Using Machine Learning to Predict Mortality for COVID-19 Patients on Day Zero in the ICU

Elham Jamshidi^{1*}, Amirhossein Asgary^{2*}, Nader Tavakoli^{3*}, Alireza Zali⁴, Hadi Esmaily⁵, Seyed Hamid Jamaldini⁶, Amir Daaee⁷, Amirhesam Babajani⁸, Mohammad Ali Sendani Kashi⁹, Masoud Jamshidi¹⁰, Sahand Jamal Rahi¹¹⁺, Nahal Mansouri^{1,12+}

¹ Division of Pulmonary Medicine, Department of Medicine, Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Lausanne, Switzerland.

² Department of Biotechnology, College of Sciences, University of Tehran, Tehran, Iran.

³ Trauma and Injury Research Center, Iran University of Medical Sciences, Tehran, Iran.

⁴ Functional Neurosurgery Research Center, Shohada Tajrish Comprehensive Neurosurgical Center of Excellence, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

⁵ Department of Clinical Pharmacy, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

⁶ Department of Genetic, Faculty of Advanced Science and Technology, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran.

⁷ School of Mechanical Engineering, Sharif University of Technology, Tehran, Iran.

⁸ Department of Pharmacology, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

⁹ Master of Business Administration (MBA)-University of Tehran, Tehran, Iran.

¹⁰ Department of Exercise Physiology, Tehran University, Iran.

¹¹ Laboratory of the Physics of Biological Systems, Institute of Physics, École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland.

¹² Swiss Institute for Experimental Cancer Research (ISREC), School of Life Sciences, École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland.

*: These authors contributed equally.

+: Contact:

Nahal Mansouri (nahal.mansouri@chuv.ch)

Sahand Jamal Rahi (sahand.rahi@epfl.ch)

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Abbreviations list:

ACE2 (Angiotensin-Converting Enzyme 2), AI (Artificial Intelligence), BUN (Blood Urea Nitrogen), COVID-19 (coronavirus disease of 2019), CIC (clinical impact curve), Cr (Creatinine), CRP (C reactive protein), DC (decision curve), ICU (Intensive care unit), INR (International Normalized Ratio), IFN (interferon), IL-6 (Interleukin 6), IQR (interquartile range), KS (Kolmogorov- Smirnov), LR (Logistics regression), LIME (local interpretable model-agnostic explanation), LIME-SP (local interpretable model-agnostic explanation submodular-pick), ML (Machine learning), MCH (mean corpuscular hemoglobin), MCV (mean corpuscular volume), RF (Random forest), RDW (Red blood cell distribution width), ROC (receiver operating characteristic curve), RT-PCR (reverse transcription-polymerase chain reaction), WBC (white blood cells count)

Abstract

Rationale

Given the expanding number of COVID-19 cases and the potential for upcoming waves of infection, there is an urgent need for early prediction of the severity of the disease in intensive care unit (ICU) patients to optimize treatment strategies.

Objectives

Early prediction of mortality using machine learning based on typical laboratory results and clinical data registered on the day of ICU admission.

Methods

We studied retrospectively 263 COVID-19 ICU patients. To find parameters with the highest predictive values, Kolmogorov-Smirnov and Pearson chi-squared tests were used. Logistic regression and random forest (RF) algorithms were utilized to build classification models. The impact of each marker on the RF model predictions was studied by implementing the local interpretable model-agnostic explanation technique (LIME-SP).

Results

Among 66 documented parameters, 15 factors with the highest predictive values were identified as follows: gender, age, blood urea nitrogen (BUN), creatinine, international normalized ratio (INR), albumin, mean corpuscular volume, white blood cell count, segmented neutrophil count, lymphocyte count, red cell distribution width (RDW), and mean cell hemoglobin along with a

history of neurological, cardiovascular, and respiratory disorders. Our RF model can predict patients outcomes with a sensitivity of 70% and a specificity of 75%.

Conclusions

The most decisive variables in our model were increased levels of BUN, lowered albumin levels, increased creatinine, INR, and RDW along with gender and age. Complete blood count parameters were also crucial for some patients. Considering the importance of early triage decisions, this model can be a useful tool in COVID-19 ICU decision-making.

Keywords

SARS-CoV-2; artificial intelligence; ICU.

Introduction

COVID-19 has currently affected more than 82 million people worldwide and caused more than 1.8 million deaths ¹. Complications are more common among elderly patients and people with preexisting conditions, and the rate of intensive care unit (ICU) admission is substantially higher in these groups ^{2,3}.

ICU admissions rely on the critical care capacity of the health care system. Iran, which is the primary testbed for this study, was one of the first countries hit by COVID-19. The ICU admission rate is around 32% of all hospitalizations, and the ICU mortality rate is about 39% ⁴. With the potential upcoming waves of COVID-19 infections, these numbers are expected to rise, leading to shortages of ICU beds and critical management equipment. There is also the risk of a global shortage of effective medical supplies, making the judicious use of these medications a top priority for healthcare systems.

An individual-based prediction model is essential for tailoring treatment strategies and would aid in expanding our insights into the pathogenesis of COVID-19. A number of risk assessment scores are available to predict the severity of different diseases in ICU patients ⁵. Predictors of the need for intensive respiratory or vasopressor support in patients with COVID-19 and of mortality in COVID-19 patients with pneumonia have been put forth ^{6,7}. To date, no general mortality prediction scores have been available for COVID-19 patients presenting to the ICU, irrespective of the patients' clinical presentation. Additionally, both the existing risk scales rely on parameters measured by health care providers such as blood pressure, respiratory rate, and oxygen saturation, which are subject to human error and operator bias especially under the

challenging and stressful conditions when the numbers of COVID-19 patients surge⁸. Thus, it remains vital to develop more unbiased risk assessment tools that can predict the most likely outcomes for individual patients with COVID-19.

