

Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results

Jiangpeng Wu^{1,2}, Pengyi Zhang^{1,2}, Liting Zhang³, Wenbo Meng⁴, Junfeng Li³, Chongxiang Tong⁵, Yonghong Li⁶, Jing Cai⁵, Zengwei Yang⁵, Jinhong Zhu⁷, Meie Zhao⁸, Huirong Huang⁹, Xiaodong Xie^{10*}, Shuyan Li^{1,2*}

¹ State Key Laboratory of Applied Organic Chemistry, Lanzhou University, Lanzhou 730000, China

² College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou 730000, China

³ Department of Infectious Disease, the First Hospital of Lanzhou University, Lanzhou, 730020, China

⁴ Department of Special Minimally Invasive Surgery, the First Hospital of Lanzhou University, Lanzhou, 730020, China

⁵ The Pulmonary Hospital of Lanzhou, Lanzhou, 730050, China

⁶ NHC Key Laboratory of Diagnosis and Therapy of Gastrointestinal Tumor, Gansu Provincial Hospital, Lanzhou, 730020, China

⁷ Clinical laboratory, the First Hospital of Lanzhou University, Lanzhou 730020, China

⁸ Clinical laboratory, the First People's Hospital of Lanzhou City, Lanzhou, 730050, China

⁹ Department of Respiratory Medicine, the Second Hospital of Lanzhou University, Lanzhou, 730030, China

¹⁰ Key Laboratory of Preclinical Study for New Drug of Gansu Province, School of Basic Medical Sciences, Lanzhou University, Lanzhou, 730020, China

Correspondence to:

Dr Shuyan Li, State Key Laboratory of Applied Organic Chemistry, Lanzhou University, Lanzhou 730000, China lishuyan@lzu.edu.cn

Dr Xiaodong Xie, Key Laboratory of Preclinical Study for New Drug of Gansu Province, School of Basic Medical Sciences, Lanzhou University, Lanzhou, 730000, China xdxie@lzu.edu.cn

Abstract

Since the sudden outbreak of coronavirus disease 2019 (COVID-19), it has rapidly evolved into a momentous global health concern. Due to the lack of constructive information on the pathogenesis of COVID-19 and specific treatment, it highlights the importance of early diagnosis and timely treatment. In this study, 11 key blood indices were extracted through random forest algorithm to build the final assistant discrimination tool from 49 clinical available blood test data which were derived by commercial blood test equipments. The method presented robust outcome to accurately identify COVID-19 from a variety of suspected patients with similar CT information or similar symptoms, with accuracy of 0.9795 and 0.9697 for the cross-validation set and test set, respectively. The tool also demonstrated its outstanding performance on an external validation set that was completely independent of the modeling process, with sensitivity, specificity, and overall accuracy of 0.9512, 0.9697, and 0.9595, respectively. Besides, 24 samples from overseas infected patients with COVID-19 were used to make an in-depth clinical assessment with accuracy of 0.9167. After multiple verification, the reliability and repeatability of the tool has been fully evaluated, and it has the potential to develop

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

proposed tool is well-suited to carry out preliminary assessment of suspected patients and help them to get timely treatment and quarantine suggestion. The assistant tool is now available online at http://lishuyan.lzu.edu.cn/COVID2019_2/.

Funding This work was supported by the Fundamental Research Funds for the Central Universities (lzujbky-2020-sp11) and the Gansu Provincial COVID-19 Science and Technology Major Project, China.

Introduction

Coronavirus disease 2019 (COVID-19)¹ can induce emerging acute respiratory illness by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which is classified by International Committee on Taxonomy of Viruses (ICTV).² This enveloped RNA virus is the seventh member of the coronavirus family to infect humans belonging to lineage B of the genus betacoronavirus through deep sequencing and phylogenetic analysis.³ The epidemic pneumonia has spread quickly and become a global public health concern.⁴ After making careful assessment for COVID-19, the epidemic is classified as a pandemic due to the alarming spread of the infection and the severity of the disease in March 11, 2020.⁵ A large amount of practice and research show that the common clinical symptoms of the pneumonia are fever, cough, myalgia or fatigue.⁶ And the characteristics of chest radiography or computed tomography (CT) are multiple mottling or ground-glass opacity.⁷

