BMJ Open

Threshold values of the biomarkers predictive of COVID-19 severity

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-044500
Article Type:	Original research
Date Submitted by the Author:	07-Sep-2020
Complete List of Authors:	Statsenko, Yauhen; United Arab Emirates University College of Medicine and Health Sciences, Radiology Al Zahmi, Fatmah ; Mediclinic Parkview Hospital, Neurology; Mohammed Bin Rashid University Of Medicine and Health Sciences Habuza, Tetiana; UAE University College of Information Technology, Department of Computer science Gorkom, Klaus; United Arab Emirates University College of Medicine and Health Sciences, Radiology Zaki, Nazar; United Arab Emirates University,
Keywords:	COVID-19, BIOTECHNOLOGY & BIOINFORMATICS, INFECTIOUS DISEASES, Respiratory infections < THORACIC MEDICINE, Information technology < BIOTECHNOLOGY & BIOINFORMATICS, Biochemistry < NATURAL SCIENCE DISCIPLINES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Threshold values of the biomarkers predictive of COVID-19 severity

Yauhen Statsenko[a], Fatmah Al Zahmi[b], Tetiana Habuza[c], Klaus Neidl Van Gorkom[a], Nazar Zaki[c]

[a]*College of Medicine and Health Sciences, United Arab Emirates University, P.O. Box 17666, Al Ain, United Arab Emirates*
[b]*Mediclinic Middle East Parkview hospital, P.O. Box 51122, Dubai, United Arab Emirates*
[c]*College of Information Technology, United Arab Emirates University, P.O. Box 15551, Al Ain, United Arab Emirates*

Background: Despite the necessity, there is no reliable biomarker to predict disease severity and prognosis of COVID-19 patients. The currently published prediction models are not fully applicable to clinical use.

Objectives: 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.

Methods: The study cohort included all consecutive patients admitted to Dubai Mediclinic Parkview hospital from February to May 2020 with COVID-19 confirmed by the polymerase chain reaction. The challenge of finding the cut-off thresholds was the imbalanced dataset (e.g., the disproportion in the number of patients admitted to ICU versus non-severe cases). Therefore, we customized supervised ML algorithm in terms of threshold value used to predict worsening.

Results: With the default thresholds returned by the ML estimator, the performance of the models was low. It was improved by setting the cut-off level to the 25^{th} percentile for lymphocyte count and the 75^{th} - for other features. The study justified the following threshold values of the laboratory tests done at the admission: lymphocyte count lower than $2.59x10^9$/L, and the upper levels for total bilirubin - 11.9 umol/L, ALT - 43 U/L, AST - 32 U/L, D-Dimer - 0.7 ug/mL, APTT - 39.9 sec, CK - 247 U/L, CRP - 14.3 mg/L, LDH - 246 U/L, Troponin - 0.037 ng/mL, Ferritin - 498 ng/mL, Fibrinogen - 446 mg/dL.

Conclusion: 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).

Keywords: COVID-19 pandemic, coronavirus, severity, biomarkers, threshold values, infectious disease

Strength and limitations of the study

- The research is based on a unique study cohort that is representative of the entire population because of the National Standard that required all patients with confirmed COVID-19 to be admitted to acute care hospitals regardless of their symptoms or illness severity.
- To distinguish the patients with the confirmed COVID-19 who may worsen while being treated, we justified threshold values of the laboratory tests done at the admission.
- The prediction of the future deterioration by the neural network is reliable even with the top three valuable laboratory tests (APTT, CRP, and Fibrinogen) being used for training (AUC 0.86;CI 0.486 - 0.884; p<0.001).
- The limitation of the study was the imbalanced dataset (e.g., the disproportion in the number of patients admitted to ICU versus non-severe cases).
- Machine learning shows high performance with each laboratory test taken as a predictor. The prognosis is almost accurate (AUC 0.998) in the model based on the combination of all the tests with demographic and clinical characteristics.

Abbreviations

ALT – alanine aminotransferase
AST - aminotransferase
AUC - area under the curve
CI - confidence interval
hs-CRP - high-sensitivity C-reactive protein
ICU - intensive care unit
IL - interleukin
ML - machine learning
NN - neural network
PC - precision-recall
PCR - polymerase chain reaction
PR - precision-recall
RNA - ribonucleic acid
ROC - receiver operating characteristic
SARS-CoV-2 - severe acute respiratory syndrome-related coronavirus 2
SOB - shortness of breath
TNF - tumor necrosis factor

Definitions

Mild level of COVID-19 severity - nonpneumonia and mild pneumonia.

Severe level of COVID-19 severity - dyspnea, respiratory frequency ≥30/min, blood oxygen saturation ≤93%, the partial pressure of arterial oxygen to fraction of inspired oxygen ratio <300, and/or lung infiltrates >50% within 24 to 48 hours.

Critical level of COVID-19 severity - respiratory, septic shock, and/or multiple organ dysfunction or failure.

1. Introduction

Despite the necessity, there is no reliable prognostic biomarker to predict disease severity and prognosis of COVID-19 patients [1]. Studies on COVID-19 have built up several types of prediction models. These have been the models designed to indicate the disease risk in the general population, the diagnostic models based on medical imaging, and the prognostic models. Unfortunately, these models have had some limitations that have precluded their use in clinical practice [2].

*Corresponding author. Tel.: +971 3 713 7124; E-mail address: e.a.statsenko@uaeu.ac.ae (Y. Statsenko).

1.1. Models using laboratory findings as the inputs

Researchers tried to establish the role of laboratory findings in the diagnosis of COVID-19 [3], i.e., they showed that the severe cases of COVID-19 were associated with D-dimer level over 0.28μg/L, interleukin (IL) 6 level over 24.3pg/mL [3], and LDH activity with an upper limit cut-off in the range of 240-255U/L [4]). However, the use of these laboratory parameters with the above mentioned cut-off values was limited for the following reasons. First, these studies were conducted on severe forms of the disease. There was limited research done on patients who were asymptomatic or had mild disease [3, 5]. Second, the whole spectrum of the regularly used clinical laboratory data is unavailable for non-severe patients. Thus, the published papers add justification on the diagnostic utility of separate laboratory findings, instead of working out reliable diagnostic criteria for a set of them.

Gong and colleagues [6] have generated a tool for the early prediction of severe COVID-19 pneumonia out of the following data: age, serum lactate dehydrogenase activity, C-reactive protein, the coefficient of variation of red blood cell distribution width, blood urea nitrogen, direct bilirubin, lower albumin. The resulting performance was not high (sensitivity 77.5%, specificity 78.4%) [6]. Supposedly, this is because the dataset used as the input consists of exceptionally the age and laboratory findings.

In another model, the inputs included basic information, symptoms, and the results of laboratory tests. After the feature selection, the number of key features was set to just three laboratory results: LDH, lymphocytes, and high-sensitivity C-reactive protein (hs-CRP). The model was trained with the follow-up studies of the general, severe, and critical patients [1]. By feeding ML algorithm with the results obtained at the time of admission and in follow-up studies, the authors worked out a decision rule to predict patients at the highest risk. However, physicians are interested in the early prediction of the disease outcomes, and it is highly disputable that the model will not loose its predictive potential if being applied exceptionally to the data received on admission.

We believe that a more accurate model can be built based on the simultaneous interpretation of laboratory results, clinical data, and physical examination findings (e.g., BMI, body temperature, respiratory rate) at the time of presentation. The analysis utilizing a machine learning algorithm could provide an accurate prediction of the disease severity.

1.2. Data used by clinicians for stratifying risks

Clinicians routinely use physical examination findings and laboratory parameters for risk stratification of their patients, some of which may be repeated to monitor progression. We believe that threshold values should be re-adjusted for a disease being treated, rather than having one threshold for all pathologies.

Inflammatory markers. There is evidence that IL-6, tumor necrosis factor-α do not indicate the level of COVID-19 progression [7]. Some markers of inflammation are elevated in the serum of COVID-19 patients compared to the healthy people, i.e., the serum SARS-CoV-2 viral load (RNAaemia) is closely correlated with drastically elevated interleukin 6 levels in critically ill COVID-19 patients [8]. However, there is no significant difference between severe and mild groups [7]. In contrast to this, the indicators are reflective in the progression of the diseases caused by other coronaviruses (e.g., MERS, SARS) [9]. This may be explained by the huge amino acid differences in viral proteins of distinct coronaviruses. Even with different MERS-CoV strains, common cytokine signaling by TNF and IL-1α results in the differential expression of innate immune genes [10].

Ferritin. Ferritin is a marker of iron storage. However, it is also an acute-phase reactant, the level of which elevates in processes of acute inflammation, whether infectious or non-infectious. Marked elevations have been reported in cases of COVID-19 infection [11].

D-Dimer. A common finding in most COVID-19 patients is high D-dimer levels (over 0.28μg/L), which are associated with a worse prognosis [12, 3].

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].

APTT. In a study conducted in February 2020, the levels of APTT as well as WBC, lymphocytes, AST, ALT, and creatinine, were not significantly different between severe and mild patients [3]. At the same time, other researchers showed no significant difference in APTT in survivors versus non-survivors [13]. According to the results of another study published in March 2020, no significant difference in APTT values were found in the severe cohort of patients versus the non-severe one [6]. The results obtained in another study in April in Italy were the same [12]. The common limitation of these early studies was a small sample size. Finally, a meta-analysis justified that the elevation of D-Dimer, rather than prothrombin time and APTT, reflects the progression of COVID-19 toward an unfavorable outcome [14].

LDH and CK. Increased levels of the enzymes may reflect the level of the organ damage in a systemic disease [15, 4]. Reasonably, they may serve as biomarkers for COVID-19 progression.

CRP. In the early stage of COVID-19, CRP levels are positively correlated with the diameter of lung lesions and severe presentation [16].

Liver enzymes and total bilirubin. COVID-19 leads to elevated liver biochemistries (e.g., the level of AST, ALT, GGT, total bilirubin) in over 50% of patients at admission. AST - dominant aminotransferase elevation reflects the disease severity and true hepatic injury [17, 18].

2. Objectives

We decided to identify predictive biomarkers of COVID-19 severity and to justify the threshold values of them. Hypothetically, the absolute values of the biomarkers at the admission to the clinics could provide physicians with an accurate prognosis on the future worsening of the patient that

would require the transfer of the individual to the intensive care unit (ICU). Getting a reliable tool for such a prognosis will support decision making and logistical planning in clinics.

To address the objective, we designed a set of the following tasks:

- to study the linear separability of the laboratory findings values in patients with confirmed COVID-19 who were transferred to ICU versus non-severe cases of the disease, and to make the comparative analysis of the ICU department cases (both the deceased and survived cohorts) with other patients with COVID-19.
- to identify the risk factors by selecting the most valuable features for predicting the deterioration that would require the transfer of the patient to ICU.
- to work out the threshold criteria for the major clinical data for the early identification of the patients with a high risk of being transferred to ICU.
- to identify the accuracy of the prediction of the patient's deterioration by the machine learning algorithm and by a set of the newly created threshold values of the laboratory and clinical findings.

3. Materials and methods

3.1. Study sample

The study sample includes all the patients with a diagnosis of COVID-19 verified by the polymerase chain reaction (PCR) in Dubai Mediclinic from 24th February to 1st July 2020. Using this sample meets the intention of the study: to allow for the early prognostic stratification.

The inclusion criteria are as follows: age 18 years or older; inpatient admission; SARS-CoV-2 positive real-time reverse-transcriptase polymerase chain reaction (PCR) from nasopharyngeal swabs only, at our site. Those patients who met the inclusion criteria for our studies were included in the study sample. All the patients were discharged at the time of writing the paper.

The remarkable feature of our study is that at the beginning of the pandemic, all the COVID-19 verified by PCR were hospitalized in the Mediclinic even if they did not present any symptoms. We observed many mild and asymptomatic forms of the disease, with all the required spectrum of analyses being conducted. All patients who were hospitalized stayed in Dubai Mediclinic until they were afebrile for more than 72 h and had SpO2 value non less than 94%. They were discharged after two consecutive negative PCR tests for COVID-19, more than 24 h apart.

3.2. Methods used

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.

To make the comparative analysis of features with regard to the transfer to ICU, we utilized a set of non-parametric tests. The relationships involving two variables were assessed with the Mann-Whitney U test or Kruskal-Wallis test for the continuous features, and with Fisher's Exact test or Chi-square test for the quantitative ones. Data were expressed as *IQR*, *mean* ± *std* or number of cases, and their percentage.

To address the second task, we used a set of different methods. First, we trained the NN ML model on each variable separately. We assessed their statistical significance against chance performance to come up with laboratory data cut-off levels, which may be considered as bookmakers of severe course of the disease. We calculated 95% CI for ROC and PR AUC scores with the bootstrap technique and p-values with permutation tests.

Second, we used ML tree-based methods (AdaBoost, Gradient Boosting, Random Forest, and Extra Trees) to check if there were unique patterns within the data that could unambiguously identify the event of transferring the patient to ICU from the data obtained at the admission. The list of features used as predictors is displayed at the top of Appendix A. To assess the importance of the variables, we ranked all features concerning their impurity-based predictive potential by averaging all ranking scores among classifiers.

To tackle the third task, we used a threshold moving technique [19] or a heuristically chosen percentile-based cut-off level along with supervised ML classification model (NN). The problem of predicting the transfer to ICU had a severe class imbalance. Therefore, we needed to focus on the performance of the classifier on the minority *class* (admitted to ICU patients). The ROC AUC was used as a measure to find the optimal threshold for the ROC curve for each significant laboratory finding. These threshold values allowed us to find the optimal cut-off level for each laboratory test results.

To evaluate the classifier output quality, we trained several ML classification models using a stratified 10-fold cross-validation technique to generalize the models to the true rate error. For each fold, we used 90% of the data to train the model and then tested it with the rest 10%. The decision matrices built on the test dataset for all folds were combined and used to calculate the performance metrics.

4. Results

4.1. Comparison of the ICU vs. non-ICU patients

The problem of predicting an event of being admitted to ICU has a severe class imbalance. Therefore, we need to focus on the performance of the classifier on the minority class (admitted to ICU patients).

We look at the linear separability of the groups of numerical data composed from the laboratory findings values with regard to their quartiles. In Figure 1, boxplots for the laboratory findings data are presented with the red dashed line that marks the 75^{th} percentile for the subjects who were not transferred to ICU. The assumption is to use the third quartile (Q3) start point value as the threshold if there is clear separability between ICU and non-ICU groups. In each diagram in Figure 1, the red

line indicates the 75th percentile for not admitted to the ICU group. The exception is the diagram for the lymphocyte count, where it stands for the 25th percentile.

The results of the comparative analysis of features with regard to the transfer to ICU and the final outcomes of the disease are presented in Table 1. We excluded from further analysis the laboratory findings that didn't have a significant difference in the distribution of two groups. Therefore, we considered the list of 13 variables: WBC, lymphocyte count, total bilirubin, ALT, AST, D-Dimer, APTT, CK, CRP, LDH, Troponin, Ferritin, and Fibrinogen at admission.

4.2. Feature ranking with regard to ML model performance

The features of the dataset listed in Appendix A were ranked with four tree-based ML classifiers (e.g., Random Forest, AdaBoost, Gradient Boosting, and ExtraTrees). Averaged values of impurity-based attribute ranks were calculated as the mean of rank values for the algorithms mentioned above (see Figure 2 in Appendix A). The evaluation of the performance of the classifiers is in Figure 3 in Appendix A.

4.3. The cut-off levels of the laboratory findings

To come up with laboratory data cut-off levels, which may be considered as biomarkers of the severe course of the disease, we trained the NN ML model on each variable separately and assessed their statistical significance against chance performance. We calculated 95% CI for ROC and PR AUC scores with the bootstrap technique and p-values with permutation tests (see Table 2).

From Table 2, there is a significant difference between the performance of the model in terms of ROC AUC and the performance at the chance level. High-performance measures were obtained for APTT, CRP, and Fibrinogen values, so we also built the classification model based on the combination of these three features.

ML models were trained in the 10-folds stratified cross-validation manner and then ROC curves were built for the test data (combined from all 10 folds) as it is presented in Figure 4 at Appendix B.

To improve the model's efficiency and choose the cut-off value set for some laboratory findings data, we used a threshold moving technique along with a supervised ML classification model (NN).

The ML estimator assigns threshold values for interpreting probabilities. The default threshold returned by the estimator to class labels is 0.5, however, when the dataset is imbalanced, tuning this hyperparameter can improve the model's efficiency by finding the optimal threshold. This is crucial when the importance of predicting the positive class (admitted to ICU) out- weigh true negative predictions. Performance metrics calculated for all laboratory features with regard to the optimal threshold value are presented in Table 3. The table displays the sensitivity and specificity values obtained after applying the threshold moving technique. The optimal cut-off value returned by the technique is shown in the appropriate column.

Looking at the boxplots presented in Figure 1 we also decided to check the performance of the model when the cut-off level is set to the 25th percentile for lymphocyte count (values lower than or equal to the chosen level were set to 1, or 0 otherwise) and 75th for the other features (values higher or equal to the cut-off limit were set to 1, or 0 otherwise). The performance of the models with regard to the aforementioned cut-off levels is presented in Table 3.

The performance of the logistic regression model built on the binary data by applying the cut-off level for the threshold moving technique is shown in Figure 5a, for the percentiles' cut-off levels - in Figure 5b.

4.4. The performance of the classification models

All the features mentioned in Appendix A were used as models' predictors. The ranking scores of the predictors are listed in Table 4. The performance of the applied ML algorithms trained with stratified 10-folds cross-validation technique is presented in Tables 5 and 6.

5. Discussion

With the ML approach, we justify the cut-off thresholds for the major laboratory tests regularly done at admission.

The disproportion in the number of patients admitted to ICU versus non-severe cases was challenging. Therefore, we customized the ML algorithm in terms of threshold value used to predict worsening. For each laboratory findings feature, we fit the model to the training dataset using 10-fold cross-validation, then predicted the probabilities on the test dataset, and finally found the optimal threshold value which maximizes the ROC AUC measure.

By using the optimized threshold values (marked in bold font in Table 3), one can predict the supposed deterioration of the patient out of the initial findings at presentation. It is notable that some of the thresholds are close to the normal reference values, but not completely, i.e. the cut-off for CRP is 3 times bigger than the top reference value. It is challenging to interpret why the cut-offs for WBC and total bilirubin level is within the normal ranges for the indicators.

For better prediction, it is recommended that several biomarkers are analyzed concomitantly. A combination of three most valuable ones, if being feed to the deployed ML algorithm, provides a reliable prognosis.

6. Conclusion

- By comparing the data for the patients who were transported to ICU versus those who did not worsen throughout the hospitalization we selected a set of laboratory findings with the significant differences at the admission to the clinics. The variables were used as the predictors to build up the classification model. The performance of the models was low, with the default thresholds returned by the ML estimator, we improved it by setting the cut-off level to the 25th percentile for lymphocyte count and the 75th - for other features.

Table 1: The comparison of the patients hospitalized to intensive care unit with regard to the COVID-19 outcomes

		All patients				ICU patients		
		Total	Not admitted to ICU	Admitted to ICU	$p_{2\text{-}3}$	Dead	Discharged	$p_{4\text{-}5}$
		n_1=560	n_2=488 (87.14%)	n_3=72 (12.86%)		n_4=15 (20.83%)	n_5=57 (79.17%)	
Age		39.0[33.0-49.0]	38.0+11.97	51.0+13.08	**<0.0001**	46.0+12.56	62.0+11.01	**<0.0018**
Gender	female	189 (33.75%)	**175 (35.86%)***	**14 (19.44%)***	**<0.0072**	8 (14.04%)	6 (40.0%)	*0.06*
	male	371 (66.25%)	**313 (64.14%)***	**58 (80.56%)***		49 (85.96%)	9 (60.0%)	
Comorbidities	count	0.0[0.0-1.0]	0.0+1.04	1.0+1.22	**<0.0002**	1.0+1.15	0.0+1.45	0.4072
Current smoking		36 (6.43%)	34 (6.97%)	2 (2.78%)	0.2984	2(3.51%)		
Chronic cardiac disease		20 (3.57%)	15 (3.07%)	5 (6.94%)	0.1611	4 (7.02%)	1 (6.67%)	
Hypertension		115 (20.54%)	92 (18.85%)	23 (31.94%)	**<0.018**	18 (31.58%)	5 (33.33%)	1
Asthma		38 (6.79%)	31 (6.35%)	7 (9.72%)	0.3121	6 (10.53%)	1 (6.67%)	
Chronic kidney disease		7 (1.25%)	5 (1.02%)	2 (2.78%)		1 (1.75%)	1 (6.67%)	
Diabetes		98 (17.5%)	71 (14.55%)	27 (37.5%)	**<0.0001**	21 (36.84%)	6 (40.0%)	1
Active malignant cancer		6 (1.07%)	4 (0.82%)	2 (2.78%)		1 (1.75%)	1 (6.67%)	
BMI	adm	27.0[23.92-30.44]	26.84+5.44	28.0+4.54	**<0.01**	27.82+4.7	31.14+0.48	0.2575
Body temperature,°C	adm	37.0[37.0-37.9]	37.0+0.63	38.0+0.97	**<0.0001**	38.0+0.97	38.0+0.98	0.3925
HR BPM	adm	85.0[78.0-95.0]	84.5+12.32	94.5+19.97	**<0.0001**	95.0+20.93	85.0+15.3	0.1589
SBP	adm	124.0[114.0-135.0]	123.0+16.51	126.0+17.31	0.2092	129.0+16.29	120.0+20.58	0.2122
DBP	adm	78.0[70.0-84.0]	78.0+10.92	75.0+10.1	**<0.0208**	75.0+9.46	75.0+12.05	0.4254
RR /min	adm	18.0[18.0-18.0]	18.0+1.56	25.0+6.74	**<0.0001**	24.0+6.95	28.0+5.62	0.1336
SOFA score	adm	0.0[0.0-0.0]	0.0+0.75	3.0+2.85	**<0.0001**	3.0+2.42	4.0+3.69	**<0.0275**
WBC $x10^9/L$	adm	5.8[4.5-7.2]	5.65±2.68	7.35±5.21	**<0.0001**	7.4±5.34	7.0±4.68	0.3801
	min	5.5[4.1-7.2]	5.5±7.72	7.0±6.68	**<0.0008**	7.2±6.93	5.5±5.38	*0.0775*
Platelet $x10^9/L$	adm	224.0[180.25-272.0]	224.5+78.42	222.0+82.13	0.4102	225.0+86.02	196.0+57.76	*0.0516*
	min	224.0[178.0-272.0]	226.0+79.7	197.0+123.27	**<0.0049**	202.0+116.33	102.0+84.42	**<0.0001**
Lymphocyte $x10^9/L$	adm	1.56[1.06-2.1]	1.66+0.76	0.81+2.97	**<0.0001**	0.83+3.32	0.73+0.64	0.4806
	min	1.49[0.89-2.09]	1.6+0.8	0.49+3.64	**<0.0001**	0.5+4.07	0.38+0.62	0.1412
T.bilirubin (umol/L)	adm	9.0[6.0-12.6]	8.6+5.24	11.0+9.17	**<0.0001**	11.0+8.6	13.0+11.03	0.4094
	peak	9.85[6.5-14.38]	9.0+6.55	16.3+37.25	**<0.0001**	16.0+17.77	25.0+68.93	0.1412
ALT (U/L)	adm	28.0[17.25-47.75]	27.0+34.84	39.0+38.04	**<0.0001**	39.0+39.5	41.0+31.76	0.4889
	peak	32.0[19.0-67.75]	28.5+50.05	102.5+7266.58	**<0.0001**	99.0+114.51	289.0+15305.74	**<0.0495**
AST (U/L)	adm	24.0[18.0-36.22]	23.0+24.3	47.0+30.9	**<0.0001**	46.0+30.35	63.0+32.56	0.3722
	peak	25.5[19.0-44.0]	24.0+29.8	82.5+914.01	**<0.0001**	79.0+69.77	200.0+1715.26	**<0.0009**
D-Dimer (ug/L)	adm	0.4[0.2-0.6]	0.3+0.72	1.15+3.13	**<0.0001**	1.1+2.96	1.4+3.62	0.1638
	peak	0.4[0.3-0.7]	0.3+0.73	2.6+7.56	**<0.0001**	1.6+6.37	18.0+7.12	**<0.0001**
APTT (sec)	adm	37.4[35.0-41.05]	37.2+4.65	40.0+23.0	**<0.0014**	39.0+19.65	41.0+31.76	0.1429
	peak	38.0[35.15-42.35]	37.4+5.14	47.0+44.56	**<0.0001**	45.0+38.41	63.0+54.06	**<0.0005**
Creatinine (umol/L)	adm	76.1 [67.0-89.0]	75.4+27.52	80.5+54.62	*0.0767*	81.0+50.84	76.0+66.53	0.4448
	peak	78.0[67.78-91.0]	76.2+27.74	86.5+98.51	**<0.0001**	83.0+69.12	196.0+130.29	**<0.0003**
CK (U/L)	adm	106.0[66.0-173.0]	99.0+529.25	173.0+1168.65	**<0.0001**	174.0+1278.56	152.0+561.74	0.2269
	peak	109.5[66.75-199.75]	100.0+536.11	391.0+10621.26	**<0.0001**	391.0+11963.38	370.0+563.66	0.4855
CRP (mg/L)	adm	5.8[1.75-27.0]	4.2+32.27	101.0+105.14	**<0.0001**	102.0+102.19	100.0+115.53	0.4367
	peak	6.5[1.9-50.65]	4.8+45.93	157.5+113.35	**<0.0001**	143.0+108.72	219.0+115.19	**<0.0191**
LDH (U/L)	adm	192.0[159.0-264.0]	181.0+80.08	445.0+267.95	**<0.0001**	432.5+284.01	480.0+199.68	0.2706
	peak	194.0[160.0-280.0]	182.0+83.76	538.0+1232.13	**<0.0001**	490.5+302.93	1925.0+2039.83	**<0.0001**
Troponin (ng/mL)	adm	0.0[0.0-0.0]	0.0+0.15	0.0+1.31	**<0.0001**	0.0+0.04	0.0+2.73	*0.0598*
	peak	0.0[0.0-0.0]	0.0+0.18	0.04+1.85	**<0.0001**	0.0+0.26	0.36+3.66	**<0.0001**
Ferritin (ng/mL)	adm	216.7[84.5-475.5]	181.95+876.92	725.0+2282.55	**<0.0001**	882.0+2480.17	612.0+1214.49	0.3036
	peak	230.0[89.95-595.5]	196.5+1530.13	2258.0+9784.72	**<0.0001**	2063.5+4781.9	4669.0+15029.77	**<0.0014**
Fibrinogen (mg/dL)	adm	396.0[330.0-529.5]	377.0+187.31	610.0+199.71	**<0.0001**	612.0+204.96	567.0+179.01	0.3104
	peak	405.0[331.25-554.0]	380.0+130.61	700.0+735.07	**<0.0001**	701.0+816.38	692.0+252.63	0.1613
Clinical severity	asymp/mild	431 (76.96%)	**431 (88.32%)***	**0 (0.0%)***	**<0.0001**			
	severe	83 (14.82%)	**54 (11.07%)***	**29 (40.28%)***		**29 (50.88%)***	**0 (0.0%)***	**<0.0002**
	critical	46 (8.21%)	**3(0.61%)***	**43 (59.72%)***		**28 (49.12%)***	**15 (100.0%)***	
Ethnicity	White	60(10.71%)	53 (10.86%)	7 (9.72%)	0.1102	7 (12.28%)	0 (0.0%)	**<0.0219**
	S. Asians	244 (43.57%)	206 (42.21%)	38 (52.78%)		28 (49.12%)	10 (66.67%)	
	M. Easterns	148 (26.43%)	**136 (27.87%)***	**12 (16.67%)***		7 (12.28%)	5 (33.33%)	
	E.Asians	94 (16.79%)	79(16.19%)	15 (20.83%)		**15 (26.32%)***	**0 (0.0%)***	
	Others	14 (2.5%)	14 (2.87%)	0 (0.0%)				
Onset to hospitalization days		14.0[8.0-19.0]	12.0+7.07	22.0+16.5	**<0.0001**	21.0+17.72	27.5+10.25	0.1336
Onset to positive PCR days		2.0[1.0-5.0]	2.0+3.89	5.0+4.97	**<0.0001**	5.0+5.01	4.0+4.79	0.3425
High-risk group patients		41 (7.32%)	3 (0.61%)	38 (52.78%)	**<0.0001**	24(42.11%)	14 (93.33%)	**<0.0003**
Discharged alive		545 (97.32%)	488 (100.0%)	57 (79.17%)	**<0.0001**	57 (100.0%)		**<0.0001**
Length of stay in clinics		7.0[3.0-12.25]	6.0+8.25	16.0+16.08	**<0.0001**	16.0+17.34	23.0+9.97	0.1521
Duration of viral shedding		10.0[6.0-14.0]	10.5+5.64	8.0+9.04	0.0714	8.0+9.05	13.0+8.65	0.1304
Need for supplementary O2		82 (14.64%)	23 (4.71%)	59 (81.94%)	**<0.0001**	46 (80.7%)	13 (86.67%)	0.7229
Any complication		123 (21.96%)	53 (10.86%)	70 (97.22%)	**<0.0001**	55 (96.49%)	15 (100.0%)	1
ARDS		76 (13.57%)	7 (1.43%)	69 (95.83%)	**<0.0001**	54 (94.74%)	15 (100.0%)	1
Liver dysfunction		54 (9.64%)	23 (4.71%)	31 (43.06%)	**<0.0001**	23 (40.35%)	8 (53.33%)	0.3944

Table 2: Statistical significance of ROC AUC

No	Feature	AUC	CI		p-value
1	WBC	0.5727	[0.427	0.573]	0.035
2	Lymphocyte	0.5881	[0.474	0.588]	0.01
3	Total bilirubin	0.5573	[0.443	0.557]	0.077
4	ALT	0.5057	[0.482	0.538]	0.331
5	AST	0.4882	[0.399	0.595]	0.828
6	D-Dimer	0.6151	[0.5	0.615]	0.004
7	APTT	0.7534	[0.219	0.755]	<0.001
8	CK	0.6918	[0.6	0.725]	<0.001
9	CRP	0.8194	[0.798	0.822]	<0.001
10	LDH	0.5652	[0.515	0.644]	0.072
11	Troponin	0.6088	[0.5	0.609]	0.008
12	Ferritin	0.6973	[0.616	0.74]	<0.001
13	Fibrinogen	0.7704	[0.718	0.771]	<0.001
APTT+CRP+Fibrinogen		0.8618	[0.486	0.884]	<0.001
All together		0.9019	[0.812	0.902]	<0.001

- To distinguish the patients with the confirmed COVID-19 who may worsen while being treated we justified the following threshold values of the laboratory tests done at the admission: lymphocyte count lower than $2.59x10^9$/L, and the upper levels for total bilirubin - 11.9 umol/L, ALT 43 U/L, AST - 32 U/L, D-Dimer - 0.7 ug/mL, APTT - 39.9 sec, CK - 247 U/L, CRP - 14.3 mg/L, LDH - 246 U/L, Troponin - 0.037 ng/mL, Ferritin - 498 ng/mL, Fibrinogen - 446 mg/dL.
- 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), it is comparable with the model trained with all the tests (AUC 0.90; CI 0.812 - 0.902; p<0.001).