Recent advances in artificial intelligence (AI) technology for disease screening show promise as computer-aided diagnosis and prediction tools⁹⁻¹². In the era of COVID-19, AI has played an important role in the early diagnosis of infection, contact tracing, and drug and vaccine development¹³. Thus, AI represents a useful technology for the management of COVID-19 patients with the potential to reduce the mortality rate of this disease. Nevertheless, an AI tool for making standardized and accurate predictions of outcome in COVID-19 patients with severe disease is missing.

Beyond the general benefits of data-driven decision-making, the pandemic has also exposed the need for computational assistance to health care providers, who under the rush of severely ill patients may make mistakes in judgement^{8,14,15}. Stressful conditions and burnout in health care providers can reduce their clinical performance, and a lack of accurate judgment can lead to increased mortality rates^{16,17}. Artificial intelligence can help healthcare professionals determine who needs a critical level of care more precisely. Indeed, the effective use of AI could mitigate the severity of this outbreak.

Here, we propose a personalized machine-learning (ML) method for predicting mortality in COVID-19 patients based on routinely available laboratory and clinical data at the date of ICU admission.

Methods

Data Resources

This retrospective study includes data from 263 adult patients with COVID-19 infection confirmed through reverse transcription-polymerase chain reaction (RT-PCR). They were admitted to ICUs at different hospitals in Tehran, Iran between February 19 and May 1, 2020. The study was performed after approval by Iran University of Medical Sciences Ethics Committee (approval ID: IR.IUMS.REC.1399.595)

Statistical Analysis & Feature Selection

On the day of the ICU admission, 66 parameters were assessed for each patient including 11 demographic characteristics (e.g., age and gender), past medical history and comorbidities (including nine different preexisting conditions), and 55 laboratory biomarkers. These parameters are listed in Table 1. 69% of measurements were reported on the day of admission, 27% were reported one day after, and 4% were reported within two days of ICU admission because of sampling limitations and laboratory practice. We excluded patients whose laboratory data were obtained more than two days after the date of admission to the ICU.

The aim was to predict patient survival. For the selection of parameters with the highest predictive value, under the null hypothesis of distributions being the same between the two groups, the two-sample Kolmogorov-Smirnov test (KS), shown in Supplementary Figure 1, was used for numerical parameters (age and laboratory biomarkers), and the Pearson chi-squared test (χ^2), shown in Supplementary Figure 2, was used for categorical parameters (e.g., gender and

comorbidities). To reduce the model's feature space dimensionality, only parameters with the highest predictive values were used in the modeling, leading to more robustness and generalizability of the model.¹⁸

Data Preprocessing

Data processing was carried out in four steps: First, because of incomplete laboratory data and in order to reduce difficulties associated with missing values, 235 patients having data for at least seven out of ten biomarkers were selected. Second, samples were randomly separated into 10 independent sets with stratification over outcomes for 10-fold cross-validation to ensure the generalizability of the models¹⁹. Of the 10 subsets, a single subset was retained as a validation set for model testing and the remaining nine subsets were used as training data. The cross-validation process was then iterated ten times with each of the 10 subsets being used as the validation data exactly once. Third, numerical parameters were standardized by scaling the features to mean zero and unit variance. Lastly, missing biomarker values were imputed using the k-nearest neighbor (k-NN) algorithm, and a binary indicator of missingness for each biomarker was added to the dataset^{18,20}. Standardization and imputation were performed separately on each cross-validation iteration by using training set samples.

Machine Learning Model

Logistic regression (LR) and random forest (RF) methods were used to build classification models using the Python scikit-learn package²¹. The performance of each method on training and validation sets in each cross-validation iteration was compared using a receiver operating characteristic curve (ROC), which is shown in Figure 1. To prevent overfitting in the training

process, the LR model was trained with an L2 regularization factor equal to one, and the RF was forced to hold more than 10% of samples in each of its terminal leaves^{22,23}. To find the most influential parameters in the LR model prediction, we used regression coefficients, which is shown in the Supplementary Material. Using the local interpretable model-agnostic explanation submodular-pick (LIME-SP) method, we identified different patterns among the whole feature space in the RF model²⁴. The LIME-SP method can interpret the model's predictions in different parts of the feature space by modeling a subset of model predictions in the feature space around the sample with the help of linear models that are more interpretable. In our study, LIME-SP was performed on 100 random samples to find six submodules with the most disparity in their selected markers, as shown in Figure 2. To identify meaningful clinical differences between patients, seven parameters with the highest predictive values were derived from each submodule.

[Figure 1]

[Figure 2]

Results

The median age of patients was 69 years with an interquartile range (IQR) of 54-78. The minimum and maximum ages were 20 and 98 years, respectively. One hundred fifty-three patients (65.1%) were men, and 82 (34.9%) were women. One hundred five (39.9%) were discharged from the ICU after recovery and 158 (60.1%) patients lost their lives. The most frequent comorbidities among the patients were hypertension, diabetes, and cardiovascular disorders in 94, 92, and 86 patients, respectively. Among the 158 deceased patients, neurological

disorders were the most prevalent comorbidity (42 patients, 84%). The statistical analysis and the availability of each parameter in our dataset are summarized in Table 1.

[Table 1]

In the RF model, the optimum point between overfitting and efficiency was found by selecting 10 laboratory biomarkers out of 55 with the lowest KS p -values and three out of nine comorbidities with the lowest χ^2 p -values, besides demographic characteristics.

