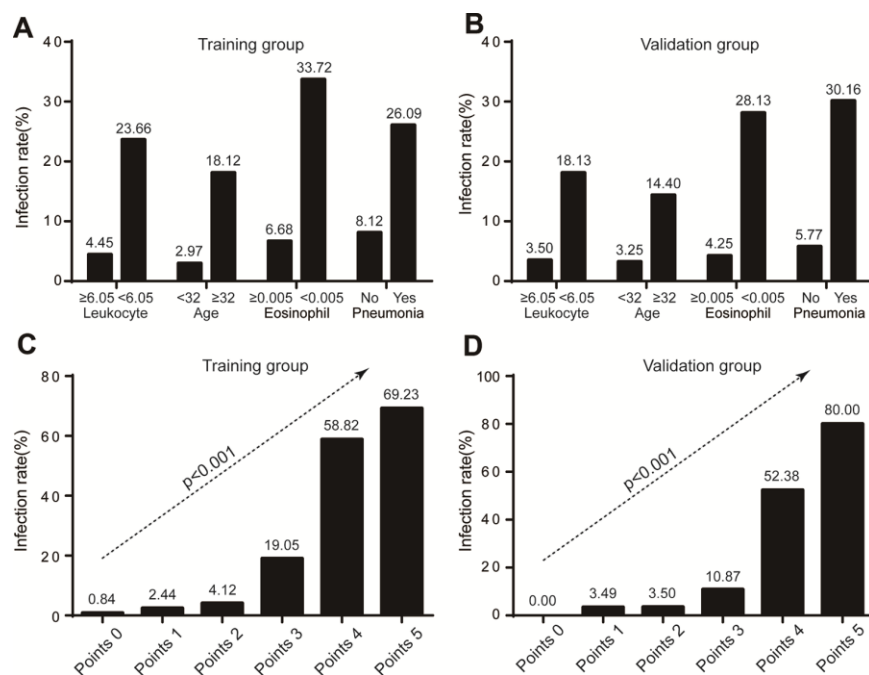


Figure 3. Infection rate in risk stratification. (A) Infection rate stratified by leukocyte, age, eosinophil, and radiological evidence of pneumonia in training group; (B) Infection rate stratified by leukocyte, age, eosinophil, and radiological evidence of pneumonia in validation group; (C) Infection rate according to COVID-19-REAL score in training group; (D) Infection rate according to COVID-19-REAL score in validation group.

Table 3. Multivariate analyses of indicators for SARS-CoV-2 infection in training group.

Variable	OR (95% CI)	β Coefficient (95% CI)	P-value	Point score
Age (years)				
<32 (n = 236)	1	1		
≥ 32 (n = 287)	8.63 (3.60 - 20.64)	2.16 (1.28- 3.03)	<0.001	2
Eosinophil (10⁹/L)				
>0.005 (n = 437)	1	1		
≤ 0.005 (n = 86)	4.92 (2.50 - 9.69)	1.59 (0.94 - 2.27)	<0.001	1
Leucocyte (10⁹/L)				
>6.05 (n =337)	1	1		
≤ 6.05 (n =186)	6.23 (3.14 - 12.35)	1.83 (1.14 - 2.51)	<0.001	1
Radiological evidence				
No Pneumonia(n = 431)	1	1		
Pneumonia(n = 92)	3.73 (1.83 - 7.62)	1.32 (0.60 – 2.03)	<0.001	1

Abbreviations: SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; CT: chest computed tomography scan; CI: confidence interval; OR: odds ratio.

Table 4. Performances of the risk stratification algorithm in the diagnosis of SARS-CoV-2 infection in training and validation groups.

Group	AUROC (95% CI)	Specificity	Sensitivity	Positive PV	Negative PV
Training group	0.863 (0.813-0.912)	0.778	0.831	0.322	0.973
Validation group	0.871 (0.816-0.925)	0.772	0.818	0.259	0.978

Abbreviations: SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; CI: confidence interval; Positive PV: positive predictive value; Negative PV: negative predictive value.

