

# Detection of COVID-19 by Machine Learning Using Routine Laboratory Tests

Hikmet Can Çubukçu, MD,<sup>1,\*</sup> Deniz İlhan Topcu, MD, PhD,<sup>2,\*</sup> Nilüfer Bayraktar, MD,<sup>2</sup> Murat Gülşen, MD,<sup>3</sup> Nuran Sarı, MD,<sup>4</sup> and Ayşe Hande Arslan, MD<sup>4</sup>

From the <sup>1</sup>Interdisciplinary Stem Cells and Regenerative Medicine, Ankara University Stem Cell Institute, Ankara, Turkey; Departments of <sup>2</sup>Medical Biochemistry and <sup>4</sup>Infectious Diseases and Clinical Microbiology, Başkent University Faculty of Medicine, Ankara, Turkey; and <sup>3</sup>Autism, Special Mental Needs and Rare Diseases Department, Turkish Ministry of Health, Ankara, Turkey.

## ABSTRACT

**Objectives:** The present study aimed to develop a clinical decision support tool to assist coronavirus disease 2019 (COVID-19) diagnoses with machine learning (ML) models using routine laboratory test results.

**Methods:** We developed ML models using laboratory data (n = 1,391) composed of six clinical chemistry (CC) results, 14 CBC parameter results, and results of a severe acute respiratory syndrome coronavirus 2 real-time reverse transcription–polymerase chain reaction as a gold standard method. Four ML algorithms, including random forest (RF), gradient boosting (XGBoost), support vector machine (SVM), and logistic regression, were used to build eight ML models using CBC and a combination of CC and CBC parameters. Performance evaluation was conducted on the test data set and external validation data set from Brazil.

**Results:** The accuracy values of all models ranged from 74% to 91%. The RF model trained from CC and CBC analytes showed the best performance on the present study's data set (accuracy, 85.3%; sensitivity, 79.6%; specificity, 91.2%). The RF model trained from only CBC parameters detected COVID-19 cases with 82.8% accuracy. The best performance on the external validation data set belonged to the SVM model trained from CC and CBC parameters (accuracy, 91.18%; sensitivity, 100%; specificity, 84.21%).

**Conclusions:** ML models presented in this study can be used as clinical decision support tools to contribute to physicians' clinical judgment for COVID-19 diagnoses.

## INTRODUCTION

Coronavirus disease 2019 (COVID-19) was first seen in China after severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in December 2019. The disease became a global health problem and was declared a pandemic on March 3, 2020, by the World Health Organization.<sup>1</sup>

Clinical chemistry (CC), hematology, coagulation, and specific protein test results of patients with COVID-19 changed during the illness. Zhang et al<sup>2</sup> reported that patients with COVID-19 could have lymphopenia (75.4%) and eosinopenia (52.9%). In this study, high C-reactive protein, procalcitonin, and D-dimer were associated with disease severity. Lippi and Plebani<sup>3</sup> stated that procalcitonin levels increased in patients with COVID-19 who had bacterial coinfection. In a meta-analysis study from Lippi et al,<sup>4</sup> thrombocytopenia was

## KEY POINTS

- The combined utility of routine laboratory results using machine learning (ML) models can provide precious opportunities to assist coronavirus disease 2019 (COVID-19) diagnosis.
- The ML models can be used as clinical decision support tools to contribute to physicians' clinical judgment about COVID-19 diagnosis.
- The accuracy values of all models ranged from 74% to 91%. The best performance on the external validation data set belonged to the support vector model trained from clinical chemistry and CBC parameters (accuracy, 91.18%).

## KEY WORDS

SARS-CoV-2; COVID-19; Machine learning; Artificial intelligence; Laboratory tests

*Am J Clin Pathol* May 2022;157:758-766  
[HTTPS://DOI.ORG/10.1093/AJCP/AQAB187](https://doi.org/10.1093/AJCP/AQAB187)

Received: May 20, 2021

Accepted: September 22, 2021

Advance publication: November 17, 2021

Corresponding author: Hikmet Can Çubukçu; [hikmetcancubukcu@gmail.com](mailto:hikmetcancubukcu@gmail.com).

\*First authors.

This article is available for CME credit. Go to [academic.oup.com/ajcp/pages/journal\\_cme](https://academic.oup.com/ajcp/pages/journal_cme) to see the latest articles. The complete catalog of journal CME courses can be found at [store.ascp.org](https://store.ascp.org).

associated with increased mortality risk in patients with COVID-19. Furthermore, higher neutrophil, D-dimer, prothrombin time, alanine aminotransferase (ALT), lactate dehydrogenase (LDH), total bilirubin, high sensitive troponin I, and lower lymphocyte and albumin levels were detected in patients with COVID-19 who were hospitalized in an intensive care unit (ICU).<sup>5</sup> Prothrombin time, fibrin degradation products, and D-dimer were higher in patients who died of COVID-19 pneumonia than in surviving patients.<sup>6</sup> The recommended laboratory tests and their changes during COVID-19 were extensively described by the International Federation of Clinical Chemistry (IFCC) and Laboratory Medicine Taskforce on COVID-19.<sup>7</sup> Although the literature covered which analytes were valuable in diagnosis, monitoring, and estimating prognosis, it did not quantitatively reflect the overall contribution and use of analytes in SARS-CoV-2 detection concerning the accuracy, sensitivity, and specificity.

Real-time reverse transcription–polymerase chain reaction (rRT-PCR) is a gold standard method that detects the SARS-CoV-2 RNA.<sup>8</sup> Errors originating from the preanalytical phase, such as improper handling and transportation of specimens, contamination, inadequate sample quality, the presence of PCR inhibitors, and misidentifications, lead to false-negative test results.<sup>9–11</sup> Thus, a negative test result cannot exclude infection where strong clinical suspicion exists.<sup>12</sup>

Machine learning (ML) models using routine laboratory results can provide valuable tools that support clinical decisions for detecting COVID-19 cases. The present study aimed to detect SARS-CoV-2 cases with high accuracy, sensitivity, and specificity with ML models using routine laboratory test results.

## MATERIALS AND METHODS

### Study Population

The present study was approved by the Ministry of Health (form No. 2020-05-28T154351) and the Başkent University Institutional Review Board (project No. KA20-169).