The selected numerical parameters for modeling were as follows: age, blood urea nitrogen (BUN), serum creatinine level (Cr), international normalized ratio (INR), serum albumin, mean corpuscular volume (MCV), red cell distribution width (RDW), mean corpuscular hemoglobin (MCH), white blood cell count (WBC), segmented neutrophil count, and lymphocyte count. In addition, selected categorical parameters were gender and a history of neurological, respiratory, and cardiovascular diseases. The distributions of selected numerical (age and biomarkers) and categorical (gender and preexisting conditions) variables are shown in Supplementary Figure 4 and Figure 5, respectively.

Based on the ROC curves of the models (Figure 1), the RF model outperformed the LR model in both training and validation sets and had superior efficiency. The better performance of RF could be explained by the complexity of the effects of COVID-19 and the varied etiologies for the deterioration of COVID-19 patients due to which the nonlinear characteristics of the RF model made it a more suitable option for predictions than the linear LR model. The RF model can predict a patient's outcome with a sensitivity of 70% and a specificity of 75%, while the sensitivity for the LR model was 65% and the specificity was 70%.

By using the LIME technique, variables that provide the most information on the probability of each patient's death were identified.

Among the six submodules identified with the highest disparity among 100 patients, albumin, BUN, and RDW were present in five of them. Age, MCH, and creatinine were present in four of the abovementioned submodules. This points out the importance of these measurements in the recorded parameters. Additionally, BUN (in three of these submodules), RDW (in two submodules), and age (in one submodule), were the most decisive ones.

This model could predict a patient's outcome reliably (AUC between 80 to 85) over a 15-day period, as shown in Figure 3. The mortality rate was highest between zero and four days. Given that the model was designed for first-day ICU admissions, moving away from this day reduced the accuracy of the predictions and the efficacy of the LIME method for clinical interventions, as expected.

[Figure 3]

To evaluate the clinical capability of the model, the decision curve (DC), and the clinical impact curve (CIC) were investigated ²⁵. The DC framework measures the clinical “net benefit” for the prediction model relative to the current treatment strategy for all or no patients. Net benefit is measured over a spectrum of threshold probabilities, defined as the minimum disease risk at which further intervention is required. Based on the DC, CIC, and on the assumption of the same interventions for high-risk patients, our model indicated a superior or equal net benefit within a wide range of risk thresholds and patient outcomes, as shown in Figure 4.

[Figure 4]

Discussion

The aim of this study was developing an interpretable ML model to predict the mortality rate of COVID-19 patients at the time of admission to the ICU. To the best of our knowledge, this is the first study to develop a predictive model of mortality in patients with severe COVID-19 infection at such an early stage using routine laboratory results and demographic characteristics.

The most decisive parameters based on the two-sample KS test were, in decreasing order of importance, increased BUN, Cr, INR, MCV, WBC, segmented neutrophils count, RDW, MCH, and decreased albumin and lymphocyte levels. Moreover, based on a χ^2 test, age, gender, and a history of neurological, cardiovascular, and respiratory disorders were identified as parameters with high predictive values.

A number of studies have investigated the risk factors affecting COVID-19 infections. Elevated inflammatory cytokines such as interleukin-6 (IL-6), granulocyte colony-stimulating factor (G-CSF), interferon gamma-induced protein 10 (IP-10), and interferon (IFN)- γ have been suggested as poor prognostic factors for COVID-19 patients^{26–28}. These markers, however, are not usually used as predictors of the severity of disease in clinical practice. Although using these cytokines in modeling may enable a more accurate prediction of the severity of COVID-19 infection, doing so impedes the model's clinical application, as most of the cytokines are not routinely checked at presentation to the ICU. In contrast, all 10 laboratory biomarkers identified in our model are commonly measured and are available to most clinical laboratories. Thus, the DC and CIC analyses indicated the notable clinical benefit of our model especially in a situation characterized by resource scarcity.

Only patients who had at least seven of the 10 selected biomarkers have been included in the training phase of the modeling and missing parameters were imputed using k-NN based on the data. As can be seen in the models' ROC curve, the RF algorithm outperformed LR in predicting the outcome. This difference is mainly due to the non-linear correlation between variables, manifesting the complexity of the problem.

The application of the LIME-SM method allowed us to determine a patient-specific marker set that each patient's prognosis is based on. This technique explains the predictions by perturbing the input of data samples and evaluating the effects. The output of LIME is a list of features, reflecting each feature's contribution to a given prediction. Understanding the 'reasoning' of the ML model is crucial for increasing physicians' confidence in selecting treatments based on the prognosis scores. Using the LIME method, the significance of variables with high predictive value was determined for each prediction made for an individual. The evaluation of the variables in the individual's personalized prediction can lead to supportive measures and help determine treatment strategies according to the interpretation of the individual prognosis.

As severe COVID-19 may result from various underlying etiologies, our model can help categorize patients into groups with distinct clinical prognosis, thus allowing personalized treatments. In addition to targeted therapies, the differentiation between patients may reveal disease mechanisms that coincide or that occur under specific preexisting conditions. Future cohort studies could explore these assumptions with increased sample sizes.

In this study, hypoalbuminemia and renal function were identified as the main factors with high predictive values for the model. These findings are in agreement with recent results showing that hypoalbuminemia is an indicator of poor prognosis for COVID-19 patients²⁹. It is well

documented that endogenous albumin is the primary extracellular molecule responsible for regulating the plasma redox state among plasma antioxidants³⁰. Moreover, it has been shown that albumin downregulates the expression of the angiotensin-converting enzyme 2 (ACE2) which may explain the association of hypoalbuminemia with severe COVID-19³¹. Intravenous albumin therapy has been shown to improve multiple organ functions³². Therefore, early treatment with human albumin in severe cases of COVID-19 patients before the drop in albumin levels might have positive outcomes and needs to be further investigated.