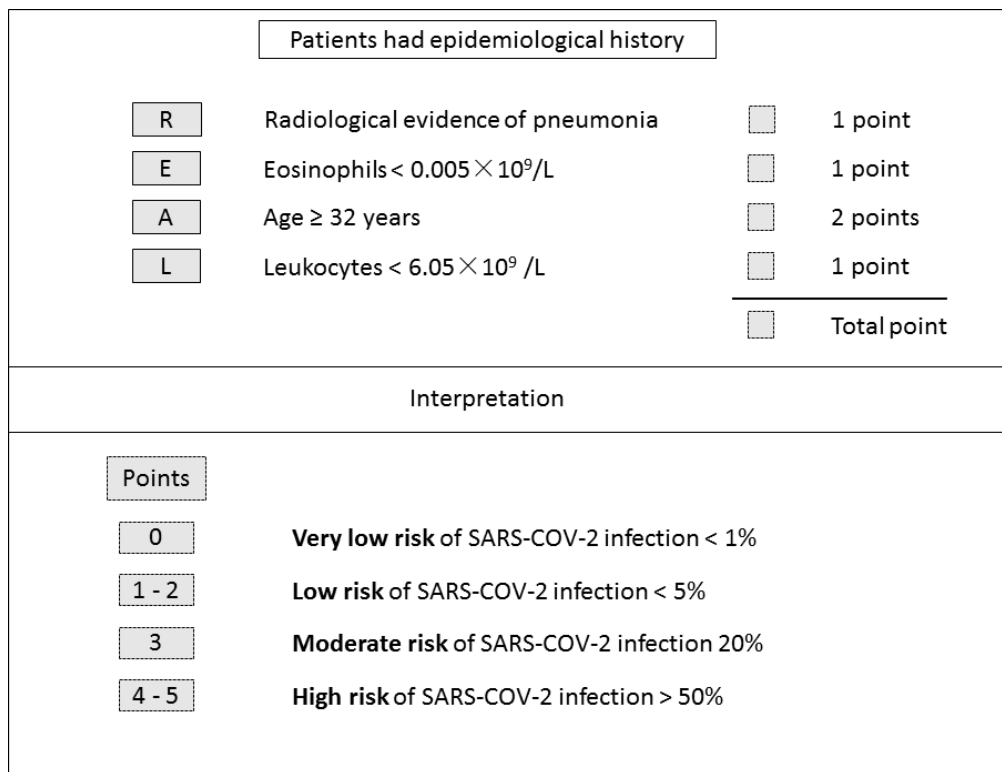


Figure 4. COVID-19-REAL model for risk stratification of SARS-CoV-2 infection.

There are limitations in current study. Our training and validation data comes from China; their applicability to Western populations must be separately evaluated. The results were obtained from people over 12 years of age, and may not be applicable to younger people. Only routine tests including hsCRP, radiological image, and blood cell count were performed, and other hematological indicators including liver and kidney function are lacking.

In conclusion, this study provides a simple, practical, and robust screening model (COVID-19-REAL) to identify high risk populations for SARS-CoV-2 infection. This prediction model will help reduce the burden on hospitals in pandemic areas and help them allocate resources more rationally.

MATERIALS AND METHODS

Patients

Suspect cases of COVID-19 with age ≥ 13 years with an epidemiological history were included from fever clinics of the First Affiliated Hospital, College of Medicine, Zhejiang University and Taizhou Enze Medical Center (Group), Enze Hospital, between 23 January 2020 and 5 February 2020. All suspected cases received sequencing or RT-PCR assay for SARS-CoV-2. According to National Health Commission, an epidemiological history of COVID-19 is defined as follows: within 14 days before the onset of the disease (1) there were tourism or residence histories of Wuhan or its surrounding areas, or other communities with confirmed cases; (2) there were contacts with confirmed cases of COVID-19; (3) there were contacts with suspected cases (having fever or respiratory symptoms) from Wuhan or its surrounding areas, or other communities with confirmed cases; (4) one confirmed case was found in an enclosed environment (such as a family house, a construction site, an office, etc.), with one or more cases of fever/respiratory tract infection re found at the same time

