PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Prediction of COVID-19 severity using laboratory findings on admission: informative values, thresholds, ML model performance.
AUTHORS	Statsenko, Yauhen; Al Zahmi, Fatmah; Habuza, Tetiana; Gorkom, Klaus; Zaki, Nazar

VERSION 1 – REVIEW

REVIEWER	Robert Balshaw Centre for Healthcare Innovation University of Manitoba Canada
REVIEW RETURNED	09-Oct-2020

GENERAL COMMENTS	This paper presents an analysis of a cohort of patients who have presented as testing positive for SARS-CoV-2 and are followed in hospital until they can recover (confirmed test negative) OR are admitted to the ICU. The authors examine conventional lab and clinical parameters in an attempt to identify. This cohort has desirable characteristics for this work, with adequate size (n=560) and number of events (72 ICU admissions). However, the authors have inadequate documented their analytic methods and results. See for example, the TRIPOD guidelines (https://www.equator-network.org/reporting-guidelines/tripod-statement/) I hope that they will revise and resubmit. Specific comments: + The sample size (n=560) and the number of admissions to ICU (n=72) is not mentioned until Table 1. This must be more prominent in abstract and results. + Timing of samples for lab assessments. Insufficient attention has been paid to the timing of the lab assessments. There is no explanation until the hints buried in table 1 that this has been considered at all. (e.g., "adm", "min", "peak"). No explanation at all is provided in table 2, not even the barely informative labels from table 1. + Further, which measures were used when multiple assessments were made? The only mention of this appears to be that "Clinicians routinely use physical examination findings and laboratory parameters ... some of which may be repeated to monitor

	progression." How was this handled? I would guess they have used the first values available (at admission), but this is neither clear nor is the time window for these assessments specified. + A CONSORT Flow Diagram would helpful to assess the generalizability of these results. Critically, were *any* patients not tested nasally? Were any pts not tested despite symptoms? Were any not admitted despite positive or inconclusive test results? Were any excluded because their test was *not* nasopharyngeal? What about emergency admissions for other reasons (MI, stroke, pneumonia, COPD, etc.) + what is the improvement in ROC when a NN is trained for each individual predictor vs. the observed values of the individual predictor? Were they different? + Methods / Table 2: what test was used to assess the p-value for the AUC? The p-values here should correspond to those in table 1, assuming that the Mann-Whitney U test has been used in table 1. U and AUC are related arithmetically. https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test#Area-under-curve_(AUC)_statistic_for_ROC_curves + How did the ranking of the importance of the predictors evaluated by "averaging all ranking scores among classifiers" differ from the ranking based, e.g., on the Mann-Whitney U statistic? + Please describe the "predictive performance" of the model rather than the "classifier output quality". The methods described for "output quality" are adequate for this task. + Please compare the performance of the neural network trained with all the tests (what time points??) against a multiple logistic regression model fit* as a basic point of reference (* fit using both maximum likelihood and LASSO). + Below table 2, reference is made to "laboratory tests done at the admission"? Why is this the only place where a choice of assessments (min, peak, adm) is mentioned? Does this mean that Table 2 and all other descriptions referent to labs done "at the admission"? + Why are "precision recall curves" presented in the appendix when only ROC curves are mentioned elsewhere. Why switch to precision, recall, F1 score and support without explanation. Please choose a terminology and stick with it. + Please provide some sense of the complexity of the "three-layer dense NN." + what Sens/Spec levels are achieved by each of the lab test thresholds described in the conclusion? It is not clear how these thresholds were established. Youden's index? Max Sen at fixed Spec? Some sort of decision-theoretic balance of false positive and false negative error severities? Minor observations: + Abbreviations not defined: e.g., ARDS.

	+ The explanation of methods used for the third task are not clear. How was the AUC used to determine a threshold it is an integral of the ROC curve? + There is no need for percentages to be reported to 2 decimal places (e.g., 14.55%). This would apply to the text and figures as well as to the tables. + Table 1: what units for "duration of viral shedding"? How was this assessed?