**FIGURE 1** schematically shows the data collection process and ML model development.

The demographic information and laboratory results of patients with COVID-19 detected by rRT-PCR were obtained from the laboratory information system. Data from patients with COVID-19 were collected from April 2020 to November 2020. Only patient data consisting of concomitant CC and CBC test results within a 48-hour window from rRT-PCR testing were selected. CC test parameters comprised total bilirubin, ALT, aspartate aminotransferase (AST), LDH, creatinine, and C-reactive protein (CRP). CBC parameters included eosinophils, monocytes, lymphocytes, red cell distribution width, platelets, mean corpuscular hemoglobin, leukocytes, mean corpuscular volume, mean corpuscular hemoglobin concentration, neutrophils, hemoglobin, hematocrit, basophils, and RBCs. Patients with multiple PCR results were considered follow-up patients, and only the first rRT-PCR results of these patients were included in the study. In line with the Republic of Turkey Ministry of Health's decision, only symptomatic patients were tested by rRT-PCR for SARS-CoV-2. In our hospital, patients with suspected COVID-19 are almost

always admitted to the COVID-19 outpatient clinic. Thus, only outpatient cases were included in the present study. The patient data from January 2018 to November 2019 were recruited from the laboratory information system to establish a PCR-negative group. The following exclusion criteria were then used for PCR-positive and PCR-negative groups: inpatients, patients with missing corresponding laboratory test parameters, and patients younger than 18 years and older than 65 years, as shown in **FIGURE 1**. The data belonged to patients admitted to Başkent University Ankara Hospital, which is a single-center university hospital in Ankara, Turkey. The present study's data did not include a COVID-19 vaccinated individual.

We obtained the public data set from the Israelita Albert Einstein Hospital in São Paulo, Brazil,<sup>13</sup> for external validation of ML models. The São Paulo data set (n = 5,644) included SARS-CoV-2 rRT-PCR results and routine laboratory results of patients admitted to the abovementioned hospital from March 8, 2020, to April 3, 2020. The data set was divided into two subsets comprising concomitant CC and CBC (n = 34) results and only CBC (n = 513) results. The CBC data set had imbalanced SARS-CoV-2 results (positive SARS-CoV-2 results, 75; negative SARS-CoV-2 results, 438). Therefore, we randomly sampled 75 patients' data whose SARS-CoV-2 results were negative. The final CBC data set comprised 150 patients' results (positive SARS-CoV-2 results, 75; negative SARS-CoV-2 results, 75), as shown in **FIGURE 1**.

### Clinical Laboratory Analyses

Serum AST, ALT, total bilirubin, creatinine, LDH, and CRP levels were measured using the Abbott Alinity c analyzer (Abbott Diagnostics). The following methods were used for CC tests: the enzymatic NADH method without pyridoxal-5'-phosphate for AST and ALT. The IFCC recommended the lactate to pyruvate forward reaction method for LDH, the diazo reaction for total bilirubin, the kinetic alkaline picrate method for creatinine, and the immunoturbidimetric method for CRP. CBC analyses were performed using Abbott CELL DYN Ruby hematology analyzers (Abbott Diagnostics).

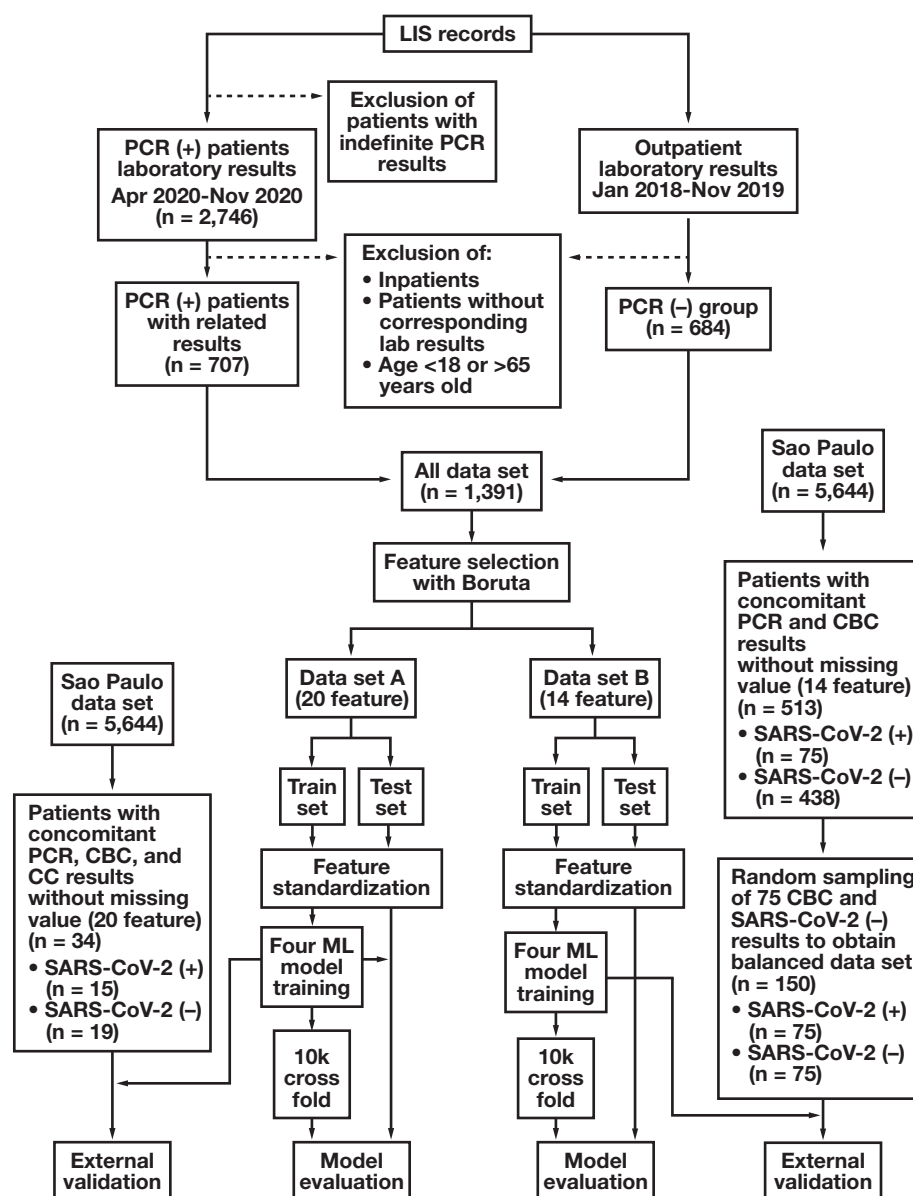
The rRT-PCR analysis was conducted on a Rotor-Gene Q rRT-PCR quantification system (Qiagen) with a Diagnovital HS SARS-CoV-2 rRT-PCR kit (A1 Life Sciences) that targets *ORF1ab* and *N* genes.