Figure 1: Variation of laboratory findings values in the ICU cohort (orange boxplot) versus the non-ICU cohort of patients (blue boxplot).

References

[13] L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, M. Zhang, et al., An interpretable mortality prediction model for covid-19 patients, Nature Machine Intelligence (2020) 1–6.

[14] L. Wynants, B. Van Calster, M. M. Bonten, G. S. Collins, T. P. Debray, M. De Vos, M. C. Haller, G. Heinze, K. G. Moons, R. D. Riley, et al., Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal, bmj 369 (2020).

[15] Y. Gao, T. Li, M. Han, X. Li, D. Wu, Y. Xu, Y. Zhu, Y. Liu, X. Wang, L. Wang, Diagnostic utility of clinical laboratory data determinations for patients with the severe covid-19, Journal of medical virology (2020).

[16] B. M. Henry, G. Aggarwal, J. Wong, S. Benoit, J. Vikse, M. Plebani, G. Lippi, Lactate dehydrogenase levels predict coronavirus disease 2019 (covid-19) severity and mortality: A pooled analysis, The American Journal of Emergency Medicine (2020).

[17] R. Zhou, F. Li, F. Chen, H. Liu, J. Zheng, C. Lei, X. Wu, Viral dynamics in asymptomatic patients with covid-19, International Journal of Infectious Diseases (2020).

[18] J. Gong, J. Ou, X. Qiu, Y. Jie, Y. Chen, L. Yuan, J. Cao, M. Tan, W. Xu, F. Zheng, et al., A tool to early predict severe 2019-novel coronavirus pneumonia (covid-19): a multicenter study using the risk nomogram in wuhan and guangdong, china, medRxiv (2020).

[19] H.-Y. Zheng, M. Zhang, C.-X. Yang, N. Zhang, X.-C. Wang, X.-P. Yang, X.-Q. Dong, Y.-T. Zheng, Elevated exhaustion levels and reduced functional diversity of t cells in peripheral blood may predict severe progression in covid-19 patients, Cellular & molecular immunology 17 (2020) 541–543.

[1] X. Chen, B. Zhao, Y. Qu, Y. Chen, J. Xiong, Y. Feng, D. Men, Q. Huang, Y. Liu, B. Yang, et al., Detectable serum sars-cov-2 viral load (rnaaemia) is closely correlated with drastically elevated interleukin 6 (il-6) level in critically ill covid-19 patients, Clinical Infectious Diseases (2020).

[2] W. H. Mahallawi, O. F. Khabour, Q. Zhang, H. M. Makhdoum, B. A. Suliman, Mers-cov infection in humans is associated with a pro-inflammatory th1 and th17 cytokine profile, Cytokine 104 (2018) 8–13.

[3] C. Selinger, J. Tisoncik-Go, V. D. Menachery, S. Agnihothram, G. L. Law, Chang, S. M. Kelly, P. Sova, R. S. Baric, M. G. Katze, Cytokine systems approach demonstrates differences in innate and pro-inflammatory host responses between genetically distinct mers-cov isolates, BMC genomics 15 (2014) 1161.

[4] Kappert, A. Jahic´, R. Tauber, Assessment of serum ferritin as a biomarker in covid-19: bystander or participant? insights by comparison with other infectious and non-infectious diseases, Biomarkers (2020) 1–36.

[5] Spiezia, A. Boscolo, F. Poletto, L. Cerruti, I. Tiberio, E. Campello, P. Navalesi, P. Simioni, Covid-19-related severe hypercoagulability in patients admitted to intensive care unit for acute respiratory failure, Throm- bosis and haemostasis 120 (2020) 998.

[6] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, et al., Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china, Jama 323 (2020) 1061–1069.

[7] D. Bashash, H. Abolghasemi, S. Salari, M. Olfatifar, P. Eshghi, M. E. Akbari, Elevation of d-dimer, but not pt and aptt, reflects the progression of covid-19 toward an unfavorable outcome: A meta-analysis, Iranian Journal of Blood & Cancer (2020) 47–53.

[8] L. Quartuccio, A. Sonaglia, D. McGonagle, M. Fabris, M. Peghin, D. Pecori, A. De Monte, T. Bove, F. Curcio, F. Bassi, et al., Profiling covid-19 pneumonia progressing into the cytokine storm syndrome: results from a single italian centre study on tocilizumab versus standard of care, Journal of Clinical Virology (2020) 104444.

[9] W. Ling, C-reactive protein levels in the early stage of covid-19, Medecine et maladies infectieuses (2020).

[10] P. P. Bloom, E. A. Meyerowitz, Z. Reinus, M. Daidone, J. Gustafson, Y. Kim, E. Schaefer, R. T. Chung, Liver biochemistries in hospitalized patients with covid-19, Hepatology (2020).

[11] N. Ali, K. Hossain, Liver injury in severe covid-19 infection: current insights and challenges, Expert review of gastroenterology & hepatology (2020).

[12] Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, F. Herrera, Learning from imbalanced data sets, Springer, 2018.

7. Acknowledgments

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. The authors would like to acknowledge Mediclinic Parkview Hospital (Dubai, UAE) and UAE University (Al Ain, UAE) for the support provided and the facilities used for conducting this research. We also wish to express our gratitude to the dedication and commitment of our healthcare staff and our patients, without whom this research would not be possible.

8. Author contributions statement

All authors contributed to the creation of the article as follows: all of them contributed to the conceptual idea of the paper equally; FA and YS formulated the objectives; FA collected the dataset; YS wrote the manuscript; TH proposed the methodology of the study, and performed the statistical analysis, prepared the figures and tables for data presentation and illustration, TL, KG, NZ contributed to the literature review and data analysis.

Table 3: Justification of the cut-off levels for the laboratory findings

No	Feature	Normal values	Threshold moving technique				Percentile level		
			Cut-off	Sensitivity	Specificity		Cut-off	Sensitivity	Specificity
1	WBC (x109/L)	4.0 - 11.0	45	0.6	0.5		**7**	0.5278	0.75
2	Lymphocytes (x109/L)	1 - 4.8	0.3	0.43	0.62		**1.24**	0.7778	0.75
3	T. bilirubin (umol/L)	3.4 - 20.5	37	0.54	0.43		**11.9**	0.4861	0.7439
4	ALT (U/L)	0 - 55	435	0.29	0.68		**43**	0.4583	0.7439
5	AST (U/L)	5 - 34	400	0.53	0.46		**32**	0.7639	0.7418
6	D-Dimer (ug/mL)	0.0 - 0.5	15	0.35	0.7		**0.7**	0.7222	0.7234
7	APTT (sec)	28.0 - 40.0	180	0.57	0.71		**39.9**	0.5139	0.7336
8	CK (U/L)	30.0 - 200.0	4808	0.54	0.63		**247**	0.4028	0.6619
9	CRP (mg/L)	0.0 - 5.0	400	0.6	0.79		**14.3**	0.9306	0.75
10	LDH (U/L)	125 - 243	1778	0.21	0.88		**246**	0.8889	0.6537
11	Troponin (ng/mL)	<0.03	11	0.33	0.75		**0.037**	0.2361	0.7172
12	Ferritin (ng/mL)	21.8 - 274.6	14025	0.35	0.82		**498**	0.6667	0.75
13	Fibrinogen (mg/dL)	200-400	3030	0.33	0.89		**446**	0.8611	0.4939

Data were analyzed and interpreted by the authors, who also reviewed the manuscript and vouch for the accuracy and completeness of the data and for the adherence of the study to the protocol.

9. Ethical Approval

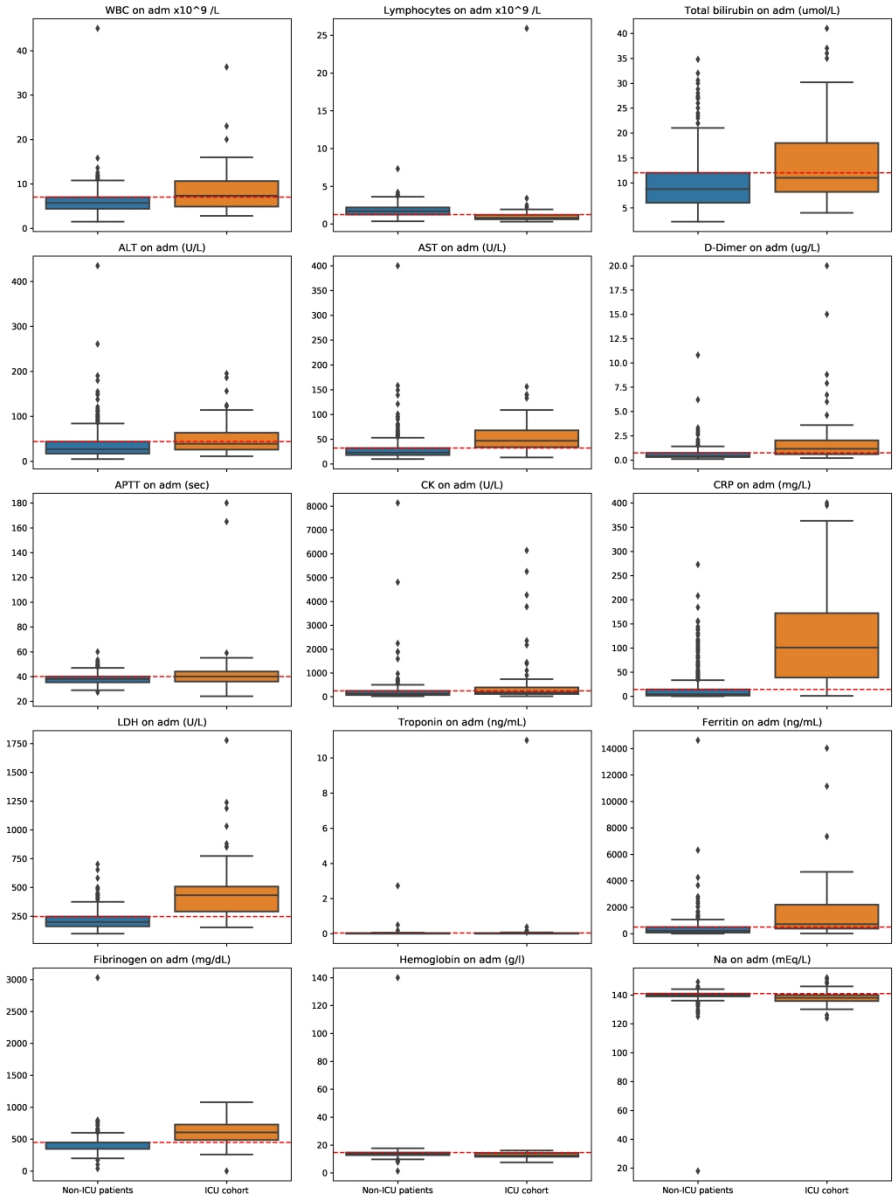
The study got an ethical review by Dubai Scientific Research Ethics Committee (DSREC), Dubai Health Authority, protocol number DSREC-05/2020_25) and was approved for the retrospective analysis of the data obtained as a standard of care. No potentially identifiable personal information is presented in the study.

10. Patient and public involvement

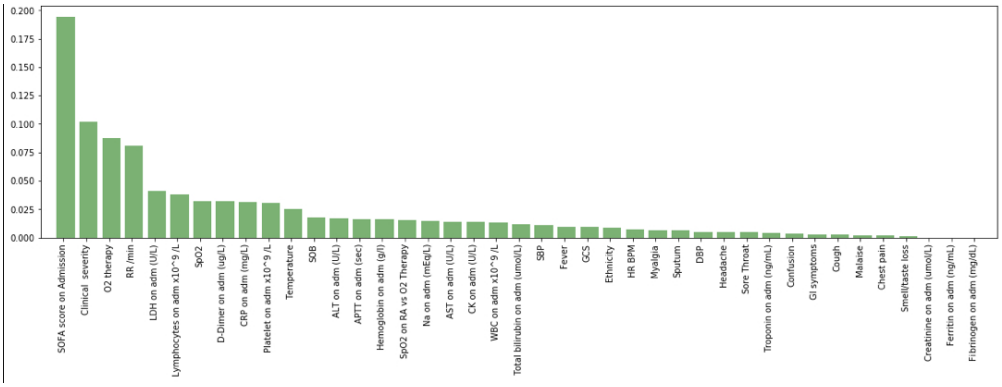
No patient involved. The data were collected retrospectively from the medical record system.

11. Data availability statement

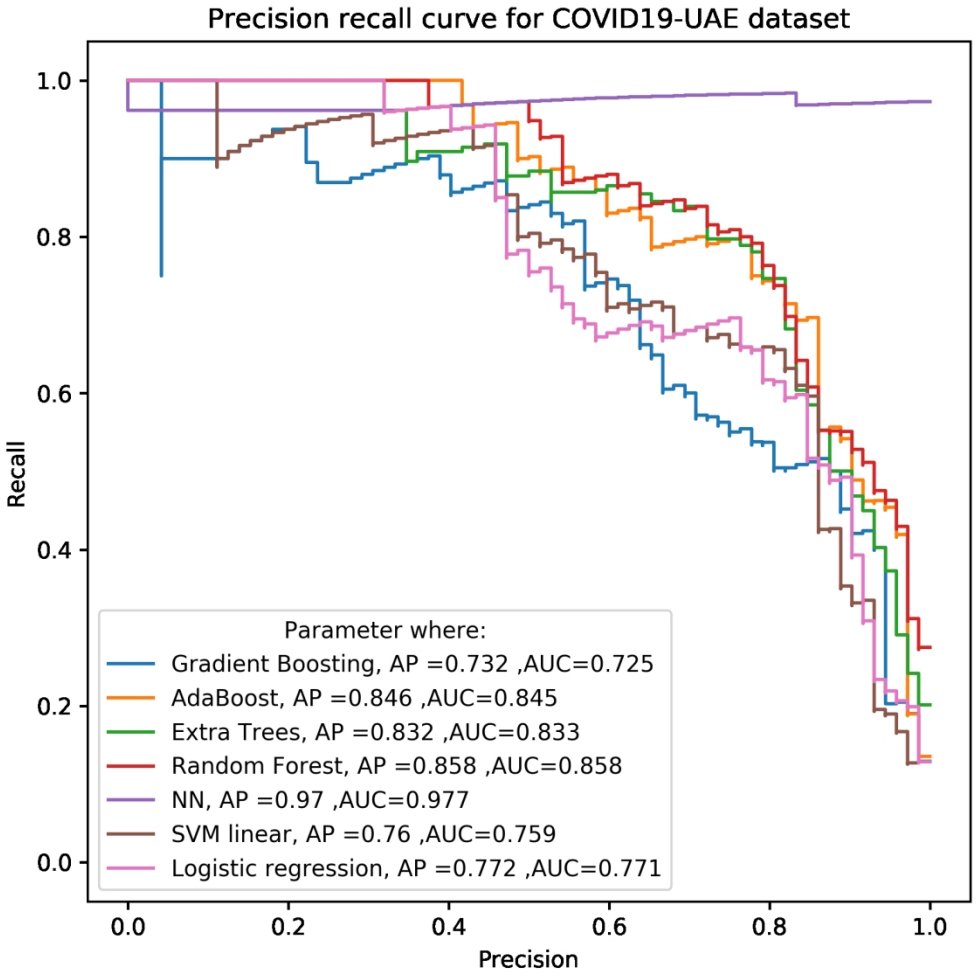
Generated Statement: The datasets generated for this study are available on request at the site of **Big Data Analytics Center** at https://bi-dac.com



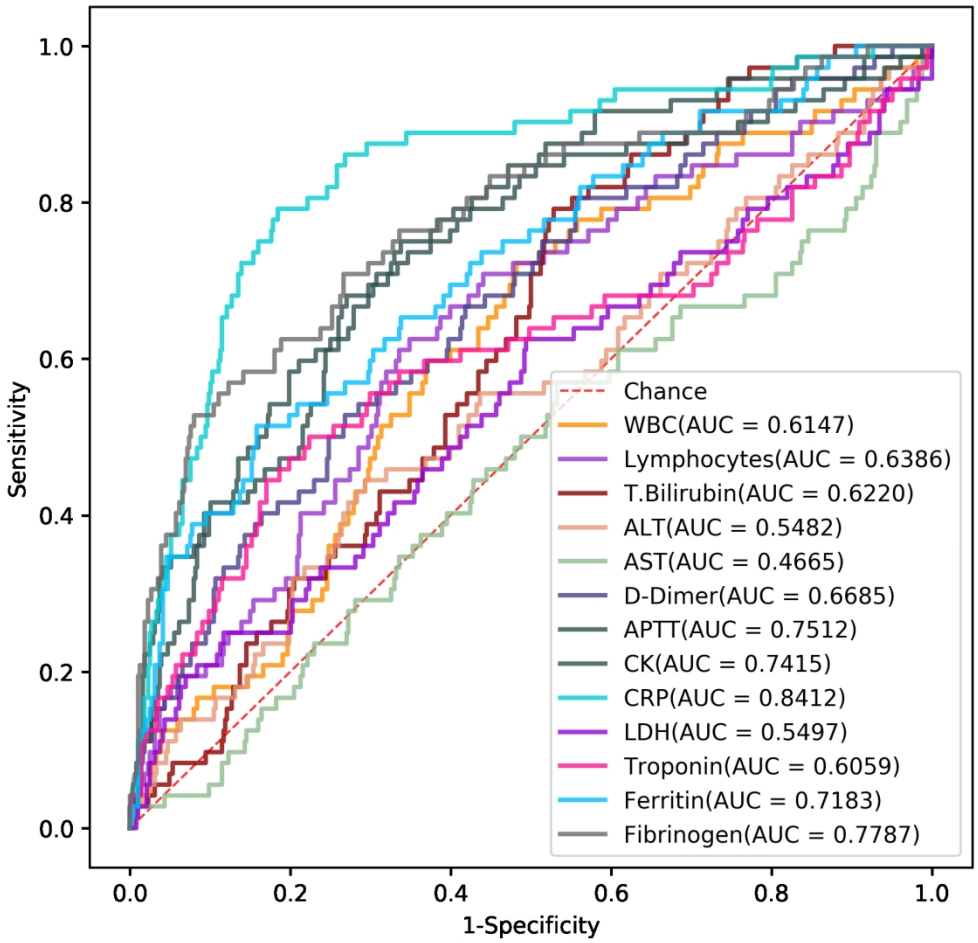
Variation of laboratory findings values in the IC cohort (orange boxplot) versus the non-ICU cohort of patients (blue boxplot)



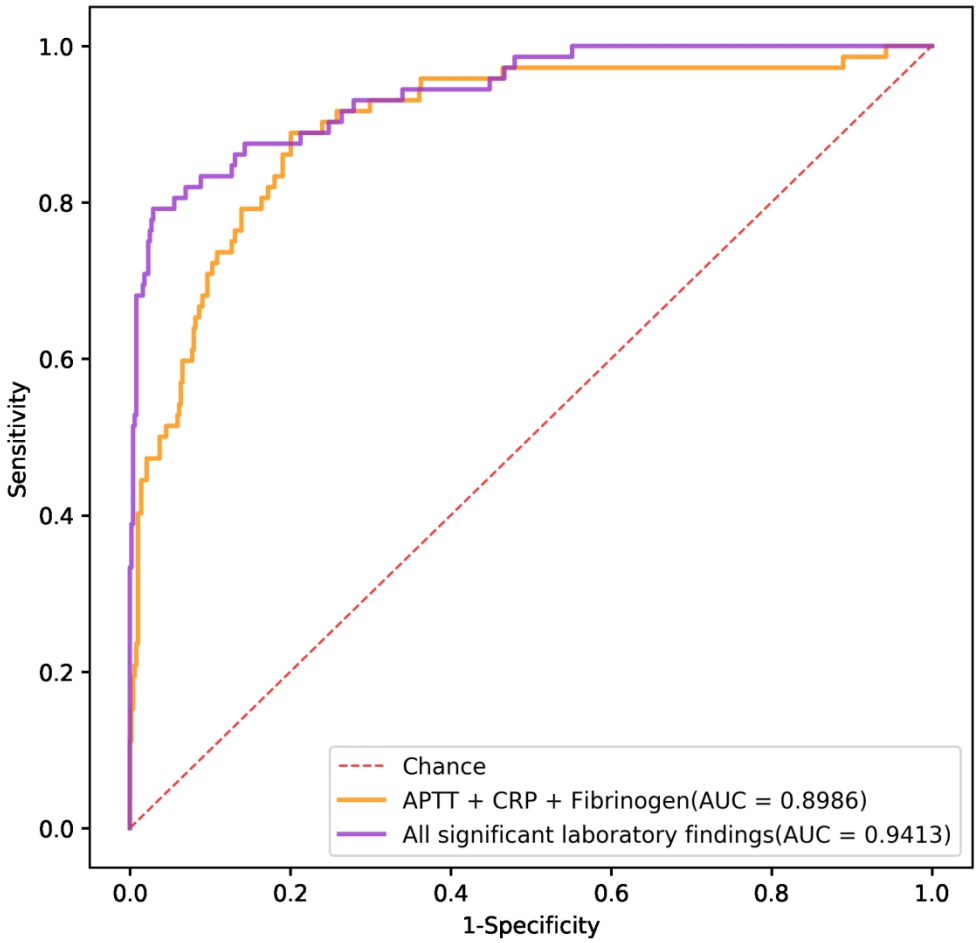
Feature selection for predicting whether a patient is going to be transferred to ICU



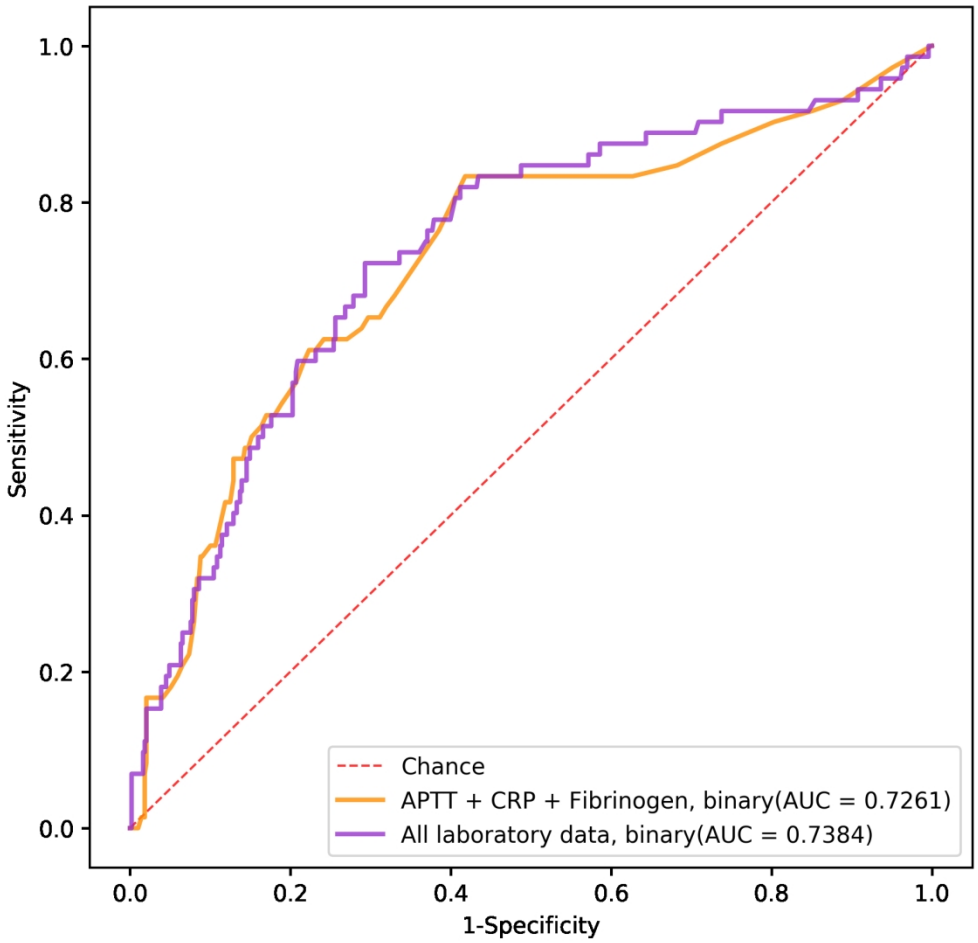
The performance of the employed NN classification method



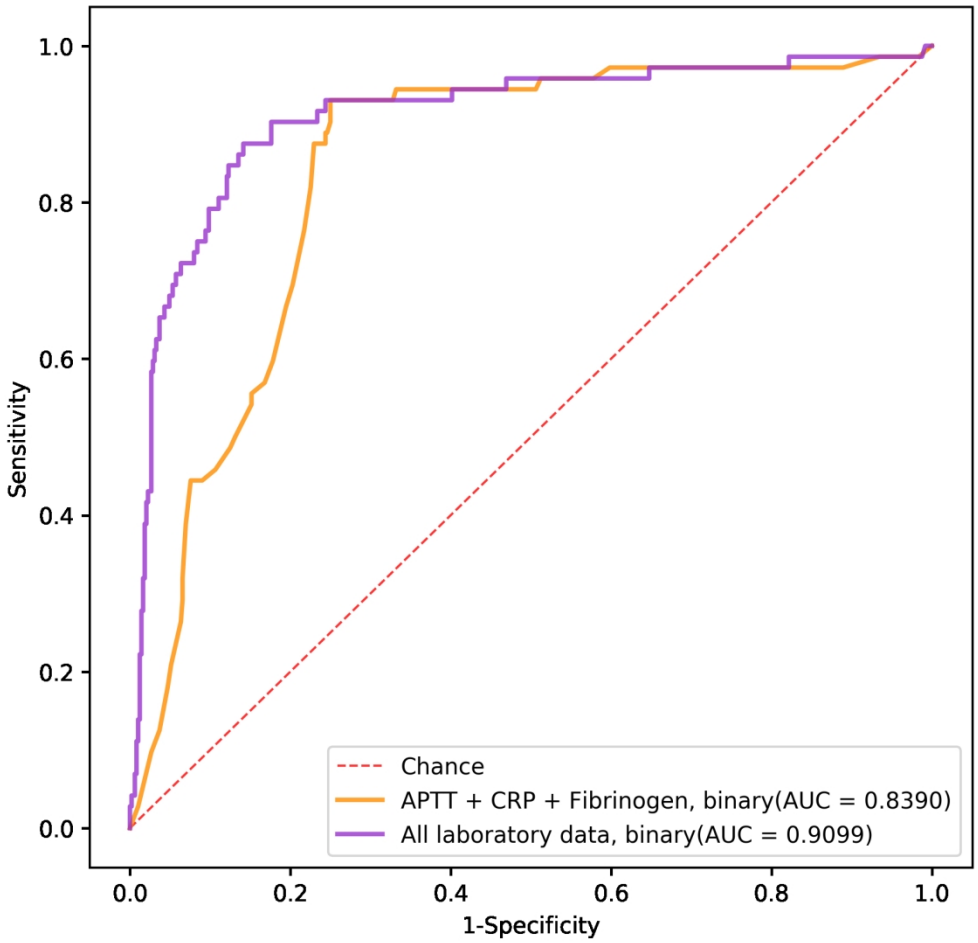
ROC curves for the laboratory tests used as input to NN separately (a)



and in the combination (b).