REVIEWER	Akira Sato University of Tsukuba, Japan
REVIEW RETURNED	20-Oct-2020

GENERAL COMMENTS	The purpose of this study aimed to identify predictive biomarkers of COVID-19 severity and to justify the threshold values of them for the stratification of the risk of deterioration that would require the transfer to ICU. The authors customized supervised ML algorithm in terms of threshold value used to predict worsening. They concluded that the performance of the neural network trained with top valuable tests (APTT, CRP, and Fibrinogen) is admissible (AUC 0.86; CI 0.486 - 0.884; p <0.001) and comparable with the model trained with all the tests (AUC 0.90; CI 0.812 - 0.902; p<0.001). The author's manuscripts are actual and clinically relevant. However, several issues should be considered to assess the results in this paper. My comments are related to the following points: 1) First of all, the severity of pneumonia on admission may affect prognosis but is not included in the analysis. Please explain why you did not include the X-ray information (presence or severity of pneumonia) in this analysis. 2) The performance of the neural network to predict the future deterioration out of the top three valuable tests (APTT, CRP, and Fibrinogen) is admissible (AUC 0.86; CI 0.486 - 0.884; p <0.001). What is the predictive sensitivity and specificity of ICU admission with these three combinations? 3) There is no description about specific treatment for corona infection after hospitalization. Please specify the drug used for corona infection. 4) Please discuss more about this result with previously reported and useful biomarkers like D-dimer.

VERSION 1 – AUTHOR RESPONSE

No	Task	Comment
	Reviewer: 1	

1	The authors have inadequate documented their analytic methods and results. See for example, the TRIPOD guidelines (https://www.equator-network.org/reporting-guidelines/tripod-statement/) ● Along with your revised manuscript, please include a copy of the TRIPOD checklist indicating the page/line numbers of your manuscript where the relevant information can be found (http://www.equator-network.org/reporting-guidelines/tripod-statement/)	By following the recommendations we improved the documentation of the analytic methods and results. Along with the revised manuscript, we submit a copy of the TRIPOD checklist indicating the page, line number, and the subsections of the manuscripts where the relevant information can be found. The line numbers can be checked in the 'TRIPOD_llineNumbers.pdf' file.
2	The sample size (n=560) and the number of admissions to ICU (n=72) is not mentioned until Table 1. This must be more prominent in abstract and results	In the second draft of the manuscript, the sample size is given in the abstract and in the Results section.
3	Timing of samples for lab assessments. Insufficient attention has been paid to the timing of the lab assessments. There is no explanation until the hints buried in table 1 that this has been considered at all. (e.g., "adm", "min", "peak"). No explanation at all is provided in table 2, not even the barely informative labels from table 1.	Thank you so much for the comment. To make the data presentation more clear, we changed the labels of the tables. We also put a legend at the bottom of Table 1 to point out at which time moment either this or that data was acquired.

3	Further, which measures were used when multiple assessments were made? The only mention of this appears to be that "Clinicians routinely use physical examination findings and laboratory parameters ... some of which may be repeated to monitor progression." How was this handled? I would guess they have used the first values available (at admission), but this is neither clear nor is the time window for these assessments specified.	Sorry for the confusing formulation that we used in the first draft. The cited sentence was in the Introduction section, and it didn't correspond directly to the study we did. No to mix up the readers, we made the following changes. First, to make the statement with the hypothesis of the study clear, we corrected the beginning of **subsection 1.2**: " Clinicians routinely use physical examination findings and laboratory parameters for risk stratification and hospital resources management. Commonly, each laboratory test kit has the only cut-off value to segregate the normal status from a pathology. We believe that threshold values should be re-adjusted for each disease rather than used as a common cut-off value for all pathologies. As a standard of care, baseline blood tests and inflammatory markers are obtained on admission to the hospital. The proper approach for the risk assessment should allow physicians to forecast the patient's future worsening out of the initial findings on admission. This is what we intend to do by applying a machine learning approach to the predictors routinely used in clinical practice. There are some promising data for the following set of prognostic biomarkers of COVID-19 severity." Second, in **subsection 3.2,** there is one more statement on the timeframe for the sample collection. "To address the first task, we studied the separability of laboratory findings values at the admission to Dubai Mediclnic concerning the future transfer of the patient to the ICU department." Third, we edited the table labels for a clear presentation of the idea of the study. "Table 1. Comparison of the patients hospitalized to intensive care unit concerning the COVID-19 outcomes: comorbidities, the result of physical examination on admission, laboratory findings on admission and deterioration (e.g., peak or minimal values), ethnicity, and disease course features" "Table 2. Statistical significance of ROC AUC for predicting the transfer to ICU out of the laboratory findings on admission" "Table 3. Justification of the cut-off levels for the admission values of laboratory findings to predict the transfer to ICU"