Furthermore, increased levels of BUN and Cr are observed in our study, which is an indication of kidney damage. An abrupt loss of kidney function in COVID-19 is strongly associated with increased mortality and morbidity³³. There are multiple mechanisms supporting this association^{34,35}.

One of the findings of this study is the identification of RDW (a measure of the variability of the sizes of RBCs) as an influential parameter. This result is in line with recently published reports³⁶. Elevated RDW, known as anisocytosis, reflects a higher heterogeneity in erythrocyte sizes caused by erythrocyte maturation and degradation abnormalities. Several studies have found that elevated RDW is associated with inflammatory markers in the blood such as IL-6, tumor necrosis factor- α , and CRP, which is common in severely ill Covid-19 patients³⁷. These inflammatory markers could disrupt the erythropoiesis by directly suppressing erythroid precursors, promoting apoptosis of precursor cells, and reducing the bioavailability of iron for hemoglobin synthesis.

Yan et al. recently identified LDH, lymphocyte, and high-sensitivity C-reactive protein (hs-CRP) as predictors of mortality in COVID-19 patients during their hospitalization. The blood results of

hospitalized patients on different days after the initial ICU admission were used for their model³⁸. Since our goal was the prediction of mortality risk as early as possible for ICU patients, this limited us to using only the laboratory results on day zero, in contrast. For patients with severe COVID-19 infection, early decision making is critical for successful clinical management. Additionally, laboratory results from other days may not always become available. We also identified lymphocyte count as a predictor of mortality, as in the previous study; however, CRP levels and LDH did not reach statistical significance.

Although IL-6 has been found to be a good predictor of disease severity by other studies, it did not reach statistical significance in our model³⁹. IL-6 had a considerable KS statistical value, but because of the high number of missing values, its *p*-value was not significant compared to other markers. The fact that IL-6 is not always measured upon ICU admission is precisely why it is not suitable for our purposes.

The missingness indicator of some markers in both LR and RF models has an impact on the predictions based on the regression coefficient and LIME, which can be the result of the model compensating for the imputation error. However, the missingness indicator may also indicate the existence of bias in biomarker reporting⁴⁰. Such biases (e.g., sampling bias) are an inevitable part of retrospective studies. They can be addressed using domain-adaptation techniques such as correlation alignment (CORAL) in future studies using additional data^{41,42}. Another limitation of this study may be the lack of an objective criterion for ICU admission. Moreover, different treatment strategies can change the survival outcome for patients who may have had similar profiles when admitted to the ICU. In future studies, the accuracy of this model may be further improved by adding chest imaging data and by using a larger dataset. Possible targets for our ML

framework include the prediction of other crucial information such as the patients' need for mechanical ventilation, the occurrence of cytokine release syndrome, the severity of acute respiratory disease syndrome, the cause of death, and the right treatment strategy.

In conclusion, we evaluated 66 parameters in COVID-19 patients at the time of ICU admission. Of those parameters, 15 metrics with the highest prediction values were identified: gender, age, BUN, Cr, INR, albumin, MCV, RDW, MCH, WBC, segmented neutrophil count, lymphocyte count, and past medical history of neurological, respiratory, and cardiovascular disorders. In addition, by using the LIME-SP method, we identified different submodules clarifying distinct clinical manifestations of severe COVID-19. The ML model trained in this study could help clinicians determine rapidly which patients are likely to have worse outcomes, and given the limited resources and reliance on supportive care allow physicians to make more informed decisions.

Declaration of competing interest

The authors reported no potential conflict of interest.

Data availability statement

The data that support the findings of this study are available from the corresponding authors upon request.

References

1. WHO Coronavirus Disease (COVID-19) Dashboard [Internet]. [cited 2020 Nov 16];Available from: <https://covid19.who.int/>
2. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA [Internet] 2020;Available from: <http://dx.doi.org/10.1001/jama.2020.6775>
3. Liu K, Chen Y, Lin R, Han K. Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients [Internet]. Journal of Infection. 2020;80(6):e14–8. Available from: <http://dx.doi.org/10.1016/j.jinf.2020.03.005>
4. Abate SM, Ahmed Ali S, Mantfardo B, Basu B. Rate of Intensive Care Unit admission and outcomes among patients with coronavirus: A systematic review and Meta-analysis. PLoS One [Internet] 2020;15(7):e0235653. Available from: <http://dx.doi.org/10.1371/journal.pone.0235653>
5. Rapsang AG, Shyam DC. Scoring systems in the intensive care unit: A compendium. Indian J Crit Care Med [Internet] 2014;18(4):220–8. Available from: <http://dx.doi.org/10.4103/0972-5229.130573>
6. Fan G, Tu C, Zhou F, et al. Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study. Eur Respir J [Internet] 2020 [cited 2020 Aug 14];Available from: <https://erj.ersjournals.com/content/early/2020/07/06/13993003.02113-2020.abstract>
7. Su Y, Tu G-W, Ju M-J, et al. Comparison of CRB-65 and quick sepsis-related organ failure assessment for predicting the need for intensive respiratory or vasopressor support in patients with COVID-19. J Infect [Internet] 2020;Available from: <http://dx.doi.org/10.1016/j.jinf.2020.05.007>
8. Hartzband P, Groopman J. Physician Burnout, Interrupted. N Engl J Med [Internet] 2020;382(26):2485–7. Available from: <http://dx.doi.org/10.1056/NEJMp2003149>
9. Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. JAMA [Internet] 2018;320(21):2199–200. Available from: <http://dx.doi.org/10.1001/jama.2018.17163>
10. Stead WW. Clinical Implications and Challenges of Artificial Intelligence and Deep Learning [Internet]. JAMA. 2018;320(11):1107–8. Available from: <http://dx.doi.org/10.1001/jama.2018.11029>
11. Ghafouri-Fard S, Taheri M, Omrani MD, Daaee A, Mohammad-Rahimi H, Kazazi H. Application of Single-Nucleotide Polymorphisms in the Diagnosis of Autism Spectrum