The performance of the 10 folds cross-validation model trained on binary data with the threshold moving technique returned by the ML estimator (a)



and with the cut-off level set to the 25th percentile for lymphocyte count and 75th for the other features (b)

Appendices

A. ML classification models and feature selection.

The variables used to build up the model:

- *physical examination on admission*: temperature, HR BPM, SBP, DBP, RR /min. SpO_2, SpO_2 on RA vs. O_2 Therapy, GCS, SOFA score
- *symptoms on admission*: cough, sputum, sore throat, chest pain, SOB, fever, headache, confusion, having any gastrointestinal symptoms (e.g., nausea, vomiting, diarrhea), myalgia, malaise, loss of smell or taste.
- *laboratory findings on admission*: the count of WBC, platelet, and lymphocyte; the concentration of hemoglobin, total bilirubin, D-Dimer, creatinine, sodium, C-reactive protein, troponin, ferritin, fibrinogen; the activity of ALT, AST, CK, LDH; APTT.

Feature selection:

To check if there are unique patterns within the data that can unambiguously identify if the patient is going to be transferred to the intensive care unit, we utilized ML algorithms.

To assess the importance of the features fed to the ML models as predictors of admitted to ICU patients, we employed four ensemble tree-based estimators such as AdaBoost, Gradient Boosting, Random Forest, and Extra Trees. These models were trained on the whole dataset and used to rank the features in ascending order concerning their predictive potential. Figure 2 and Table 4 display the averaged values of impurity-based attribute ranks, where the average for each feature is calculated as the mean of rank values for the four ML methods mentioned above.

Table 4: Ranking scores of the variables selected for predicting ethnicity

Score	Feature	Score	Feature	Score	Feature	Score	Feature
0.19429	SOFA score	0.02520	Temperature	0.01164	Total bilirubin	0.00466	Sore Throat
0.10168	Clinical severity	0.01748	SOB	0.01135	SBP	0.00445	Troponin
0.08745	O2 therapy	0.01712	ALT	0.00983	Fever	0.00367	Confusion
0.08061	RR/min	0.01623	APTT	0.00969	GCS	0.00309	GI symptoms
0.04127	LDH	0.01595	Hemoglobin	0.00896	Ethnicity	0.00287	Cough
0.03829	Lymphocytes	0.01545	SpO_2 on RA vs O_2 Therapy	0.00732	HR BPM	0.00188	Malaise
0.03223	SpO2	0.01505	Na	0.00637	Myalgia	0.00186	Chest pain
0.03212	D-Dimer	0.01383	AST	0.00633	Sputum	0.00141	Smell/taste loss
0.03125	CRP	0.01382	CK	0.00524	DBP	0.00000	Creatinine
0.03067	Platelet	0.01360	WBC	0.00513	Headache	0.00000	Ferritin
						0.00000	Fibrinogen

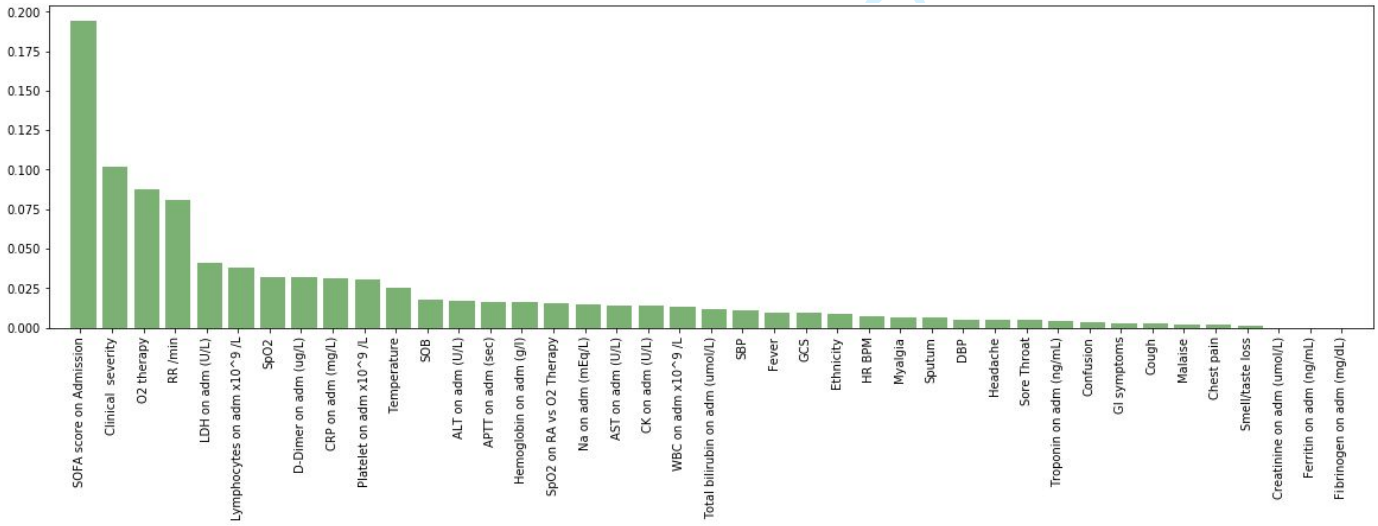


Figure 2: Feature selection for predicting whether a patient is going to be transferred to ICU

Table 5: Confusion matrix to assess the accuracy of classification with a three-layer dense NN model to predict the severity of the disease

		Predicted	
		Not admitted to ICU	Admitted to ICU
Actual	Not admitted to ICU	485	3
	Admitted to ICU	0	72

Table 6: Classification metrics of the NN model to predict the event of being transferred to ICU

	Precision	Recall	F1 score	Support
Not admitted to ICU	1.00	0.99	1.00	488
Admitted to ICU	0.96	1.00	0.98	72
accuracy			0.99	560
macro avg	0.98	1.00	0.99	560
weighted avg	0.99	0.99	0.99	560

Prediction of the admission to ICU. To evaluate the classifier output quality we trained several ML classification models using a stratified 10-fold cross-validation technique to generalize the models to the true rate error. For each fold, we used 90% of the data to train the model and then tested it on the rest 10%. The decision matrices built on the test dataset for all folds were combined and used to calculate the performance metrics. The best performance measures were obtained with a three-layer fully connected NN.

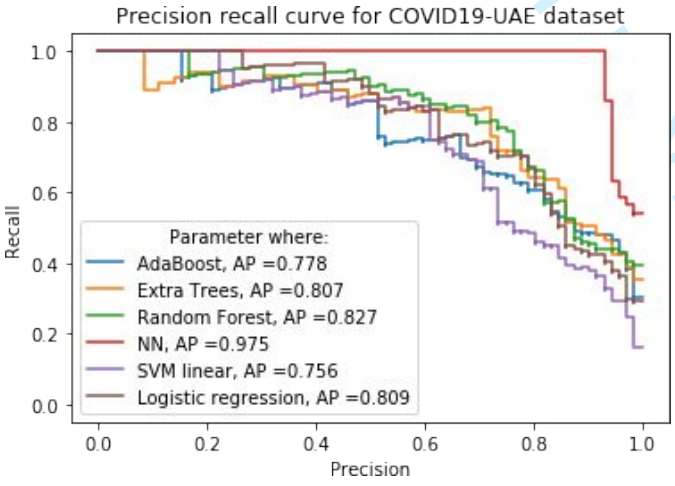


Figure 3: The performance of the employed NN classification method

B. ROC curves for laboratory tests used as input to NN.

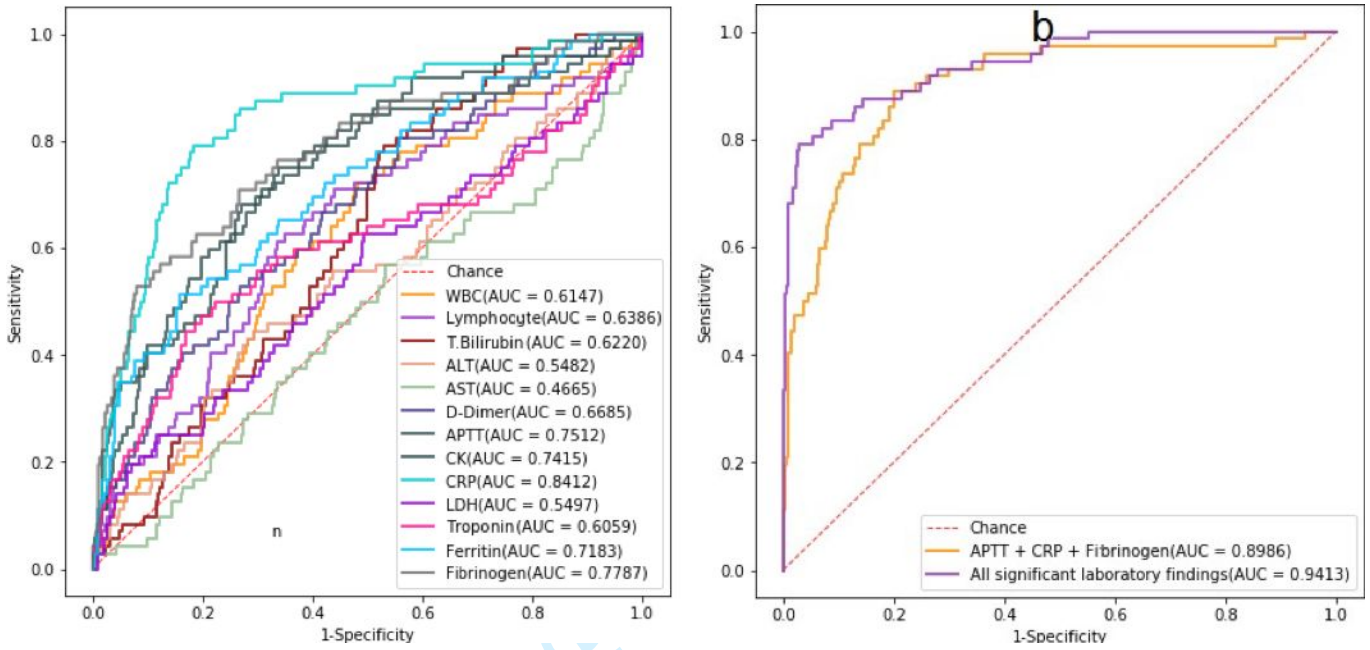


Figure 4: ROC curves for the laboratory tests used as input to NN separately (a) and in the combination (b). The models are trained with 10 folds cross-validation.

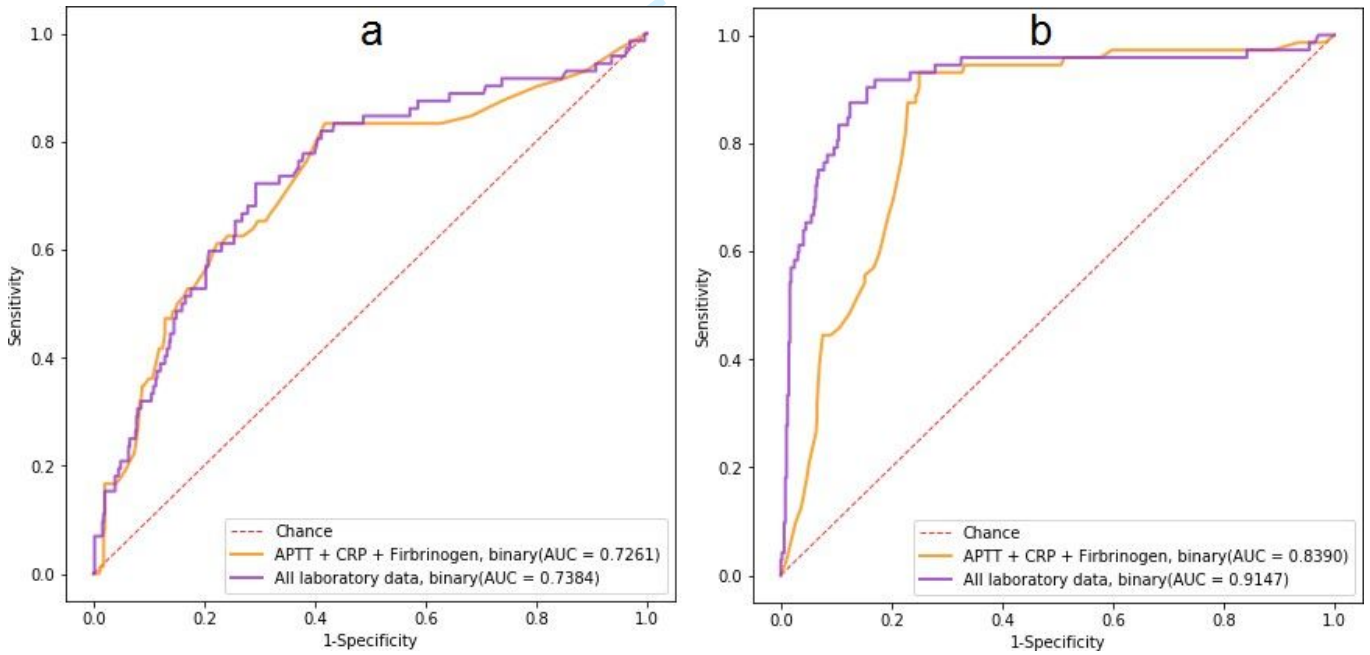


Figure 5: The performance of the 10 folds cross-validation model trained on binary data with the threshold moving technique returned by the ML estimator (a), and with the cut-off level set to the 25^{th} percentile for lymphocyte count and 75^{th} for the other features (b).

BMJ Open

Prediction of COVID-19 severity out of laboratory findings on admission: informative values, thresholds, ML model performance.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-044500.R1
Article Type:	Original research
Date Submitted by the Author:	30-Nov-2020
Complete List of Authors:	Statsenko, Yauhen; United Arab Emirates University College of Medicine and Health Sciences, Radiology Al Zahmi, Fatmah ; Mediclinic Parkview Hospital, Neurology; Mohammed Bin Rashid University Of Medicine and Health Sciences Habuza, Tetiana; UAE University College of Information Technology, Department of Computer science Gorkom, Klaus; United Arab Emirates University College of Medicine and Health Sciences, Radiology Zaki, Nazar; United Arab Emirates University,
Primary Subject Heading:	Health informatics
Secondary Subject Heading:	Infectious diseases, Research methods, Medical management, Intensive care
Keywords:	COVID-19, BIOTECHNOLOGY & BIOINFORMATICS, INFECTIOUS DISEASES, Respiratory infections < THORACIC MEDICINE, Information technology < BIOTECHNOLOGY & BIOINFORMATICS, Biochemistry < NATURAL SCIENCE DISCIPLINES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Prediction of COVID-19 severity out of laboratory findings on admission: informative values, thresholds, ML model performance.

Yauhen Statsenko[a], Fatmah Al Zahmi[b], Tetiana Habuza[c], Klaus Neidl Van Gorkom[a], Nazar Zaki[c]

[a]*College of Medicine and Health Sciences, United Arab Emirates University, P.O. Box 17666, Al Ain, United Arab Emirates*
[b]*Mediclinic Middle East Parkview hospital, P.O. Box 51122, Dubai, United Arab Emirates*
[c]*College of Information Technology, United Arab Emirates University, P.O. Box 15551, Al Ain, United Arab Emirates*

Abstract

Background: Despite the necessity, there is no reliable biomarker to predict disease severity and prognosis of COVID-19 patients. The currently published prediction models are not fully applicable to clinical use.

Objectives: To identify predictive biomarkers of COVID-19 severity and to justify their threshold values for the stratification of the risk of deterioration that would require transferring to ICU.

Methods: The study cohort (560 subjects) included all consecutive patients admitted to Dubai Mediclinic Parkview hospital from February to May 2020 with COVID-19 confirmed by the polymerase chain reaction. The challenge of finding the cut-off thresholds was the unbalanced dataset (e.g., the disproportion in the number of 72 patients admitted to ICU versus 488 non-severe cases). Therefore, we customized supervised ML algorithm in terms of threshold value used to predict worsening.

Results: With the default thresholds returned by the ML estimator, the performance of the models was low. It was improved by setting the cut-off level to the 25^{th} percentile for lymphocyte count and the 75^{th} - for other features.

The study justified the following threshold values of the laboratory tests done on admission: lymphocyte count lower than 2.59 $x10^9$/L, and the upper levels for total bilirubin - 11.9 umol/L, ALT - 43 U/L, AST - 32 U/L, D-Dimer - 0.7 mg/L, APTT - 39.9 sec, CK - 247 U/L, CRP - 14.3 mg/L, LDH - 246 U/L, Troponin - 0.037 ng/mL, Ferritin - 498 ng/mL, Fibrinogen - 446 mg/dL.

Conclusion: 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).

Keywords: COVID-19 pandemic , coronavirus, severity, biomarkers, threshold values, infectious disease

Strength and limitations of the study

- The research is based on a unique study cohort that is representative of the entire population because of the National Standard that required all patients with confirmed COVID-19 to be admitted to acute care hospitals regardless of their symptoms or illness severity.
- To distinguish the patients with the confirmed COVID-19 who may worsen while treated, we justified the threshold values of the laboratory tests done on admission.
- The prediction of the future deterioration by the neural network is reliable even with the top three valuable laboratory tests (APTT, CRP, and Fibrinogen) used for training (AUC 0.86; CI 0.486 - 0.884; p<0.001).
- The limitation of the study was the unbalanced dataset (e.g., the disproportion in the number of patients admitted to ICU versus non-severe cases).

Abbreviations

ALT - alanine aminotransferase
AST - aminotransferase
ARDS - acute respiratory distress syndrome
AUC - area under the curve
BMI - body mass index
CI - confidence interval
CoV - coronavirus
GCS - Glasgow coma scale
hs-CRP - high-sensitivity C-reactive protein
ICU - intensive care unit
IL - interleukin
MERS - Middle East respiratory syndrome
ML - machine learning
NN - neural network
PC - precision-recall
PCR - polymerase chain reaction
PR - precision-recall
RNA - ribonucleic acid
ROC - receiver operating characteristic
RR - respiratory rate

*Corresponding author. Tel.: +971 3 713 7124; E-mail address: e.a.statsenko@uaeu.ac.ae (Y. Statsenko).

November 27, 2020

SARS-CoV-2 - severe acute respiratory syndrome-related coronavirus 2
SOB - shortness of breath
SOFA - Sequential organ failure assessment
TNF - tumor necrosis factor

Definitions

Mild level of COVID-19 severity - nonpneumonia and mild pneumonia.

Severe level of COVID-19 severity - dyspnea, respiratory frequency ≥30/min, blood oxygen saturation ≤93%, the partial pressure of arterial oxygen to fraction of inspired oxygen ratio <300, and/or lung infiltrates >50% within 24 to 48 hours.

Critical level of COVID-19 severity - respiratory, septic shock, and/or multiple organ dysfunction or failure.

1. Introduction

Despite the necessity, there is no reliable prognostic biomarker to predict disease severity and prognosis of COVID-19 patients [1]. Studies on COVID-19 have built up several types of prediction models. These have been the models designed to indicate the disease risk in the general population, the diagnostic models based on medical imaging, and the prognostic models. Unfortunately, these models have had some limitations that have precluded their use in clinical practice [2].

1.1. Models using laboratory findings as the inputs

Researchers tried to establish the role of laboratory findings in the diagnosis of COVID-19 [3]. They showed that the severe cases of COVID-19 were associated with D-dimer level over 0.28μg/L, interleukin (IL) 6 level over 24.3pg/mL [3], and LDH activity with an upper limit cut-off in the range of 240- 255U/L [4]). However, the use of these laboratory parameters with the above mentioned cut-off values was limited for the following reasons. First, these studies were conducted on severe forms of the disease. Limited research was done on patients who were asymptomatic or had mild disease [3, 5]. Second, the whole spectrum of the regularly used clinical laboratory data is unavailable for non-severe patients. Thus, the published papers add justification on the diagnostic utility of separate laboratory findings, instead of working out reliable diagnostic criteria for a set of them.

Gong and colleagues [6] have generated a tool for the early prediction of severe COVID-19 pneumonia out of the following data: age, serum lactate dehydrogenase activity, C-reactive protein, the coefficient of variation of red blood cell distribution width, blood urea nitrogen, direct bilirubin, lower albumin. The resulting performance was not high (sensitivity 77.5%, specificity 78.4%) [6]. Supposedly, this is because the dataset used as the input consists of exceptionally the age and laboratory findings.

In another model, the inputs included basic information, symptoms, and the results of laboratory tests. After the feature selection, the number of key features was set to just three laboratory results: LDH, lymphocytes, and high-sensitivity C-reactive protein (hs-CRP). The model was trained with the follow-up studies of the general, severe, and critical patients [1]. By feeding ML algorithm with the results obtained at the time of admission and in follow-up studies, the authors worked out a decision rule to predict patients at the highest risk. However, physicians are interested in the early prediction of the disease outcomes, and it is highly disputable that the model will not loose its predictive potential if applied exceptionally to the data received on admission.

We believe that a more accurate model can be built based on the simultaneous interpretation of laboratory results, clinical data, and physical examination findings (e.g., BMI, body temperature, respiratory rate) at the time of presentation. The analysis utilizing a machine learning algorithm could provide an accurate prediction of the disease severity.

1.2. Data used by clinicians for stratifying risks

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.

Inflammatory markers. There is evidence that IL-6, tumor necrosis factor-α do not indicate the level of COVID-19 progression [7]. Some markers of inflammation are elevated in the serum of COVID-19 patients compared to the healthy people, i.e., the serum SARS-CoV-2 viral load (RNAaemia) is closely correlated with drastically elevated interleukin 6 levels in critically ill COVID-19 patients [8]. However, there is no significant difference between severe and mild groups [7]. In contrast to this, the indicators are reflective in the progression of the diseases caused by other coronaviruses (e.g., MERS, SARS) [9]. This may be explained by the huge amino acid differences in viral proteins of distinct coronaviruses. Even with different MERS-CoV strains, common cytokine signaling by TNF and IL-1α results in the differential expression of innate immune genes [10].

Ferritin. Ferritin is a marker of iron storage. However, it is also an acute-phase reactant, the level of which elevates in processes of acute inflammation, whether infectious or non-infectious. Marked elevations have been reported in cases of COVID-19 infection [11].

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 interests 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. This is why 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, D-Dimer level of >2.14mg/L predicted in-hospital mortality with a sensitivity of 88.2% and specificity of 71.3% [18]. Another 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].

APTT. In a study conducted in February 2020, the levels of APTT as well as WBC, lymphocytes, AST, ALT, and creatinine, differed negligibly between severe and mild patients [3]. At the same time, other researchers showed inconsequential distinction in APTT in survivors versus non-survivors [20]. According to the results of another study published in March 2020, no significant difference in APTT values were found in the cohort of severe cases versus the non-severe one [6]. The results obtained in another study in April in Italy were the same [12]. The common limitation of these early studies was a small sample size. Finally, a meta-analysis justified that the elevation of D-Dimer, rather than prothrombin time and APTT, reflects the progression of COVID-19 toward an unfavorable outcome [21].

LDH and CK. Increased levels of the enzymes may reflect the level of the organ damage in a systemic disease [22, 4]. Reasonably, they may serve as biomarkers for COVID-19 progression.

CRP. In the early stage of COVID-19, CRP levels are positively correlated with the diameter of lung lesions and severe presentation [23].

Liver enzymes and total bilirubin. COVID-19 leads to elevated liver biochemistries (e.g., the level of AST, ALT, GGT, total bilirubin) in over 50% of patients on admission. AST - dominant aminotransferase elevation reflects the disease severity and true hepatic injury [24, 25].

2. Objectives

We decided to identify predictive biomarkers of COVID-19 severity and to justify their threshold values. Hypothetically, the absolute values of the biomarkers on admission to the clinics could provide physicians with an accurate prognosis on the future worsening of the patient that would require transferring the individual to the intensive care unit (ICU). Getting a reliable tool for such a prognosis will support decision making and logistical planning in clinics.

To address the objective, we designed a set of the following tasks:

- to study the linear separability of the laboratory findings values in patients with confirmed COVID-19 who were transferred to ICU versus non-severe cases of the disease, and to make the comparative analysis of the ICU department cases (both the deceased and survived cohorts) with other patients with COVID-19.
- to identify the risk factors by selecting the most valuable features for predicting the deterioration that would require transferring patient to ICU.
- to work out the threshold criteria for the major clinical data for the early identification of the patients with a high risk of being transferred to ICU.
- to identify the accuracy of the prediction of the patient's deterioration by the machine learning algorithm and by a set of the newly created threshold values of the laboratory and clinical findings.

3. Materials and methods

3.1. Study design and sample

We did a retrospective analysis of the clinical data obtained as a standard of primary and secondary care. The study sample included all the consecutive patients admitted to Dubai Mediclinic from 24th February to 1st July 2020 who fit the criteria of eligibility mentioned above (560 cases totally). Using this sample met the intention of the study: to allow for the early prognostic stratification.

The inclusion criteria were as follows: age 18 years or older; inpatient admission; SARS-CoV-2 positive real-time reverse-transcriptase polymerase chain reaction (PCR) from nasopharyngeal swabs only, at our site. Those patients who met the inclusion criteria for our studies were included in the study sample. All the patients were discharged at the time of writing the paper.

The remarkable feature of our study is that at the beginning of the pandemic, all the COVID-19 verified by PCR were hospitalized in the Mediclinic even if they did not present any

symptoms. We observed many mild and asymptomatic forms of the disease, with all the required spectrum of analyses being conducted. All patients who were hospitalized stayed in Dubai Mediclinic until they were afebrile for more than 72 h and had SpO2 value non less than 94%.

We assessed the duration of viral shedding as the number of days from the disease onset when the diagnosis was confirmed (e.g., the first positive PCR test) to the first negative PCR test [26]. 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 disease symptoms, the specimen collection continued on a daily basis until two subsequent negative PCR tests for COVID-19 more than 24 h apart. In the case of the mild disease course, patients might be transported to isolation facilities before being discharged home (see the flow chart diagram in Figure 1). If the facilities were run by Mediclinic, we had their follow up PCR results. For those patients who went to other isolation facilities not connected to Mediclinic, we couldn't study the duration of viral shedding (the data are missing for 27 out of 560 patients).

The treatment was administered in full accordance with "National Guidelines for Clinical Management and Treatment of COVID-19". The indications for the supportive oxygen therapy were (a) the oxygen saturation level below 94%, (b) the respiratory rate (RR) above 30 breaths per minute (c) both of them. In case of suspicion of superimposed bacterial pneumonia physicians ordered empirical broad-spectrum antibiotics. The administration of the antiviral and antimalarial drugs followed the national guidelines [27].

3.2. Patient and public involvement

No patient involved. The data were collected retrospectively from the medical record system.