The patient-selection process is shown in Figure 1. The COVID-19 cases were all confirmed by sequencing or RT-PCR assay [9]. The RT-PCR was mainly performed using a commercial kit for SARS-CoV-2 detection (BoJie, Shanghai, China) which was approved by China Food and Drug Administration. We excluded patients with HIV infection, cancer, organ transplantation, stroke, active tuberculosis, severe and critical COVID-19 patients according to the National Health Commission [17], and suspected cases without confirmed laboratory evidence until 10 February 2020. The study was approved by the Ethics Committee of the First

Affiliated Hospital, College of Medicine, Zhejiang University, and complied with the ethical guidelines of the Declaration of Helsinki. The researchers only analyzed anonymous data, so informed consent was waived. Age, gender, laboratory assessments consisting of hsCRP, complete blood count, and radiological images were obtained from electronic medical records. Radiological evidence of pneumonia was defined as lung consolidation and/or ground-glass opacity [20]. The images were reviewed independently by two radiologists, and if there were disagreements, a third radiologist would perform further examination.

Statistical analysis

Continuous variables were expressed as medians and interquartile range (IQR), and were compared by t-test or Mann–Whitney U-test. Chi-squared test or Fisher's exact test was used to compare categorical variables and expressed as percentages. Generalized linear models with a logit link were used to test the association between age and the risk of COVID-19 infection. Univariate and multivariate analyses were performed to identify indicators of COVID-19 patients. Variables with $P < 0.1$ in a univariate analysis were then included in a forward stepwise regression model. A score for the final model was developed by rounding the coefficients of the logit model. Predicted and observed risk was calculated for each score. The area under receiver operating characteristic (AUROC) curve was used to assess the accuracy of different scores in diagnosis power. Internal validation was performed using a bootstrap procedure with 500 bootstrapped samples. The Youden's index was used to determine the optimal cut-off level for predicting clinical outcomes. All statistical analysis was performed by Statistical Package for the Social Sciences version 19.0 (International Business Machines Corporation, Armonk, NY) and R version 3.4 (R Foundation, Vienna, Austria). All tests were two tailed and $P < 0.05$ was considered to indicate statistical significance.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (No. 2016YFA 0101001).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, et al, and