According to these valuable experience and achievements, the seventh edition of a new coronavirus pneumonia diagnosis and treatment plan has jointly updated by two Chinese departments. The official guidelines emphasize that the use of quantitative real-time PCR (qRT-PCR) assays for the detection of viral nucleic acid to diagnose the infected patients is still one of the main reference standards.⁸ Although RNA of this coronavirus is the prime marker that can be detected in the early stage of infection, the high requirements of its testing technology and condition limit its positive rate to as about 30-50%.⁹ This situation may imply that many patients with COVID-19 cannot be identified in time.¹⁰ The situation may delay the isolation and treatment of infected patients and accelerate the spread of the epidemic due to the characteristic of human-to-human transmission.¹¹ Apart from this, some objective factors, such as complicated operation, time-consuming, easy to pollute, high cost and so on, further hinder the technology from exerting its due advantages on emergency prevention and control conditions.¹² Many clinicians strongly recommended the use of CT imaging in severe epidemic areas, however, the tardy method would present obvious abnormalities when the lung produced inflammatory or the tissue generated lesions.¹³ At present, the combined detection of immunoglobulin M (IgM) and immunoglobulin G (IgG) antibody provides a fast and simple screening method, but the two substances are impossible to be detected in the window period of about two weeks.^{14,15} So this method can only be used as an effective supplement and auxiliary diagnosis of nucleic acid detection.¹⁶

Until now, the quick diagnosis of COVID-19 patients in a large scale is still a major problem for global health concern.¹⁷ Except for the laboratory-confirmed COVID-19 patients, there are a large number of suspected patients and close contacts who need to be confirmed whether they are infected or not.¹⁸ In this case, it is difficult to screen and identify the cases during the incubation period or subclinical infections, they have great potential to become super spreaders.¹⁹ Although timely isolation measures can effectively cut down the spread of the virus, early diagnosis can enable patients to get more active treatment at the initial stage of presentation.²⁰ Therefore, the early identification and timely management of COVID-19 patients is of great importance due to no specific treatment.²¹

The most common laboratory findings, including hematological and biochemical parameters, play a prominent part in the initial screening for COVID-19.²² In the present work, we use machine learning algorithm to process these valuable parameters in order to realize the progress of accurate and rapid identification for the emerging pneumonia. The proposed assistant discrimination tool is more convenient and practical than nucleic acid detection and antibody detection, which can accelerate the recognition of suspected patients and thus reduce the risk of transmission. Moreover, the deep relationship between these important reference targets and COVID-19 could help to monitor the progress of the disease and to guide useful strategies for clinicians. Finally, we aim to design a simple and effective pre-warning system to help proactively initial screening for suspected patients which will contribute to the outbreak control and public health.

Materials and Methods

Source of materials

In this study, a total of 253 samples from 169 suspected patients were collected from multiple sources, including Lanzhou Pulmonary Hospital, the First Hospital of Lanzhou University, Lanzhou University Second Hospital, the First People's Hospital of Lanzhou City, and Gansu Provincial Hospital. Data from routine laboratory assessments with 49 parameters for each sample included 24 routine hematological parameters and 25 biochemical parameters. General distribution information about the datasets can be listed in Table 1 and the detailed information about the patients, including sex, age, and 49 parameters are listed in Table S1. 105 consecutive samples from 27 patients with confirmed COVID-19 admitted to Lanzhou Pulmonary Hospital in Gansu Province were considered as positive samples and randomly divided into training set, test set, and external validation set. The virus nucleic acid assays (qRT-PCR) of throat swab and sputum samples were detected for all these patients to make a definite diagnosis. The diagnostic evidence for COVID-19 is based on World Health Organization interim guidance. Except for the COVID-19 samples, the remaining samples, which were collected from patients with similar symptoms or similar radiologic characteristic with COVID-19, including patients with common pneumonia, tuberculosis and lung cancer, were treated as negative samples. These patients were diagnosed by at least two professional experts based on comprehensive examination results. Among them, pneumonia samples were collected from multiple sources. The pneumonia samples in training set and test set were collected from the First Hospital of Lanzhou University, the First People's Hospital of Lanzhou City, and Gansu Provincial Hospital. The pneumonia samples in external validation set were from Lanzhou University Second Hospital. The tuberculosis and lung cancer samples were also collected from Lanzhou University Second Hospital. Clinical data collection from patients was approved by Ethics Committee of the First Hospital of Lanzhou University which is the leading department for this project.