3.3. Methods used

To address the first task, we studied the separability of laboratory findings values on admission to Dubai Mediclnic concerning the future transfer of the patient to the ICU department. To carry out the comparative analysis of features with regard to transferring to ICU, we utilized a set of non-parametric tests. The relationships involving two variables were assessed with the Mann-Whitney U test or Kruskal-Wallis test for the continuous features, and with Fisher's Exact test or Chi-square test for the quantitative ones. The data were expressed as *IQR*, *median* ± *std* or number of cases, and their percentage. The missing data for the comparative analysis were treated with the complete-case analysis method.

To address the second task, we used a set of different methods. First, we trained the NN ML model on each variable separately. To come up with laboratory data cut-off levels, which may be considered as bookmakers of severe course of the disease, we assessed their statistical significance against chance performance. We calculated 95% CI for ROC and ROC AUC scores with the bootstrap technique and p-values with permutation tests.

Second, we used ML tree-based methods (AdaBoost, Gradient Boosting, Random Forest, and Extra Trees) to check if there were unique patterns within the data that could unambiguously identify the event of transferring the patient to ICU from the data obtained on admission. For the list of features used as predictors see Appendix A. To assess the importance of the variables, we ranked all features concerning their impurity-based predictive potential. For ranking, we utilized a set of classifiers and then averaged all the received scores. Missing data in all ML models were replaced by the mean or median values with regard to the continuous or quantitative feature respectively utilizing single imputation method.

To tackle the third task, we used two approaches: a threshold moving technique (Youden's index) [28] and a heuristically chosen percentile-based cut-off level. The problem of predicting the transfer to ICU had a severe class imbalance. Therefore, we needed to focus on the performance of the classifier on the minority class (admitted to ICU patients). The sensitivity and specificity of the supervised ML classification model (NN) were used to evaluate the quality of the chosen optimal thresh- old for each important laboratory finding.

To evaluate the classifier output quality, we trained several ML classification models using a stratified 10-fold cross-validation technique to generalize the models to the true rate error. For each fold, we used 90% of the data to train the model and then tested it with the rest 10%. The decision matrices built on the test dataset for all folds were combined and used to calculate the performance metrics.

4. Results

4.1. Comparison of the ICU vs. non-ICU patients

The problem of predicting admission to ICU has a severe class imbalance (488 vs 72). Therefore, we need to focus on the performance of the classifier on the minority class (the patients admitted to ICU).

We look at the linear separability of the groups of numerical data composed from the laboratory findings values with regard to their quartiles. In Figure 2, boxplots for the laboratory findings data are presented with the red dashed line that marks the 75^{th} percentile for the subjects that were not transferred to ICU. The assumption is to use the third quartile (Q3) start point value as the threshold if there is separability between ICU and non- ICU groups. In diagrams in Figure 2, the red line indicates the 75^{th} percentile for not admitted to the ICU group. The exception is the diagram for the lymphocyte count, where it stands for the 25^{th} percentile.

The results of the comparative analysis of features with regard to transferring to ICU and the final outcomes of the disease are presented in Table 1. We excluded from further analysis the laboratory findings that did not significantly differ in the distribution of two groups. Therefore, we considered the list of 13 variables: WBC, lymphocyte count, total bilirubin, ALT, AST, D-Dimer, APTT, CK, CRP, LDH, troponin, ferritin, and fibrinogen on admission.

4.2. Feature ranking with regard to ML model performance

The features of the dataset listed in Appendix A were

ranked with four tree-based ML classifiers (e.g., Random Forest, AdaBoost, Gradient Boosting, and ExtraTrees). Tree-based models provide measures of feature importances. The classifiers are based on the mean decrease in impurity (MDI). The impurity is quantified by the splitting criterion of the decision trees. Averaged values of impurity-based attribute ranks were calculated as the mean of rank values for the algorithms mentioned above (see Appendix Figure 1).

4.3. The cut-off levels of the laboratory findings

To come up with laboratory data cut-off levels, which may be considered as biomarkers of the severe course of the disease, we trained the NN ML model on each variable separately and assessed their statistical significance against chance performance. We calculated 95% CI for ROC and PR AUC scores with the bootstrap technique and p-values with permutation tests (see Table 2).

From Table 2, there is a notable difference between the performance of the model in terms of ROC AUC and the performance at chance level. High-performance measures were obtained for APTT, CRP, and Fibrinogen values (sensitivity and specificity are 0.9877 and 0.4028 respectively). It rises to 0.9754 and 0.75 respectively for all thirteen significant tests. So we used the performance of the classification model based on the combination of these three and thirteen features.

First we trained the ML model on the data of one lab feature in the 10-folds stratified cross-validation manner. Then we built ROC for the test data of all 10 folds (see diagrams in Appendix Figure 3).

We trained the ML model on the data taken from only one lab feature in the 10-folds stratified cross-validation manner and then built ROC and for the test data (combined from all 10 folds) as it is presented in Appendix Figure 3.

To improve the model's efficiency and choose the cut-off value set for some laboratory findings data, we used a threshold moving technique along with a supervised ML classification model (NN).

The ML estimator assigns threshold values for interpreting probabilities. The default threshold returned by the estimator to class labels is 0.5. However, when the dataset is unbalanced, tuning this hyperparameter can improve the model's efficiency by finding the optimal threshold. This is crucial when the importance of predicting the positive class (admitted to ICU) outweighs true negative predictions. Performance metrics calculated for all laboratory features with regard to the optimal threshold value are presented in Table 3. The table displays the sensitivity, specificity, and AUC values obtained after applying the threshold moving technique. We marked in bold the AUC values which are higher than the ones displayed in Appendix Figure 3a. The optimal cut-off value returned by the technique is shown in the appropriate column.

Looking at the boxplots in Figure 2 we decided to check whether the performance of the model is good if we applied thresholds in the following manner. For lymphocyte count, we set the cut-off level to the 25^{th} percentile (values lower than or equal to the chosen level were set to 1, or 0 otherwise). For the other features we set the thresholds to the 75^{th} percentile (values higher or equal to the cut-off limit were set to 1 or 0 otherwise). The performance of the models with regard to the aforementioned cut-off levels is presented in Table 3.

Appendix Figure 4a shows the performance of the logistic regression model built on the binary data by applying the cut-off level for the threshold moving technique. Appendix Figure 4b. illustrates the same information for the percentile's cut-off levels.

4.4. The performance of the classification models

The applied ML algorithms were trained with stratified 10-folds cross-validation technique. The predictors used are listed in Appendix Table 1. 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 Appendix Figure 2 and Appendix Table 2. It displays all 560 test points concatenated from test (actual and predicted) label values for each fold. Appendix Tables 3-4 show the performance metrics obtained by the NN model with the highest output quality. Appendix Figure 3 displays ROC curves and AUC for the NN model with different variables, observed on admission, as predictors. Appendix Figure 4 illustrates the quality of the performance for the binary data obtained by using the threshold moving or percentile-based heuristic approach.

5. Discussion

5.1. Severity of the disease course in SARS-CoV-2 infection

There are different risk factors for COVID-19 severity. Finding and justifying them are the issues of the ongoing studies because of the persistence of the viral infection. In research on the severe respiratory illness for COVID-19, the authors justified the age above 65 years as a predictor of clinical outcomes of interest [29]. The data we received support this fact. In the same study the authors showed inconsistent results regarding the race of the patient. In the univariate model, the race was a non-significant predictor of the disease severity, however it turned out to be significant in the multivariate prediction. We did not find ethnic differences between ICU and non-ICU cohorts, but observed a notable difference in the outcome of the disease within these groups (e.g., discharged vs. deceased patients). According to other studies, age is the largest contributor to risk of death for SARS-CoV-2, the impact of the race or ethnicity on the disease course remains not fully understood. The researchers have difficulty adjusting the samples for comorbidities as physicians did not examine all the patients thoroughly before the disease [30, 31]. Presumably, the same limitations account for disparities between the studies in which the authors try to consider comorbidities (e.g., asthma, diabetes, hypertension, chronic kidney disease, etc.) as risk factors. To overcome the limitation, we decided to base the prediction on the laboratory findings on admission. They are standardized and unambiguously interpretable.

5.2. Biomarkers of the deterioration of the patients

It is common sense that people with unmanaged chronic

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

		All patients				ICU patients			Missing values, count
		Total n_1=560	Not admitted to ICU n_2=488 (87.14%)	Admitted to ICU n_3=72 (12.86%)	p_{2-3}	Dead n_4=15 (20.83%)	Discharged n_5=57 (79.17%)	p_{4-5}	
Age		39.0[33.0-49.0]	38.0±11.97	51.0±13.08	**<0.0001**	46.0±12.56	62.0±11.01	**<0.0018**	
Gender	female	189 (33.75%)	**175 (35.86%)***	**14 (19.44%)***	<0.0072	8 (14.04%)	6 (40.0%)	0.06	
	male	371 (66.25%)	**313 (64.14%)***	**58 (80.56%)***		49 (85.96%)	9 (60.0%)		
Comorbidities	count	0.0[0.0-1.0]	0.0±1.04	1.0±1.22	**<0.0002**	1.0±1.15	0.0±1.45	0.4072	
Current smoking		36 (6.43%)	34 (6.97%)	2 (2.78%)	0.2984	2 (3.51%)			
Chronic cardiac disease		20 (3.57%)	15 (3.07%)	5 (6.94%)	0.1611	4 (7.02%)	1 (6.67%)		
Hypertension		115 (20.54%)	92 (18.85%)	23 (31.94%)	**<0.018**	18 (31.58%)	5 (33.33%)	1	
Asthma		38 (6.79%)	31 (6.35%)	7 (9.72%)	0.3121	6 (10.53%)	1 (6.67%)		
Chronic kidney disease		7 (1.25%)	5 (1.02%)	2 (2.78%)		1 (1.75%)	1 (6.67%)		
Diabetes		98 (17.5%)	71 (14.55%)	27 (37.5%)	**<0.0001**	21 (36.84%)	6 (40.0%)	1	
Active malignant cancer		6 (1.07%)	4 (0.82%)	2 (2.78%)		1 (1.75%)	1 (6.67%)		
BMI	adm	27.0[23.92-30.44]	26.84±5.44	28.0±4.54	**<0.01**	27.82±4.7	31.14±0.48	0.2575	278
Body temperature, °C	adm	37.0[37.0-37.9]	37.0±0.63	38.0±0.97	**<0.0001**	38.0±0.97	38.0±0.98	0.3925	
HR BPM	adm	85.0[78.0-95.0]	84.5±12.32	94.5±19.97	**<0.0001**	95.0±20.93	85.0±15.3	0.1589	
SBP	adm	124.0[114.0-135.0]	123.0±16.51	126.0±17.31	0.2092	129.0±16.29	120.0±20.58	0.2122	
DBP	adm	78.0[70.0-84.0]	78.0±10.92	75.0±10.1	**<0.0208**	75.0±9.46	75.0±12.05	0.4254	
RR /min	adm	18.0[18.0-18.0]	18.0±1.56	25.0±6.74	**<0.0001**	24.0±6.95	28.0±5.62	0.1336	
SOFA score	adm	0.0[0.0-0.0]	0.0±0.75	3.0±2.85	**<0.0001**	3.0±2.42	4.0±3.69	**<0.0275**	4
WBC, $x10^9/L$	adm	5.8[4.5-7.2]	5.65±2.68	7.35±5.21	**<0.0001**	7.4±5.34	7.0±4.68	0.3801	3
	min	5.5[4.1-7.2]	5.5±7.72	7.0±6.68	**<0.0008**	7.2±6.93	5.5±5.38	0.0775	3
Platelet, $x10^9/L$	adm	224.0[180.25-272.0]	224.5±78.42	222.0±82.13	0.4102	225.0±86.02	196.0±57.76	0.0516	2
	min	224.0[178.0-272.0]	226.0±79.7	197.0±123.27	**<0.0049**	202.0±116.33	102.0±84.42	**<0.0001**	2
Lymphocyte, $x10^9/L$	adm	1.56[1.06-2.1]	1.66±0.76	0.81±2.97	**<0.0001**	0.83±3.32	0.73±0.64	0.4806	3
	min	1.49[0.89-2.09]	1.6±0.8	0.49±3.64	**<0.0001**	0.5±4.07	0.38±0.62	0.1412	3
T.bilirubin, umol/L	adm	9.0[6.0-12.6]	8.6±5.24	11.0±9.17	**<0.0001**	11.0±8.6	13.0±11.03	0.4094	11
	peak	9.85[6.5-14.38]	9.0±6.55	16.3±37.25	**<0.0001**	16.0±17.77	25.0±68.93	0.1412	10
ALT, U/L	adm	28.0[17.25-47.75]	27.0±34.84	39.0±38.04	**<0.0001**	39.0±39.5	41.0±31.76	0.4889	10
	peak	32.0[19.0-67.75]	28.5±50.05	102.5±7266.58	**<0.0001**	99.0±114.51	289.0±15305.74	**<0.0495**	10
AST, U/L	adm	24.0[18.0-36.22]	23.0±24.3	47.0±30.9	**<0.0001**	46.0±30.35	63.0±32.56	0.3722	10
	peak	25.5[19.0-44.0]	24.0±29.8	82.5±914.01	**<0.0001**	79.0±69.77	200.0±1715.26	**<0.0009**	10
D-Dimer, mg/L	adm	0.4[0.2-0.6]	0.3±0.72	1.15±3.13	**<0.0001**	1.1±2.96	1.4±3.62	0.1638	86
	peak	0.4[0.3-0.7]	0.3±0.73	2.6±7.56	**<0.0001**	1.6±6.37	18.0±7.12	**<0.0001**	86
APTT, sec	adm	37.4[35.0-41.05]	37.2±4.65	40.0±23.0	**<0.0014**	39.0±19.65	41.0±31.76	0.1429	73
	peak	38.0[35.15-42.35]	37.4±5.14	47.0±44.56	**<0.0001**	45.0±38.41	63.0±54.06	**<0.0005**	73
Creatinine, umol/L	adm	76.1[67.0-89.0]	75.4±27.52	80.5±54.62	0.0767	81.0±50.84	76.0±66.53	0.4448	6
	peak	78.0[67.78-91.0]	76.2±27.74	86.5±98.51	**<0.0001**	83.0±69.12	196.0±130.29	**<0.0003**	6
CK, U/L	adm	106.0[66.0-173.0]	99.0±529.25	173.0±1168.65	**<0.0001**	174.0±1278.56	152.0±561.74	0.2269	126
	peak	109.5[66.75-199.75]	100.0±536.11	391.0±10621.26	**<0.0001**	391.0±11963.38	370.0±563.66	0.4855	125
CRP, mg/L	adm	5.8[1.75-27.0]	4.2±32.27	101.0±105.14	**<0.0001**	102.0±102.19	100.0±115.53	0.4367	5
	peak	6.5[1.9-50.65]	4.8±45.93	157.5±113.35	**<0.0001**	143.0±108.72	219.0±115.19	**<0.0191**	5
LDH, U/L	adm	192.0[159.0-264.0]	181.0±80.08	445.0±267.95	**<0.0001**	432.5±284.01	480.0±199.68	0.2706	95
	peak	194.0[160.0-280.0]	182.0±83.76	538.0±1232.13	**<0.0001**	490.5±302.93	1925.0±2039.83	**<0.0001**	95
Troponin, ng/mL	adm	0.0[0.0-0.0]	0.0±0.15	0.0±1.31	**<0.0001**	0.0±0.04	0.0±2.73	0.0598	135
	peak	0.0[0.0-0.0]	0.0±0.18	0.04±1.85	**<0.0001**	0.0±0.26	0.36±3.66	**<0.0001**	135
Ferritin, ng/mL	adm	216.7[84.5-475.5]	181.95±876.92	725.0±2282.55	**<0.0001**	882.0±2480.17	612.0±1214.49	0.3036	53
	peak	230.0[89.95-595.5]	196.5±1530.13	2258.0±9784.72	**<0.0001**	2063.5±4781.9	4669.0±15029.77	**<0.0014**	53
Fibrinogen, mg/dL	adm	396.0[330.0-529.5]	377.0±187.31	610.0±199.71	**<0.0001**	612.0±204.96	567.0±179.01	0.3104	153
	peak	405.0[331.25-554.0]	380.0±130.61	700.0±735.07	**<0.0001**	701.0±816.38	692.0±252.63	0.1613	153
Clinical severity	asymp/mild	431 (76.96%)	**431 (88.32%)***	**0 (0.0%)***					
	severe	83 (14.82%)	**54 (11.07%)***	**29 (40.28%)***	**<0.0001**	**29 (50.88%)***	**0 (0.0%)***	**<0.0002**	
	critical	46 (8.21%)	**3 (0.61%)***	**43 (59.72%)***		**28 (49.12%)***	**15 (100.0%)***		
Ethnicity	White	60 (10.71%)	53 (10.86%)	7 (9.72%)		7 (12.28%)	0 (0.0%)		
	S.Asians	244 (43.57%)	206 (42.21%)	38 (52.78%)		28 (49.12%)	10 (66.67%)		
	M.Easterns	148 (26.43%)	**136 (27.87%)***	**12 (16.67%)***	0.1102	7 (12.28%)	5 (33.33%)	**<0.0219**	
	E.Asians	94 (16.79%)	79 (16.19%)	15 (20.83%)		**15 (26.32%)***	**0 (0.0%)***		
	Others	14 (2.5%)	14 (2.87%)	0 (0.0%)					
Onset to hospitalization days		14.0[8.0-19.0]	12.0±7.07	22.0±16.5	**<0.0001**	21.0±17.72	27.5±10.25	0.1336	72
Onset to positive PCR days		2.0[1.0-5.0]	2.0±3.89	5.0±4.97	**<0.0001**	5.0±5.01	4.0±4.79	0.3425	72
High-risk group patients		41 (7.32%)	3 (0.61%)	38 (52.78%)	**<0.0001**	24 (42.11%)	14 (93.33%)	**<0.0003**	
Discharged alive		545 (97.32%)	488 (100.0%)	57 (79.17%)	**<0.0001**	57 (100.0%)		**<0.0001**	
Length of stay in clinics		7.0[3.0-12.25]	6.0±8.25	16.0±16.08	**<0.0001**	16.0±17.34	23.0±9.97	0.1521	94
Duration of viral shedding, days		10.0[6.0-14.0]	10.5±5.64	8.0±9.04	0.0714	8.0±9.05	13.0±8.65	0.1304	28
Need for supplementary O_2		82 (14.64%)	23 (4.71%)	59 (81.94%)	**<0.0001**	46 (80.7%)	13 (86.67%)	0.7229	
Any complication		123 (21.96%)	53 (10.86%)	70 (97.22%)	**<0.0001**	55 (96.49%)	15 (100.0%)	1	
ARDS		76 (13.57%)	7 (1.43%)	69 (95.83%)	**<0.0001**	54 (94.74%)	15 (100.0%)	1	
Liver dysfunction		54 (9.64%)	23 (4.71%)	31 (43.06%)	**<0.0001**	23 (40.35%)	8 (53.33%)	0.3944	

* **adm** - data on admission; **min** - the minimal levels; **peak** - the peak levels

Table 2: Statistical significance of ROC AUC for predicting transfer to ICU out of the laboratory findings on admission

No	Feature	AUC	CI		p-value
1	AST	0.4882	[0.399	0.595]	0.828
2	ALT	0.5057	[0.482	0.538]	0.331
3	Total bilirubin	0.5573	[0.443	0.557]	0.077
4	LDH	0.5652	[0.515	0.644]	0.072
5	WBC	0.5727	[0.427	0.573]	0.035
6	Lymphocyte	0.5881	[0.474	0.588]	0.01
7	Troponin	0.6088	[0.5	0.609]	0.008
8	D-Dimer	0.6151	[0.5	0.615]	0.004
9	CK	0.6918	[0.6	0.725]	<0.001
10	Ferritin	0.6973	[0.616	0.74]	<0.001
11	APTT	0.7534	[0.219	0.755]	<0.001
12	Fibrinogen	0.7704	[0.718	0.771]	<0.001
13	CRP	0.8194	[0.798	0.822]	<0.001
APTT + CRP + Fibrinogen		0.8618	[0.486	0.884]	<0.001
All together		0.9019	[0.812	0.902]	<0.001

conditions are more vulnerable to severe outcomes. High sensitive laboratory findings are a reliable tool for assessing pathologies of these kinds. Reasonably, these findings may serve as predictors of the disease progression.

As it comes from feature selection, LDH activity is the laboratory finding that has maximal informative value for the prediction of worsening of the patient (see Appendix Table 1). This keeps up with the results of a pooled analysis that show an association of elevated LDH values with a 6-fold increase in odds of developing severe disease. Notably, the LDH cutoff in the included studies ranged from 240 to 253.2 U/L. The threshold value for the LDH activity in our study is 246 U/L which is close to the median of the range [4]. It is also known to be a predictor of worse outcomes in inpatients [32]. In our study, LDH is the top rank predictor of disease severity, CK levels have a medium informativeness. Both of them are unspecific biomarkers of energy deficiency and hypoxia. The levels of CRP have an expectedly high predictive value as they reflect the activity of an inflammatory process.

The concentration of D-Dimer seems to be a more promising biomarker of COVID-19 severity because of the endothelial dysfunction mechanism which is specific for this viral infection (see Subsection 1.2). For the same reason, APTT is an interesting predictor for SARS-CoV-2 infected patients. Therefore, recent studies justified the coagulation indicators on admission (e.g., D-Dimer, APTT, prothrombin time, and fibrinogen) as significant indicators of severe COVID-19 course [33].

From Appendix Table 1, fibrinogen values are not predictive of disease severity. The explanation to this discrepancy is many missing values for this indicator in our database. As it is seen from Table 1, the total number of 153 cases (27%) were missing. We had to replace them with the mean values to perform the multivariate prediction with the tree based model. The replacement decreased the real prognostic value, which was expected to be high. In contrast to this, the univariate model based on fibrinogen levels had the best classifying metrics compared to other predictors. Its ROC AUC value is 0.7704 (see Table 2).

5.3. Threshold criteria for the major clinical data

With the ML approach, we justify the cut-off thresholds for the major laboratory tests regularly done on admission.

The disproportion in the number of patients admitted to ICU versus non-severe cases was challenging. Therefore, we customized the ML algorithms in terms of threshold values used to predict worsening. For each laboratory findings feature, we (1) fit the model to the training dataset using 10-fold cross-validation, (2) predicted the probabilities on the test dataset, (3) found the optimal threshold value which maximizes the ROC AUC measure.

The optimized threshold values (marked in bold font in Table 3) can be used to predict the supposed deterioration of the patient from the initial findings at presentation. Some of the thresholds are close to the normal reference values, but not completely. For instance, the cut-off for CRP is 3 times bigger than the top reference value. The cut-offs that we found for WBC and total bilirubin are within the range of normal values for these laboratory findings. That is why it is challenging to interpret them.

The prediction based on C-reactive protein with ROC AUC equal to 0.8403 proved to be most accurate. A meta-analysis done by other authors showed that possibility to predict mortality for COVID-19 out of CRP with the same level of accuracy (ROC AUC 0.84) [17]. Unfortunately, they do not state clearly the time point for collecting the samples.

In our study the performance of the disease severity prediction based on the coagulation indicators was not so high (e.g., D-Dimer 0.7228; Fibrinogen 0.6774). However, it almost equals the results of ROC analyses for mortality risk by other authors who received AUCs value of 0.742 for D-Dimer on admission and 0.643 for AAPT on admission [33]. Other authors reached even better performance for the prediction of in-hospital mortality based on D-Dimer on admission (AUC 0.85).

Despite the similarities in performance metrics, the studies cannot be compared as they are based on different inclusion criteria, study cohorts, and threshold values found. In general, our findings support the idea of other researchers to use laboratory findings on admission for risk stratification. Moreover, they encourage the further studies to implement new biomarkers into prognostic models along with the proven ones [17].

5.4. The multivariable prediction of the severity of COVID-19

For better prediction, it is recommended that several biomarkers are analyzed concomitantly. A combination of three and thirteen most valuable ones, if fed to the deployed ML algorithm, provide a reliable prognosis. From Appendix Figure 2 it is clearly seen that there is a separability pattern within all variables used to build the predictive model. When we rank the features in accordance with their importance, most laboratory findings variables are listed at the top (see Appendix Table 1). It also helps to justify the threshold values, presented in this study.

6. Limitations

There are several limitations in the current study. First, the

Table 3: Justification of the cut-off levels for the admission values of laboratory findings to predict transferring to ICU

No	Feature	Normal values	Threshold moving technique				Percentile level			
			Cut-off	Sensitivity	Specificity	AUC	Cut-off	Sensitivity	Specificity	AUC*
1	WBC (x10^9/L)	4.0 - 11.0	45	0.6	0.5	0.5486	**7**	0.5278	0.75	**0.6389**
2	Lymphocytes (x10^9/L)	1 - 4.8	0.3	0.43	0.62	0.5267	**1.24**	0.7778	0.75	**0.7639**
3	T. bilirubin (umol/L)	3.4 - 20.5	37	0.54	0.43	0.4880	**11.9**	0.4861	0.7439	**0.6150**
4	ALT (U/L)	0 - 55	435	0.29	0.68	0.4880	**43**	0.4583	0.7439	**0.6011**
5	AST (U/L)	5 - 34	400	0.53	0.46	0.4944	**32**	0.7639	0.7418	**0.7528**
6	D-Dimer (mg/L)	0.0 - 0.5	15	0.35	0.7	0.5261	**0.7**	0.7222	0.7234	**0.7228**
7	APTT (sec)	28.0 - 40.0	180	0.57	0.71	0.6413	**39.9**	0.5139	0.7336	0.6237
8	CK (U/L)	30.0 - 200.0	4808	0.54	0.63	0.5864	**247**	0.4028	0.6619	0.5323
9	CRP (mg/L)	0.0 - 5.0	400	0.6	0.79	0.6921	**14.3**	0.9306	0.75	0.8403
10	LDH (U/L)	125 - 243	1778	0.21	0.88	0.5427	**246**	0.8889	0.6537	**0.7713**
11	Troponin (ng/mL)	<0.03	11	0.33	0.75	0.5427	**0.037**	0.2361	0.7172	0.4767
12	Ferritin (ng/mL)	21.8 - 274.6	14025	0.35	0.82	0.5824	**498**	0.6667	0.75	0.7083
13	Fibrinogen (mg/dL)	200-400	3030	0.33	0.89	0.6124	**446**	0.8611	0.4939	0.6774

* The AUC values marked in bold are higher than the ones displayed in Appendix Figure 3a.

dataset is unbalanced. Therefore, we customized the supervised ML algorithm in terms of the threshold value used to predict worsening. Second, the severity and mortality of the included patients might not be representative of the community because of the latent course of the mild and asymptomatic cases. Third, the population of Dubai is specific in terms of unequal age distribution and ethnic heterogeneity. However, one may consider the last feature as a strength because we can generalize the results to the world population. Forth, though other clinical examinations (e.g., diagnostic imaging) could provide additional information, we limited the predictors of disease deterioration to laboratory findings. None the less, this was enough to build up an ML algorithm with good performance. The concomitant analysis of the top three valuable biomarkers on admission provided a reliable prognosis without radiological predictors. Another advantage of the choice we made is the high applicability of study results into practice. The justified cut-off thresholds for the laboratory tests are easy to use on admission to the hospital.

7. Conclusion

- By comparing the data for the patients who were transported to ICU with those who did not worsen throughout the hospitalization we selected a set of laboratory findings with the significant differences on admission to the clinics. The variables were used as the predictors to build up the classification model. The performance of the models was low, with the default thresholds returned by the ML estimator, we improved it by setting the cut-off level to the 25^{th} percentile for lymphocyte count and the 75^{th} - for other features.

- To distinguish the patients with the confirmed COVID-19 who may worsen while treated we justified the following threshold values of the laboratory tests done on admission: lymphocyte count lower than 2.59x10^9/L, and the upper levels for total bilirubin - 11.9 umol/L, ALT - 43 U/L, AST - 32 U/L, D-Dimer - 0.7 mg/L, APTT - 39.9 sec, CK - 247 U/L, CRP - 14.3 mg/L, LDH - 246 U/L, Troponin - 0.037 ng/mL, Ferritin - 498 ng/mL, Fibrinogen - 446 mg/dL.

- 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). It is comparable with the model trained with all the tests (AUC 0.90; CI 0.812 - 0.902; p<0.001).

References

[1] L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, M. Zhang, et al., An interpretable mortality prediction model for covid-19 patients, Nature Machine Intelligence (2020) 1–6.