- China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* 2020; 382:727–33.
<https://doi.org/10.1056/NEJMoa2001017>
PMID:[31978945](https://pubmed.ncbi.nlm.nih.gov/31978945/)
2. Cabrini L, Landoni G, Zangrillo A. Minimise nosocomial spread of 2019-nCoV when treating acute respiratory failure. *Lancet.* 2020; 395:685.
[https://doi.org/10.1016/S0140-6736\(20\)30359-7](https://doi.org/10.1016/S0140-6736(20)30359-7)
PMID:[32059800](https://pubmed.ncbi.nlm.nih.gov/32059800/)
 3. Patel A, Jernigan DB, and 2019-nCoV CDC Response Team. Initial public health response and interim clinical guidance for the 2019 novel coronavirus outbreak — United States, December 31, 2019–February 4, 2020. *MMWR Morb Mortal Wkly Rep.* 2020; 69:140–46.
<https://doi.org/10.15585/mmwr.mm6905e1>
PMID:[32027631](https://pubmed.ncbi.nlm.nih.gov/32027631/)
 4. Lu CW, Liu XF, Jia ZF. 2019-nCoV transmission through the ocular surface must not be ignored. *Lancet.* 2020; 395:e39.
[https://doi.org/10.1016/S0140-6736\(20\)30313-5](https://doi.org/10.1016/S0140-6736(20)30313-5)
PMID:[32035510](https://pubmed.ncbi.nlm.nih.gov/32035510/)
 5. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet.* 2020; 395:689–97.
[https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
PMID:[32014114](https://pubmed.ncbi.nlm.nih.gov/32014114/)
 6. Qiu J. Covert coronavirus infections could be seeding new outbreaks. *Nature.* 2020. [Epub ahead of print].
<https://doi.org/10.1038/d41586-020-00822-x>
PMID:[32203376](https://pubmed.ncbi.nlm.nih.gov/32203376/)
 7. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science.* 2020; 368:489–93.
<https://doi.org/10.1126/science.abb3221>
PMID:[32179701](https://pubmed.ncbi.nlm.nih.gov/32179701/)
 8. To KK, Tsang OT, Leung WS, Tam AR, Wu TC, Lung DC, Yip CC, Cai JP, Chan JM, Chik TS, Lau DP, Choi CY, Chen LL, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect Dis.* 2020; 20:565–74.
[https://doi.org/10.1016/S1473-3099\(20\)30196-1](https://doi.org/10.1016/S1473-3099(20)30196-1)
PMID:[32213337](https://pubmed.ncbi.nlm.nih.gov/32213337/)
 9. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020; 395:497–506.
[https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
PMID:[31986264](https://pubmed.ncbi.nlm.nih.gov/31986264/)
 10. Pang J, Wang MX, Ang IY, Tan SH, Lewis RF, Chen JI, Gutierrez RA, Gwee SX, Chua PE, Yang Q, Ng XY, Yap RK, Tan HY, et al. Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): a systematic review. *J Clin Med.* 2020; 9:623.
<https://doi.org/10.3390/jcm9030623>
PMID:[32110875](https://pubmed.ncbi.nlm.nih.gov/32110875/)
 11. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA.* 2020. [Epub ahead of print].
<https://doi.org/10.1001/jama.2020.4683>
PMID:[32203977](https://pubmed.ncbi.nlm.nih.gov/32203977/)
 12. Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q, Xie G, Lin S, Wang R, Yang X, Chen W, Wang Q, Zhang D, et al. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January–March 2020: retrospective cohort study. *BMJ.* 2020; 369:m1443.
<https://doi.org/10.1136/bmj.m1443>
PMID:[32317267](https://pubmed.ncbi.nlm.nih.gov/32317267/)
 13. Xiao SY, Wu Y, Liu H. Evolving status of the 2019 novel coronavirus infection: Proposal of conventional serologic assays for disease diagnosis and infection monitoring. *J Med Virol.* 2020; 92:464–467.
<https://doi.org/10.1002/jmv.25702>
PMID:[32031264](https://pubmed.ncbi.nlm.nih.gov/32031264/)
 14. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. *Radiology.* 2020; 296:E41-45.
<https://doi.org/10.1148/radiol.2020200343>
PMID:[32049601](https://pubmed.ncbi.nlm.nih.gov/32049601/)
 15. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA.* 2020; 323:1061–69.
<https://doi.org/10.1001/jama.2020.1585>
PMID:[32031570](https://pubmed.ncbi.nlm.nih.gov/32031570/)
 16. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet.* 2020; 395:507–13.
[https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)
PMID:[32007143](https://pubmed.ncbi.nlm.nih.gov/32007143/)
 17. Wu J, Liu J, Zhao X, Liu C, Wang W, Wang D, Xu W, Zhang C, Yu J, Jiang B, Cao H, Li L. Clinical characteristics of imported cases of COVID-19 in Jiangsu province: a multicenter descriptive study. *Clin Infect Dis.* 2020. [Epub ahead of print].