Internal quality control materials ran every 12 hours at two levels for CC analytes (Technopath Clinical Diagnostics) and every 8 hours at three levels for CBC parameters (Abbott Diagnostics) in our laboratory. Our laboratory also enrolled in a monthly external quality control program (Randox Quality Control) for all analytes. Total error values are given in [Supplementary Materials](#) (all supplemental materials can be found at *American Journal of Clinical Pathology* online).

Our laboratory followed health quality standards determined by the Republic of Turkey Ministry of Health.

### Development and Evaluation of ML Models

In this study's context, two data sets were formed using patient records. Data set A contained both CC and CBC results, and data set B consisted of only CBC results **FIGURE 1**. The data sets' results were used as input variables, with rRT-PCR results the target variable.



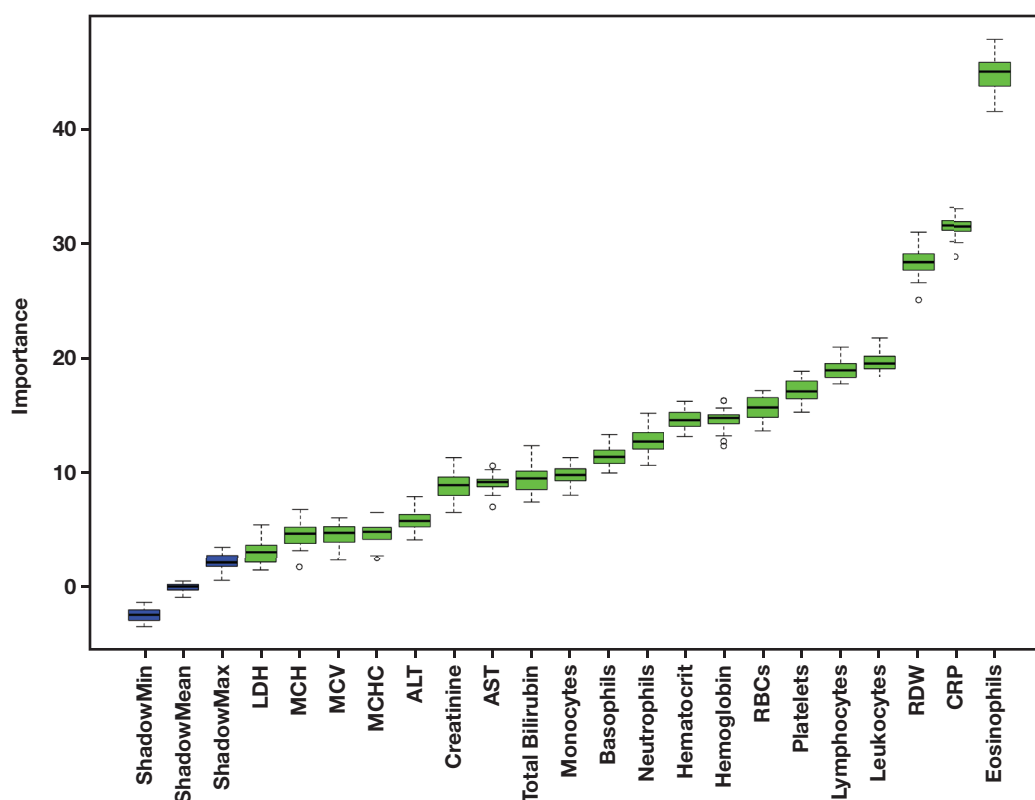
**FIGURE 1** Study design. CC, clinical chemistry; LIS, laboratory information system; ML, machine learning; PCR, polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

The Boruta feature selection method was applied to detect the model's noninformative or redundant features.<sup>14</sup> The Boruta method created a corresponding shadow for each attribute, whose values were obtained by shuffling the original attributes' values across properties. Then, the algorithm trained a random forest (RF) model to evaluate the importance of each data set feature. The importance measure of the real feature was compared with a threshold value using z scores. The threshold was determined dynamically using a binomial distribution. Finally, importance was classified into three classes: discard (red), speculative (blue), and keep (green) to identify important features.<sup>14</sup> Every parameter's importance was confirmed, as given in **FIGURE 2**.

ML methods, including RF, XGBoost, logistic regression, and support vector machine (SVM), were used to predict SARS-CoV-2 results. The RF classifier model was constructed using 200 decision trees, entropy for information gain, and other default parameters

based on the scikit-learn 0.24.1 package.<sup>15,16</sup> The XGBoost classifier model was built using 100 decision trees and other default parameters based on the scikit-learn 0.24.1 package.<sup>16,17</sup> The support vector classifier<sup>18</sup> and logistic regression models were implemented using default parameters based on the scikit-learn 0.24.1 package.<sup>16</sup> Hyperparameter optimization was not performed for any model.

The first ML construction process was to split all data into training and test data sets using 80% and 20% of the overall data, respectively. Then, the data sets were standardized by z score transformation. The training data set was used for ML models' construction and 10-fold cross-validation<sup>19</sup> to assess the proposed models' performance. Finally, model performances were independently evaluated using the test data set with accuracy, sensitivity, specificity, *F* score values, receiver operating characteristic (ROC) curve analysis, and  $\kappa$  statistics.<sup>20</sup> Furthermore, the ML models'



**FIGURE 2** Boruta feature selection plot. Green indicates important features, and blue represents tentative features. ALT, alanine aminotransferase; AST, aspartate aminotransferase; CRP, C-reactive protein; LDH, lactate dehydrogenase; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RDW, red cell distribution width.

performances were assessed on the São Paulo data set using accuracy, sensitivity, specificity, *F* score values, and ROC curve analysis for external validation.