[2] L. Wynants, B. Van Calster, M. M. Bonten, G. S. Collins, T. P. Debray, M. De Vos, M. C. Haller, G. Heinze, K. G. Moons, R. D. Riley, et al., Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal, bmj 369 (2020).

[3] Y. Gao, T. Li, M. Han, X. Li, D. Wu, Y. Xu, Y. Zhu, Y. Liu, X. Wang, L. Wang, Diagnostic utility of clinical laboratory data determinations for patients with the severe covid-19, Journal of medical virology (2020).

[4] B. M. Henry, G. Aggarwal, J. Wong, S. Benoit, J. Vikse, M. Plebani, G. Lippi, Lactate dehydrogenase levels predict coronavirus disease 2019 (covid-19) severity and mortality: A pooled analysis, The American Jour- nal of Emergency Medicine (2020).

[5] R. Zhou, F. Li, F. Chen, H. Liu, J. Zheng, C. Lei, X. Wu, Viral dynamics in asymptomatic patients with covid-19, International Journal of Infectious Diseases (2020).

[6] J. Gong, J. Ou, X. Qiu, Y. Jie, Y. Chen, L. Yuan, J. Cao, M. Tan, W. Xu, F. Zheng, et al., A tool to early predict severe 2019-novel coronavirus pneumonia (covid-19): a multicenter study using the risk nomogram in wuhan and guangdong, china, medRxiv (2020).

[7] H.-Y. Zheng, M. Zhang, C.-X. Yang, N. Zhang, X.-C. Wang, X.-P. Yang, X.-Q. Dong, Y.-T. Zheng, Elevated exhaustion levels and reduced functional diversity of t cells in peripheral blood may predict severe progression in covid-19 patients, Cellular & molecular immunology 17 (2020) 541–543.

[8] X. Chen, B. Zhao, Y. Qu, Y. Chen, J. Xiong, Y. Feng, D. Men, Q. Huang, Liu, B. Yang, et al., Detectable serum sars-cov-2 viral load (rnaaemia) is closely correlated with drastically elevated interleukin 6 (il-6) level in critically ill covid-19 patients, Clinical Infectious Diseases (2020).

[9] W. H. Mahallawi, O. F. Khabour, Q. Zhang, H. M. Makhdoum, B. A. Suliman, Mers-cov infection in humans is associated with a pro-inflammatory th1 and th17 cytokine profile, Cytokine 104 (2018) 8–13.

[10] C. Selinger, J. Tisoncik-Go, V. D. Menachery, S. Agnihothram, G. L. Law, J. Chang, S. M. Kelly, P. Sova, R. S. Baric, M. G. Katze, Cytokine systems approach demonstrates differences in innate and pro-inflammatory host

responses between genetically distinct mers-cov isolates, BMC genomics 15 (2014) 1161.
[11] K. Kappert, A. Jahic´, R. Tauber, Assessment of serum ferritin as a biomarker in covid-19: bystander or participant? insights by comparison with other infectious and non-infectious diseases, Biomarkers (2020) 1–36.
[12] L. Spiezia, A. Boscolo, F. Poletto, L. Cerruti, I. Tiberio, E. Campello, P. Navalesi, P. Simioni, Covid-19-related severe hypercoagulability in patients admitted to intensive care unit for acute respiratory failure, Thrombosis and haemostasis 120 (2020) 998.
[13] N. Tang, D. Li, X. Wang, Z. Sun, Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia, Journal of thrombosis and haemostasis 18 (2020) 844–847.
[14] D. O. Griffin, A. Jensen, M. Khan, J. Chin, K. Chin, J. Saad, R. Parnell, C. Awwad, D. Patel, Pulmonary embolism and increased levels of d-dimer in patients with coronavirus disease, Emerging infectious diseases 26 (2020) 1941.
[15] I. Leonard-Lorant, X. Delabranche, F. Severac, J. Helms, C. Pauzet, O. Collange, F. Schneider, A. Labani, P. Bilbault, S. Moliere, et al., Acute pulmonary embolism in covid-19 patients on ct angiography and relationship to d-dimer levels, Radiology (2020).
[16] G. Lippi, E. J. Favaloro, D-dimer is associated with severity of coronavirus disease 2019: a pooled analysis, Thrombosis and haemostasis 120 (2020) 876.
[17] I. Huang, R. Pranata, M. A. Lim, A. Oehadian, B. Alisjahbana, C-reactive protein, procalcitonin, d-dimer, and ferritin in severe coronavirus disease-2019: a meta-analysis, Therapeutic advances in respiratory disease 14 (2020) 1753466620937175.
[18] Y. Yao, J. Cao, Q. Wang, Q. Shi, K. Liu, Z. Luo, X. Chen, S. Chen, K. Yu, Z. Huang, et al., D-dimer as a biomarker for disease severity and mortality in covid-19 patients: a case control study, Journal of intensive care 8 (2020) 1–11.
[19] N. Tang, H. Bai, X. Chen, J. Gong, D. Li, Z. Sun, Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy, Journal of thrombosis and haemostasis 18 (2020) 1094–1099.
[20] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, et al., Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china, Jama 323 (2020) 1061–1069.
[21] D. Bashash, H. Abolghasemi, S. Salari, M. Olfatifar, P. Eshghi, M. E. Akbari, Elevation of d-dimer, but not pt and aptt, reflects the progression of covid-19 toward an unfavorable outcome: A meta-analysis, Iranian Journal of Blood & Cancer (2020) 47–53.
[22] L. Quartuccio, A. Sonaglia, D. McGonagle, M. Fabris, M. Peghin, D. Pecori, A. De Monte, T. Bove, F. Curcio, F. Bassi, et al., Profiling covid-19 pneumonia progressing into the cytokine storm syndrome: results from a single italian centre study on tocilizumab versus standard of care, Journal of Clinical Virology (2020) 104444.
[23] W. Ling, C-reactive protein levels in the early stage of covid-19, Medecine et maladies infectieuses (2020).
[24] P. P. Bloom, E. A. Meyerowitz, Z. Reinus, M. Daidone, J. Gustafson, A. Y. Kim, E. Schaefer, R. T. Chung, Liver biochemistries in hospitalized patients with covid-19, Hepatology(2020).
[25] N. Ali, K. Hossain, Liver injury in severe covid-19 infection: current insights and challenges, Expert review of gastroenterology & hepatology (2020).
[26] N. Lee, P. K. Chan, D. S. Hui, T. H. Rainer, E. Wong, K.-W. Choi, G. C. Lui, B. C. Wong, R. Y. Wong, W.-Y. Lam, et al., Viral loads and duration of viral shedding in adult patients hospitalized with influenza, The Journal of infectious diseases 200 (2009) 492–500.
[27] National Emergency Crisis and Disasters Management Authority, National guidelines for clinical management and treatment of covid-19- version 4.1, https://www.dha.gov.ae/en/HealthRegulation/Documents/National_Guidelines_of_COVID_19_1st_June_2020.pdf, 2020. Accessed 01/08/2020.
[28] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, F. Herrera Learning from imbalanced data sets, Springer, 2018.
[29] E. V. Robilotti, N. E. Babady, P. A. Mead, T. Rolling, R. Perez-Johnston, M. Bernardes, Y. Bogler, M. Caldararo, C. J. Figueroa, M. S. Glickman, et al., Determinants of covid-19 disease severity in patients with cancer, Nature medicine 26 (2020) 1218–1223.
[30] K. Ravi, Ethnic disparities in covid-19 mortality: are comorbidities to blame?, The Lancet 396 (2020) 22.
[31] P. Baqui, I. Bica, V. Marra, A. Ercole, M. van Der Schaar, Ethnic and regional variations in hospital mortality from covid-19 in brazil: a cross-sectional observational study, The Lancet Global Health 8 (2020) e1018–e1026.
[32] A. Erez, O. Shental, J. Z. Tchebiner, M. Laufer-Perl, A. Wasserman, T. Sella, H. Guzner-Gur, Diagnostic and prognostic value of very high serum lactate dehydrogenase in admitted medical patients, Isr Med As- soc J 16 (2014) 439–443.
[33] H. Long, L. Nie, X. Xiang, H. Li, X. Zhang, X. Fu, H. Ren, W. Liu, Q. Wang, Q. Wu, D-dimer and prothrombin time are the significant indicators of severe covid-19 and poor prognosis, BioMed research international 2020 (2020).

8. Acknowledgments

The authors would like to acknowledge UAE University (Al Ain, UAE) and Mediclinic Parkview Hospital (Dubai, UAE) for the support provided and the facilities used for conducting this research. We also wish to express our gratitude to the dedication and commitment of our healthcare staff and our patients, without whom this research would not be possible.

9. Funding statement

This study was supported by the United Arab Emirates University StartUp grant (fund code 31M442) to cover the publication fees.

10. Author contributions statement

All authors contributed to the creation of the article as follows: all of them contributed to the conceptual idea of the paper equally; FA and YS formulated the objectives; FA collected the dataset; YS wrote the manuscript; TH proposed the methodology of the study, and performed the statistical analysis, prepared the figures and tables for data presentation and illustration, TL, KG, NZ contributed to the literature review and data analysis.

The data were analyzed and interpreted by the authors, who also reviewed the manuscript and vouch for the accuracy and completeness of the data and for the adherence of the study to the protocol.

11. Ethical Approval

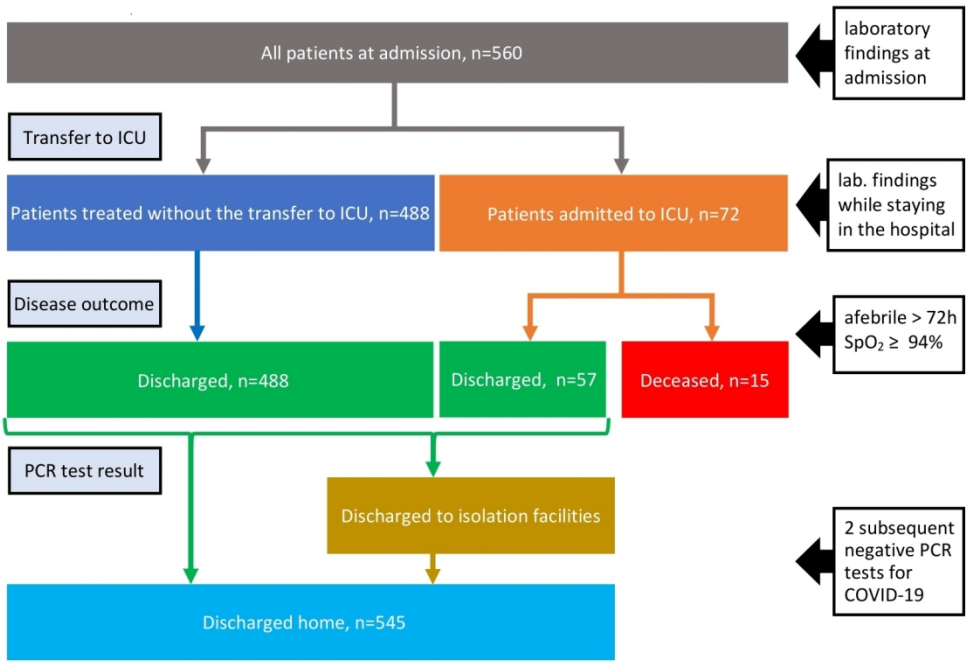
The study got an ethical review by Dubai Scientific Research Ethics Committee (DSREC), Dubai Health Authority, protocol No DSREC-05/2020_25) and was approved for the retrospective analysis of the data obtained as a standard of care. No potentially identifiable personal information is presented in the study.

12. Data availability statement

Generated Statement: The datasets generated for this study are available upon request at the site of **Data Analytics Group** at https://bi-dac.com

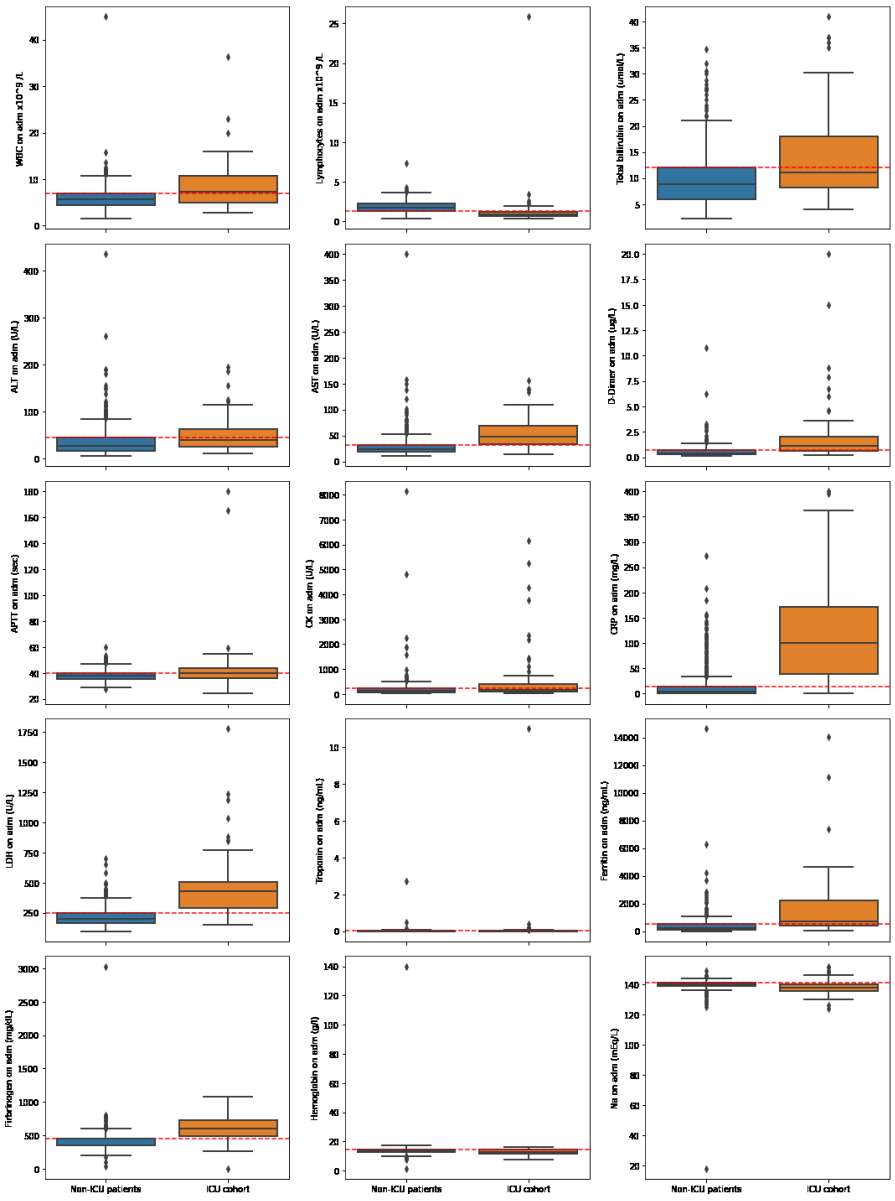
Figures

Figure 1. The flow of patients with COVID-19 in Dubai Mediclinic.
Figure 2. Variation of laboratory findings values in the ICU cohort (orange boxplot) versus the non-ICU cohort of patients (blue boxplot).
Appendix Figure 1. Feature selection for predicting whether a patient is going to be transferred to ICU.
Appendix Figure 2. The performance of the employed NN classification method.
Appendix Figure 3. ROC curves for the laboratory tests used as input to NN separately (a) and in the combination (b). The models are trained with 10 folds cross-validation.
Appendix Figure 4. The performance of the 10 folds cross-validation logistic regression model trained on binary data with the threshold moving technique returned by the ML estimator (a), and with the cut-off level set to the 25th percentile for lymphocyte count and 75th for the other features (b)

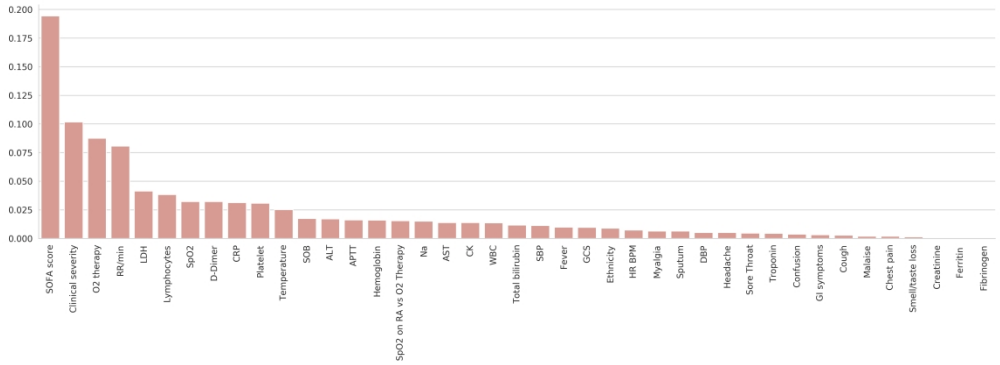


The flow of patients with COVID-19 in Dubai Mediclinic.

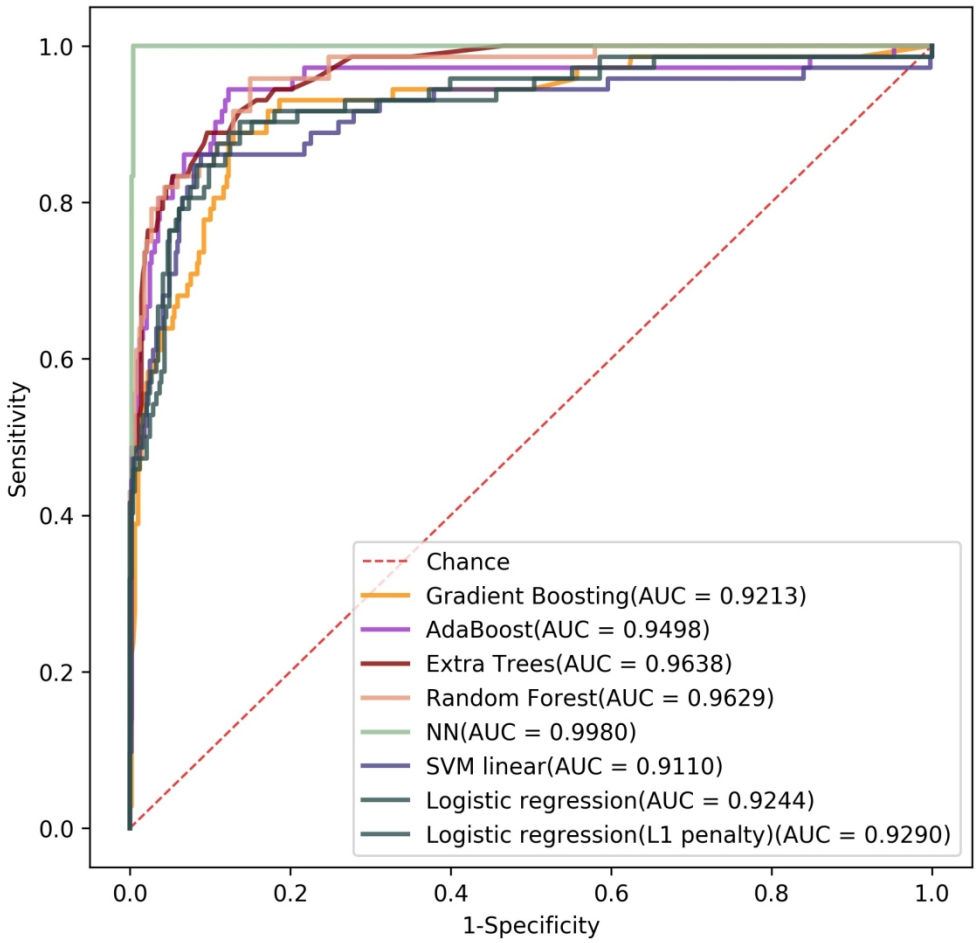
436x307mm (96 x 96 DPI)



Variation of laboratory findings values in the ICU cohort (orange boxplot) versus the non-ICU cohort of patients (blue boxplot).

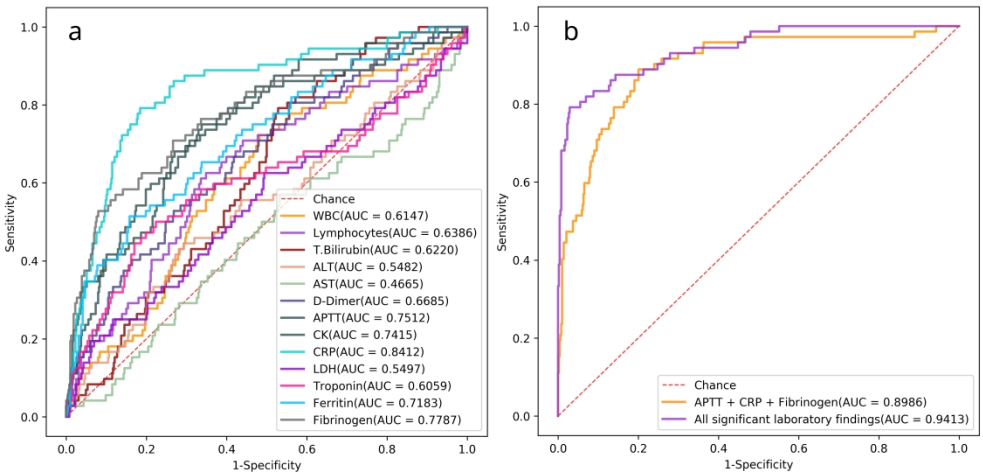


Feature selection for predicting whether a patient is going to be transferred to ICU.



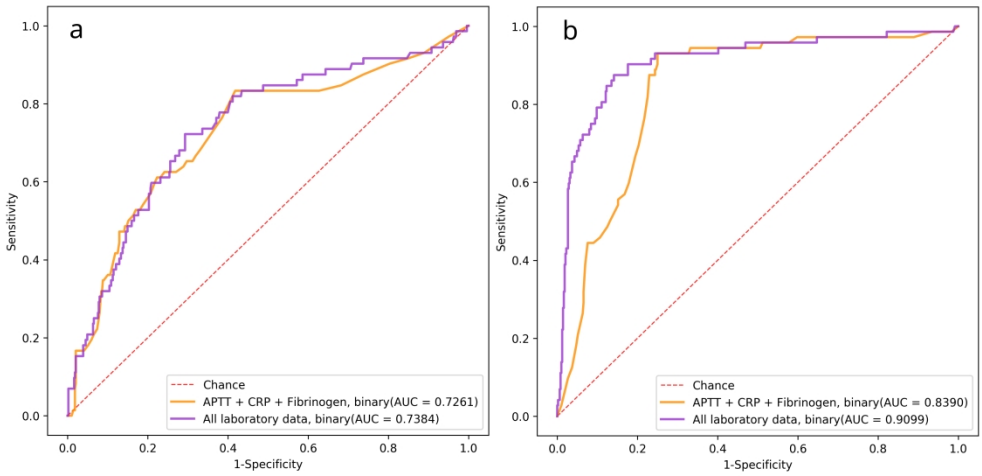
The performance of the employed NN classification method.

245x236mm (191 x 191 DPI)



ROC curves for the laboratory tests used as input to NN separately (a) and in the combination (b). The models are trained with 10 folds cross-validation.

312x150mm (300 x 300 DPI)



The performance of the 10 folds cross-validation logistic regression model trained on binary data with the threshold moving technique returned by the ML estimator (a), and with the cut-off level set to the 25th percentile for lymphocyte count and 75th for the other features (b).

312x150mm (300 x 300 DPI)

Appendix

A. ML classification models and feature selection.

The variables used to build up the model:

- *physical examination on admission*: temperature, HR BPM, SBP, DBP, RR /min. SpO_2, SpO_2 on RA vs. O_2 Therapy, GCS, SOFA score
- *symptoms on admission*: cough, sputum, sore throat, chest pain, SOB, fever, headache, confusion, having any gastrointestinal symptoms (e.g., nausea, vomiting, diarrhea), myalgia, malaise, loss of smell or taste.
- *laboratory findings on admission*: the count of WBC, platelet, and lymphocyte; the concentration of hemoglobin, total bilirubin, D-Dimer, creatinine, sodium, C-reactive protein, troponin, ferritin, fibrinogen; the activity of ALT, AST, CK, LDH; APTT.

Feature selection:

To check if there are unique patterns within the data that can unambiguously identify if the patient is going to be transferred to the intensive care unit, we utilized ML algorithms.

To assess the importance of the features fed to the ML models as predictors of admitted to ICU patients, we employed four ensemble tree-based estimators such as AdaBoost, Gradient Boosting, Random Forest, and Extra Trees. These models were trained on the whole dataset and used to rank the features in ascending order concerning their predictive potential. Figure 1 and Table 1 display the averaged values of impurity-based attribute ranks, where the average for each feature is calculated as the mean of rank values for the four ML methods mentioned above.

Table 1: Ranking scores of the variables selected for predicting the disease severity

Score	Feature	Score	Feature	Score	Feature	Score	Feature
0.19429	SOFA score	0.02520	Temperature	0.01164	Total bilirubin	0.00466	Sore Throat
0.10168	Clinical severity	0.01748	SOB	0.01135	SBP	0.00445	Troponin
0.08745	O2 therapy	0.01712	ALT	0.00983	Fever	0.00367	Confusion
0.08061	RR/min	0.01623	APTT	0.00969	GCS	0.00309	GI symptoms
0.04127	LDH	0.01595	Hemoglobin	0.00896	Ethnicity	0.00287	Cough
0.03829	Lymphocytes	0.01545	SpO2 on RA vs O2 Therapy	0.00732	HR BPM	0.00188	Malaise
0.03223	SpO2	0.01505	Na	0.00637	Myalgia	0.00186	Chest pain
0.03212	D-Dimer	0.01383	AST	0.00633	Sputum	0.00141	Smell/taste loss
0.03125	CRP	0.01382	CK	0.00524	DBP	0.00000	Creatinine
0.03067	Platelet	0.01360	WBC	0.00513	Headache	0.00000	Ferritin
						0.00000	Fibrinogen

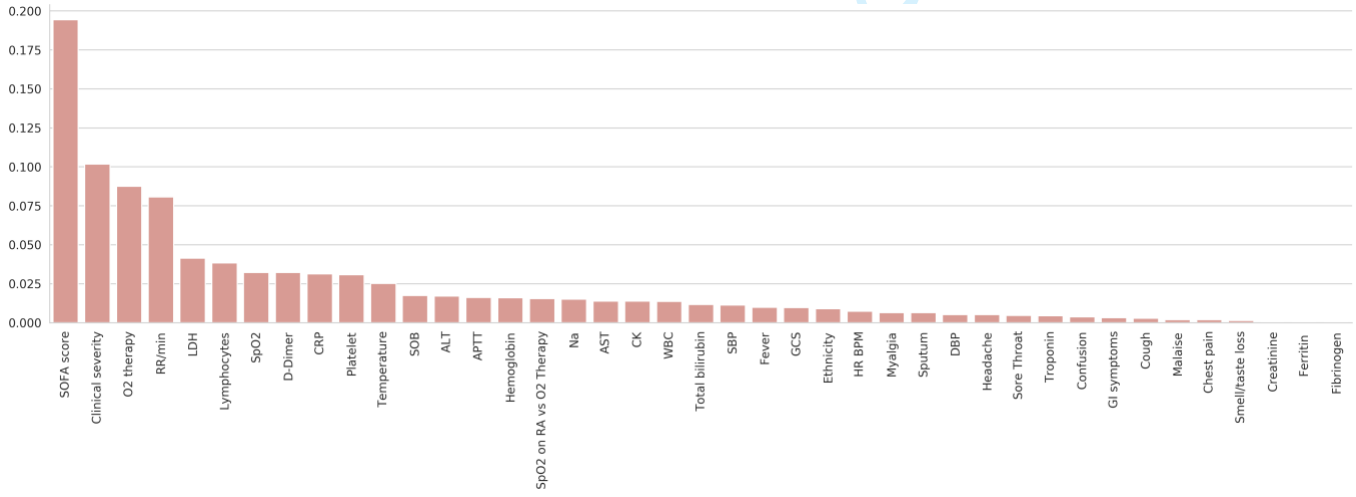


Figure 1: Feature selection for predicting whether a patient is going to be transferred to ICU

Table 2: Specificity and sensitivity of the ML model applied to the all features as predictors of the severity of the disease

ML model	Specificity	Sensitivity
Gradient Boosting	0.5972	0.9734
AdaBoost	0.6667	0.9775
Extra Trees	0.7361	0.9693
Random Forest	0.75	0.9795
NN	0.9938	1.0
SVM linear	0.6806	0.9508
Logistic regression	0.6667	0.952
Logistic regression (L1 penalty)	0.7083	0.959

Table 3: Confusion matrix to assess the accuracy of classification with a three-layer dense NN model to predict the severity of the disease

		Predicted	
		Not admitted to ICU	Admitted to ICU
Actual	Not admitted to ICU	485	3
	Admitted to ICU	0	72

Prediction of transferring to ICU. We utilized three-layer fully connected NN with the following configuration of hidden layers (35, 30, 10) and with the stochastic gradient descent optimizer. The learning rate hyperparameter of the model was assigned to 0.1. The model was also regularized using L2 penalty with 0.0001 alpha value. NN was trained for maximum 100 epochs or before converged. Convergence implies that the loss function is not improving by at least 0.0001 for 10 consecutive iterations.