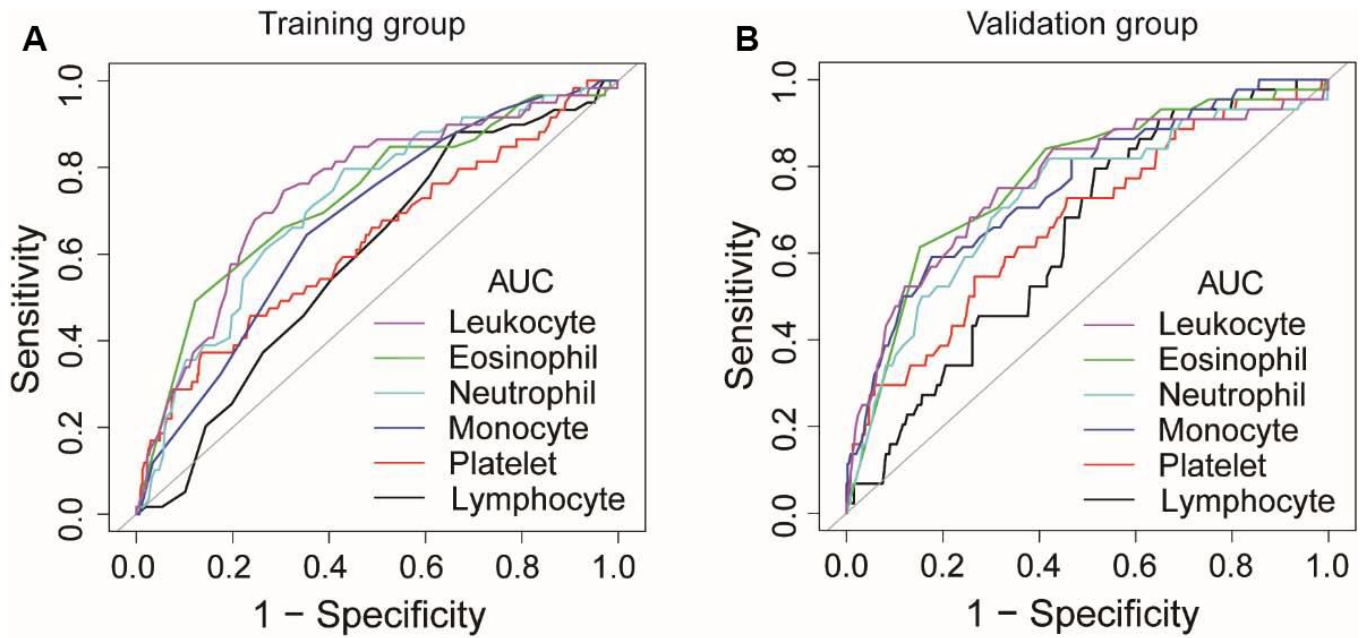
<https://doi.org/10.1093/cid/ciaa199>

PMID:[32109279](https://pubmed.ncbi.nlm.nih.gov/32109279/)