Data preprocessing, implementation of artificial intelligence models, and statistical analyses were conducted using Python 3.7.6<sup>21</sup> and R statistical software 3.6.0.<sup>22</sup> The Python codes of the present study's steps are given in our GitHub account (<https://github.com/hikmetc/COVID-19-AI>).

## RESULTS

The demographic features and laboratory measurements of the present study's population are summarized in **TABLE 1**. The PCR-positive and PCR-negative groups were matched by sex; however, the PCR-positive group's median age was lower than that of the PCR-negative group (42 vs 48,  $P < .001$ ). While there was no statistical difference detected for LDH (186 vs 187 IU/L), the most significant difference among laboratory parameters was the eosinophils count ( $0.04$  vs  $0.13 \times 10^3/\mu\text{L}$ ; effect size, 0.62). Furthermore, the COVID-19–positive and COVID-19–negative groups differed in terms of pulmonary and gastrointestinal symptoms **TABLE 1**.

The São Paulo data set did not include analyzer, sex, race, or clinical information. The data set's laboratory results consisted of standardized values, and the patients' ages were given as quantiles. Therefore, the exact values of the laboratory results and ages were unknown due to the São Paulo data set's inherent characteristics. Positive

and negative cases in the São Paulo data set were determined based on rRT-PCR testing. See **TABLE 2** for patients' admission information.

The ML models' performance for COVID-19 detection on our data set is given in **TABLE 3**. Models trained from CBC and CC analytes represented better performance than the models trained from CBC analytes alone. The moderate agreements (Cohen's  $\kappa$ , 0.6–0.7) were observed between all models' predicted results and rRT-PCR results. The ML models' accuracy on the present study's data set ranged from 80% to 85%, and the area under the curve (AUC) values of all models were higher than 0.8. On the other hand, in the external validation, the ML models' accuracies ranged from 74% to 91%. Interestingly, sensitivity values of ML models were higher in the external validation data set than in the present study's data set. Moreover, found sensitivity values were 100% on the external validation data set for all ML models trained from CC and CBC results, as shown in **TABLE 4**.

The RF model trained from CC and CBC analytes showed the best performance among the models (accuracy, 85.30%; specificity, 91.24%; sensitivity, 79.58%; positive predictive value, 90.40%; AUC, 0.925) on our data set, as shown in **TABLE 3** and illustrated in **FIGURE 3**. However, in external validation, the SVM model's performance was better than other ML models, as given in **TABLE 4** and **FIGURE 4**.

## DISCUSSION

We developed eight ML models to predict SARS-CoV-2 results. The ML models trained from CC and CBC results showed better

**TABLE 1** Demographic Features and Laboratory Measurements of the Present Study Population<sup>a</sup>

Characteristic	COVID-19 Positive	COVID-19 Negative	P Value	Effect Size
Age, median (IQR), y	41.9 (20.4)	48.4 (15.4)	<.001 <sup>b</sup>	0.23 <sup>c</sup>
Sex, No. (%)				
Female	341 (48.2)	340 (49.7)	.582 <sup>d</sup>	0.01 <sup>e</sup>
Male	366 (51.8)	344 (50.3)		
White race, No. (%)	707 (100)	684 (100)		
Turkish ethnicity, No. (%)	707 (100)	684 (100)		
ALT, median (IQR), IU/L	21.00 (18.00)	18.00 (14.00)	<.001 <sup>b</sup>	0.17 <sup>c</sup>
AST, median (IQR), IU/L	21.00 (11.00)	17.00 (9.00)	<.001 <sup>b</sup>	0.24 <sup>c</sup>
Basophils, median (IQR), × 10 <sup>3</sup> /μL	0.05 (0.04)	0.06 (0.04)	<.001 <sup>b</sup>	0.28 <sup>c</sup>
CRP, median (IQR), mg/L	6.60 (12.85)	2.00 (4.50)	<.001 <sup>b</sup>	0.38 <sup>c</sup>
Eosinophils, median (IQR), × 10 <sup>3</sup> /μL	0.04 (0.06)	0.13 (0.12)	<.001 <sup>b</sup>	0.62 <sup>c</sup>
Hemoglobin, median (IQR), g/dL	14.50 (2.00)	13.60 (2.22)	<.001 <sup>b</sup>	0.32 <sup>c</sup>
Hematocrit, median (IQR), %	43.50 (5.80)	40.80 (6.10)	<.001 <sup>b</sup>	0.32 <sup>c</sup>
Creatinine, median (IQR), mg/dL	0.86 (0.25)	0.78 (0.21)	<.001 <sup>b</sup>	0.24 <sup>c</sup>
LDH, median (IQR), IU/L	186.00 (62.00)	187.00 (56.00)	.343 <sup>b</sup>	0.03 <sup>c</sup>
Lymphocytes, median (IQR), × 10 <sup>3</sup> /μL	1.47 (0.88)	2.01 (1.03)	<.001 <sup>b</sup>	0.37 <sup>c</sup>
Leukocytes, median (IQR), × 10 <sup>3</sup> /μL	5.48 (2.25)	7.03 (3.23)	<.001 <sup>b</sup>	0.40 <sup>c</sup>
MCH, median (IQR), pg	29.30 (2.00)	29.05 (2.70)	.006 <sup>b</sup>	0.09 <sup>c</sup>
MCHC, median (IQR), %	33.30 (1.20)	33.20 (1.20)	.005 <sup>b</sup>	0.09 <sup>c</sup>
MCV, median (IQR), fL	87.90 (5.45)	87.10 (6.60)	.020 <sup>b</sup>	0.07 <sup>c</sup>
Monocytes, median (IQR), × 10 <sup>3</sup> /μL	0.53 (0.30)	0.51 (0.25)	.130 <sup>b</sup>	0.05 <sup>c</sup>
Neutrophils, median (IQR), × 10 <sup>3</sup> /μL	3.21 (1.94)	3.96 (2.53)	<.001 <sup>b</sup>	0.29 <sup>c</sup>
RBC, median (IQR), × 10 <sup>6</sup> /μL	4.98 (0.61)	4.70 (0.72)	<.001 <sup>b</sup>	0.28 <sup>c</sup>
RDW, median (IQR), %	13.70 (3.70)	14.40 (3.40)	<.001 <sup>b</sup>	0.28 <sup>c</sup>
Total bilirubin, median (IQR), mg/dL	0.50 (0.30)	0.57 (0.42)	<.001 <sup>b</sup>	0.12 <sup>c</sup>
Platelets, median (IQR), × 10 <sup>3</sup> /μL	200.00 (69.00)	244.00 (103.25)	<.001 <sup>b</sup>	0.37 <sup>c</sup>
Symptoms, No. (%)				
Fever	549 (77.7)	89 (13.0)	<.001 <sup>d</sup>	0.65 <sup>e</sup>
Cough	340 (48.1)	21 (3.1)	<.001 <sup>d</sup>	0.51 <sup>e</sup>
Fatigue	320 (45.3)	103 (15.1)	<.001 <sup>d</sup>	0.37 <sup>e</sup>
Myalgia	289 (40.9)	69 (10.1)	<.001 <sup>d</sup>	0.35 <sup>e</sup>
Sore throat	173 (24.5)	35 (5.1)	<.001 <sup>d</sup>	0.27 <sup>e</sup>
Headache	117 (16.5)	45 (6.6)	<.001 <sup>d</sup>	0.16 <sup>e</sup>
Diarrhea	89 (12.6)	18 (2.6)	<.001 <sup>d</sup>	0.19 <sup>e</sup>
Shortness of breath	62 (8.8)	42 (6.1)	.056 <sup>d</sup>	0.05 <sup>e</sup>
Loss of taste or smell	24 (3.4)	0 (0.0)	<.001 <sup>d</sup>	0.13 <sup>e</sup>
Stomachache	17 (2.4)	110 (16.1)	<.001 <sup>d</sup>	0.24 <sup>e</sup>
Other	63 (8.9)	199 (29.1)	<.001 <sup>d</sup>	0.26 <sup>e</sup>