- Disorders: A Preliminary Study with Artificial Neural Networks [Internet]. *Journal of Molecular Neuroscience*. 2019;68(4):515–21. Available from: <http://dx.doi.org/10.1007/s12031-019-01311-1>
12. Ghafouri-Fard S, Taheri M, Omrani MD, Daaee A, Mohammad-Rahimi H. Application of Artificial Neural Network for Prediction of Risk of Multiple Sclerosis Based on Single Nucleotide Polymorphism Genotypes. *J Mol Neurosci* [Internet] 2020;70(7):1081–7. Available from: <http://dx.doi.org/10.1007/s12031-020-01514-x>
13. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab Syndr* [Internet] 2020 [cited 2020 Aug 23];14(4):337. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7195043/>
14. Lai J, Ma S, Wang Y, et al. Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. *JAMA Netw Open* [Internet] 2020 [cited 2020 Aug 14];3(3):e203976–e203976. Available from: https://jamanetwork.com/journals/jamanetworkopen/articlepdf/2763229/lai_2020_oi_200192.pdf
15. Kannampallil TG, Goss CW, Evanoff BA, Strickland JR, McAlister RP, Duncan J. Exposure to COVID-19 patients increases physician trainee stress and burnout. *PLoS One* [Internet] 2020 [cited 2020 Aug 14];15(8):e0237301. Available from: <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0237301&type=printable>
16. Motluk A. Do doctors experiencing burnout make more errors? *CMAJ* [Internet] 2018;190(40):E1216–7. Available from: <http://dx.doi.org/10.1503/cmaj.109-5663>
17. Dyrbye LN, Awad KM, Fiscus LC, Sinsky CA, Shanafelt TD. Estimating the Attributable Cost of Physician Burnout in the United States [Internet]. *Ann. Intern. Med.* 2019;171(8):600–1. Available from: <http://dx.doi.org/10.7326/L19-0522>
18. Mladenović D. Feature Selection for Dimensionality Reduction [Internet]. *Subspace, Latent Structure and Feature Selection*. 2006;84–102. Available from: http://dx.doi.org/10.1007/11752790_5
19. Wong T-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation [Internet]. *Pattern Recognition*. 2015;48(9):2839–46. Available from: <http://dx.doi.org/10.1016/j.patcog.2015.03.009>
20. Murti DMP, Pujianto U, Wibawa AP, Akbar MI. K-Nearest Neighbor (K-NN) based Missing Data Imputation [Internet]. 2019 5th International Conference on Science in Information Technology (ICSITech). 2019;Available from: <http://dx.doi.org/10.1109/icsitech46713.2019.8987530>
21. scikit-learn [Internet]. [cited 2020 Aug 23];Available from: <https://scikit-learn.org>

22. Pedregosa F, Cauvet E, Varoquaux G, Pallier C, Thirion B, Gramfort A. Learning to Rank from Medical Imaging Data [Internet]. *Machine Learning in Medical Imaging*. 2012;234–41. Available from: http://dx.doi.org/10.1007/978-3-642-35428-1_29
23. Ng AY. Feature selection, L1 vs. L2 regularization, and rotational invariance [Internet]. *Twenty-first international conference on Machine learning - ICML '04*. 2004; Available from: <http://dx.doi.org/10.1145/1015330.1015435>
24. Ribeiro MT, Singh S, Guestrin C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier [Internet]. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM; 2016. p. 1135–44. Available from: <https://dl.acm.org/doi/10.1145/2939672.2939778>
25. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* [Internet] 2006;26(6):565–74. Available from: <http://dx.doi.org/10.1177/0272989X06295361>
26. Mehta P, McAuley DF, Brown M, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* [Internet] 2020;395(10229):1033–4. Available from: [http://dx.doi.org/10.1016/S0140-6736\(20\)30628-0](http://dx.doi.org/10.1016/S0140-6736(20)30628-0)
27. Conti P, Ronconi G, Caraffa A, et al. Induction of pro-inflammatory cytokines (IL-1 and IL-6) and lung inflammation by Coronavirus-19 (COVI-19 or SARS-CoV-2): anti-inflammatory strategies. *J Biol Regul Homeost Agents* [Internet] 2020;34(2). Available from: <http://dx.doi.org/10.23812/CONTI-E>
28. Jamshidi E, Babajani A, Soltani P, Niknejad H. Proposed Mechanisms of Targeting COVID-19 by Delivering Mesenchymal Stem Cells and Their Exosomes to Damaged Organs. *Stem Cell Reviews and Reports* [Internet] 2021; Available from: <https://doi.org/10.1007/s12015-020-10109-3>
29. Rica R de la, de la Rica R, Borges M, et al. Low albumin levels are associated with poorer outcomes in a case series of COVID-19 patients in Spain: a retrospective cohort study [Internet]. Available from: <http://dx.doi.org/10.1101/2020.05.07.20094987>
30. Taverna M, Marie A-L, Mira J-P, Guidet B. Specific antioxidant properties of human serum albumin. *Ann Intensive Care* [Internet] 2013;3(1):4. Available from: <http://dx.doi.org/10.1186/2110-5820-3-4>
31. Liu B-C, Gao J, Li Q, Xu L-M. Albumin caused the increasing production of angiotensin II due to the dysregulation of ACE/ACE2 expression in HK2 cells. *Clin Chim Acta* [Internet] 2009;403(1–2):23–30. Available from: <http://dx.doi.org/10.1016/j.cca.2008.12.015>
32. Caironi P, Gattinoni L. The clinical use of albumin: the point of view of a specialist in intensive care. *Blood Transfus* [Internet] 2009;7(4):259–67. Available from: <http://dx.doi.org/10.2450/2009.0002-09>