Table 1. Distribution and baseline characteristics of the whole data

	Training set	Test set	External validation set	Total number	Median age (IQR)	Male/Female
COVID-19	50 (12)*	15 (5)*	40 (10)*	105 (27)*	47 (33-64)	13/14
Common pneumonia	76	8	14 (8)*	98 (92)*	63 (47-75)	44/48
Tuberculosis	10	5	10	25	54 (33-68)	15/10
Lung cancer	10	5	10	25	61 (54-69)	19/6

*Note: The number inside the bracket is the total number of patients where all the serial samples were collected

Machine learning method

Random forest algorithm (RF) is a promising and well-known classifier with praiseworthy characteristics that can use multiple trees to train and predict samples, which has received extensive attention in the fields of chemometrics and bioinformatics.^{23,24} The algorithm in this study was used to establish the final assistant discrimination tool to proactively seek out patients with COVID-19. The whole process could be divided into three stepwise stages. In the first stage, all the parameters were fully employed to build the model in order to evaluate the importance of each parameter. In the second stage, while the top-ranking parameters were increased one by one, different models were built under the adjustment of RF parameters including the tree number (ntree) and the number of randomly selected features to split at each node (mtry). The adjustment range of ntree ranged from 500 to 1500 with a step of 100; Once the value of ntree was fixed, mtry spanned from 2 to 15 via a step of 1. In the last stage, based on the principle that the minimum number of top-ranking parameters could have comparable performance with the whole parameters, the appropriate number of parameters, ntree, and mtry were selected respectively. The study was executed on R package randomForest v4 6-7.

Validation method

In this work, several different validation methods are performed to ensure the reliability and reproducibility of this COVID-19 assistant identification tool.

Firstly, the internal 10-fold cross-validation which was performed on the training set was a persuasive and universal method to obtain robust performance. Secondly, the test set is totally different from the training set just to verify the discriminating ability of the tool. In addition, all modeling processes were accomplished in the training set based on 10-fold cross-validation and had no connection to the test set. Finally, after building the model with the selected optimal parameters, the external validation set was used to reconfirm the accuracy and repeatability of the tool.

There are some popular metrics to comprehensively assess the performance of the assistant tool including sensitivity (Sens), specificity (Spec), total accuracy (ACC), Matthews correlation coefficient (MCC), and area under the curve (AUC).

Results and discussion

Modeling results

It was highly important to select appropriate number of parameters for the foundation of the final discrimination tool with the advantages of convenient operation and accurate diagnosis. Based on the minimum number of top-ranking parameters without compromising performance, 11 laboratory findings were carefully selected as the final input indicators as shown in Figure 1. The final model revealed good prediction performance on the training set with ACC, MCC, and related AUC values of 0.9795, 0.9594, and 1.000, respectively. These important metrics implied that the tool integrated with 11 top-ranking parameters had the significant ability to discriminate patients with COVID-19. At the same time, the improvement of sensitivity (0.9750) and specificity (0.9909) heralded that this method had great potential as a practical clinical tool for large-scale initial screening. Although 34 parameters seemed like a sensible choice, the redundant ones didn't play an irreplaceable role in sharply improving the performance, which might make the model over-fitting and prolong the diagnosis process. So the selected parameters were of extraordinary benefit to build a cost-effective and rapid assistant tool which had commensurate ability compared with the whole parameters.