ALT, alanine aminotransferase; AST, aspartate aminotransferase; COVID-19, coronavirus disease 2019; CRP, C-reactive protein; IQR, interquartile range; LDH, lactate dehydrogenase; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RDW, red cell distribution width.

<sup>a</sup>Cohen's *w* values with 0.1, 0.3, and 0.5 correspond to small, medium, and high large effect sizes, respectively. Rank biserial correlation values with 0.1, 0.24, and 0.37 correspond to small, medium, and large effect sizes, respectively.

<sup>b</sup>Mann-Whitney *U* test.

<sup>c</sup>Rank biserial correlation.

<sup>d</sup> $\chi^2$  test.

<sup>e</sup>Cohen's *w* effect size.

performance than the models trained from only CBC results. While our RF model showed the best performance on the present study's data set (85.30% accuracy with CC and CBC parameters, 82.80% for CBC parameters), the SVM model's performance was superior

to other ML models when performed on the external validation data set (91.18% accuracy with CC and CBC parameters, 80% for CBC parameters). The accuracy values of the ML models obtained from 10-fold cross-validation were comparable to their corresponding

**TABLE 2** São Paulo Data Set: COVID-19 Outpatients and the Patients Referred to General Ward, Semi-Intensive Care Unit, or Intensive Care Unit

SARS-CoV-2	No.	Regular Ward, No. (%)	Semi-Intensive Care Unit, No. (%)	Intensive Care Unit, No. (%)	Outpatient, No. (%)
CBC data set					
Positive	75	25 (33.3)	7 (9.3)	7 (9.3)	36 (48)
Negative	75	5 (6.6)	4 (5.3)	1 (1.3)	65 (86.6)
CBC + CC data set					
Positive	15	7 (46.6)	3 (20)	3 (20)	2 (13.3)
Negative	19	5 (26.3)	2 (10.5)	1 (5.3)	11 (57.9)

CC, clinical chemistry.

test set accuracy, which demonstrated the models' generalizability. We performed external validation to show our ML models' reproducibility and found that sensitivity values were quite satisfactory, as shown in **TABLE 4**.

Several ML models using routine laboratory tests have been proposed to support COVID-19 diagnoses. Some models were built using only CBC parameters and others with extended routine laboratory parameters and other clinical features. Cabitza et al<sup>23</sup> reported that the RF model trained from CBC parameters could predict COVID-19 with 76% accuracy, 76% sensitivity, 82% specificity, and an AUC of 0.76. Another RF model trained from CBC parameters proposed by Tschöellitsch et al<sup>24</sup> showed 86% accuracy; however, the positive predictive value was only 20%. Our RF model trained from CBC parameters outperformed the previously reported models and showed 82.8% accuracy, 80.28% sensitivity, 85.4% specificity, and 85.07% positive predictive value (PPV) in the present study's data set and 79.33% accuracy, 90.67% sensitivity, 68% specificity, and 73.91% PPV in the external validation data set. In a study from Joshi et al,<sup>25</sup> the specificity and positive predictive values of logistic regression models using CBC parameters and sex information were lower than 50% and 30%, respectively. In the present study's data, our logistic regression model's specificity and positive predictive values trained from only CBC parameters were 78.10% and 79.45%, respectively, as given in **TABLE 3**.

The ML models offered by Schwab et al<sup>26</sup> were developed using 106 laboratory tests, demographics, and clinical parameters. In this study, the models' best AUC, sensitivity, and specificity values were 0.66, 75%, and 59%, respectively. Goodman-Meza et al<sup>27</sup> used a combination of seven ML models using age, sex, and broad laboratory test spectrum to predict COVID-19 diagnoses. Although their model's AUC, sensitivity, and specificity values were 0.91, 93%, and 63%, respectively, the positive predictive value was only 29%. More balanced performance characteristics were reported in a study from Yang et al.<sup>28</sup> The XGBoost model, using 27 laboratory test parameters and three demographic features, showed 85% sensitivity, 81% specificity, and 0.854 AUC. Likewise, our study also reached balanced performance characteristics for our data set and external validation data set. Our ML models used 20 CC and CBC and 14 CBC routine laboratory test parameters as input variables to predict a SARS-CoV-2 result. With only 14 CBC parameters, our ML models achieved 79.9% to 82.8% accuracy **TABLE 3**.

