

RESEARCH ARTICLE

Identification of predictive patient characteristics for assessing the probability of COVID-19 in-hospital mortality

Bartek Rajwa^{1,2*}, Md Mobasshir Arshed Naved³, Mohammad Adibuzzaman⁴, Ananth Y. Grama³, Babar A. Khan⁵, M. Murat Dundar⁶, Jean-Christophe Rochet^{2,7*}

1 Bindley Bioscience Center, Purdue University, West Lafayette, Indiana, United States of America, **2** Purdue Institute for Integrative Neuroscience, Purdue University, West Lafayette, Indiana, United States of America, **3** Dept. of Computer Science, Purdue University, West Lafayette, Indiana, United States of America, **4** Oregon Clinical and Translational Research Institute, Oregon Health and Science University, Portland, Oregon, United States of America, **5** Regenstrief Institute, Indianapolis, Indiana, United States of America, **6** Dept. of Computer and Information Science, IUPUI, Indianapolis, Indiana, United States of America, **7** Borch Dept. of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, Indiana, United States of America

* brajwa@purdue.edu (BR); jrochet@purdue.edu (J-CR)



OPEN ACCESS

Citation: Rajwa B, Naved MMA, Adibuzzaman M, Grama AY, Khan BA, Dundar MM, et al. (2024) Identification of predictive patient characteristics for assessing the probability of COVID-19 in-hospital mortality. *PLOS Digit Health* 3(4): e0000327. <https://doi.org/10.1371/journal.pdig.0000327>

Editor: Gilles Guillot, CSL Behring / Swiss Institute for Translational and Entrepreneurial Medicine (SITEM), SWITZERLAND

Received: July 15, 2023

Accepted: March 6, 2024

Published: April 23, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pdig.0000327>

Copyright: © 2024 Rajwa et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data used in this research has been provided through a data transfer

Abstract

As the world emerges from the COVID-19 pandemic, there is an urgent need to understand patient factors that may be used to predict the occurrence of severe cases and patient mortality. Approximately 20% of SARS-CoV-2 infections lead to acute respiratory distress syndrome caused by the harmful actions of inflammatory mediators. Patients with severe COVID-19 are often afflicted with neurologic symptoms, and individuals with pre-existing neurodegenerative disease have an increased risk of severe COVID-19. Although collectively, these observations point to a bidirectional relationship between severe COVID-19 and neurologic disorders, little is known about the underlying mechanisms. Here, we analyzed the electronic health records of 471 patients with severe COVID-19 to identify clinical characteristics most predictive of mortality. Feature discovery was conducted by training a regularized logistic regression classifier that serves as a machine-learning model with an embedded feature selection capability. SHAP analysis using the trained classifier revealed that a small ensemble of readily observable clinical features, including characteristics associated with cognitive impairment, could predict in-hospital mortality with an accuracy greater than 0.85 (expressed as the area under the ROC curve of the classifier). These findings have important implications for the prioritization of clinical measures used to identify patients with COVID-19 (and, potentially, other forms of acute respiratory distress syndrome) having an elevated risk of death.

Author summary

Gaining insight into the patient attributes that are indicative of poor outcomes resulting from infections caused by highly lethal viruses such as SARS-CoV-2 is of paramount importance for hospitals and healthcare systems in order to adequately anticipate and

agreement (DTA) between the Regenstrief Institute, 1101 West 10th Street, Indianapolis, IN 46202, and Purdue University, West Lafayette, IN 47907. The data has been de-identified in compliance with 45 C.F.R. § 164.51. The title to the data is solely owned and remains with the Regenstrief Institute, and it has been offered to Purdue University as a service to the research community. As per the agreement, Purdue researchers are obligated to restrict access to and processing of the data to authorized personnel or employees who are required to process the data to conduct their work in connection with the purposes outlined in the DTA. Third-party researchers or scientists who wish to access these data for their own research purposes are encouraged to directly contact the Regenstrief Institute via the <https://www.regenstrief.org/data-request/> form. Such requests will be assessed in accordance with the Institute's data-sharing guidelines, which demonstrate their commitment to supporting wider scientific collaboration while maintaining rigorous standards of data privacy and security.

Funding: BR, BAK and JCR gratefully acknowledge funding from the COMMIT (COvid-19 unMet Medical needs and associated research exTension) Program of the Gilead Foundation (grant number IN-US-983-6060). The sponsor had no involvement in the study's design, data collection and analysis, decision to publish, or in the preparation of the manuscript. Information about the Gilead Foundation is available at <https://www.gilead.com/purpose/gilead-foundation>.