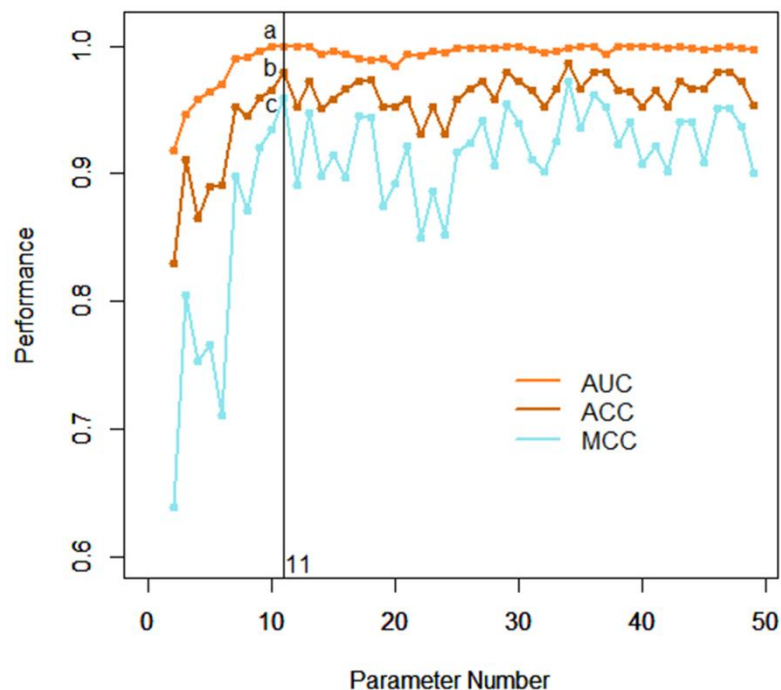


Figure 1: Performance of the various models

With the gradual increase of top-ranking parameters, the performance of different models is obtained based on the 10-fold cross-validation results in the training set (a=1 000; b=0 9795; c=0 9594).

The above data raised the hope that the tool, which combined machine learning algorithm and laboratory parameters, would play a prominent part in screening patients with COVID-19. The test set was completely independent of the entire modeling process and was adopted to prove the stability and reliability of this method again. As shown in Figure 2, the performance of the method in the test set was consistent with the training set with an AUC of 0 9926, achieving a sensitivity of 1.0000 and a specificity of 0 9444. It confirmed that the discrimination tool had the ability to deal with a large number of suspected patients with COVID-19, which might greatly simplify the laboratory blood test process and provide timely treatment for the confirmed patients.

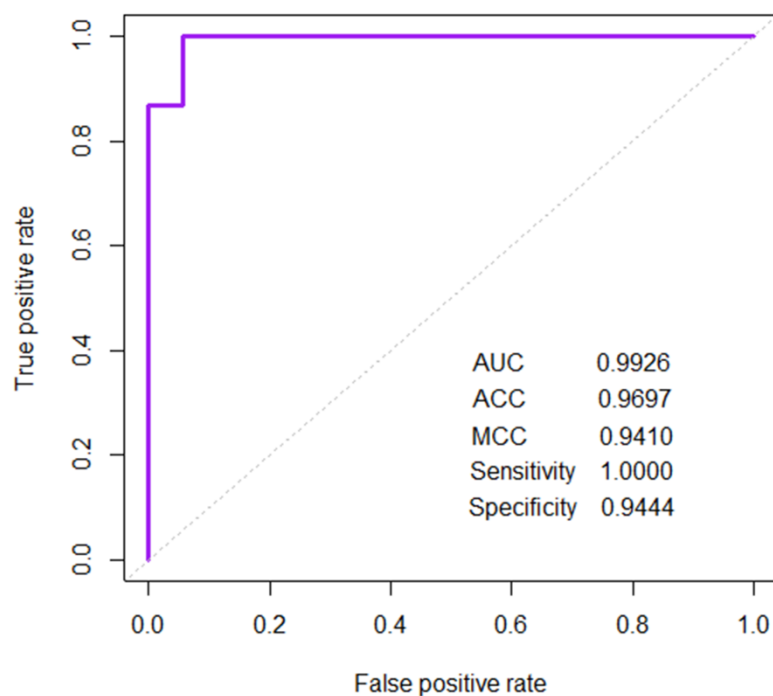


Figure 2: The ROC curve and discrimination performance for the test set

After the specific top-ranking parameter are selected, the performance of the built model on the test set is measured by some metrics.