While the highest value of the ML models' sensitivities was 81.67% in the present study's data set, the ML models' sensitivities

ranged from 85.33% to 100% in the external validation, as shown in **TABLE 4**. Our data set comprised outpatients whose ages were younger than 65 years. On the other hand, the external validation data set included outpatients and patients referred to the regular ward, semi-ICU, and ICU, as given in **TABLE 2**. Therefore, even if the exact ages were not available, the external validation data set probably included more severe cases considering admission characteristics. Furthermore, severe cases in the external validation data set might have had more abnormal laboratory results than the present study's outpatients. Thus, it can be suggested that ML models can distinguish true positives from false negatives more efficiently when the population has more severe cases, eventually leading to improved sensitivity.

The importance of input parameters in COVID-19 prediction differs among studies. Some studies reported that leukocytes,<sup>24</sup> arterial lactic acid,<sup>26</sup> and LDH<sup>23,27,28</sup> were the most important parameters. However, consistent with the present study, Plante et al<sup>29</sup> found eosinophils as the most important feature. Eosinophil count is lower in patients with COVID-19,<sup>30</sup> which is related to the disease's severity.<sup>31</sup> Moreover, it was demonstrated that eosinophils were recruited in lung tissue during severe viral respiratory infections.<sup>32</sup> Similarly, the present study found eosinophils are the most important feature, followed by CRP, as illustrated in the Boruta plot **FIGURE 2**.

A single routine laboratory test cannot accurately detect a COVID-19 case.<sup>33</sup> However, the combined utility of routine laboratory tests using ML methods can contribute to more accurate COVID-19 diagnoses, as shown in the present study. The turnaround time for routine laboratory tests is relatively shorter than that for rRT-PCR tests. For example, in our laboratory the mean turnaround times for CBC, CC, and rRT-PCR tests were 22 minutes, 41 minutes, and 8 hours, respectively. Furthermore, the total cost of routine laboratory tests is cheaper than rRT-PCR. While routine laboratory parameters' total reagent cost in the present study was only ~\$1 per analysis, the rRT-PCR's reagent market cost was \$5. Therefore, we can infer that ML models using routine laboratory test results can serve as cheaper and quicker clinical decision support tools for COVID-19 diagnoses.

When our ML models were first developed, their performances were evaluated using laboratory results from a single-centered university hospital. A limitation of the study was that our PCR-positive and PCR-negative groups in our data set were not age matched, even though they comprised 18- to 65-year-old adults.

**TABLE 3** Performances of Machine Learning Models for COVID-19 Prediction

Parameter/Model	PPV, <sup>a</sup> %	Sensitivity, <sup>a</sup> %	Specificity, <sup>a</sup> %	F1 Score, <sup>a</sup> %	Accuracy, <sup>a</sup> %	AUC <sup>a</sup>	Validation Accuracy, <sup>b</sup> Mean ± SD	Cohen's κ <sup>a</sup>
CBC + CC								
Random forest	90.40	79.58	91.24	84.64	85.30	0.925	84.98 ± 3.06	0.707
XGBoost	91.82	71.13	93.43	80.16	82.08	0.887	85.53 ± 2.10	0.643
SVM	87.02	80.28	87.59	83.52	83.87	0.890	81.83 ± 2.85	0.678
Logistic regression	80.42	80.99	79.56	80.70	80.29	0.883	80.39 ± 2.44	0.606
CBC								
Random forest	85.07	80.28	85.40	82.61	82.80	0.881	83.63 ± 3.10	0.656
XGBoost	83.21	76.76	83.94	79.85	80.29	0.867	82.73 ± 3.39	0.606
SVM	81.69	81.69	81.02	81.69	81.36	0.866	80.84 ± 2.94	0.627
Logistic regression	79.45	81.69	78.10	80.56	79.93	0.874	79.67 ± 2.08	0.598

AUC, area under the curve; CC, clinical chemistry; PPV, positive predictive value; SVM, support vector machine.

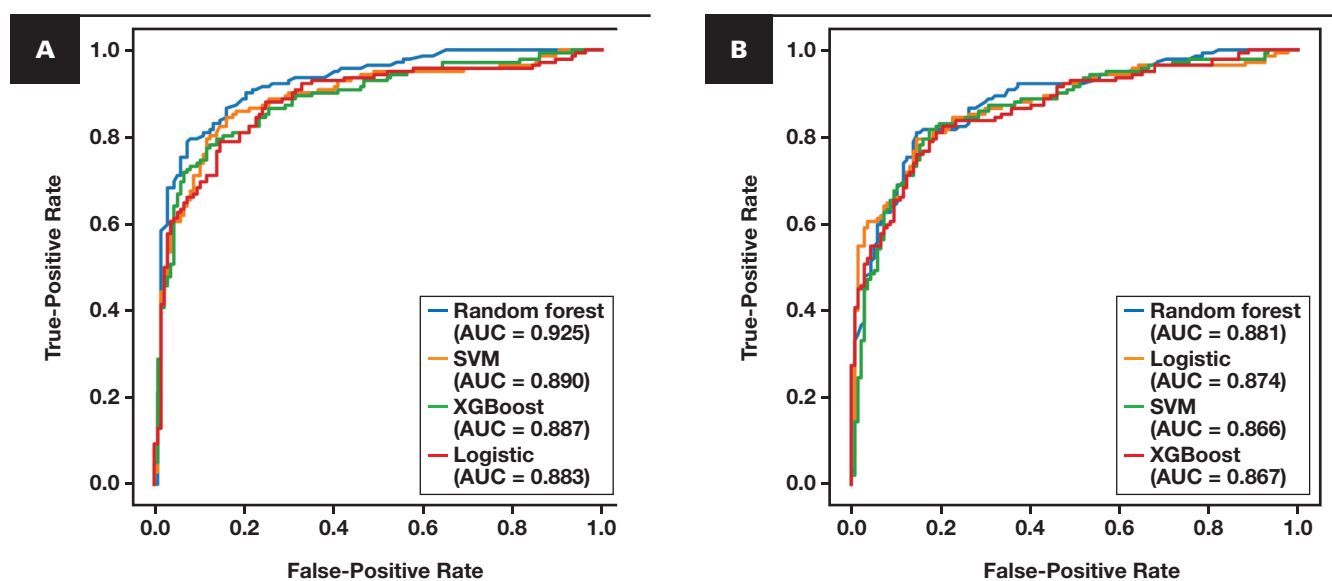
<sup>a</sup>Measures obtained from the evaluation of the models' performance on the test set.