4	A CONSORT Flow Diagram would helpful to assess the generalizability of these results. Critically, were *any* patients not tested nasally? Were any pts not tested despite symptoms? Were any not admitted despite positive or inconclusive test results? Were any excluded because their test was *not* nasopharyngeal? What about emergency admissions for other reasons (MI, stroke, pneumonia, COPD, etc.)	In the second version of the manuscript, we put a clear explanation of the time frame for data acquisition. Now we state clearly a single time point at which we collect the data for the prediction (e.g., on admission). Also, there is a single event that we try to predict (e.g., the transfer of the patient to ICU, it could happen at any time while staying in the hospital). We presume that this provides an unambiguous understanding of the study design. We prepared a consort flow diagram as well (see Figure 1). The inclusion criteria were applied to each case with no exception. This was also true for the verification of the diagnose ("SARS-CoV-2 positive real-time reverse-transcriptase polymerase chain reaction from nasopharyngeal swabs only, at our site"). This means that there was no patient who was not tested. - Were *any* patients not tested nasally? No, all patients had nasopharyngeal swabs only - that is the only test offered by our hospital for SARS-COV-2 PCR - Were any pts not tested despite symptoms? It's unlikely, as we had a low threshold to test patients, some patients come for testing only because someone at their workplace was diagnosed with COVID-19 even if they did not have close contact with them, and they will get tested, so anyone with symptoms would have been tested. - Were any not admitted despite positive or inconclusive test results? Patients with positive PCR were definitely admitted regardless of their symptom severity. Patients who were highly suspected (either because of classic symptoms of anosmia or classic CT chest findings) were also admitted and treated as positive cases; however, we excluded these cases from our cohort. - Were any excluded because their test was *not* nasopharyngeal? Not applicable, as all our PCR tests were Nasopharyngeal samples. - What about emergency admissions for other reasons (MI, stroke, pneumonia, COPD, etc.)? All patients admitted during the study period got a SARS-COV-2 PCR by nasopharyngeal sample regardless of the cause of admission (surgeries, labor, MI, etc.).

5	what is the improvement in ROC when a NN is trained for each individual predictor vs. the observed values of the individual predictor? Were they different?	We added **a column for AUC values to Table 3** to show the AUC for each individual predictor utilizing threshold moving or percentile-based heuristic technique. The AUC values, which are higher than the AUC obtained for the observed values, are highlighted in bold.
6	Methods / Table 2: what test was used to assess the p-value for the AUC? The p-values here should correspond to those in table 1, assuming that the Mann-Whitney U test has been used in table 1. U and AUC are related arithmetically. https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test#Area-under-curve_(AUC)_statistic_for_ROC_curves	We calculated 95% CI for AUC scores with the bootstrap technique (1000 times) and p-values with permutation tests (1000 times). A permutation test was used to test against chance performance. The test technically went over all permutations of our observation sequence and evaluated our AUC with the permuted target values. It is clear that U test evaluates the null hypothesis that it is "equally likely that a randomly selected value from one sample will be less than or greater than a randomly selected value from a second sample.", whereas AUC reflects a similar probability, that a randomly chosen positive case will receive a higher score from our model than a randomly chosen negative case. However, different classifiers can put predicted labels in a different order; therefore, the ROC AUC scores and their p-values may vary. Meanwhile, we did not calculate U statistics out of the predicted values (in this case, p-values should be identical) but rather from the initial database; that is why p-values in Tables 1 and 2 do not correspond to each other.
7	How did the ranking of the importance of the predictors evaluated by "averaging all ranking scores among classifiers" differ from the ranking based, e.g., on the Mann-Whitney U statistic?	To address the reviewer's questions, we provided the following details in subsection 4.2 of the revised manuscript: We used 4 tree-based models to rank the features. Tree-based models provide a measure of feature importance based on the mean decrease in impurity (MDI). Impurity is quantified by the splitting criterion of the decision trees (Gini, Entropy, or Mean Squared Error). Therefore, we ranked features using Random Forest, AdaBoost, Gradient Boosting, and Extra Trees classifiers and then averaged all ranking scores among the aforementioned four classifiers. Mann-Whitney U statistic determines if two independent samples were selected from populations having the same mean rank, working with one variable at a time.