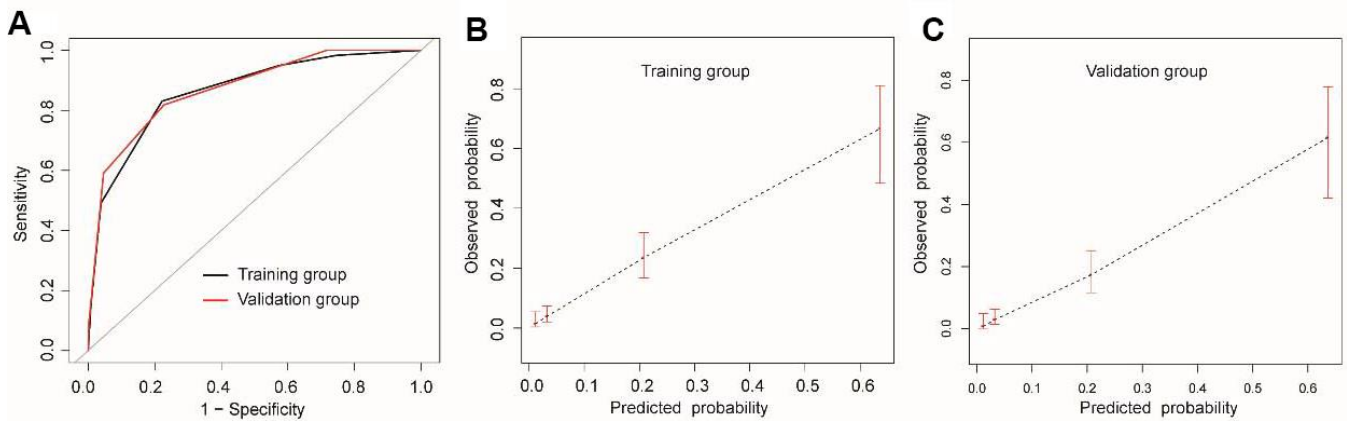
18. Yo CH, Hsieh PS, Lee SH, Wu JY, Chang SS, Tasi KC, Lee CC. Comparison of the test characteristics of procalcitonin to c-reactive protein and leukocytosis for the detection of serious bacterial infections in children presenting with fever without source: a systematic review and meta-analysis. *Ann Emerg Med*. 2012; 60:591–600.
<https://doi.org/10.1016/j.annemergmed.2012.05.027>
PMID:[22921165](https://pubmed.ncbi.nlm.nih.gov/22921165/)
19. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY, Xing X, Xiang N, Wu Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*. 2020; 382:1199–207.
<https://doi.org/10.1056/NEJMoa2001316>
PMID:[31995857](https://pubmed.ncbi.nlm.nih.gov/31995857/)
20. Song F, Shi N, Shan F, Zhang Z, Shen J, Lu H, Ling Y, Jiang Y, Shi Y. Emerging 2019 novel coronavirus (2019-nCoV) pneumonia. *Radiology*. 2020; 295:210–17.
<https://doi.org/10.1148/radiol.2020200274>
PMID:[32027573](https://pubmed.ncbi.nlm.nih.gov/32027573/)
21. Liu J, Zheng X, Tong Q, Li W, Wang B, Sutter K, Trilling M, Lu M, Dittmer U, Yang D. Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV. *J Med Virol*. 2020; 92:491–94.
<https://doi.org/10.1002/jmv.25709>
PMID:[32056249](https://pubmed.ncbi.nlm.nih.gov/32056249/)
22. Li YH, Hu CY, Wu NP, Yao HP, Li LJ. Molecular characteristics, functions, and related pathogenicity of MERS-CoV proteins. *Engineering (Beijing)*. 2019; 5:940–47.
<https://doi.org/10.1016/j.eng.2018.11.035>
PMID:[32288963](https://pubmed.ncbi.nlm.nih.gov/32288963/)
23. Ahmed AE, Al-Jahdali H, Alshukairi AN, Alaqeel M, Siddiq SS, Alsaab H, Sakr EA, Alyahya HA, Alandonisi MM, Subedar AT, Aloudah NM, Baharoon S, Alsalamah MA, et al. Early identification of pneumonia patients at increased risk of middle east respiratory syndrome coronavirus infection in Saudi Arabia. *Int J Infect Dis*. 2018; 70:51–56.
<https://doi.org/10.1016/j.ijid.2018.03.005>
PMID:[29550445](https://pubmed.ncbi.nlm.nih.gov/29550445/)
24. Chu YK, Ali GD, Jia F, Li Q, Kelvin D, Couch RC, Harrod KS, Hutt JA, Cameron C, Weiss SR, Jonsson CB. The SARS-CoV ferret model in an infection-challenge study. *Virology*. 2008; 374:151–63.
<https://doi.org/10.1016/j.virol.2007.12.032>
PMID:[18234270](https://pubmed.ncbi.nlm.nih.gov/18234270/)
25. Gralinski LE, Sheahan TP, Morrison TE, Menachery VD, Jensen K, Leist SR, Whitmore A, Heise MT, Baric RS. Complement activation contributes to severe acute respiratory syndrome coronavirus pathogenesis. *mBio*. 2018; 9:e01753–18.
<https://doi.org/10.1128/mBio.01753-18>
PMID:[30301856](https://pubmed.ncbi.nlm.nih.gov/30301856/)
26. Yasui F, Kai C, Kitabatake M, Inoue S, Yoneda M, Yokochi S, Kase R, Sekiguchi S, Morita K, Hishima T, Suzuki H, Karamatsu K, Yasutomi Y, et al. Prior immunization with severe acute respiratory syndrome (SARS)-associated coronavirus (SARS-CoV) nucleocapsid protein causes severe pneumonia in mice infected with SARS-CoV. *J Immunol*. 2008; 181:6337–48.
<https://doi.org/10.4049/jimmunol.181.9.6337>
PMID:[18941225](https://pubmed.ncbi.nlm.nih.gov/18941225/)
27. Du Y, Tu L, Zhu P, Mu M, Wang R, Yang P, Wang X, Hu C, Ping R, Hu P, Li T, Cao F, Chang C, et al. Clinical features of 85 fatal cases of COVID-19 from Wuhan. A retrospective observational study. *Am J Respir Crit Care Med*. 2020; 201:1372–79.
<https://doi.org/10.1164/rccm.202003-0543OC>
PMID:[32242738](https://pubmed.ncbi.nlm.nih.gov/32242738/)
28. Sun Y, Koh V, Marimuthu K, Ng OT, Young B, Vasoo S, Chan M, Lee VJ, De PP, Barkham T, Lin RT, Cook AR, Leo YS. Epidemiological and clinical predictors of COVID-19. *Clin Infect Dis*. 2020. [Epub ahead of print].
<https://doi.org/10.1093/cid/ciaa322>
PMID:[32211755](https://pubmed.ncbi.nlm.nih.gov/32211755/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. AUROC of leukocyte, monocyte, lymphocyte, eosinophil, neutrophil and platelets in COVID-19 diagnosis. (A) training group; (B) validation group.