<sup>b</sup>Performance measure obtained from 10-fold cross-validation.

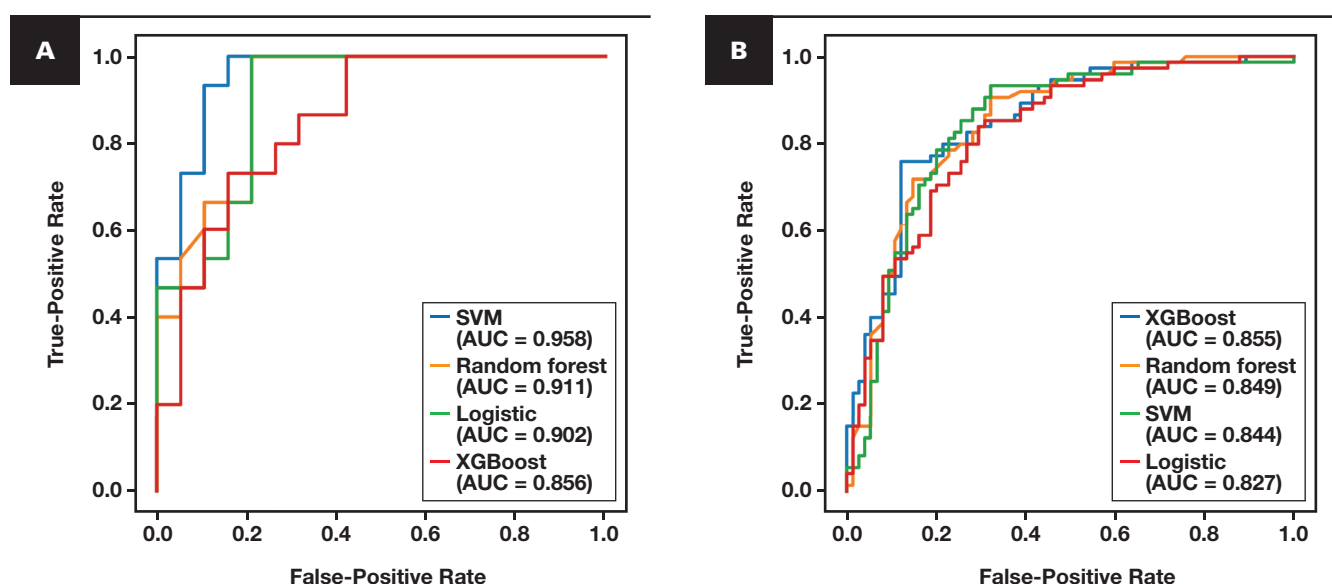
**TABLE 4** Performances of Machine Learning Models for COVID-19 Prediction on the São Paulo Data Set

Parameter/Model	PPV, %	Sensitivity, %	Specificity, %	F1 Score, %	Accuracy, %	AUC	Cohen's κ
CBC + CC							
Random forest	68.18	100.00	63.16	81.08	79.41	0.911	0.602
XGBoost	65.22	100.00	57.89	78.95	76.47	0.856	0.548
SVM	83.33	100.00	84.21	90.91	91.18	0.958	0.825
Logistic regression	71.43	100.00	68.42	83.33	82.35	0.902	0.657
CBC							
Random forest	73.91	90.67	68.00	81.44	79.33	0.849	0.587
XGBoost	69.57	85.33	62.67	76.65	74.00	0.855	0.480
SVM	73.68	93.33	66.67	82.35	80.00	0.844	0.600
Logistic Regression	68.37	89.33	58.67	77.46	74.00	0.827	0.480

AUC, area under the curve; CC, clinical chemistry; PPV, positive predictive value; SVM, support vector machine.



**FIGURE 3** Receiver operating characteristic curves of the machine learning (ML) models. **A**, ML models built using clinical chemistry and CBC parameters. **B**, ML models built using CBC parameters. AUC, area under the curve; SVM, support vector machine.



**FIGURE 4** Receiver operating characteristic curves of the machine learning (ML) models on São Paulo data set. **A**, ML models built using clinical chemistry and CBC parameters. **B**, ML models built using CBC parameters. AUC, area under the curve; SVM, support vector machine.

In addition, the COVID-19–positive group was sicker than the historic outpatient group regarding pulmonary symptoms, as inferred from **TABLE 1**. On the other hand, stomachaches and other symptoms were more prevalent in the control group **TABLE 1**. Furthermore, the study patients' comorbidities were not available. Hence, we could not confirm that our PCR-positive and PCR-negative groups were congruent, which may have affected the initial performance evaluation. Second, while our PCR-negative group's data were gathered from January 2018 to November 2019, the PCR-positive group included patients admitted from April 2020 to November 2020. Therefore, our PCR-positive group's data did not include the winter season, which may have confounded the study's outcome. Third, due to the reliance on rRT-PCR results, our study might have omitted false positives that came from rRT-PCR testing. In addition, our data set may have included predominantly mild/moderate cases due to the inclusion of outpatients whose ages were younger than 65 years. On the other hand, patients in the PCR-positive group had at least one symptom during admission. Thus, the present study's data did not include severe and asymptomatic patients with COVID-19. For these reasons, we externally evaluated our ML models' performance on the public data set from the Israelita Albert Einstein Hospital in São Paulo. The São Paulo data set also lacked detailed clinical-demographical characteristics and the exact laboratory results due to the standardized values presented for the sake of anonymity. However, the São Paulo data set included patients referred to the regular ward, semi-ICU, and ICU in addition to outpatients. Thus, we could externally evaluate our ML model's performance on the data set with more severe cases. Nevertheless, our ML models showed a satisfactory performance on the external validation data set. While the present study's data set includes only true negatives in the PCR-negative group, the external validation data set may contain false-negative

results due to the reliance on rRT-PCR tests. Hence, the external validation performance should be interpreted with this limitation. Furthermore, ML models presented in this study still need to be validated in asymptomatic cases.