To evaluate the classifier output quality, we trained several ML classification models using a stratified 10-fold cross-validation technique to generalize the models to the true rate error. For each fold, we used 90% of the data to train the model and then tested it on the rest 10%.

The decision matrices built on the test dataset for all folds were combined and used to calculate the performance metrics. The best performance measures were obtained with a three-layer fully connected NN.

Table 4: Classification metrics of the NN model to predict the event of being transferred to ICU

	Recall	Precision	F1 score	Support
Not admitted to ICU	1.00	0.99	1.00	488
Admitted to ICU	0.96	1.00	0.98	72
accuracy			0.99	560
macro average	0.98	1.00	0.99	560
weighted average	0.99	0.99	0.99	560

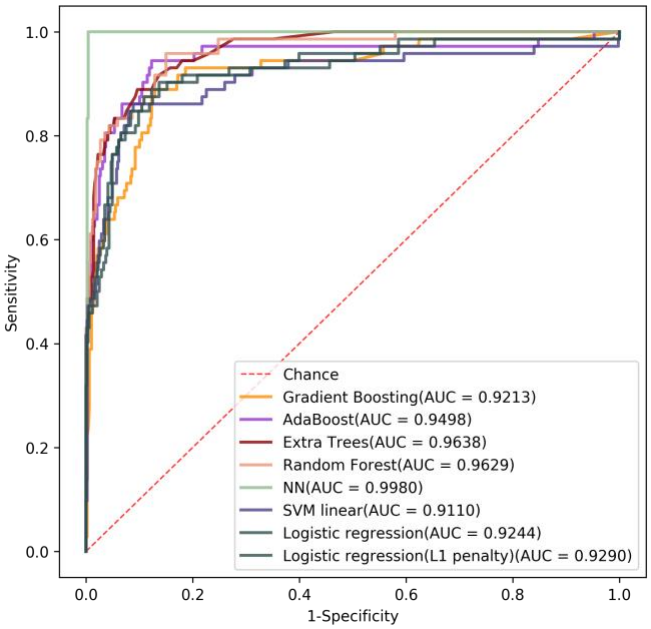


Figure 2: The performance of the employed NN classification method.

B. ROC curves for laboratory tests used as input to NN.

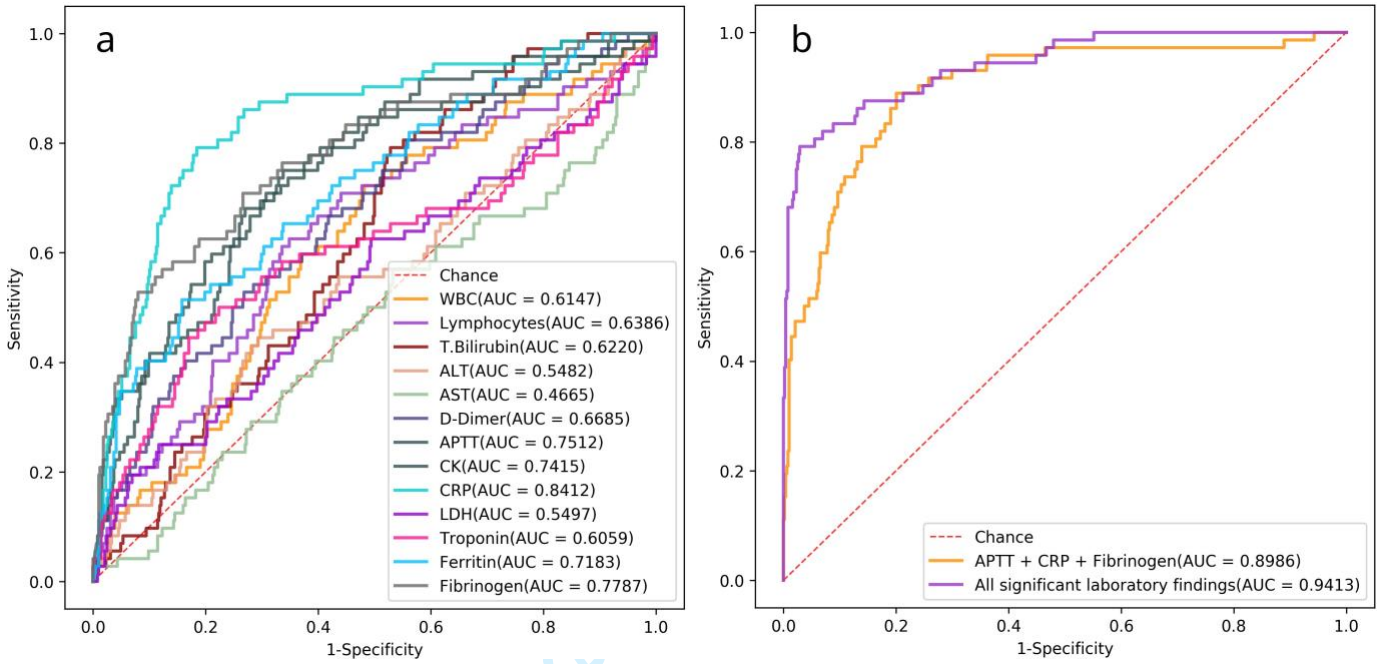


Figure 3: ROC curves for the laboratory tests used as input to NN separately (a) and in the combination (b). The models are trained with 10 folds cross-validation.

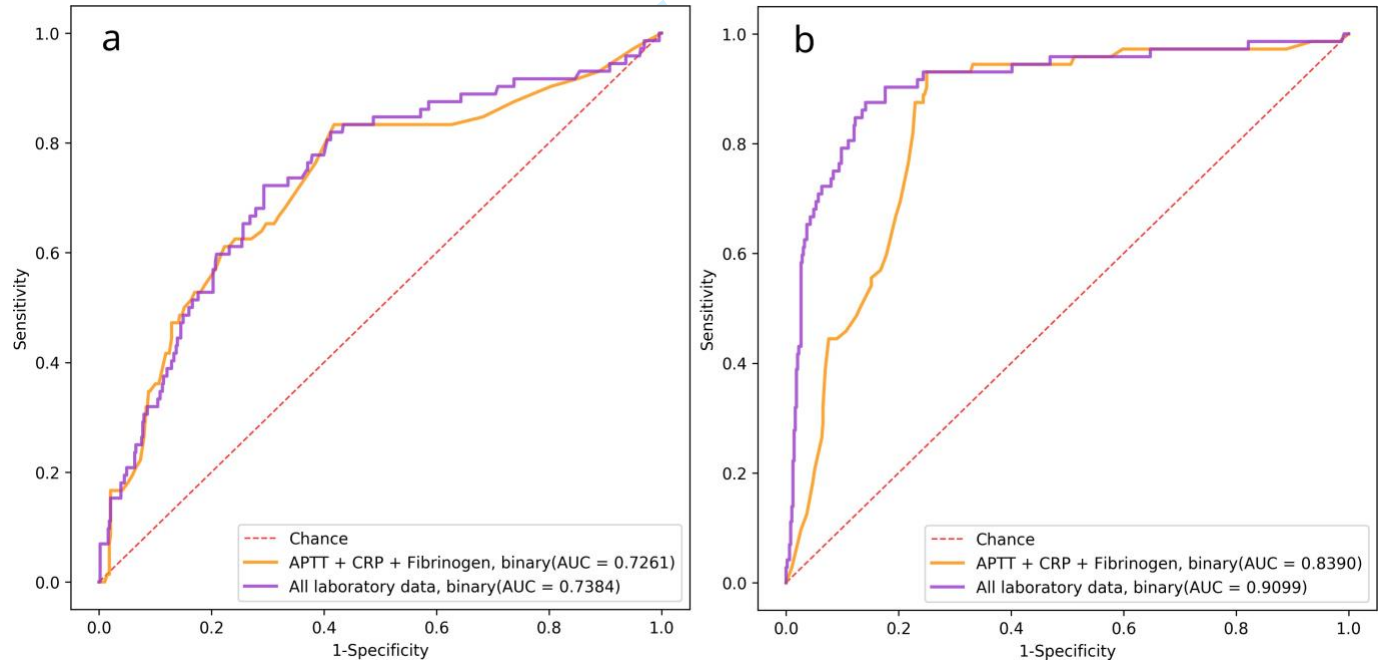


Figure 4: The performance of the 10 folds cross-validation logistic regression model trained on binary data with the threshold moving technique returned by the ML estimator (a), and with the cut-off level set to the 25^{th} percentile for lymphocyte count and 75^{th} for the other features (b).

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1 (Title)
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	1 (Abstract)
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	2/61 - 3/204 (sec. 1.1-1.2)
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	3/205-231 (sec. 2)
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	3/234-235 (sec. 3.1)
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	3/235-238 (sec. 3.1)
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	3/235 (sec. 3.1)
	5b	D;V	Describe eligibility criteria for participants.	3/241-246 (sec. 3.1)
	5c	D;V	Give details of treatments received, if relevant.	4/272-4/280 (sec. 3.1)
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	4/314-322 (sec. 3.2)
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	not applicable
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	4/295-296 (sec. 3.2) 12/758-770 (App. A)
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	not applicable
Sample size	8	D;V	Explain how the study size was arrived at.	3/235-238 (sec. 3.1)
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	4/291-293 , 4/310-313 (sec. 3.2)
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	4/282-284 (sec. 3.2)
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	4/295-301, 4/313-322 (sec. 3.2)
	10c	V	For validation, describe how the predictions were calculated.	4/323-329 (sec. 3.2)
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	4/299-301, 4/319-320 (sec. 3.2)
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	4/314-316 (sec. 3.2)
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	not applicable
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	4/323-329 (sec. 3.2)
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	3/248-4/271 (sec. 4.1) 5 (Figure 1)
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	7 (Table 1)
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	4/336-354 (sec. 4.1) 6 (Figure 2) 7 (Table 1)
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	4/332-335 (sec. 4.1)
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	-
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	13/785-793 (App. A)
	15b	D	Explain how to the use the prediction model.	5/381-392 (sec. 4.3)
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	5/370-380 (sec. 4.3); 8 (Table 2)
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	5/393-406 (sec. 4.3) 9 (Table 3)
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	9/553-10/571 (sec. 6)
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	9/504-9/534 (sec. 5.3)
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	8/437-461 (sec. 5.1)
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	9/537-541 (sec. 5.3)
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	11/753-755 (sec. 12)
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	11/729-731 (sec. 9)

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

BMJ Open

Prediction of COVID-19 severity using laboratory findings on admission: informative values, thresholds, ML model performance.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-044500.R2
Article Type:	Original research
Date Submitted by the Author:	30-Jan-2021
Complete List of Authors:	Statsenko, Yauhen; United Arab Emirates University College of Medicine and Health Sciences, Radiology Al Zahmi, Fatmah ; Mediclinic Parkview Hospital, Neurology; Mohammed Bin Rashid University Of Medicine and Health Sciences Habuza, Tetiana; UAE University College of Information Technology, Department of Computer science Gorkom, Klaus; United Arab Emirates University College of Medicine and Health Sciences, Radiology Zaki, Nazar; United Arab Emirates University,
Primary Subject Heading:	Health informatics
Secondary Subject Heading:	Infectious diseases, Research methods, Medical management, Intensive care, Respiratory medicine
Keywords:	COVID-19, BIOTECHNOLOGY & BIOINFORMATICS, INFECTIOUS DISEASES, Respiratory infections < THORACIC MEDICINE, Information technology < BIOTECHNOLOGY & BIOINFORMATICS, Biochemistry < NATURAL SCIENCE DISCIPLINES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Prediction of COVID-19 severity using laboratory findings on admission: informative values, thresholds, ML model performance.

Yauhen Statsenko[a], Fatmah Al Zahmi[b], Tetiana Habuza[c], Klaus Neidl Van Gorkom[a], Nazar Zaki[c]

[a]*College of Medicine and Health Sciences, United Arab Emirates University, P.O. Box 17666, Al Ain, United Arab Emirates*
[b]*Mediclinic Middle East Parkview hospital, P.O. Box 51122, Dubai, United Arab Emirates*
[c]*College of Information Technology, United Arab Emirates University, P.O. Box 15551, Al Ain, United Arab Emirates*

Abstract

Background: Despite the necessity, there is no reliable biomarker to predict disease severity and prognosis of COVID-19 patients. The currently published prediction models are not fully applicable to clinical use.

Objectives: To identify predictive biomarkers of COVID-19 severity and to justify their threshold values for the stratification of the risk of deterioration that would require transferring to ICU.

Methods: The study cohort (560 subjects) included all consecutive patients admitted to Dubai Mediclinic Parkview hospital from February to May 2020 with COVID-19 confirmed by the polymerase chain reaction. The challenge of finding the cut-off thresholds was the unbalanced dataset (e.g., the disproportion in the number of 72 patients admitted to ICU versus 488 non-severe cases). Therefore, we customized supervised ML algorithm in terms of threshold value used to predict worsening.

Results: With the default thresholds returned by the ML estimator, the performance of the models was low. It was improved by setting the cut-off level to the 25^{th} percentile for lymphocyte count and the 75^{th} - for other features.

The study justified the following threshold values of the laboratory tests done on admission: lymphocyte count lower than 2.59 $x10^9$/L, and the upper levels for total bilirubin - 11.9 umol/L, ALT - 43 U/L, AST - 32 U/L, D-Dimer - 0.7 mg/L, APTT - 39.9 sec, CK - 247 U/L, CRP - 14.3 mg/L, LDH - 246 U/L, Troponin - 0.037 ng/mL, Ferritin - 498 ng/mL, Fibrinogen - 446 mg/dL.

Conclusion: 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).

Keywords: COVID-19 pandemic , coronavirus, severity, biomarkers, threshold values, infectious disease

Strength and limitations of the study

- The research is based on a unique study cohort that is representative of the entire population because of the National Standard that required all patients with confirmed COVID-19 to be admitted to acute care hospitals regardless of their symptoms or illness severity.
- To distinguish the patients with the confirmed COVID-19 who may worsen while treated, we justified the threshold values of the laboratory tests done on admission.
- The prediction of the future deterioration by the neural network is reliable even with the top three valuable laboratory tests (APTT, CRP, and Fibrinogen) used for training (AUC 0.86; CI 0.486 - 0.884; p<0.001).
- The limitation of the study was the unbalanced dataset (e.g., the disproportion in the number of patients admitted to ICU versus non-severe cases).

*Corresponding authors.
Tel.: +971 3 713 7124; E-mail: e.a.statsenko@uaeu.ac.ae (Y. Statsenko);
Tel.:+971 4 416 8615; E-mail: fatmah.alzahmi@mediclinic.ae; (F. Al Zahmi).

Abbreviations

ALT - alanine aminotransferase
AST - aspartate aminotransferase
ARDS - acute respiratory distress syndrome
AUC - area under the curve
BMI - body mass index
CI - confidence interval
CK - creatine kinase
CoV - coronavirus
GCS - Glasgow coma scale
hs-CRP - high-sensitivity C-reactive protein
ICU - intensive care unit
IL - interleukin
LDH - lactate dehydrogenase
MERS - Middle East respiratory syndrome
ML - machine learning
NN - neural network
PC - precision-recall
PCR - polymerase chain reaction
RNA - ribonucleic acid
ROC - receiver operating characteristic
RR - respiratory rate

November 27, 2020

SARS-CoV-2 - severe acute respiratory syndrome-related coronavirus 2
SOB - shortness of breath
SOFA - Sequential organ failure assessment
TNF - tumor necrosis factor

Definitions

Mild level of COVID-19 severity - nonpneumonia and mild pneumonia.

Severe level of COVID-19 severity - dyspnea, respiratory frequency ≥30/min, blood oxygen saturation ≤93%, the partial pressure of arterial oxygen to fraction of inspired oxygen ratio <300, and/or lung infiltrates >50% within 24 to 48 hours.

Critical level of COVID-19 severity - respiratory, septic shock, and/or multiple organ dysfunction or failure.

1. Introduction

Despite the necessity, there is no reliable prognostic biomarker to predict disease severity and prognosis of COVID-19 patients [1]. Studies on COVID-19 have built up several types of prediction models. These have been the models designed to indicate the disease risk in the general population, the diagnostic models based on medical imaging, and the prognostic models. Unfortunately, these models have had some limitations that have precluded their use in clinical practice [2].

1.1. Models using laboratory findings as the inputs

Researchers tried to establish the role of laboratory findings in the diagnosis of COVID-19 [3]. They showed that the severe cases of COVID-19 were associated with D-dimer level over 0.28μg/L, interleukin (IL) 6 level over 24.3pg/mL [3], and LDH activity with an upper limit cut-off in the range of 240- 255U/L [4]). However, the use of these laboratory parameters with the above mentioned cut-off values was limited for the following reasons. First, these studies were conducted on severe forms of the disease. Limited research was done on patients who were asymptomatic or had mild disease [3, 5]. Second, the whole spectrum of the regularly used clinical laboratory data is unavailable for non-severe patients. Thus, the published papers add justification on the diagnostic utility of separate laboratory findings, instead of working out reliable diagnostic criteria for a set of them.

Gong and colleagues [6] have generated a tool for the early prediction of severe COVID-19 pneumonia out of the following data: age, serum lactate dehydrogenase activity, C-reactive protein, the coefficient of variation of red blood cell distribution width, blood urea nitrogen, direct bilirubin, lower albumin. The resulting performance was not high (sensitivity 77.5%, specificity 78.4%) [6]. Supposedly, this is because the dataset used as the input consists of exceptionally the age and laboratory findings.

In another model, the inputs included basic information, symptoms, and the results of laboratory tests. After the feature selection, the number of key features was set to just three laboratory results: LDH, lymphocytes, and high-sensitivity C-reactive protein (hs-CRP). The model was trained with the follow-up studies of the general, severe, and critical patients [1]. By feeding ML algorithm with the results obtained at the time of admission and in follow-up studies, the authors worked out a decision rule to predict patients at the highest risk. However, physicians are interested in the early prediction of the disease outcomes, and it is highly disputable that the model will not loose its predictive potential if applied exceptionally to the data received on admission.

We believe that a more accurate model can be built based on the simultaneous interpretation of laboratory results, clinical data, and physical examination findings (e.g., BMI, body temperature, respiratory rate) at the time of presentation. The analysis utilizing a machine learning algorithm could provide an accurate prediction of the disease severity.

1.2. Data used by clinicians for stratifying risks

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.

Inflammatory markers. There is evidence that IL-6, tumor necrosis factor-α do not indicate the level of COVID-19 progression [7]. Some markers of inflammation are elevated in the serum of COVID-19 patients compared to the healthy people, i.e., the serum SARS-CoV-2 viral load (RNAaemia) is closely correlated with drastically elevated interleukin 6 levels in critically ill COVID-19 patients [8]. However, there is no significant difference between severe and mild groups [7]. In contrast to this, the indicators are reflective in the progression of the diseases caused by other coronaviruses (e.g., MERS, SARS) [9]. This may be explained by the huge amino acid differences in viral proteins of distinct coronaviruses. Even with different MERS-CoV strains, common cytokine signaling by TNF and IL-1α results in the differential expression of innate immune genes [10].

Ferritin. Ferritin is a marker of iron storage. However, it is also an acute-phase reactant, the level of which elevates in processes of acute inflammation, whether infectious or non-infectious. Marked elevations have been reported in cases of COVID-19 infection [11].

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 interests 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. This is why 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, D-Dimer level of >2.14mg/L predicted in-hospital mortality with a sensitivity of 88.2% and specificity of 71.3% [18]. Another 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].

APTT. In a study conducted in February 2020, the levels of APTT as well as WBC, lymphocytes, AST, ALT, and creatinine, differed negligibly between severe and mild patients [3]. At the same time, other researchers showed inconsequential distinction in APTT in survivors versus non-survivors [20]. According to the results of another study published in March 2020, no significant difference in APTT values were found in the cohort of severe cases versus the non-severe one [6]. The results obtained in another study in April in Italy were the same [12]. The common limitation of these early studies was a small sample size. Finally, a meta-analysis justified that the elevation of D-Dimer, rather than prothrombin time and APTT, reflects the progression of COVID-19 toward an unfavorable outcome [21].

LDH and CK. Increased levels of the enzymes may reflect the level of the organ damage in a systemic disease [22, 4]. Reasonably, they may serve as biomarkers for COVID-19 progression.

CRP. In the early stage of COVID-19, CRP levels are positively correlated with the diameter of lung lesions and severe presentation [23].

Liver enzymes and total bilirubin. COVID-19 leads to elevated liver biochemistries (e.g., the level of AST, ALT, GGT, total bilirubin) in over 50% of patients on admission. AST - dominant aminotransferase elevation reflects the disease severity and true hepatic injury [24, 25].

2. Objectives

We decided to identify predictive biomarkers of COVID-19 severity and to justify their threshold values. Hypothetically, the absolute values of the biomarkers on admission to the clinics could provide physicians with an accurate prognosis on the future worsening of the patient that would require transferring the individual to the intensive care unit (ICU). Getting a reliable tool for such a prognosis will support decision making and logistical planning in clinics.

To address the objective, we designed a set of the following tasks:

- to study the linear separability of the laboratory findings values in patients with confirmed COVID-19 who were transferred to ICU versus non-severe cases of the disease, and to make the comparative analysis of the ICU department cases (both the deceased and survived cohorts) with other patients with COVID-19.
- to identify the risk factors by selecting the most valuable features for predicting the deterioration that would require transferring patient to ICU.
- to work out the threshold criteria for the major clinical data for the early identification of the patients with a high risk of being transferred to ICU.
- to identify the accuracy of the prediction of the patient's deterioration by the machine learning algorithm and by a set of the newly created threshold values of the laboratory and clinical findings.

3. Materials and methods

3.1. Study design and sample

We did a retrospective analysis of the clinical data obtained as a standard of primary and secondary care. The study sample included all the consecutive patients admitted to Dubai Mediclinic from 24th February to 1st July 2020 who fit the criteria of eligibility mentioned below (560 cases totally). Using this sample met the intention of the study: to allow for the early prognostic stratification.

The inclusion criteria were as follows: age 18 years or older; inpatient admission; SARS-CoV-2 positive real-time reverse-transcriptase polymerase chain reaction (PCR) from nasopharyngeal swabs only, at our site. Those patients who met the inclusion criteria for our studies were included in the study sample. All the patients were discharged at the time of writing the paper.

The remarkable feature of our study is that at the beginning of the pandemic, all the COVID-19 verified by PCR were hospitalized in the Mediclinic even if they did not present any

symptoms. We observed many mild and asymptomatic forms of the disease, with all the required spectrum of analyses being conducted. All patients who were hospitalized stayed in Dubai Mediclinic until they were afebrile for more than 72 h and had SpO2 value not less than 94%.

We assessed the duration of viral shedding as the number of days from the disease onset when the diagnosis was confirmed (e.g., the first positive PCR test) to the first negative PCR test [26]. 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 disease symptoms, the specimen collection continued on a daily basis until two subsequent negative PCR tests for COVID-19 more than 24 h apart. In the case of the mild disease course, patients might be transported to isolation facilities before being discharged home (see the flow chart diagram in Figure 1). If the facilities were run by Mediclinic, we had their follow up PCR results. For those patients who went to other isolation facilities not connected to Mediclinic, we couldn't study the duration of viral shedding (the data are missing for 27 out of 560 patients).

The treatment was administered in full accordance with "National Guidelines for Clinical Management and Treatment of COVID-19". The indications for the supportive oxygen therapy were (a) the oxygen saturation level below 94%, (b) the respiratory rate (RR) above 30 breaths per minute (c) both of them. In case of suspicion of superimposed bacterial pneumonia physicians ordered empirical broad-spectrum antibiotics. The administration of the antiviral and antimalarial drugs followed the national guidelines [27].

3.2. Patient and public involvement

No patient involved. The data were collected retrospectively from the medical record system.

3.3. Methods used

To address the first task, we studied the separability of laboratory findings values on admission to Dubai Mediclnic concerning the future transfer of the patient to the ICU department. To carry out the comparative analysis of features with regard to transferring to ICU, we utilized a set of non-parametric tests. The relationships involving two variables were assessed with the Mann-Whitney U test or Kruskal-Wallis test for the continuous features, and with Fisher's Exact test or Chi-square test for the quantitative ones. The data were expressed as *IQR*, *median* ± *std* or number of cases, and their percentage. The missing data for the comparative analysis were treated with the complete-case analysis method.

To address the second task, we used a set of different methods. First, we trained the NN ML model on each variable separately. To come up with laboratory data cut-off levels, which may be considered as bookmakers of severe course of the disease, we assessed their statistical significance against chance performance. We calculated 95% CI for ROC and ROC AUC scores with the bootstrap technique and p-values with permutation tests.

Second, we used ML tree-based methods (AdaBoost, Gradient Boosting, Random Forest, and Extra Trees) to check if there were unique patterns within the data that could unambiguously identify the event of transferring the patient to ICU from the data obtained on admission. For the list of features used as predictors see Appendix A. To assess the importance of the variables, we ranked all features concerning their impurity-based predictive potential. For ranking, we utilized a set of classifiers and then averaged all the received scores. Missing data in all ML models were replaced by the mean or median values with regard to the continuous or quantitative feature respectively utilizing single imputation method.

To tackle the third task, we used two approaches: a threshold moving technique (Youden's index) [28] and a heuristically chosen percentile-based cut-off level. The problem of predicting the transfer to ICU had a severe class imbalance. Therefore, we needed to focus on the performance of the classifier on the minority class (admitted to ICU patients). The sensitivity and specificity of the supervised ML classification model (NN) were used to evaluate the quality of the chosen optimal thresh- old for each important laboratory finding.

To evaluate the classifier output quality, we trained several ML classification models using a stratified 10-fold cross-validation technique to generalize the models to the true rate error. For each fold, we used 90% of the data to train the model and then tested it with the rest 10%. The decision matrices built on the test dataset for all folds were combined and used to calculate the performance metrics.

4. Results

4.1. Comparison of the ICU vs. non-ICU patients

The problem of predicting admission to ICU has a severe class imbalance (488 vs 72). Therefore, we need to focus on the performance of the classifier on the minority class (the patients admitted to ICU).

We look at the linear separability of the groups of numerical data composed from the laboratory findings values with regard to their quartiles. In Figure 2, boxplots for the laboratory findings data are presented with the red dashed line that marks the 75^{th} percentile for the subjects that were not transferred to ICU. The assumption is to use the third quartile (Q3) start point value as the threshold if there is separability between ICU and non- ICU groups. In diagrams in Figure 2, the red line indicates the 75^{th} percentile for not admitted to the ICU group. The exception is the diagram for the lymphocyte count, where it stands for the 25^{th} percentile.

The results of the comparative analysis of features with regard to transferring to ICU and the final outcomes of the disease are presented in Table 1. We excluded from further analysis the laboratory findings that did not significantly differ in the distribution of two groups. Therefore, we considered the list of 13 variables: WBC, lymphocyte count, total bilirubin, ALT, AST, D-Dimer, APTT, CK, CRP, LDH, troponin, ferritin, and fibrinogen on admission.

4.2. Feature ranking with regard to ML model performance

The features of the dataset listed in Appendix A were

ranked with four tree-based ML classifiers (e.g., Random Forest, AdaBoost, Gradient Boosting, and ExtraTrees). Tree-based models provide measures of feature importances. The classifiers are based on the mean decrease in impurity (MDI). The impurity is quantified by the splitting criterion of the decision trees. Averaged values of impurity-based attribute ranks were calculated as the mean of rank values for the algorithms mentioned above (see Appendix Figure 1). The classification performance is seen in Appendix Figure 2.