8	+ Please describe the "predictive performance" of the model rather than the "classifier output quality". The methods described for "output quality" are adequate for this task.	"The performance of the classification models" subsection is rewritten as follows: All the applied ML algorithms trained with stratified 10-folds cross-validation technique. The performance of the classification models such as Gradient Boosting, AdaBoost, ExtraTrees, Random Forest, NN, Logistic regression with and without L1 regularization is presented in Figure 3 in Appendix A. It displays all 560 test points concatenated from test (actual and predicted) label values for each fold. Tables 5-7 are composed of the performance metrics obtained by the NN model with the highest output quality. Figure 4 displays the ROC curves and AUC for the NN model with different variables, observed at admission, as predictors. In contrast, Figure 5 illustrates the performance's quality for the binary data obtained by using the threshold moving or percentile-based heuristic approach.
9	+ Please compare the performance of the neural network trained with all the tests (what time points??) against a multiple logistic regression model fit* as a basic point of reference (* fit using both maximum likelihood and LASSO).	Figure 3 is updated by adding logistic regression models with maximum likelihood estimation and with L1 regularization. The NN shows higher performance in terms of AUC. All models are trained with the data points obtained on admission.
10	+ Below table 2, reference is made to "laboratory tests done at the admission"? Why is this the only place where a choice of assessments (min, peak, adm) is mentioned? Does this mean that Table 2 and all other descriptions referent to labs done "at the admission"?	Thank you for the comment. We totally agree that the name of the variable "SOFA on admission" looked confusing as other variables were collected at the same time moment as well. In the second draft of the manuscript. we removed "on admission" from this reference.
11	+ Why are "precision recall curves" presented in the appendix when only ROC curves are mentioned elsewhere. Why switch to precision, recall, F1 score and support without explanation. Please choose a terminology and stick with it.	We apologize for this mistake and any inconvenience it may cause. Figure 3 is changed to the ROC curves.
12	+ Please provide some sense of the complexity of the "three-layer dense NN."	We utilized three-layer NN with the following configuration of hidden layers (35,30,10) and the stochastic gradient descent optimizer. The learning rate hyperparameter of the model was assigned to 0.1. The model was also regularized using an L2 penalty with a 0.0001 alpha value. NN was trained for a maximum of 250 epochs or before converged. By convergence here, we mean when the loss

		function is not improving by at least 0.0001 for 10 consecutive iterations.
13	+ what Sens/Spec levels are achieved by each of the lab test thresholds described in the conclusion? It is not clear how these thresholds were established. Youden's index? Max Sen at fixed Spec? Some sort of decision-theoretic balance of false positive and false negative error severities?	The sensitivity and specificity of each lab test were specified in Table 3. We used two approaches to establish the threshold values, the first one was based on Youden's index (Threshold moving technique column in Table 3), the second one - on the percentile level (setting the cut-off level to the 25th percentile for lymphocyte count and the 75th - for other features). The correspondent explanation is now in the 'Methods used' section. Please, kindly look at the paragraph which starts with the words "To tackle the third task... "
14	Abbreviations not defined: e.g., ARDS.	After checking them up we added the explanation to the following abbreviations: ARDS - acute respiratory distress syndrome CoV - coronavirus BMI - body mass index GCS - Glasgow coma scale MERS - Middle East respiratory syndrome RR - respiratory rate SOFA - Sequential organ failure assessment
15	The explanation of methods used for the third task are not clear. How was the AUC used to determine a threshold it is an integral of the ROC curve?	We apologize for the unclearly formulated methods in the third task. It is reformulated to explain that we used two approaches: (1) a Youden's index (or the threshold moving technique) which is used in conjunction with ROC (the index is defined for all points of a ROC curve, and the maximum value of the index is used as a criterion for selecting the optimum cut-off point), (2) a heuristically chosen percentile-based cut-off level (setting the cut-off level to the 25th percentile for lymphocyte count and the 75th - for other features).
16	There is no need for percentages to be reported to 2 decimal places (e.g., 14.55%). This would apply to the text and figures as well as to the tables.	Let us argue that this does not affect the paper's readability while showing the reported percentages' accuracy. Therefore, we kindly ask you to leave the percentage values as they are.