Application of this identification tool on external validation set

Under the guidance of the outstanding performance, the final model was built based on the sophisticated combination of 11 selected parameters. According to the performance requirements, the value of cutoff was adjusted to 0.43. The external validation set with 11 laboratory parameters also didn't participate in any modeling process and was only used to reconsider the reliability of the model, which could provide the factual foundation for future clinical applications. Among these samples, 39 of 40 samples with COVID-19, 13 of 14 samples with common pneumonia, 10 of 10 samples with tuberculosis, and 9 of 10 samples with lung cancer were accurately identified. Overall, the sensitivity reached $39/(39+1+1) = 0.9512$, the specificity reached $32/(32+1) = 0.9697$, and the total accuracy reached 0.9595. These results indicated that the assistant tool might have a high resolution on distinguishing patients with COVID-19 from the complex disease condition. It suggested that the technology had great potential to develop into a powerful assistant tool for large-scale detection of the highly infectious disease.

Key blood indices that were highly related with COVID-19

Only the integration of 11 laboratory parameters could achieve high-level performance to accurately identify COVID-19. The reason was not only the powerful ability of machine learning algorithm in medicine, but also the inextricable connection between the selected parameters and the epidemic pneumonia. Once the virus invaded the body, the composition of the blood would change. The abnormalities sometimes cannot be observed directly from these routine laboratory blood tests, but the

parameters could play a material impact on identifying the disease after processed by machine learning. Subsequently the laboratory findings were analyzed from the statistical perspective and the detailed information of the findings was listed in Table 2S, including the abbreviations and reference ranges. Except for glucose (GLU) and magnesium (Mg), almost all the parameters manifested the significant differences between COVID-19 and general pneumonia in Figure 3. It also indicated the parameters were highly correlated to the infection of SARS-CoV-2 and the powerful ability of the tool to effectively distinguish patients with COVID-19 and patients with pneumonia. In fact, many abnormal changes of laboratory parameters had been widely reported for important clinical references, including total bilirubin (TBIL), GLU, creatinine (CREA), lactate dehydrogenase (LDH), creatine kinase isoenzyme (CK-MB), and kalium (K).^{22,25,26} For example, the descriptive study of 99 cases with COVID-19 in Wuhan showed that 51 (52%) patients with the elevated level of GLU and 75 (76%) patients with the elevated level of LDH, which was consistent with our results.²⁷ The remaining parameters, including total protein (TP), calcium (Ca), magnesium (Mg), platelet distribution width (PDW), and basophil (BA#), were rarely received extensive attention due to the lack of information on the clinical characteristics of affected patient. However, they played an irreplaceable part in the random forest algorithm and had great potential for diagnostic markers with future clinical practice. Indeed, the existence of differences was not a decisive factor to select critical parameters for machine learning algorithms, it was helpful to promote the tool to quickly identify COVID-19 infection and provide a sufficient consciousness for dynamically monitoring the process of the disease.