33. Cheng Y, Luo R, Wang K, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* [Internet] 2020;97(5):829–38. Available from: <http://dx.doi.org/10.1016/j.kint.2020.03.005>
34. Fu Q, Colgan SP, Shelley CS. Hypoxia: The Force that Drives Chronic Kidney Disease. *Clin Med Res* [Internet] 2016;14(1):15–39. Available from: <http://dx.doi.org/10.3121/cmr.2015.1282>
35. Batlle D, Soler MJ, Sparks MA, et al. Acute Kidney Injury in COVID-19: Emerging Evidence of a Distinct Pathophysiology. *J Am Soc Nephrol* [Internet] 2020; Available from: <http://dx.doi.org/10.1681/ASN.2020040419>
36. Foy BH, Carlson JCT, Reinertsen E, et al. Elevated RDW is Associated with Increased Mortality Risk in COVID-19 [Internet]. Available from: <http://dx.doi.org/10.1101/2020.05.05.20091702>
37. Agarwal S. Red cell distribution width, inflammatory markers and cardiorespiratory fitness: results from the National Health and Nutrition Examination Survey. *Indian Heart J* [Internet] 2012;64(4):380–7. Available from: <http://dx.doi.org/10.1016/j.ihj.2012.06.006>
38. Yan L, Zhang H-T, Goncalves J, et al. An interpretable mortality prediction model for COVID-19 patients [Internet]. *Nature Machine Intelligence*. 2020;2(5):283–8. Available from: <http://dx.doi.org/10.1038/s42256-020-0180-7>
39. Elevated levels of IL-6 and CRP predict the need for mechanical ventilation in COVID-19. *J Allergy Clin Immunol* [Internet] 2020 [cited 2020 Aug 23];146(1):128–136.e4. Available from: <http://dx.doi.org/10.1016/j.jaci.2020.05.008>
40. Horton NJ, Lipsitz SR, Parzen M. A Potential for Bias When Rounding in Multiple Imputation [Internet]. *The American Statistician*. 2003;57(4):229–32. Available from: <http://dx.doi.org/10.1198/0003130032314>
41. Sun B, Feng J, Saenko K. Correlation Alignment for Unsupervised Domain Adaptation [Internet]. *Domain Adaptation in Computer Vision Applications*. 2017;153–71. Available from: http://dx.doi.org/10.1007/978-3-319-58347-1_8
42. Iii HD, Daume H III, Marcu D. Domain Adaptation for Statistical Classifiers [Internet]. *Journal of Artificial Intelligence Research*. 2006;26:101–26. Available from: <http://dx.doi.org/10.1613/jair.1872>

Table 1: Characteristics of intensive care unit patients with COVID-19 in our data.

	Survived (N=105)		Died (N=158)		Total (number)	χ^2 Statistics	χ^2 p-value
	Number (%)	Available data (%)	Number (%)	Available data (%)			
Gender		105 (100)		158 (100)		1.70	0.19
Male	63 (36.4)	..	110 (64.6)	..	173	..	..
Female	42 (46.7)	..	48 (53.3)	..	90	..	..
Comorbidity	..	105 (100)	..	158 (100)	..	..	..
Autoimmune Disorder	2 (33.3)	..	4 (66.7)	..	6	0.10	0.74
Cancer	6 (42.9)	..	8 (57.1)	..	14	0.05	0.82
Cardiovascular Disorder	25 (29.1)	..	61 (70.9)	..	86	4.22	0.04
Diabetes Mellitus	35 (38.0)	..	57 (62.0)	..	92	0.13	0.71
Thrombosis	2 (40.0)	..	3 (60.0)	..	5	0.0003	0.99
Hypertension	32 (34.0)	..	62 (66.0)	..	94	1.35	0.24
Hepatic Failure	2 (40.0)	..	3 (60.0)	..	5	0.0007	0.99
Neurological Disorder	8 (16.0)	..	42 (84.0)	..	50	11.93	<0.001
Respiratory Disorder	7 (24.1)	..	22 (75.9)	..	29	3.01	0.08
..	Median (IQR)	Available data (%)	Median (IQR)	Available data (%)	Normal range	KS Statistics	KS p-value
Age (years)	58.0 (47.0 - 73.0)	105 (100)	72.5 (64.0 - 80.75)	158 (100)	..	0.35	<0.001
pH	7.42 (7.375 - 7.457)	87 (82)	7.4 (7.33 - 7.441)	129 (81)	7.31-7.41	0.18	0.05
pCO₂ (mm Hg)	38.4 (34.8 - 45.1)	87 (82)	40.2 (33.9 - 47.1)	125 (79)	35-40	0.09	0.66
pO₂ (mm Hg)	37.05 (25.1 - 57.425)	86 (81)	39.9 (26.975 - 56.65)	124 (78)	42-51	0.08	0.81
HCO₃ (meq·L)	25.5 (22.825 - 28.575)	86 (81)	24.2 (21.2 - 27.55)	123 (77)	22-26	0.15	0.14
O₂ Saturation (%)	72.7 (48.3 - 89.2)	85 (80)	73.5 (50.2 - 88.95)	123 (77)	-2.0-2.0	0.08	0.87
Base Excess (mEq/L)	2.2 (-0.55 - 4.65)	87 (82)	0.6 (-3.1 - 3.275)	126 (79)	..	0.18	0.06
Total Buffer Base (mEq/L)	49.1 (46.65 -	87 (82)	47.5 (43.75 -	126 (79)	..	0.20	0.01