4.3. The cut-off levels of the laboratory findings

To come up with laboratory data cut-off levels, which may be considered as biomarkers of the severe course of the disease, we trained the NN ML model on each variable separately and assessed their statistical significance against chance performance. We calculated 95% CI for ROC and AUC scores with the bootstrap technique and p-values with permutation tests (see Table 2).

From Table 2, there is a notable difference between the performance of the model in terms of ROC AUC and the performance at chance level. High-performance measures were obtained for APTT, CRP, and Fibrinogen values (sensitivity and specificity are 0.9877 and 0.4028 respectively). It rises to 0.9754 and 0.75 respectively for all thirteen significant tests. So we used the performance of the classification model based on the combination of these three and thirteen features.

First we trained the ML model on the data of one lab feature in the 10-folds stratified cross-validation manner. Then we built ROC for the test data of all 10 folds (see diagrams in Appendix Figure 3).

We trained the ML model on the data taken from only one lab feature in the 10-folds stratified cross-validation manner and then built ROC and for the test data (combined from all 10 folds) as it is presented in Appendix Figure 3.

To improve the model's efficiency and choose the cut-off value set for some laboratory findings data, we used a threshold moving technique along with a supervised ML classification model (NN).

The ML estimator assigns threshold values for interpreting probabilities. The default threshold returned by the estimator to class labels is 0.5. However, when the dataset is unbalanced, tuning this hyperparameter can improve the model's efficiency by finding the optimal threshold. This is crucial when the importance of predicting the positive class (admitted to ICU) outweighs true negative predictions. Performance metrics calculated for all laboratory features with regard to the optimal threshold value are presented in Table 3. The table displays the sensitivity, specificity, and AUC values obtained after applying the threshold moving technique. We marked in bold the AUC values which are higher than the ones displayed in Appendix Figure 3a. The optimal cut-off value returned by the technique is shown in the appropriate column.

Looking at the boxplots in Figure 2 we decided to check whether the performance of the model is good if we applied thresholds in the following manner. For lymphocyte count, we set the cut-off level to the 25^{th} percentile (values lower than or equal to the chosen level were set to 1, or 0 otherwise). For the other features we set the thresholds to the 75^{th} percentile (values higher or equal to the cut-off limit were set to 1 or 0 otherwise). The performance of the models with regard to the aforementioned cut-off levels is presented in Table 3.

Appendix Figure 4a shows the performance of the logistic regression model built on the binary data by applying the cut-off level for the threshold moving technique. Appendix Figure 4b. illustrates the same information for the percentile's cut-off levels.

4.4. The performance of the classification models

The applied ML algorithms were trained with stratified 10-folds cross-validation technique. The predictors used are listed in Appendix Table 1. 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 Appendix Figure 2 and Appendix Table 2. It displays all 560 test points concatenated from test (actual and predicted) label values for each fold. Appendix Tables 3-4 show the performance metrics obtained by the NN model with the highest output quality. Appendix Figure 3 displays ROC curves and AUC for the NN model with different variables, observed on admission, as predictors. Appendix Figure 4 illustrates the quality of the performance for the binary data obtained by using the threshold moving or percentile-based heuristic approach.

5. Discussion

5.1. Severity of the disease course in SARS-CoV-2 infection

There are different risk factors for COVID-19 severity. Finding and justifying them are the issues of the ongoing studies because of the persistence of the viral infection. In research on the severe respiratory illness for COVID-19, the authors justified the age above 65 years as a predictor of clinical outcomes of interest [29]. The data we received support this fact. In the same study the authors showed inconsistent results regarding the race of the patient. In the univariate model, the race was a non-significant predictor of the disease severity, however it turned out to be significant in the multivariate prediction. We did not find ethnic differences between ICU and non-ICU cohorts, but observed a notable difference in the outcome of the disease within these groups (e.g., discharged vs. deceased patients). According to other studies, age is the largest contributor to risk of death for SARS-CoV-2, the impact of the race or ethnicity on the disease course remains not fully understood. The researchers have difficulty adjusting the samples for comorbidities as physicians did not examine all the patients thoroughly before the disease [30, 31]. Presumably, the same limitations account for disparities between the studies in which the authors try to consider comorbidities (e.g., asthma, diabetes, hypertension, chronic kidney disease, etc.) as risk factors. To overcome the limitation, we decided to base the prediction on the laboratory findings on admission. They are standardized and unambiguously interpretable.

5.2. Biomarkers of the deterioration of the patients

It is common sense that people with unmanaged chronic

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

		All patients				ICU patients			Missing values, count
		Total n_1=560	Not admitted to ICU n_2=488 (87.14%)	Admitted to ICU n_3=72 (12.86%)	p_{2-3}	Dead n_4=15 (20.83%)	Discharged n_5=57 (79.17%)	p_{4-5}	
Age		39.0[33.0-49.0]	38.0±11.97	51.0±13.08	**<0.0001**	46.0±12.56	62.0±11.01	**<0.0018**	
Gender	female	189 (33.75%)	**175 (35.86%)***	**14 (19.44%)***	<0.0072	8 (14.04%)	6 (40.0%)	0.06	
	male	371 (66.25%)	**313 (64.14%)***	**58 (80.56%)***		49 (85.96%)	9 (60.0%)		
Comorbidities	count	0.0[0.0-1.0]	0.0±1.04	1.0±1.22	**<0.0002**	1.0±1.15	0.0±1.45	0.4072	
Current smoking		36 (6.43%)	34 (6.97%)	2 (2.78%)	0.2984	2 (3.51%)			
Chronic cardiac disease		20 (3.57%)	15 (3.07%)	5 (6.94%)	0.1611	4 (7.02%)	1 (6.67%)		
Hypertension		115 (20.54%)	92 (18.85%)	23 (31.94%)	**<0.018**	18 (31.58%)	5 (33.33%)	1	
Asthma		38 (6.79%)	31 (6.35%)	7 (9.72%)	0.3121	6 (10.53%)	1 (6.67%)		
Chronic kidney disease		7 (1.25%)	5 (1.02%)	2 (2.78%)		1 (1.75%)	1 (6.67%)		
Diabetes		98 (17.5%)	71 (14.55%)	27 (37.5%)	**<0.0001**	21 (36.84%)	6 (40.0%)	1	
Active malignant cancer		6 (1.07%)	4 (0.82%)	2 (2.78%)		1 (1.75%)	1 (6.67%)		
BMI	adm	27.0[23.92-30.44]	26.84±5.44	28.0±4.54	**<0.01**	27.82±4.7	31.14±0.48	0.2575	278
Body temperature, ˚C	adm	37.0[37.0-37.9]	37.0±0.63	38.0±0.97	**<0.0001**	38.0±0.97	38.0±0.98	0.3925	
HR BPM	adm	85.0[78.0-95.0]	84.5±12.32	94.5±19.97	**<0.0001**	95.0±20.93	85.0±15.3	0.1589	
SBP	adm	124.0[114.0-135.0]	123.0±16.51	126.0±17.31	0.2092	129.0±16.29	120.0±20.58	0.2122	
DBP	adm	78.0[70.0-84.0]	78.0±10.92	75.0±10.1	**<0.0208**	75.0±9.46	75.0±12.05	0.4254	
RR /min	adm	18.0[18.0-18.0]	18.0±1.56	25.0±6.74	**<0.0001**	24.0±6.95	28.0±5.62	0.1336	
SOFA score	adm	0.0[0.0-0.0]	0.0±0.75	3.0±2.85	**<0.0001**	3.0±2.42	4.0±3.69	**<0.0275**	4
WBC, $x10^9/L$	adm	5.8[4.5-7.2]	5.65±2.68	7.35±5.21	**<0.0001**	7.4±5.34	7.0±4.68	0.3801	3
	min	5.5[4.1-7.2]	5.5±7.72	7.0±6.68	**<0.0008**	7.2±6.93	5.5±5.38	0.0775	3
Platelet, $x10^9/L$	adm	224.0[180.25-272.0]	224.5±78.42	222.0±82.13	0.4102	225.0±86.02	196.0±57.76	0.0516	2
	min	224.0[178.0-272.0]	226.0±79.7	197.0±123.27	**<0.0049**	202.0±116.33	102.0±84.42	**<0.0001**	2
Lymphocyte, $x10^9/L$	adm	1.56[1.06-2.1]	1.66±0.76	0.81±2.97	**<0.0001**	0.83±3.32	0.73±0.64	0.4806	3
	min	1.49[0.89-2.09]	1.6±0.8	0.49±3.64	**<0.0001**	0.5±4.07	0.38±0.62	0.1412	3
T.bilirubin, umol/L	adm	9.0[6.0-12.6]	8.6±5.24	11.0±9.17	**<0.0001**	11.0±8.6	13.0±11.03	0.4094	11
	peak	9.85[6.5-14.38]	9.0±6.55	16.3±37.25	**<0.0001**	16.0±17.77	25.0±68.93	0.1412	10
ALT, U/L	adm	28.0[17.25-47.75]	27.0±34.84	39.0±38.04	**<0.0001**	39.0±39.5	41.0±31.76	0.4889	10
	peak	32.0[19.0-67.75]	28.5±50.05	102.5±7266.58	**<0.0001**	99.0±114.51	289.0±15305.74	**<0.0495**	10
AST, U/L	adm	24.0[18.0-36.22]	23.0±24.3	47.0±30.9	**<0.0001**	46.0±30.35	63.0±32.56	0.3722	10
	peak	25.5[19.0-44.0]	24.0±29.8	82.5±914.01	**<0.0001**	79.0±69.77	200.0±1715.26	**<0.0009**	10
D-Dimer, mg/L	adm	0.4[0.2-0.6]	0.3±0.72	1.15±3.13	**<0.0001**	1.1±2.96	1.4±3.62	0.1638	86
	peak	0.4[0.3-0.7]	0.3±0.73	2.6±7.56	**<0.0001**	1.6±6.37	18.0±7.12	**<0.0001**	86
APTT, sec	adm	37.4[35.0-41.05]	37.2±4.65	40.0±23.0	**<0.0014**	39.0±19.65	41.0±31.76	0.1429	73
	peak	38.0[35.15-42.35]	37.4±5.14	47.0±44.56	**<0.0001**	45.0±38.41	63.0±54.06	**<0.0005**	73
Creatinine, umol/L	adm	76.1[67.0-89.0]	75.4±27.52	80.5±54.62	0.0767	81.0±50.84	76.0±66.53	0.4448	6
	peak	78.0[67.78-91.0]	76.2±27.74	86.5±98.51	**<0.0001**	83.0±69.12	196.0±130.29	**<0.0003**	6
CK, U/L	adm	106.0[66.0-173.0]	99.0±529.25	173.0±1168.65	**<0.0001**	174.0±1278.56	152.0±561.74	0.2269	126
	peak	109.5[66.75-199.75]	100.0±536.11	391.0±10621.26	**<0.0001**	391.0±11963.38	370.0±563.66	0.4855	125
CRP, mg/L	adm	5.8[1.75-27.0]	4.2±32.27	101.0±105.14	**<0.0001**	102.0±102.19	100.0±115.53	0.4367	5
	peak	6.5[1.9-50.65]	4.8±45.93	157.5±113.35	**<0.0001**	143.0±108.72	219.0±115.19	**<0.0191**	5
LDH, U/L	adm	192.0[159.0-264.0]	181.0±80.08	445.0±267.95	**<0.0001**	432.5±284.01	480.0±199.68	0.2706	95
	peak	194.0[160.0-280.0]	182.0±83.76	538.0±1232.13	**<0.0001**	490.5±302.93	1925.0±2039.83	**<0.0001**	95
Troponin, ng/mL	adm	0.0[0.0-0.0]	0.0±0.15	0.0±1.31	**<0.0001**	0.0±0.04	0.0±2.73	0.0598	135
	peak	0.0[0.0-0.0]	0.0±0.18	0.04±1.85	**<0.0001**	0.0±0.26	0.36±3.66	**<0.0001**	135
Ferritin, ng/mL	adm	216.7[84.5-475.5]	181.95±876.92	725.0±2282.55	**<0.0001**	882.0±2480.17	612.0±1214.49	0.3036	53
	peak	230.0[89.95-595.5]	196.5±1530.13	2258.0±9784.72	**<0.0001**	2063.5±4781.9	4669.0±15029.77	**<0.0014**	53
Fibrinogen, mg/dL	adm	396.0[330.0-529.5]	377.0±187.31	610.0±199.71	**<0.0001**	612.0±204.96	567.0±179.01	0.3104	153
	peak	405.0[331.25-554.0]	380.0±130.61	700.0±735.07	**<0.0001**	701.0±816.38	692.0±252.63	0.1613	153
Clinical severity	asymp/mild	431 (76.96%)	**431 (88.32%)***	**0 (0.0%)***	**<0.0001**				
	severe	83 (14.82%)	**54 (11.07%)***	**29 (40.28%)***		**29 (50.88%)***	**0 (0.0%)***	**<0.0002**	
	critical	46 (8.21%)	**3 (0.61%)***	**43 (59.72%)***		**28 (49.12%)***	**15 (100.0%)***		
Ethnicity	White	60 (10.71%)	53 (10.86%)	7 (9.72%)	0.1102	7 (12.28%)	0 (0.0%)	**<0.0219**	
	S.Asians	244 (43.57%)	206 (42.21%)	38 (52.78%)		28 (49.12%)	10 (66.67%)		
	M.Easterns	148 (26.43%)	**136 (27.87%)***	**12 (16.67%)***		7 (12.28%)	5 (33.33%)		
	E.Asians	94 (16.79%)	79 (16.19%)	15 (20.83%)		**15 (26.32%)***	**0 (0.0%)***		
	Others	14 (2.5%)	14 (2.87%)	0 (0.0%)					
Onset to hospitalization days		14.0[8.0-19.0]	12.0±7.07	22.0±16.5	**<0.0001**	21.0±17.72	27.5±10.25	0.1336	72
Onset to positive PCR days		2.0[1.0-5.0]	2.0±3.89	5.0±4.97	**<0.0001**	5.0±5.01	4.0±4.79	0.3425	72
High-risk group patients		41 (7.32%)	3 (0.61%)	38 (52.78%)	**<0.0001**	24 (42.11%)	14 (93.33%)	**<0.0003**	
Discharged alive		545 (97.32%)	488 (100.0%)	57 (79.17%)	**<0.0001**	57 (100.0%)		**<0.0001**	
Length of stay in clinics		7.0[3.0-12.25]	6.0±8.25	16.0±16.08	**<0.0001**	16.0±17.34	23.0±9.97	0.1521	94
Duration of viral shedding, days		10.0[6.0-14.0]	10.5±5.64	8.0±9.04	0.0714	8.0±9.05	13.0±8.65	0.1304	28
Need for supplementary O_2		82 (14.64%)	23 (4.71%)	59 (81.94%)	**<0.0001**	46 (80.7%)	13 (86.67%)	0.7229	
Any complication		123 (21.96%)	53 (10.86%)	70 (97.22%)	**<0.0001**	55 (96.49%)	15 (100.0%)	1	
ARDS		76 (13.57%)	7 (1.43%)	69 (95.83%)	**<0.0001**	54 (94.74%)	15 (100.0%)	1	
Liver dysfunction		54 (9.64%)	23 (4.71%)	31 (43.06%)	**<0.0001**	23 (40.35%)	8 (53.33%)	0.3944	

* **adm** - data on admission; **min** - the minimal levels; **peak** - the peak levels

Table 2: Statistical significance of ROC AUC for predicting transfer to ICU out of the laboratory findings on admission

No	Feature	AUC	CI		p-value
1	AST	0.4882	[0.399	0.595]	0.828
2	ALT	0.5057	[0.482	0.538]	0.331
3	Total bilirubin	0.5573	[0.443	0.557]	0.077
4	LDH	0.5652	[0.515	0.644]	0.072
5	WBC	0.5727	[0.427	0.573]	0.035
6	Lymphocyte	0.5881	[0.474	0.588]	0.01
7	Troponin	0.6088	[0.5	0.609]	0.008
8	D-Dimer	0.6151	[0.5	0.615]	0.004
9	CK	0.6918	[0.6	0.725]	<0.001
10	Ferritin	0.6973	[0.616	0.74]	<0.001
11	APTT	0.7534	[0.219	0.755]	<0.001
12	Fibrinogen	0.7704	[0.718	0.771]	<0.001
13	CRP	0.8194	[0.798	0.822]	<0.001
APTT + CRP + Fibrinogen		0.8618	[0.486	0.884]	<0.001
All together		0.9019	[0.812	0.902]	<0.001

conditions are more vulnerable to severe outcomes. High sensitive laboratory findings are a reliable tool for assessing pathologies of these kinds. Reasonably, these findings may serve as predictors of the disease progression.

As it comes from feature selection, LDH activity is the laboratory finding that has maximal informative value for the prediction of worsening of the patient (see Appendix Table 1). This keeps up with the results of a pooled analysis that show an association of elevated LDH values with a 6-fold increase in odds of developing severe disease. Notably, the LDH cutoff in the included studies ranged from 240 to 253.2 U/L. The threshold value for the LDH activity in our study is 246 U/L which is close to the median of the range [4]. It is also known to be a predictor of worse outcomes in inpatients [32]. In our study, LDH is the top rank predictor of disease severity, CK levels have a medium informativeness. Both of them are unspecific biomarkers of energy deficiency and hypoxia. The levels of CRP have an expectedly high predictive value as they reflect the activity of an inflammatory process.

The concentration of D-Dimer seems to be a more promising biomarker of COVID-19 severity because of the endothelial dysfunction mechanism which is specific for this viral infection (see Subsection 1.2). For the same reason, APTT is an interesting predictor for SARS-CoV-2 infected patients. Therefore, recent studies justified the coagulation indicators on admission (e.g., D-Dimer, APTT, prothrombin time, and fibrinogen) as significant indicators of severe COVID-19 course [33].

From Appendix Table 1, fibrinogen values are not predictive of disease severity. The explanation to this discrepancy is many missing values for this indicator in our database. As it is seen from Table 1, the total number of 153 cases (27%) were missing. We had to replace them with the mean values to perform the multivariate prediction with the tree based model. The replacement decreased the real prognostic value, which was expected to be high. In contrast to this, the univariate model based on fibrinogen levels had the best classifying metrics compared to other predictors. Its ROC AUC value is 0.7704 (see Table 2).

5.3. Threshold criteria for the major clinical data

With the ML approach, we justify the cut-off thresholds for the major laboratory tests regularly done on admission.

The disproportion in the number of patients admitted to ICU versus non-severe cases was challenging. Therefore, we customized the ML algorithms in terms of threshold values used to predict worsening. For each laboratory findings feature, we (1) fit the model to the training dataset using 10-fold cross-validation, (2) predicted the probabilities on the test dataset, (3) found the optimal threshold value which maximizes the ROC AUC measure.

The optimized threshold values (marked in bold font in Table 3) can be used to predict the supposed deterioration of the patient from the initial findings at presentation. Some of the thresholds are close to the normal reference values, but not completely. For instance, the cut-off for CRP is 3 times bigger than the top reference value. The cut-offs that we found for WBC and total bilirubin are within the range of normal values for these laboratory findings. That is why it is challenging to interpret them.

The prediction based on C-reactive protein with ROC AUC equal to 0.8403 proved to be most accurate. A meta-analysis done by other authors showed that possibility to predict mortality for COVID-19 out of CRP with the same level of accuracy (ROC AUC 0.84) [17]. Unfortunately, they do not state clearly the time point for collecting the samples.

In our study the performance of the disease severity prediction based on the coagulation indicators was not so high (e.g., D-Dimer 0.7228; Fibrinogen 0.6774). However, it almost equals the results of ROC analyses for mortality risk by other authors who received AUCs value of 0.742 for D-Dimer on admission and 0.643 for AAPT on admission [33]. Other authors reached even better performance for the prediction of in-hospital mortality based on D-Dimer on admission (AUC 0.85).

Despite the similarities in performance metrics, the studies cannot be compared as they are based on different inclusion criteria, study cohorts, and threshold values found. In general, our findings support the idea of other researchers to use laboratory findings on admission for risk stratification. Moreover, they encourage the further studies to implement new biomarkers into prognostic models along with the proven ones [17].

5.4. The multivariable prediction of the severity of COVID-19

For better prediction, it is recommended that several biomarkers are analyzed concomitantly. A combination of three and thirteen most valuable ones, if fed to the deployed ML algorithm, provide a reliable prognosis. From Appendix Figure 2 it is clearly seen that there is a separability pattern within all variables used to build the predictive model. When we rank the features in accordance with their importance, most laboratory findings variables are listed at the top (see Appendix Table 1). It also helps to justify the threshold values, presented in this study.

6. Limitations

There are several limitations in the current study. First, the

Table 3: Justification of the cut-off levels for the admission values of laboratory findings to predict transferring to ICU

No	Feature	Normal values	Threshold moving technique				Percentile level			
			Cut-off	Sensitivity	Specificity	AUC	Cut-off	Sensitivity	Specificity	AUC*
1	WBC (x10^9/L)	4.0 - 11.0	45	0.6	0.5	0.5486	**7**	0.5278	0.75	**0.6389**
2	Lymphocytes (x10^9/L)	1 - 4.8	0.3	0.43	0.62	0.5267	**1.24**	0.7778	0.75	**0.7639**
3	T. bilirubin (umol/L)	3.4 - 20.5	37	0.54	0.43	0.4880	**11.9**	0.4861	0.7439	**0.6150**
4	ALT (U/L)	0 - 55	435	0.29	0.68	0.4880	**43**	0.4583	0.7439	**0.6011**
5	AST (U/L)	5 - 34	400	0.53	0.46	0.4944	**32**	0.7639	0.7418	**0.7528**
6	D-Dimer (mg/L)	0.0 - 0.5	15	0.35	0.7	0.5261	**0.7**	0.7222	0.7234	**0.7228**
7	APTT (sec)	28.0 - 40.0	180	0.57	0.71	0.6413	**39.9**	0.5139	0.7336	0.6237
8	CK (U/L)	30.0 - 200.0	4808	0.54	0.63	0.5864	**247**	0.4028	0.6619	0.5323
9	CRP (mg/L)	0.0 - 5.0	400	0.6	0.79	0.6921	**14.3**	0.9306	0.75	0.8403
10	LDH (U/L)	125 - 243	1778	0.21	0.88	0.5427	**246**	0.8889	0.6537	**0.7713**
11	Troponin (ng/mL)	<0.03	11	0.33	0.75	0.5427	**0.037**	0.2361	0.7172	0.4767
12	Ferritin (ng/mL)	21.8 - 274.6	14025	0.35	0.82	0.5824	**498**	0.6667	0.75	0.7083
13	Fibrinogen (mg/dL)	200-400	3030	0.33	0.89	0.6124	**446**	0.8611	0.4939	0.6774

* The AUC values marked in bold are higher than the ones displayed in Appendix Figure 3a.

dataset is unbalanced. Therefore, we customized the supervised ML algorithm in terms of the threshold value used to predict worsening. Second, the severity and mortality of the included patients might not be representative of the community because of the latent course of the mild and asymptomatic cases. Third, the population of Dubai is specific in terms of unequal age distribution and ethnic heterogeneity. However, one may consider the last feature as a strength because we can generalize the results to the world population. Forth, though other clinical examinations (e.g., diagnostic imaging) could provide additional information, we limited the predictors of disease deterioration to laboratory findings. None the less, this was enough to build up an ML algorithm with good performance. The concomitant analysis of the top three valuable biomarkers on admission provided a reliable prognosis without radiological predictors. Another advantage of the choice we made is the high applicability of study results into practice. The justified cut-off thresholds for the laboratory tests are easy to use on admission to the hospital.

7. Conclusion

- By comparing the data for the patients who were transported to ICU with those who did not worsen throughout the hospitalization we selected a set of laboratory findings with the significant differences on admission to the clinics. The variables were used as the predictors to build up the classification model. The performance of the models was low, with the default thresholds returned by the ML estimator, we improved it by setting the cut-off level to the 25^{th} percentile for lymphocyte count and the 75^{th} - for other features.

- To distinguish the patients with the confirmed COVID-19 who may worsen while treated we justified the following threshold values of the laboratory tests done on admission: lymphocyte count lower than 2.59x10^9/L, and the upper levels for total bilirubin - 11.9 umol/L, ALT - 43 U/L, AST - 32 U/L, D-Dimer - 0.7 mg/L, APTT - 39.9 sec, CK - 247 U/L, CRP - 14.3 mg/L, LDH - 246 U/L, Troponin - 0.037 ng/mL, Ferritin - 498 ng/mL, Fibrinogen - 446 mg/dL.

- 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). It is comparable with the model trained with all the tests (AUC 0.90; CI 0.812 - 0.902; p<0.001).

References

[1] L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, M. Zhang, et al., An interpretable mortality prediction model for covid-19 patients, Nature Machine Intelligence (2020) 1–6.

[2] L. Wynants, B. Van Calster, M. M. Bonten, G. S. Collins, T. P. Debray, M. De Vos, M. C. Haller, G. Heinze, K. G. Moons, R. D. Riley, et al., Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal, bmj 369 (2020).

[3] Y. Gao, T. Li, M. Han, X. Li, D. Wu, Y. Xu, Y. Zhu, Y. Liu, X. Wang, L. Wang, Diagnostic utility of clinical laboratory data determinations for patients with the severe covid-19, Journal of medical virology (2020).

[4] B. M. Henry, G. Aggarwal, J. Wong, S. Benoit, J. Vikse, M. Plebani, G. Lippi, Lactate dehydrogenase levels predict coronavirus disease 2019 (covid-19) severity and mortality: A pooled analysis, The American Jour- nal of Emergency Medicine (2020).

[5] R. Zhou, F. Li, F. Chen, H. Liu, J. Zheng, C. Lei, X. Wu, Viral dynamics in asymptomatic patients with covid-19, International Journal of Infectious Diseases (2020).

[6] J. Gong, J. Ou, X. Qiu, Y. Jie, Y. Chen, L. Yuan, J. Cao, M. Tan, W. Xu, F. Zheng, et al., A tool to early predict severe 2019-novel coronavirus pneumonia (covid-19): a multicenter study using the risk nomogram in wuhan and guangdong, china, medRxiv (2020).

[7] H.-Y. Zheng, M. Zhang, C.-X. Yang, N. Zhang, X.-C. Wang, X.-P. Yang, X.-Q. Dong, Y.-T. Zheng, Elevated exhaustion levels and reduced functional diversity of t cells in peripheral blood may predict severe progression in covid-19 patients, Cellular & molecular immunology 17 (2020) 541–543.

[8] X. Chen, B. Zhao, Y. Qu, Y. Chen, J. Xiong, Y. Feng, D. Men, Q. Huang, Liu, B. Yang, et al., Detectable serum sars-cov-2 viral load (rnaaemia) is closely correlated with drastically elevated interleukin 6 (il-6) level in critically ill covid-19 patients, Clinical Infectious Diseases (2020).

[9] W. H. Mahallawi, O. F. Khabour, Q. Zhang, H. M. Makhdoum, B. A. Suliman, Mers-cov infection in humans is associated with a pro-inflammatory th1 and th17 cytokine profile, Cytokine 104 (2018) 8–13.