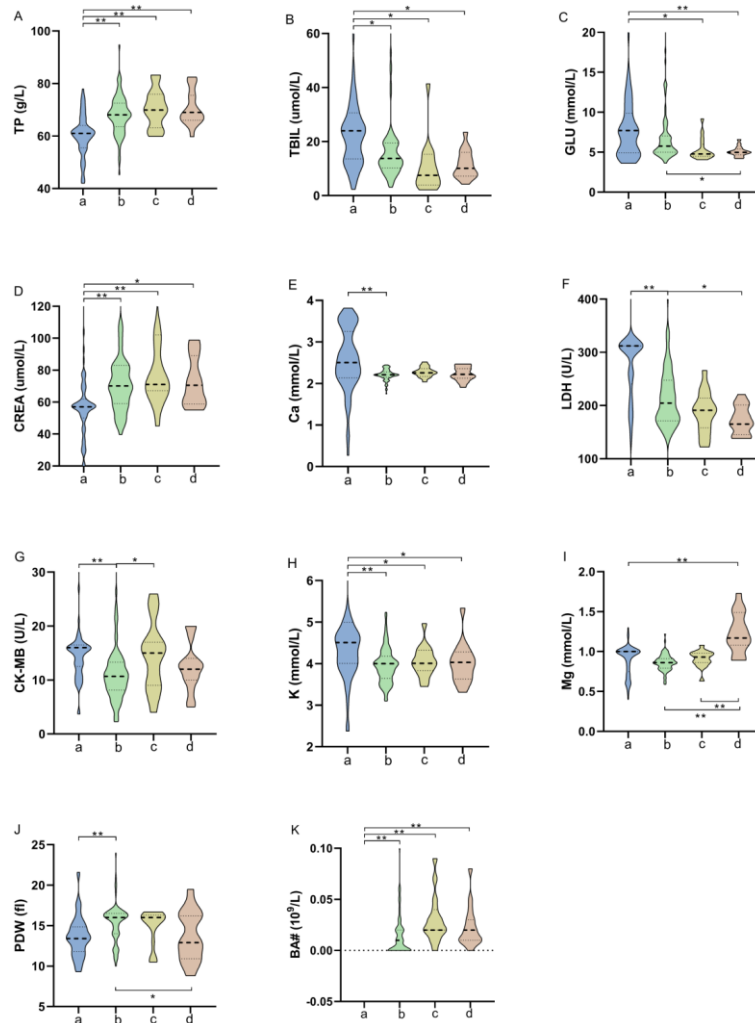


Figure 3: Differential expressed parameters in different patients' groups

The statistical variations of 11 parameters are analyzed from different groups including group of COVID-19 (a), group of common pneumonia (b), group of tuberculosis (c), and group of lung cancer (d). ** represents $p < 0.0001$ except for CLU (** $p = 0.0076$), CREA (** $p = 0.0022$ for a and b, ** $p = 0.0002$ for a and c), LDH (** $p = 0.0047$), and BA# (** $p = 0.0004$ for a and b). TBIL (* $p = 0.018$ for a and b, $p = 0.021$ for a and c, $p = 0.029$ for a and d), GLU (* $p = 0.016$ for a and c, $p = 0.037$ for b and d), CREA (* $p = 0.015$), LDH (* $p = 0.038$), CK-MB (* $p = 0.025$), K (* $p = 0.036$ for a and c, $p = 0.030$ for a and d), PDW (* $p = 0.020$).

Complementarity with other analytical techniques

Currently, suspected patients were recommended to use nucleic acid assay to effectively control the spread of the epidemic pneumonia. The complexity of the operation and the fragility of the specimens make it unsuitable for large-scale initial screening. The time-consuming etiology testing might delay the timely isolation and treatment for COVID-19 patients. And a study reported that the positive rate of the testing was only 30-50% in clinical application.^{9,28} The sequencing technology of the viral had elevated the accuracy of detection results, but it needed high-quality laboratory requirement which could only be provided for supplementary evidence. Rapid antibody detection often lagged behind the nucleic acid detection due to the presence of the window period.¹⁴ After the virus invaded, the dynamic

balance was often broken and the composition was changed in the blood. The timely responses might be used for the identification of the disease as long as they could be captured by machine learning algorithms. And the assistant tool was not subject to complex operation and prolonged waiting because the conventional laboratory blood tests were the mature technology that could be fully automated. It was convenient to operate the large-scale initial screening and help the confirmed patients provide appropriate treatment.

Web server of the assistant tool

Finally, a user-friendly website service was built to experience and test the assistant tool online at http://lishuyan.lzu.edu.cn/COVID2019_2/. The 11 selected parameters and more detailed information were listed in Figure 4. The users only needed to input the value of specified parameters in the corresponding text box based on the instructions of the interface, and then clicked the “Submit” button. After calculation and analysis of the tool, the results page would display whether this sample was classified as a COVID-19 patient with the probability of prediction. It was vital to check that the units of laboratory parameters were consistent with those on the main page. Importantly, the tool could only be applied to test this method at present, and it is insufficient to evolve the definite evidence of clinical diagnosis for directly discriminating the COVID-19 patients.