	51.75)		50.375)				
Base Excess in the Extracellular Fluid (mEq/L)	2.2 (-0.4 - 4.9)	87 (82)	0.35 (- 3.175 - 3.75)	126 (79)	..	0.21	0.01
White Blood Cells Count (x1000·mm³)	7.4 (5.0 - 11.225)	104 (99)	9.7 (7.1 - 13.45)	155 (98)	4.0-10.0	0.23	0.002
Band (%)	3.0 (2.0 - 5.5)	23 (21)	3.0 (2.0 - 6.0)	38 (24)	..	0.05	1
Segment (%)	78.0 (70.65 - 83.0)	87 (82)	82.8 (77.05 - 86.95)	119 (75)	..	0.25	0.002
Lymphocyte (%)	14.0 (10.0 - 20.225)	86 (81)	10.7 (6.85 - 15.4)	119 (75)	..	0.25	0.002
Monocyte (%)	6.0 (4.0 - 8.5)	45 (42)	5.0 (3.35 - 7.0)	59 (37)	..	0.17	0.34
Basophil (%)	0.3 (0.2 - 0.8)	13 (12)	0.1 (0.0 - 0.1)	13 (8)	..	0.61	0.01
Red Blood Cells Count (Mill·mm³)	4.335 (3.83 - 4.908)	102 (97)	4.185 (3.64 - 4.748)	154 (97)	4.2-5.4	0.12	0.28
Hemoglobin (g·dl)	12.6 (10.95 - 13.8)	103 (98)	12.2 (10.2 - 13.75)	155 (98)	12.0-16.0	0.07	0.81
Hematocrite (%)	37.0 (32.85 - 41.2)	103 (98)	36.6 (31.45 - 40.75)	155 (98)	36-46	0.06	0.93
Mean Corpuscular Volume (fL)	85.0 (81.4 - 88.65)	103 (98)	88.0 (84.65 - 91.9)	155 (98)	77-97	0.24	<0.001
Mean Corpuscular Hemoglobin (Pgm)	28.7 (26.6 - 29.85)	103 (98)	29.6 (27.8 - 30.55)	155 (98)	26-32	0.21	0.006
Mean Corpuscular Hemoglobin Concentration (%)	33.1 (32.45 - 34.4)	103 (98)	33.3 (31.95 - 34.15)	155 (98)	32-36	0.09	0.59
Platelet Count (x1000·mm³)	196.0 (151.5 - 260.0)	103 (98)	179.0 (125.0 - 255.0)	155 (98)	140-440	0.17	0.04
Red cell Distribution Width (%)	13.95 (13.2 - 14.825)	88 (83)	14.6 (13.75 - 16.0)	131 (82)	11.0-16.0	0.23	0.006

Platelet Distribution Width (FL)	12.8 (11.5 - 14.0)	85 (80)	13.2 (11.4 - 14.7)	120 (75)	10.0-17.0	0.13	0.32
Mean Platelet Volume (FL)	9.7 (9.175 - 10.5)	84 (80)	10.0 (9.3 - 10.7)	120 (75)	8.5-12.5	0.13	0.30
Platelet Larger Cell Ratio (%)	24.4 (19.85 - 29.3)	83 (79)	26.7 (21.05 - 30.825)	120 (75)	17-45	0.17	0.07
C-Reactive Protein (mg·l)	48.0 (24.0 - 48.0)	56 (53)	48.0 (48.0 - 48.0)	67 (42)	<6	0.23	0.05
Erythrocyte Sedimentation Rate (mm · hr)	42.0 (27.5 - 68.5)	55 (52)	59.0 (33.75 - 75.25)	56 (35)	<20	0.24	0.06
Albumin Level (g·dl)	3.3 (3.0 - 3.7)	48 (45)	2.9 (2.6 - 3.2)	71 (44)	3.5-5.5	0.37	<0.001
Serum Calcium Level (mg·dl)	8.8 (8.3 - 9.2)	72 (68)	8.6 (7.9 - 9.2)	97 (61)	8.6-10.6	0.13	0.37
Inorganic P Level (mg·dl)	3.3 (2.45 - 4.4)	59 (56)	4.0 (2.95 - 5.4)	87 (55)	2.5-5.0	0.20	0.08
Serum Na Level (mg·dl)	137.5 (135.0 - 140.0)	102 (97)	139.0 (135.0 - 142.0)	155 (98)	136-145	0.17	0.03
Serum K Level (mg·dl)	4.3 (3.925 - 4.6)	102 (97)	4.4 (4.0 - 4.85)	155 (98)	3.7-5.5	0.11	0.37
Serum Mg Level (mg·dl)	2.25 (2.0 - 2.5)	66 (62)	2.4 (2.0 - 2.7)	96 (60)	1.8-2.6	0.14	0.32
Uric Acid Level (mg·dl)	6.7 (4.05 - 9.0)	15 (14)	8.2 (5.95 - 9.95)	31 (19)	3.4-7.0	0.37	0.10
Fasting Plasma Glucose (mg·dl)	124.0 (105.0 - 177.0)	65 (61)	154.0 (120.5 - 246.5)	99 (62)	..	0.21	0.04
Blood Urea Nitrogen (mg·dl)	16.0 (11.25 - 22.5)	102 (97)	30.0 (21.0 - 52.5)	156 (98)	5.0-23.0	0.47	<0.001
Creatinine (mg·dl)	1.1 (0.9 - 1.4)	102 (97)	1.5 (1.2 - 2.2)	156 (98)	0.5-1.5	0.31	<0.001
Aspartate Aminotransferase (IU·L)	40.0 (29.0 - 55.0)	83 (79)	45.0 (31.5 - 82.5)	112 (70)	5.0-40.0	0.17	0.10
Alanine Aminotransferase (IU·L)	26.0 (16.0 - 38.5)	83 (79)	25.0 (18.0 - 45.0)	113 (71)	5.0-40.0	0.11	0.54
Lactate Dehydrogenase	710.0 (561.0 -	57 (54)	859.0 (623.5 -	95 (60)	225-500	0.17	0.20