[10] C. Selinger, J. Tisoncik-Go, V. D. Menachery, S. Agnihothram, G. L. Law, J. Chang, S. M. Kelly, P. Sova, R. S. Baric, M. G. Katze, Cytokine systems approach demonstrates differences in innate and pro-inflammatory host responses between genetically distinct mers-cov isolates, BMC genomics 15

(2014) 1161.
[11] K. Kappert, A. Jahic´, R. Tauber, Assessment of serum ferritin as a biomarker in covid-19: bystander or participant? insights by comparison with other infectious and non-infectious diseases, Biomarkers (2020) 1–36.
[12] L. Spiezia, A. Boscolo, F. Poletto, L. Cerruti, I. Tiberio, E. Campello, P. Navalesi, P. Simioni, Covid-19-related severe hypercoagulability in patients admitted to intensive care unit for acute respiratory failure, Thrombosis and haemostasis 120 (2020) 998.
[13] N. Tang, D. Li, X. Wang, Z. Sun, Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia, Journal of thrombosis and haemostasis 18 (2020) 844–847.
[14] D. O. Griffin, A. Jensen, M. Khan, J. Chin, K. Chin, J. Saad, R. Parnell, C. Awwad, D. Patel, Pulmonary embolism and increased levels of d-dimer in patients with coronavirus disease, Emerging infectious diseases 26 (2020) 1941.
[15] I. Leonard-Lorant, X. Delabranche, F. Severac, J. Helms, C. Pauzet, O. Collange, F. Schneider, A. Labani, P. Bilbault, S. Moliere, et al., Acute pulmonary embolism in covid-19 patients on ct angiography and relationship to d-dimer levels, Radiology (2020).
[16] G. Lippi, E. J. Favaloro, D-dimer is associated with severity of coronavirus disease 2019: a pooled analysis, Thrombosis and haemostasis 120 (2020) 876.
[17] I. Huang, R. Pranata, M. A. Lim, A. Oehadian, B. Alisjahbana, C-reactive protein, procalcitonin, d-dimer, and ferritin in severe coronavirus disease-2019: a meta-analysis, Therapeutic advances in respiratory disease 14 (2020) 1753466620937175.
[18] Y. Yao, J. Cao, Q. Wang, Q. Shi, K. Liu, Z. Luo, X. Chen, S. Chen, K. Yu, Z. Huang, et al., D-dimer as a biomarker for disease severity and mortality in covid-19 patients: a case control study, Journal of intensive care 8 (2020) 1–11.
[19] N. Tang, H. Bai, X. Chen, J. Gong, D. Li, Z. Sun, Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy, Journal of thrombosis and haemostasis 18 (2020) 1094–1099.
[20] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, et al., Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china, Jama 323 (2020) 1061–1069.
[21] D. Bashash, H. Abolghasemi, S. Salari, M. Olfatifar, P. Eshghi, M. E. Akbari, Elevation of d-dimer, but not pt and aptt, reflects the progression of covid-19 toward an unfavorable outcome: A meta-analysis, Iranian Journal of Blood & Cancer (2020) 47–53.
[22] L. Quartuccio, A. Sonaglia, D. McGonagle, M. Fabris, M. Peghin, D. Pecori, A. De Monte, T. Bove, F. Curcio, F. Bassi, et al., Profiling covid-19 pneumonia progressing into the cytokine storm syndrome: results from a single italian centre study on tocilizumab versus standard of care, Journal of Clinical Virology (2020) 104444.
[23] W. Ling, C-reactive protein levels in the early stage of covid-19, Medecine et maladies infectieuses (2020).
[24] P. P. Bloom, E. A. Meyerowitz, Z. Reinus, M. Daidone, J. Gustafson, A. Y. Kim, E. Schaefer, R. T. Chung, Liver biochemistries in hospitalized patients with covid-19, Hepatology(2020).
[25] N. Ali, K. Hossain, Liver injury in severe covid-19 infection: current insights and challenges, Expert review of gastroenterology & hepatology (2020).
[26] N. Lee, P. K. Chan, D. S. Hui, T. H. Rainer, E. Wong, K.-W. Choi, G. C. Lui, B. C. Wong, R. Y. Wong, W.-Y. Lam, et al., Viral loads and duration of viral shedding in adult patients hospitalized with influenza, The Journal of infectious diseases 200 (2009) 492–500.
[27] National Emergency Crisis and Disasters Management Authority, National guidelines for clinical management and treatment of covid-19- version 4.1, https://www.dha.gov.ae/en/HealthRegulation/Documents/National_Guidelines_of_COVID_19_1st_June_2020.pdf, 2020. Accessed 01/08/2020.
[28] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, F. Herrera Learning from imbalanced data sets, Springer, 2018.
[29] E. V. Robilotti, N. E. Babady, P. A. Mead, T. Rolling, R. Perez-Johnston, M. Bernardes, Y. Bogler, M. Caldararo, C. J. Figueroa, M. S. Glickman, et al., Determinants of covid-19 disease severity in patients with cancer, Nature medicine 26 (2020) 1218–1223.
[30] K. Ravi, Ethnic disparities in covid-19 mortality: are comorbidities to blame?, The Lancet 396 (2020) 22.
[31] P. Baqui, I. Bica, V. Marra, A. Ercole, M. van Der Schaar, Ethnic and regional variations in hospital mortality from covid-19 in brazil: a cross-sectional observational study, The Lancet Global Health 8 (2020) e1018–e1026.
[32] A. Erez, O. Shental, J. Z. Tchebiner, M. Laufer-Perl, A. Wasserman, T. Sella, H. Guzner-Gur, Diagnostic and prognostic value of very high serum lactate dehydrogenase in admitted medical patients, Isr Med As- soc J 16 (2014) 439–443.
[33] H. Long, L. Nie, X. Xiang, H. Li, X. Zhang, X. Fu, H. Ren, W. Liu, Q. Wang, Q. Wu, D-dimer and prothrombin time are the significant indicators of severe covid-19 and poor prognosis, BioMed research international 2020 (2020).

8. Acknowledgments

The authors would like to acknowledge UAE University (Al Ain, UAE) and Mediclinic Parkview Hospital (Dubai, UAE) for the support provided and the facilities used for conducting this research. We also wish to express our gratitude to the dedication and commitment of our healthcare staff and our patients, without whom this research would not be possible.

9. Funding statement

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

10. Author contributions statement

All authors contributed to the creation of the article as follows: all of them contributed to the conceptual idea of the paper equally; FA and YS formulated the objectives; FA collected the dataset; YS wrote the manuscript; TH proposed the methodology of the study, and performed the statistical analysis, prepared the figures and tables for data presentation and illustration, FA, TH, KG, NZ contributed to the literature review and data analysis.

The data were analyzed and interpreted by the authors, who also reviewed the manuscript and vouch for the accuracy and completeness of the data and for the adherence of the study to the protocol.

11. Ethical Approval

The study got an ethical review by Dubai Scientific Research Ethics Committee (DSREC), Dubai Health Authority, protocol No DSREC-05/2020_25) and was approved for the retrospective analysis of the data obtained as a standard of care. No potentially identifiable personal information is presented in the study.

12. Data availability statement

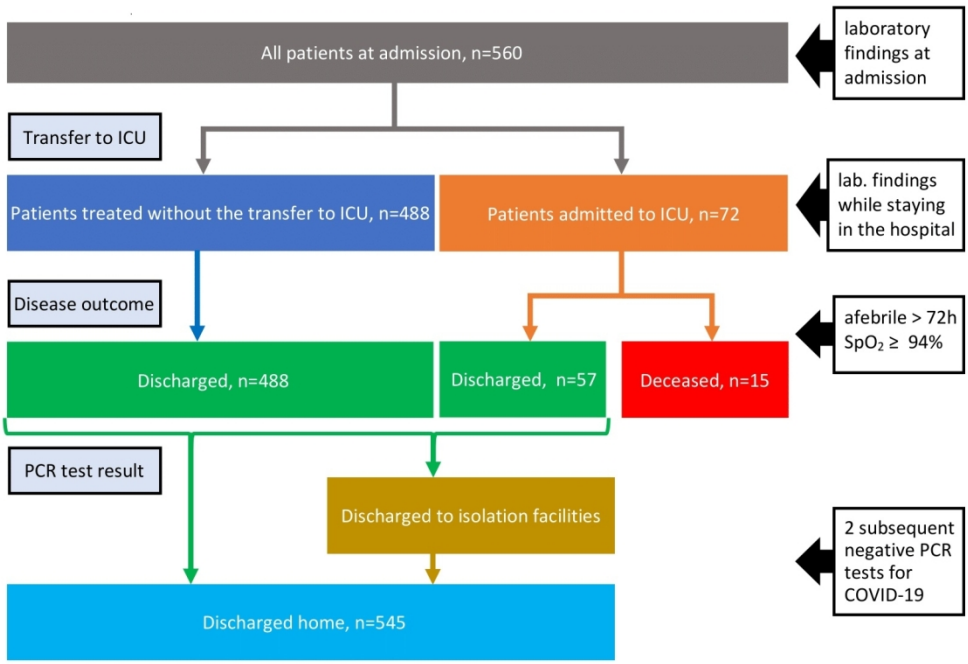
Generated Statement: The datasets generated for this study are available upon request at the site of **Data Analytics Group** at https://bi-dac.com. To assess the risk of having complications in a patient with COVID-19, one may use the ML-based free online tool at https://med-predict.com which illustrates the results of the current study.

13. Competing Interests

None declared.

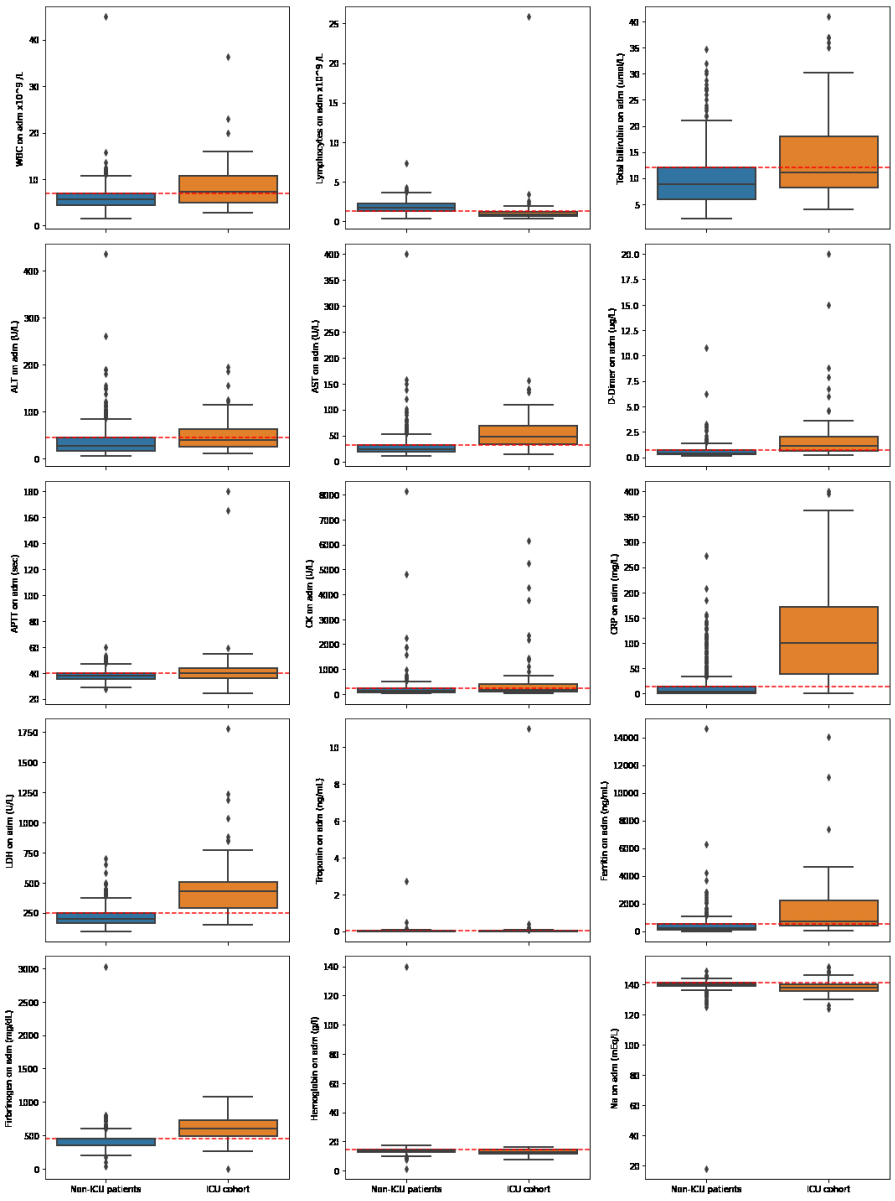
Figures

Figure 1. The flow of patients with COVID-19 in Dubai Mediclinic.
Figure 2. Variation of laboratory findings values in the ICU cohort (orange boxplot) versus the non-ICU cohort of patients (blue boxplot).
Appendix Figure 1. Feature selection for predicting whether a patient is going to be transferred to ICU.
Appendix Figure 2. The performance of the employed NN classification method.
Appendix Figure 3. ROC curves for the laboratory tests used as input to NN separately (a) and in the combination (b). The models are trained with 10 folds cross-validation.
Appendix Figure 4. The performance of the 10 folds cross-validation logistic regression model trained on binary data with the threshold moving technique returned by the ML estimator (a), and with the cut-off level set to the 25th percentile for lymphocyte count and 75th for the other features (b)



The flow of patients with COVID-19 in Dubai Mediclinic.

436x307mm (96 x 96 DPI)



Variation of laboratory findings values in the ICU cohort (orange boxplot) versus the non-ICU cohort of patients (blue boxplot).

Appendix

A. ML classification models and feature selection.

The variables used to build up the model:

- *physical examination on admission*: temperature, HR BPM, SBP, DBP, RR /min. SpO_2, SpO_2 on RA vs. O_2 Therapy, GCS, SOFA score
- *symptoms on admission*: cough, sputum, sore throat, chest pain, SOB, fever, headache, confusion, having any gastrointestinal symptoms (e.g., nausea, vomiting, diarrhea), myalgia, malaise, loss of smell or taste.
- *laboratory findings on admission*: the count of WBC, platelet, and lymphocyte; the concentration of hemoglobin, total bilirubin, D-Dimer, creatinine, sodium, C-reactive protein, troponin, ferritin, fibrinogen; the activity of ALT, AST, CK, LDH; APTT.

Feature selection:

To check if there are unique patterns within the data that can unambiguously identify if the patient is going to be transferred to the intensive care unit, we utilized ML algorithms.

To assess the importance of the features fed to the ML models as predictors of admitted to ICU patients, we employed four ensemble tree-based estimators such as AdaBoost, Gradient Boosting, Random Forest, and Extra Trees. These models were trained on the whole dataset and used to rank the features in ascending order concerning their predictive potential. Figure 1 and Table 1 display the averaged values of impurity-based attribute ranks, where the average for each feature is calculated as the mean of rank values for the four ML methods mentioned above.

Table 1: Ranking scores of the variables selected for predicting the disease severity

Score	Feature	Score	Feature	Score	Feature	Score	Feature
0.19429	SOFA score	0.02520	Temperature	0.01164	Total bilirubin	0.00466	Sore Throat
0.10168	Clinical severity	0.01748	SOB	0.01135	SBP	0.00445	Troponin
0.08745	O2 therapy	0.01712	ALT	0.00983	Fever	0.00367	Confusion
0.08061	RR/min	0.01623	APTT	0.00969	GCS	0.00309	GI symptoms
0.04127	LDH	0.01595	Hemoglobin	0.00896	Ethnicity	0.00287	Cough
0.03829	Lymphocytes	0.01545	SpO2 on RA vs O2 Therapy	0.00732	HR BPM	0.00188	Malaise
0.03223	SpO2	0.01505	Na	0.00637	Myalgia	0.00186	Chest pain
0.03212	D-Dimer	0.01383	AST	0.00633	Sputum	0.00141	Smell/taste loss
0.03125	CRP	0.01382	CK	0.00524	DBP	0.00000	Creatinine
0.03067	Platelet	0.01360	WBC	0.00513	Headache	0.00000	Ferritin
						0.00000	Fibrinogen

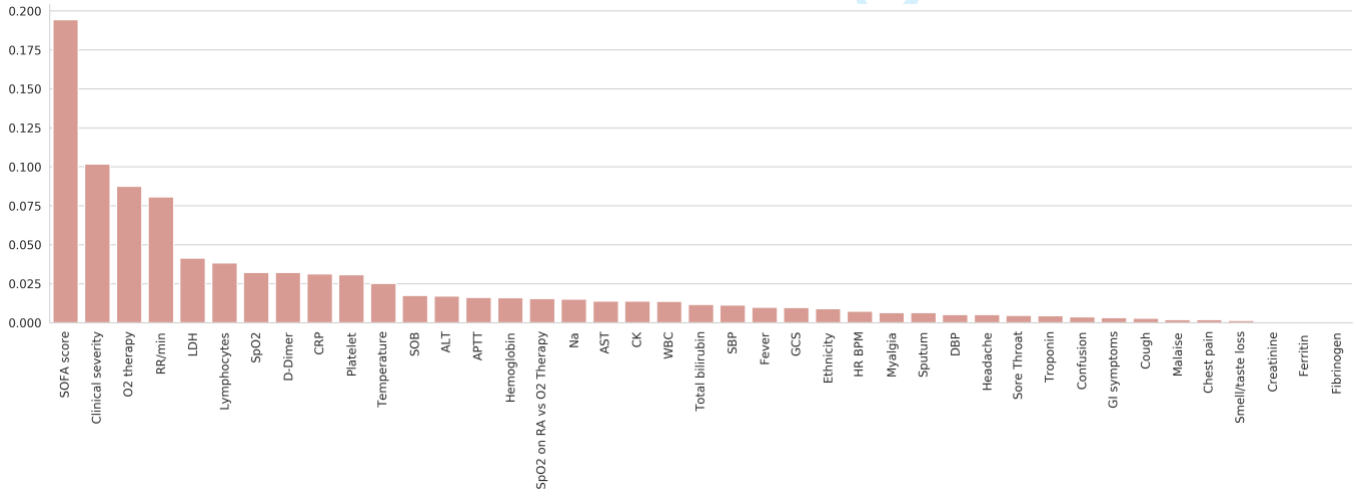


Figure 1: Feature selection for predicting whether a patient is going to be transferred to ICU

Table 2: Specificity and sensitivity of the ML model applied to the all features as predictors of the severity of the disease

ML model	Specificity	Sensitivity
Gradient Boosting	0.5972	0.9734
AdaBoost	0.6667	0.9775
Extra Trees	0.7361	0.9693
Random Forest	0.75	0.9795
NN	0.9938	1.0
SVM linear	0.6806	0.9508
Logistic regression	0.6667	0.952
Logistic regression (L1 penalty)	0.7083	0.959

Table 3: Confusion matrix to assess the accuracy of classification with a three-layer dense NN model to predict the severity of the disease

		Predicted	
		Not admitted to ICU	Admitted to ICU
Actual	Not admitted to ICU	485	3
	Admitted to ICU	0	72

Prediction of transferring to ICU. We utilized three-layer fully connected NN with the following configuration of hidden layers (35, 30, 10) and with the stochastic gradient descent optimizer. The learning rate hyperparameter of the model was assigned to 0.1. The model was also regularized using L2 penalty with 0.0001 alpha value. NN was trained for maximum 100 epochs or before converged. Convergence implies that the loss function is not improving by at least 0.0001 for 10 consecutive iterations.

To evaluate the classifier output quality, we trained several ML classification models using a stratified 10-fold cross-validation technique to generalize the models to the true rate error. For each fold, we used 90% of the data to train the model and then tested it on the rest 10%.

The decision matrices built on the test dataset for all folds were combined and used to calculate the performance metrics. The best performance measures were obtained with a three-layer fully connected NN.

Table 4: Classification metrics of the NN model to predict the event of being transferred to ICU

	Recall	Precision	F1 score	Support
Not admitted to ICU	1.00	0.99	1.00	488
Admitted to ICU	0.96	1.00	0.98	72
accuracy			0.99	560
macro average	0.98	1.00	0.99	560
weighted average	0.99	0.99	0.99	560

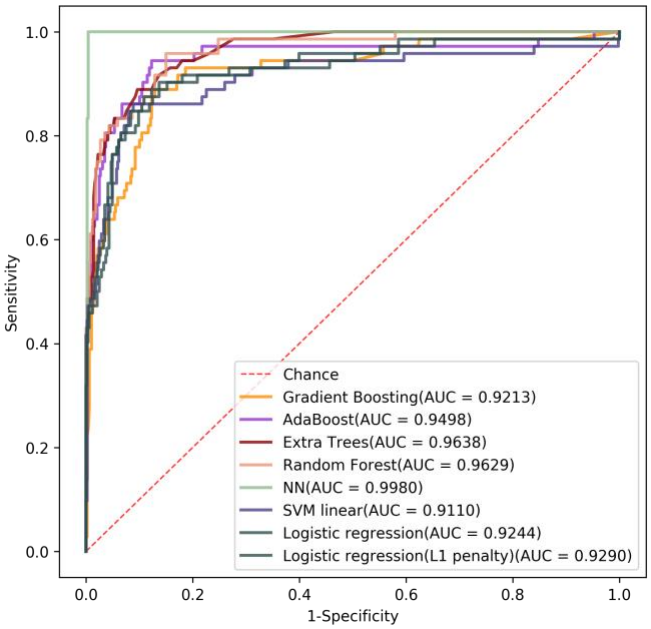


Figure 2: The performance of the employed NN classification method.

B. ROC curves for laboratory tests used as input to NN.

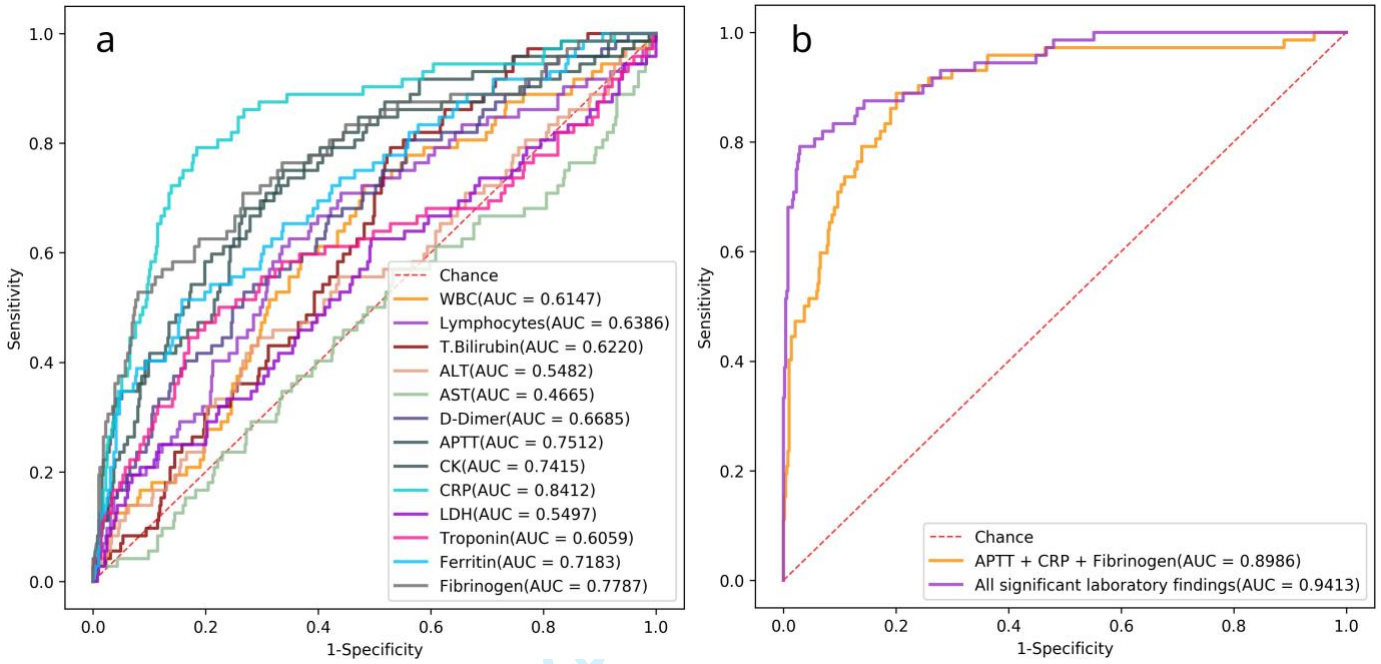


Figure 3: ROC curves for the laboratory tests used as input to NN separately (a) and in the combination (b). The models are trained with 10 folds cross-validation.

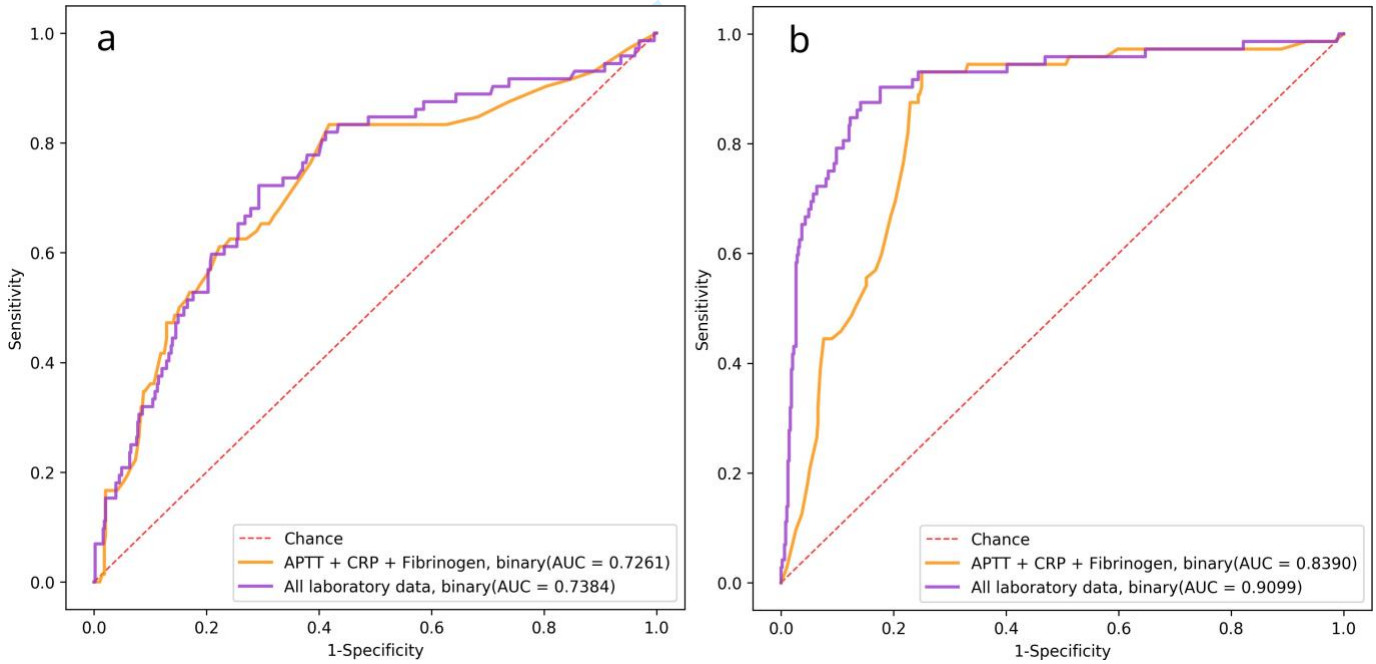


Figure 4: The performance of the 10 folds cross-validation logistic regression model trained on binary data with the threshold moving technique returned by the ML estimator (a), and with the cut-off level set to the 25^{th} percentile for lymphocyte count and 75^{th} for the other features (b).

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1 (Title)
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	1 (Abstract)
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	2/61 - 3/204 (sec. 1.1-1.2)
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	3/205-231 (sec. 2)
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	3/234-235 (sec. 3.1)
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	3/235-238 (sec. 3.1)
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	3/235 (sec. 3.1)
	5b	D;V	Describe eligibility criteria for participants.	3/241-246 (sec. 3.1)
	5c	D;V	Give details of treatments received, if relevant.	4/272-4/280 (sec. 3.1)
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	4/314-322 (sec. 3.2)
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	not applicable
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	4/295-296 (sec. 3.2) 12/758-770 (App. A)
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	not applicable
Sample size	8	D;V	Explain how the study size was arrived at.	3/235-238 (sec. 3.1)
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	4/291-293 , 4/310-313 (sec. 3.2)
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	4/282-284 (sec. 3.2)
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	4/295-301, 4/313-322 (sec. 3.2)
	10c	V	For validation, describe how the predictions were calculated.	4/323-329 (sec. 3.2)
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	4/299-301, 4/319-320 (sec. 3.2)
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	4/314-316 (sec. 3.2)
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	not applicable
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	4/323-329 (sec. 3.2)
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	3/248-4/271 (sec. 4.1) 5 (Figure 1)
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	7 (Table 1)
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	4/336-354 (sec. 4.1) 6 (Figure 2) 7 (Table 1)
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	4/332-335 (sec. 4.1)
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	-
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	13/785-793 (App. A)
	15b	D	Explain how to the use the prediction model.	5/381-392 (sec. 4.3)
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	5/370-380 (sec. 4.3); 8 (Table 2)
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	5/393-406 (sec. 4.3) 9 (Table 3)
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	9/553-10/571 (sec. 6)
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	9/504-9/534 (sec. 5.3)
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	8/437-461 (sec. 5.1)
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	9/537-541 (sec. 5.3)
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	11/753-755 (sec. 12)
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	11/729-731 (sec. 9)

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.