Chinese

COVID-19 Assistant Discrimination 2.0

A Fast and Accurate Diagnostic Tool Based on Routine Blood Examination Data

[Home](#) | [Document](#) | [Research](#) | [Contact Us](#)

Please input single sample data in the text form:

Total Protein (TP_g/L)	Total Bilirubin (TBIL_umol/L)
<input type="text" value="58"/>	<input type="text" value="5.4"/>
Glucose (GLU_mmol/L)	Creatinine (CREA_umol/L)
<input type="text" value="5.7"/>	<input type="text" value="60"/>
Calcium (Ca_mmol/L)	Lactate Dehydrogenase (LDH_U/L)
<input type="text" value="1.31"/>	<input type="text" value="263"/>
Creatine Kinase Isoenzyme (CK-MB_U/L)	Kalium (K_mmol/L)
<input type="text" value="10"/>	<input type="text" value="4.68"/>
Magnesium (Mg_mmol/L)	Platelet Distribution Width (PDW_fl)
<input type="text" value="0.9"/>	<input type="text" value="13.6"/>
Basophil (BA#_10 ⁹ /L)	
<input type="text" value="0"/>	

Or upload multiple samples (Please strictly adherence to the csv format as [sample file](#)) :

Copyright © 2020 - [Shuyan Li](#) · All Rights Reserved

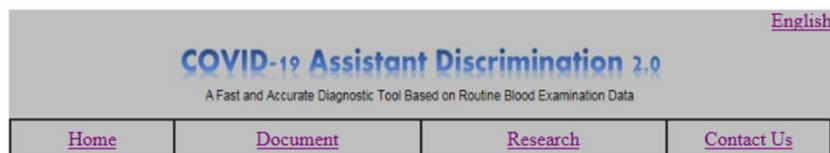
Figure 4: The main page of the COVID-19 assistant discrimination web server

Users only need to input 11 corresponding parameters according to the requirements of the above interface; the web server can display the result whether the sample is COVID-19.

Further clinical evaluation on imported cases from outside of China

After the construction of the assistant discrimination tool was completed, there are some confirmed COVID-19 cases that were imported from outside of China and were received in Gansu Province. In order to make an in-depth assessment for the built tool, 24 samples of 14 patients from Lanzhou Pulmonary Hospital were collected when first performed the routine laboratory blood tests, including 2 routine hematological parameters and 9 biochemical parameters. The prediction results were shown in Figure 5. 22 of 24 samples were accurately identified through using our designed web server with an

accuracy rate of 0.9167. Changes for related blood components could not be captured by the tool due to the early stages of virus infection, the patients with mild symptoms were not accurately diagnosed. However, the high-precision performance has greatly encouraged the confidence to make more adequately preparation for further clinical practice.



Sample No.	Results
1	COVID-19 sample with 61% probability
2	COVID-19 sample with 63% probability
3	non-COVID-19 sample with 57% probability
4	COVID-19 sample with 59% probability
5	COVID-19 sample with 61% probability
6	COVID-19 sample with 60% probability
7	COVID-19 sample with 69% probability
8	COVID-19 sample with 66% probability
9	COVID-19 sample with 77% probability
10	COVID-19 sample with 79% probability
11	COVID-19 sample with 65% probability
12	COVID-19 sample with 82% probability
13	COVID-19 sample with 64% probability
14	COVID-19 sample with 85% probability
15	COVID-19 sample with 83% probability
16	COVID-19 sample with 86% probability
17	COVID-19 sample with 76% probability
18	COVID-19 sample with 73% probability
19	COVID-19 sample with 82% probability
20	COVID-19 sample with 85% probability
21	COVID-19 sample with 80% probability
22	COVID-19 sample with 80% probability
23	non-COVID-19 sample with 69% probability
24	COVID-19 sample with 73% probability

Figure 5: Prediction results of laboratory confirmed patients with COVID-19 overseas

24 confirmed COVID-19 cases that were imported from outside of China are tested by the assistant discrimination tool.

Application domain of this tool

This method is built on data of suspected patients which presented similar CT information or similar symptoms with COVID-19 like fever, cough, myalgia or fatigue, which are followed the first triage results in hospital before the nucleic acid assay. This tool has been validated many times with good performance on these suspected patients and we have faith on its proper application on further tests to identify COVID-19 cases from other suspected cases.

However, its application on identification COVID-19 cases with atypical symptoms is not confirmed since these cases were hard to collect in the current situation. Anyway, we still believe that if more laboratory parameters are collected, machine learning algorithms may extract more valuable candidate markers to identify COVID-19 infection, even from the whole population, whether with or without typical symptoms.