(U·L)	1019·0)		1256·0)				
Creatine phosphokinase (IU·L)	233·0 (89·0 - 546·5)	59 (56)	204·0 (83·0 - 434·0)	91 (57)	24-195	0·08	0·90
Creatine Phosphokinase-MB (U·L)	30·0 (22·5 - 41·0)	35 (33)	30·0 (24·0 - 49·0)	41 (25)	5-25	0·10	0·96
Alkaline Phosphatase (IU·L)	180·0 (132·5 - 248·5)	75 (71)	193·0 (155·75 - 264·25)	96 (60)	64-306	0·12	0·46
Total Bilirubin (mg·dl)	0·7 (0·5 - 0·9)	46 (43)	0·8 (0·6 - 1·45)	71 (44)	0·2-1·2	0·25	0·04
Direct Bilirubin (mg·dl)	0·25 (0·2 - 0·3)	46 (43)	0·3 (0·2 - 0·6)	71 (44)	0·0·4	0·27	0·02
Prothrombin Time (Sec)	14·9 (14·0 - 16·5)	81 (77)	16·0 (14·3 - 18·0)	121 (76)	12·0-13·0	0·23	0·009
Prothrombin Time Activity (%)	81·0 (72·0 - 89·0)	40 (38)	73·0 (54·5 - 85·5)	51 (32)	85-100	0·29	0·03
International Normalised Ratio (Index)	1·2 (1·08 - 1·3)	81 (77)	1·3 (1·108 - 1·6)	120 (75)	1·0-1·1	0·29	<0·001
Partial Thromboplastin Time (Sec)	34·0 (30·0 - 40·0)	81 (77)	36·0 (31·0 - 45·75)	122 (77)	25-45	0·16	0·12
DDimer (ng·ml)	1513·0 (1071·0 - 2207·0)	9 (8)	1875·0 (1558·0 - 5269·0)	11 (6)	..	0·28	0·68
Interleukin-6 (pg·ml)	115·0 (76·0 - 147·25)	10 (9)	107·0 (47·0 - 301·0)	17 (10)	..	0·37	0·27
Fibrin DegenerationProduct (mg·L)	12·0 (12·0 - 12·0)	1 (0)	18·0 (18·0 - 18·0)	1 (0)	..	1	1
Troponin (ng·L)	227·5 (78·375 - 1400·5)	12 (11)	49·0 (28·2 - 1027·0)	21 (13)	..	0·27	0·53
Fibrinogen (mg·dl)	546·0 (477·5 - 727·0)	7 (6)	482·5 (226·75 - 615·75)	18 (11)	308-613	0·38	0·33
Hemoglobin A1c (%)	7·2 (6·9 - 7·5)	5 (4)	6·4 (5·5 - 7·0)	9 (5)	..	0·55	0·22

List of figures and legends.

Fig. 1. Investigation of model performance. Receiver operating characteristic curve of logistic regression (top) and random forest (bottom) models for training and validation sets for each cross-validation iteration and corresponding area under curves (AUC)s. The random forest model shows superior performance on both training and validation sets. The random forest model predicts patient outcomes with a 70% sensitivity and 75% specificity.

Fig. 2. Feature importance in random forest model. The importance of the random forest features using local interpretable model-agnostic explanation submodule-pick with six submodules. Negative values (blue) indicate favorable parameters suggesting a better prognosis, and positive values (red) indicate unfavorable parameters suggesting a worse prognosis.

Fig. 3. The relation between prediction horizon and performance. Distribution of days between intensive care unit admission and outcome (bars) and corresponding random forest model's area under the receiver operating characteristic curve scores for each bin (red line). Our model has the best performance to predict outcomes in a 15-day period.

Fig. 4. Investigation of clinical impacts and benefits of the model. Decision curve (top) and clinical impact curve (bottom) of the random forest model. The decision curve compares the net benefits of an intervention in three scenarios: for all patients, for no patients, and for high-risk patients based on the model prediction. The clinical impact curve compares the number of

patients classified as high risk and the number of patients with a poor prognosis who were classified as high risk, for each high-risk threshold.

Supplementary Fig. 1. Result of the two-sample Kolmogorov-Smirnov test for all numerical parameters. 11 out of 56 numerical features with least Kolmogorov-Smirnov p -values are used for modeling.

Supplementary Fig. 2. Result of χ^2 test for all categorical parameters. Four out of 56 categorical features with least Kolmogorov-Smirnov p -values were used for modeling.

Supplementary Fig. 3. Regression coefficients of the logistic model. Features with higher coefficient absolute value were more crucial in the logistic model prediction.

Supplementary Fig. 4. Distribution of selected numerical variables. Comparison of categorical variables between released (green) and dead (red) patients.

Supplementary Fig. 5. Distribution of selected categorical variables. Comparison of categorical variables between released and dead patients

Decision curve

Clinical impact curve

train (AUC=0.79±0.02)

validation (AUC=0.74±0.12)

train (AUC=0.87±0.00)

validation (AUC=0.81±0.08)

submodule 1

submodule 2

submodule 3

submodule 4

submodule 5

submodule 6