17	Table 1: what units for "duration of viral shedding"? How was this assessed	The duration of viral shedding is the number of days when the virus is detected in the nasopharyngeal swabs. To make this evident we mentioned the unit (e.g., days) in Table 1 while preparing the second draft of the manuscript. We also put the definition into the methodology section (page 3, lines 240-254). We used the same assessment as in a study on the influenza virus (see the reference below). All the patients hospitalized to the Mediclinics hospital were subject to the regular collection of nasopharyngeal swabs by a standard technique. Furthermore, after the patient stopped presenting symptoms, the specimen collection continued on a daily basis until two subsequent negative PCR tests for COVID-19 more than 24 h apart. For patients who were discharged early, specimen collection was arranged on an outpatient basis. The duration of viral shedding was the number of days from the disease onset when the diagnosis was confirmed to the first negative PCR test. Reference: Lee, N., Chan, P. K., Hui, D. S., Rainer, T. H., Wong, E., Choi, K. W., ... & Chu, I. M. (2009). Viral loads and duration of viral shedding in adult patients hospitalized with influenza. The Journal of infectious diseases, 200(4), 492-500.

	Reviewer: 2	
18	1) First of all, the severity of pneumonia on admission may affect prognosis but is not included in the analysis. Please explain why you did not include the X-ray information (presence or severity of pneumonia) in this analysis.	We analyzed the cases admitted to the clinics at the beginning of the pandemics. So, It was impossible to implement a common protocol of radiologic examination (e.g., chest X-rays, lung CT) for the patients. In the retrospective analysis that we did, the radiological findings were done within different time intervals after the admission and in diverse body position (e.g., prone vs supine). Severe patients were most likely to get the chest X-ray in the supine position or lung CT with no other radiogram. The acquisition parameters for the X-ray and CT examinations also varied, mainly, because some patients came from other medical facilities. Furthermore, for some patients physicians didn't order the lateral chest X-ray radiogram. For these reasons, the idea of our study was to estimate the predictive value of the laboratory findings on admission. While researching on this issue we managed to find and justify the non-radiologic predictors of worsening of the patients. We obtained the results that are suitable for future studies on forecasting the course of COVID-19. The pandemics precede and the clinics worked out standards of diagnostics. Supposedly, with the novel study cohorts, one may work our more accurate prognosis and stratify personal risk factors out of both the radiologic findings and the non-radiologic ones. The evidence on the disease progression from our study will contribute to the development of other machine learning algorithms.
19	2) The performance of the neural network to predict the future deterioration out of the top three valuable tests (APTT, CRP, and Fibrinogen) is admissible (AUC 0.86; CI 0.486 - 0.884; p <0.001). What is the predictive sensitivity and specificity of ICU admission with these three combinations?	To address the reviewer's questions, we provided the following details in subsection 4.3 of the revised manuscript: The sensitivity and specificity are 0.9877 and 0.4028, respectively, for the top three valuable tests (APTT, CRP, and Fibrinogen). It raises to 0.9754 and 0.75 respectively for all 13 significant tests.

20	3) There is no description about specific treatment for corona infection after hospitalization. Please specify the drug used for corona infection.	By following the reviewer's comments, we added the following paragraph at the end of subsection "3.1 Study sample": The treatment was done in full accordance with national guidelines. The indication for the supportive oxygen therapy was either the oxygen saturation level below 94\% or the respiratory rate (RR) above 30 breaths per minute or both of them. In case of suspicion to superimposed bacterial pneumonia, physicians ordered empirical broad-spectrum antibiotics. The administration of the antiviral and antimalarial drugs followed "National Guidelines for Clinical Management and Treatment of COVID-19" (available at https://www.dha.gov.ae/en/HealthRegulation/Documents/National_Guidelines_of_COVID_19_1st_June_2020.pdf)
21	4) Please discuss more about this result with previously reported and useful biomarkers like D-dimer.	By following the suggestion of the reviewer, we extended the literature review. We tried to elaborate on the pathophysiologic causes of rising the levels of such informative biomarkers as D-dimer and Fibrinogen in COVID-19 patients. Please, kindly see below: D-Dimer. A common finding in most COVID-19 patients is high D-dimer levels (over 0.28mg/L), which are associated with a worse prognosis [12, 3]. An exceptional interest of physicians in this biomarker comes from the fact that the vast majority of patients deceased from COVID-19 fulfilled the criteria for diagnosing the disseminated intravascular coagulation. Because of this, the incidence of pulmonary embolism in COVID-19 is high. In this condition, the D-dimer concentration will definitely rise up because it is a product of degradation of a blood clot formed out of fibrin protein [13]. Thromboembolic complications explain the association of low levels of platelets, increased levels of D-dimer, and increasing levels of prothrombin in COVID-19 [14]. Alternatively, the D-dimer level may go up as a direct consequence of SARS-CoV-2 itself [15]. Reasonably, laboratory hemostasis may provide an essential contribution to the COVID-19 prognosis and therapeutic decisions [16]. Researchers tried to forecast the severity of COVID-19 with D-dimer as a single predictor. They showed that D-dimer level>0.5mg/L had a 58% sensitivity, 69% specificity in the forecast of the disease severity [17]. In another study, a D-Dimer level of>2.14mg/L predicted in-hospital mortality with a sensitivity of 88.2% and specificity of 71.3% [18]. One more study highlighted that a D-dimer threshold of>2.66mg/L detected all patients with a pulmonary