The external validation data set used in the present study consisted of 34 and 150 concomitant patient results for CC and CBC and CBC parameters, respectively **TABLE 2**. Therefore, external validation studies of the proposed models should be performed on larger data sets to ensure the reliability of the developed ML models. On the other hand, the models' performance could be improved by incorporating additional input variables, such as medical imaging, symptoms, physical examination, vital signs, and increasing sample size. The absence of vaccinated subjects limits the present study. Moreover, none of the recent SARS-CoV-2 variants, such as Beta and Delta, were in circulation when the current study was conducted in Turkey.<sup>34</sup> The external validation data set suffered from the same limitation. Hence, ML models should be validated against vaccinated populations and more recent SARS-CoV-2 variants.

The false-negative results from rRT-PCR that originated from preanalytical errors are still an important problem in the fight against the pandemic.<sup>35</sup> We selected our PCR-negative group from prepandemic data for the ML models' training. Thus, the present study's data set was free of the false-negative results that originated from the preanalytical errors, such as inappropriate sampling, and they did not influence the ML models' development. The ML models presented in this study can be used as clinical decision support tools to contribute to physicians' clinical judgments on COVID-19 or direct them to offer repeat rRT-PCR testing in case of preanalytical error suspicions.

**Acknowledgments:** We thank Başkent University Faculty of Medicine Clinical Laboratory's personnel for their unseen endeavors in the COVID-19 pandemic.

## REFERENCES

- Sun J, He WT, Wang L, et al. COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. *Trends Mol Med*. 2020;26:483-495.
- Zhang JJ, Dong X, Cao YY, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy*. 2020;75:1730-1741.
- Lippi G, Plebani M. Procalcitonin in patients with severe coronavirus disease 2019 (COVID-19): a meta-analysis. *Clin Chim Acta*. 2020;505:190-191.
- Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis. *Clinica Chimica Acta*. 2020;506:145-148.
- 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.
- Tang N, Li D, Wang X, et al. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemost*. 2020;18:844-847.
- Thompson S, Bohn MK, Mancini N, et al; IFCC Taskforce on COVID-19. IFCC interim guidelines on biochemical/hematological monitoring of COVID-19 patients. *Clin Chem Lab Med*. 2020;58:2009-2016.
- World Health Organization (WHO). Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases: interim guidance, 14 January 2020 (No. WHO/2019-nCoV/laboratory/2020.2). World Health Organization. Available at <https://apps.who.int/iris/bitstream/handle/10665/332300/WHO-2019-nCoV-laboratory-2020.2-eng.pdf>.
- Caruana G, Croxatto A, Coste AT, et al. Diagnostic strategies for SARS-CoV-2 infection and interpretation of microbiological results. *Clin Microbiol Infect*. 2020;26:1178-1182.
- Hong KH, Lee SW, Kim TS, et al. Guidelines for laboratory diagnosis of coronavirus disease 2019 (COVID-19) in Korea. *Ann Lab Med*. 2020;40:351-360.
- Lippi G, Simundic AM, Plebani M. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). *Clin Chem Lab Med*. 2020;58:1070-1076.
- Wang P, Anderson N, Pan Y, et al. The SARS-CoV-2 outbreak: diagnosis, infection prevention, and public perception. *Clin Chem*. 2020;66:644-651.
- Data4u E. Diagnosis of COVID-19 and its clinical spectrum. <https://www.kaggle.com/einsteindata4u/covid19>. Accessed July 6, 2021.
- Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw*. 2010;36:1-13.
- Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
- Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29:1189-1232.
- Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol*. 2011;2:1-27.
- Kim J-H. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal*. 2009;53:3735-3745.
- McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22:276-282.
- Van Rossum G, Drake FL. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace; 2009.
- R Foundation for Statistical Computing. R: a language and environment for statistical computing. 2010. <http://www.R-project.org>.
- Cabitz F, Campagner A, Ferrari D, et al. Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests. *Clin Chem Lab Med*. 2020;59:421-431.
- Tschoellitsch T, Dünser M, Böck C, et al. Machine learning prediction of SARS-CoV-2 polymerase chain reaction results with routine blood tests. *Lab Med*. 2021;52:146-149.
- Joshi RP, Pejaver V, Hammarlund NE, et al. A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J Clin Virol*. 2020;129:104502.
- Schwab P, DuMont Schütte A, Dietz B, et al. Clinical predictive models for COVID-19: systematic study. *J Med Internet Res*. 2020;22:e21439.
- Goodman-Meza D, Rudas A, Chiang JN, et al. A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity. *PLoS One*. 2020;15:e0239474.
- Yang HS, Hou Y, Vasovic LV, et al. Routine laboratory blood tests predict SARS-CoV-2 infection using machine learning. *Clin Chem*. 2020;66:1396-1404.
- Plante TB, Blau AM, Berg AN, et al. Development and external validation of a machine learning tool to rule out COVID-19 among adults in the emergency department using routine blood tests: a large, multicenter, real-world study. *J Med Internet Res*. 2020;22:e24048.
- Qin C, Zhou L, Hu Z, et al. Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in Wuhan, China. *Clin Infect Dis*. 2020;71:762-768.
- Henry BM, de Oliveira MHS, Benoit S, et al. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. *Clin Chem Lab Med*. 2020;58:1021-1028.
- Percopo CM, Dyer KD, Ochkur SI, et al. Activated mouse eosinophils protect against lethal respiratory virus infection. *Blood*. 2014;123:743-752.
- Stegeman I, Ochodo EA, Guleid F, et al; Cochrane COVID-19 Diagnostic Test Accuracy Group. Routine laboratory testing to determine if a patient has COVID-19. *Cochrane Database Syst Rev*. 2020;11:CD013787.
- Ergünay K, Kaya M, Serdar M, et al. A cross-sectional overview of SARS-CoV-2 genome variations in Turkey [published online April 28, 2021]. *Turkish J Biochem*.
- Çubukçu HC, Coşkun A. False negative results and tolerance limits of SARS-CoV-2 laboratory tests. *Pathog Glob Health*. 2021;115:137-138.