Conclusion

In conclusion, the assistant tool was built with 11 top-ranking clinical available blood indices that were extracted from the random forest algorithm. This result may capture the inherent patterns of these routine parameters to make routine blood tests play an influential role on the value of alarm diseases. After several evaluations, it is of extraordinary benefit to be exploited for large-scale initial screening on suspected patients of COVID-19 to minimize the risk of the virus, especially for confirmed patients taken the timely isolation and treatment, which is currently known the most effective response. Although the proposed tool needs to be tested by more clinical practice, the method provides some new insights into for the rapid diagnosis of COVID-19 infection to cope with severe situations due to terrible characteristic of the human-to-human transmission.

Reference

1. Lu H, Stratton CW, Tang Y-W. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *J Med Virol* 2020; **92**: 401-2.
2. Gorbalenya A, Baker S, Baric R, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020; **5**: 536-44.
3. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020; **382**: 727-33.
4. WHO. Coronavirus disease (COVID-2019) situation reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (accessed Mar 1, 2020).
5. Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents* 2020; **55**: 105924.
6. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; **395**: 497-506.
7. Chung M, Bernheim A, Mei X, et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology* 2020; **295**: 202-7.
8. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 2020; **25**: 2000045.
9. Liu R, Han H, Liu F, et al. Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. *Clin Chim Acta* 2020; **505**: 172-5.
10. Xie C, Jiang L, Huang G, et al. Comparison of different samples for 2019 novel coronavirus detection by nucleic acid amplification tests. *Int J Infect Dis* 2020; **93**: 264-7.
11. Chan J, Yuan S, Kok K-H, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020; **395**: 514-23.
12. Shen M, Zhou Y, Ye J, et al. Recent advances and perspectives of nucleic acid detection for coronavirus. *J Pharm Anal* 2020; published online Mar 1, 2020. <https://doi.org/10.1016/j.jpha.2020.02.010>.
13. Xu Y-H, Dong J-H, An W-M, et al. Clinical and computed tomographic imaging features of novel coronavirus pneumonia caused by SARS-CoV-2. *J Infect* 2020; **80**: 394-400.
14. Peiris JSM, Chu CM, Cheng VCC, et al. Clinical progression and viral load in a community

- outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* 2003; **361**: 1767-72.
15. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; **579**: 270-3.
 16. Li Z, Yi Y, Luo X, et al. Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J Med Virol* 2020; published online Feb 27, 2020. <https://doi.org/10.1002/jmv.25727>.
 17. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* 2020; **395**: 470-3.
 18. Hu Z, Song C, Xu C, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci* 2020; published online Mar 4, 2020. <https://doi.org/10.1007/s11427-020-1661-4>.
 19. Lai C-C, Liu YH, Wang C-Y, et al. Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Facts and myths. *J Microbiol Immunol Infect* 2020; published online Mar 4, 2020. <https://doi.org/10.1016/j.jmii.2020.02.012>.
 20. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA* 2020; **323**: 1061-9.
 21. Gao Y, Li T, Han M, et al. Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. *J Med Virol* 2020; published online Mar 17, 2020. <https://doi.org/10.1002/jmv.25770>.
 22. Liu Y, Yang Y, Zhang C, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *Sci China Life Sci* 2020; **63**: 364-74.
 23. Breiman L. Random forests. *Machine Learning* 2001; **45**: 5-32.
 24. Petralia F, Wang P, Yang J, Tu Z. Integrative random forest for gene regulatory network inference. *Bioinformatics* 2015; **31**: i197-i205.
 25. Mo P, Xing Y, Xiao Y, et al. Clinical characteristics of refractory COVID-19 pneumonia in Wuhan, China. *Clin Infect Dis* 2020; published online Mar 16, 2020. <https://doi.org/10.1093/cid/ciaa270>.
 26. Xu X-W, Wu X-X, Jiang X-G, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *BMJ* 2020; **368**: m606.
 27. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; **395**: 507-13.
 28. Ai T, Yang Z, Hou H, et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. *Radiology* 2020; published online Feb 26, 2020. <https://doi.org/10.1148/radiol.2020200642>.