		embolus on the chest CT[15]. So, the high levels of D-dimer are a reliable prognostic biomarker of in-hospital mortality. Fibrinogen. In COVID-19 patients admitted to ICU for acute respiratory failure, the level of fibrinogen is significantly higher than in healthy controls (517±148 vs. 297±78 mg/dL)[12]. The small vessel thrombi revealed on autopsy in lungs, and other organs suggest that disseminated intravascular coagulation in COVID-19 results from severe endothelial dysfunction, driven by the cytokine storm and associated hypoxemia. As standard dose deep vein thrombosis prophylaxis cannot prevent the consumptive coagulopathy, monitoring D-dimer and fibrinogen levels are required. This will promote the early diagnostics of hypercoagulability and its treatment with direct factor Xa inhibitors [14, 19].

VERSION 2 – REVIEW

REVIEWER	Robert Balshaw Centre for Healthcare Innovation, University of Manitoba Canada
REVIEW RETURNED	16-Dec-2020

GENERAL COMMENTS	Minor issues only: + The first paragraph of Sec 3.1 mentions "... who fit the criteria of eligibility mentioned above..." but I believe the eligibility criteria are provided *below* (next paragraph). + last line of par 0 on page 4: spelling correction "... and had SpO2 value >>not<< less than 94%." + last line of par 1 of sec 4.3: what are PR AUC scores? (in "... 95% CI for ROC and PR AUC scores...")

REVIEWER	Akira Sato University of Tsukuba
REVIEW RETURNED	03-Dec-2020

GENERAL COMMENTS	There are no comments for your revision. Thank you for the change and the explanations.

VERSION 2 – AUTHOR RESPONSE

(1) The name of the article was changed into “Prediction of COVID-19 severity using laboratory findings on admission: informative values, thresholds, ML model performance.”
(2) Dr. Fatma Al Zahmi is mentioned as the second correspondent author. This is because her contribution to initiating the research and the dataset collection was substantial and she is aware of many research details.
(3) From the list of Abbreviations we excluded “PR - precision-recall” which is not used anymore in the text.
(4) In the first paragraph of Sec 3.1б we changed "...mentioned above..." to "...mentioned below...".
(5) In the last line of par 0 on page 4 we did spelling correction "... and had SpO2 value >>not<< less than 94%."
(6) In the last line of par 1 of sec 4.3: (in "... 95% CI for ROC and PR AUC scores...")
we deleted the “PR“ abbreviation, which was left by mistake from the first version of the manuscript.
(7) We updated sec 9 by putting the sentence: “This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.” It was the same at the time of the initial manuscript submission. Unfortunately, the pandemic resulted in a budget shortage, and the college stopped paying publication fees.
(8) We added one more sentence to the last line of sec 12 “To assess the risk of having complications in a patient with COVID-19, one may use the ML-based free online tool at https://med-predict.com which illustrates the results of the current study.“
(9) The legends of Figure_5 and Figure_6 (Appendix Figure 3, Appendix Figure 3 correspondingly) are corrected: the variables are listed with regards to their informational gain value, rather than in the alphabetic order as it was before. We presume that this will simplify the understanding of the article.