Supplementary Figure 2. AUROC of model in COVID-19 diagnosis (A), and Calibration chart for predicted versus observed probability (B, C) in training and validation group.

Supplementary Tables

Supplementary Table 1. Characteristics of patients with SARS-CoV-2 infected.

Characteristic	Development group	Validation group	P-value
Number	59	44	
Female	24 (40.68%)	15 (34.09%)	0.495
Age (years)	47 (38-56)	48 (35-57)	0.812
Symptom			
Fever	37 (62.71%)	32 (72.73%)	0.285
Dry cough	24 (40.68%)	18 (40.91%)	0.981
Fatigue	6 (10.17%)	4 (9.09%)	0.855
Pharyngalgia	14 (23.73%)	8 (18.18%)	0.497
Blood parameters			
Leucocyte (109/L)	5.10 (4.05-6.05)	4.50 (3.63-6.03)	0.738
hsCRP (mg/L)	10.17 (2.62-21.88)	14.80 (5.35-30.10)	0.043
Monocyte (109/L)	0.40 (0.30-0.50)	0.35 (0.27-0.52)	0.224
RBC (1012/L)	4.70 (4.25-5.01)	4.69 (4.22-5.01)	0.88
Hematocrit (%)	0.42 (0.38-0.45)	0.42 (0.38-0.44)	0.668
Lymphocyte (109/L)	1.10 (0.85-1.50)	1.10 (0.73-1.30)	0.134
MCH (pg)	30.60 (29.65-31.30)	30.15 (29.30-31.33)	0.439
MCHC (g/L)	341.00 (334.00-346.50)	341.00 (334.50-348.00)	0.939
MPV	10.40 (9.90-10.85)	10.35 (9.95-11.00)	0.707
Basophilicgranulocyte (109/L)	0.01 (0.01-0.02)	0.01 (0.00-0.01)	0.066
Eosinophil (109/L)	0.01 (0.00-0.03)	0.00 (0.00-0.02)	0.757
Hemoglobin (g/L)	144.00 (129.00-153.00)	143.00 (129.00-152.00)	0.641
PDW (%)	12.00 (11.20-13.10)	11.90 (10.73-13.03)	0.403
Platelet (109/L)	192.00 (144.50-234.00)	177.00 (140.00-226.00)	0.4
Platelet hematocrit (%)	0.20 (0.15-0.23)	0.18 (0.15-0.24)	0.453
Neutrophil (109/L)	3.40 (2.60-4.45)	3.00 (2.18-4.15)	0.981
Radiological evidence of pneumonia	24 (40.68%)	19 (43.18%)	0.799

Abbreviations: HsCRP: high-sensitivity C-reactive proteins; RBC: Red Blood Cell; MCH: mean corpuscular hemoglobin; MPV: mean platelet volume; MCHC: mean corpuscular hemoglobin concentration; PDW: Platelet distribution width; CT: chest computed tomography scan.

Supplementary Table 2. Characteristics of patients visited fever clinics.

Characteristic	Non-COVID-19 infected	COVID-19 infected	P-value
Number	452	44	
Female	217 (48.01%)	15 (34.09%)	0.077
Age (years)	31 (25-38)	48 (35-57)	<0.001
Symptom			
Fever	335 (74.12%)	32 (72.73%)	0.841
Dry cough	153 (33.85%)	18 (40.91%)	0.347
Fatigue	39 (8.63%)	4 (9.09%)	0.917
Pharyngalgia	81 (17.92%)	8 (18.18%)	0.966
Blood parameters			
Leucocyte (10 ⁹ /L)	7.20 (5.50-9.50)	4.50 (3.63-6.03)	<0.001
hsCRP (mg/L)	8.80 (2.40-22.22)	14.80 (5.35-30.10)	0.547
Monocyte (10 ⁹ /L)	0.57 (0.43-0.78)	0.35 (0.27-0.52)	<0.001
RBC (10 ¹² /L)	4.74 (4.38-5.14)	4.69 (4.22-5.01)	0.068
Hematocrit (%)	0.42 (0.40-0.46)	0.42 (0.38-0.44)	0.042
Lymphocyte (10 ⁹ /L)	1.29 (0.87-1.76)	1.10 (0.73-1.30)	0.004
MCH (pg)	30.30 (29.50-31.20)	30.15 (29.30-31.33)	0.442
MCHC (g/L)	339.00 (332.00-344.00)	341.00 (334.50-348.00)	0.046
MPV	9.95 (9.30-10.60)	10.35 (9.95-11.00)	<0.001
Basophilicgranulocyte (10 ⁹ /L)	0.02 (0.01-0.03)	0.01 (0.00-0.01)	<0.001
Eosinophil (10 ⁹ /L)	0.03 (0.01-0.09)	0.00 (0.00-0.02)	0.002
Hemoglobin (g/L)	144.00 (132.00-156.00)	143.00 (129.00-152.00)	0.177
PDW (%)	11.20 (10.10-12.45)	11.90 (10.73-13.03)	0.15
Platelet (10 ⁹ /L)	214.00 (176.00-260.00)	177.00 (139.75-226.00)	<0.001
Platelet hematocrit (%)	0.22 (0.18-0.25)	0.18 (0.15-0.24)	0.002
Neutrophil (10 ⁹ /L)	4.90 (3.50-7.22)	3.00 (2.18-4.15)	<0.001
Radiological evidence of pneumonia	44 (9.73%)	19 (43.18%)	<0.001

Abbreviations: HsCRP: high-sensitivity C-reactive proteins; RBC: Red Blood Cell; MCH: mean corpuscular hemoglobin; MPV: Mean platelet volume; MCHC: mean corpuscular hemoglobin concentration; PDW: Platelet distribution width; CT: chest computed tomography scan.