Competing interests: The authors have declared that no competing interests exist.

address forthcoming outbreaks. Approximately 20% of the infections in the initial phase of the COVID-19 pandemic resulted in severe respiratory complications triggered by inflammatory responses. Moreover, individuals with severe COVID-19, particularly those with pre-existing neurodegenerative conditions, who were at a heightened risk, frequently displayed neurologic symptoms. Although there is a documented association between severe COVID-19 and neurologic issues, the underlying causes remain elusive. In this study, we examined the electronic health records of 471 patients with severe COVID-19 to identify clinical indicators that could predict mortality. Through machine learning models, we discovered a series of clinical features, notably those associated with cognitive dysfunction, that accurately predict mortality (area under the ROC curve > 0.85). These results underscore the significance of specific clinical indicators in recognizing patients at increased risk of mortality from COVID-19, thus enabling more focused and effective healthcare strategies.

1 Introduction

1.1 COVID-19 and neurologic symptoms

Patients with severe COVID-19 warranting hospital admission present with a variety of symptoms that are carefully evaluated by admitting physicians. These patient evaluations have led to the identification of specific comorbidities linked to severe forms of COVID-19 or fatal outcomes [1–3], including neurodegenerative disorders and dementia [4–6]. Although SARS-CoV-2 infection can cause neurologic symptoms by directly affecting the central nervous system (CNS), this phenomenon has only been shown in a very small subset of patients [7]. Instead, neurologic symptoms are more likely to occur due to indirect effects involving the strong innate immune response and cytokine storm caused by SARS-CoV-2 infection.

A cytokine storm is a hyperinflammatory response to an infection caused by a sudden spike in levels of pro-inflammatory cytokines and chemokines, including IL-1, IL-2, IL-4, IL-6, IL-7, IL-8, IL-9, IL-10, IL-18, granulocyte stimulating factor (G-CSF), IP-10, monocyte chemoattractant protein (MCP)-1, MCP-3, macrophage inflammatory protein 1 (MIP-1A), cutaneous T-cell attracting chemokine (CTACK), IFN- γ , and TNF- α . In turn, this phenomenon can result in overwhelming systemic inflammation, acute respiratory distress syndrome (ARDS), and multi-organ failure [8]. Evidence suggests that a cytokine storm can trigger various neurologic symptoms, ranging from headaches, dizziness, and disorientation to convulsions or seizures [8].

1.2 Approach to prediction and feature selection

In this report, we present the results of a study aimed at establishing a link between COVID-19 symptoms observed upon patient admission (within the first 24 h of hospitalization) and the risk of patient death. We were particularly interested in the predictive value of easily observable neuropsychiatric symptoms such as disorientation, cognitive impairment, and delirium. Our strategy involved a data-driven discovery of predictors. Instead of postulating *a priori* a feasible set of clinical features likely to be associated with mortality and then testing the resulting hypotheses using a standard generalized linear model (GLM) approach, we retained all possible clinical features for the analysis. We used the disease outcome to train a supervised classifier with feature ranking and selection capability. When an explainable classifier achieved high accuracy, we queried it to determine which feature combination was responsible for its

strong performance. The final output was a set of hypotheses (that could be addressed in future intervention studies) positing that the identified predictive combinations of clinical features were causally linked to mortality or that common latent factors influenced both the mortality as well as the detected clinical characteristics.

We utilized two independent approaches to quantify the predictive power of the observations collected at the time of patient admission, or within the first 24 h of hospitalization: elastic-net regularized logistic regression (LR-ENET) and XGBoost classification [9, 10]. The use of both methods was followed by an analysis of the relative contributions of the selected features (“explanatory variables”) to the prediction via the SHapley Additive exPlanations (SHAP) approach. The SHAP factors provide a valuable tool for interpreting the significance of each clinical feature in a model. By assigning a value to each variable based on its average contribution to the outcome across all possible combinations, this method offers a comprehensive way to understand the relative importance of each feature [11–13]. This is particularly useful in complex scenarios where the features interact in intricate ways.

The described analyses led to the identification of several important predictive features, including neurologic symptoms.

1.3 Related research

Data-driven investigation of COVID-19 mortality employing machine learning tools (such as XGBoost) and feature explanation methods (SHAP values) has been demonstrated before for the processing of clinical laboratory results [14–16], for predicting death outcomes [17–19], demonstrating links between socioeconomic disparities and COVID-19 spread [20], and showing the impact of COVID-19 on mental health in self-identified Asian Indians in the USA [21]. Several other articles using SHAP were reviewed recently by Bottino et al. [22].

2 Methods and datasets

2.1 Ethics statement

This study, identified by IRB protocol #2004316653, was approved by the Indiana University Institutional Review Board. As the study procedures involved the use of a deidentified electronic medical record database, consent was not required.

2.2 Dataset description

Electronic health record (EHR) data were obtained from 471 patients with severe SARS-CoV-2 infection. The patients were admitted to the intensive care units (ICUs) of IU Health Methodist Hospital and Sidney & Lois Eskenazi Hospital, both in Indianapolis, Indiana, between March 2020, and August 2020. 399 patients were eventually discharged, whereas 72 patients died. The demographic characteristics of the cohort are shown in Table 1. 246 patients self-identified as *Black* or *African American*, and 196 patients identified as *white*. 245 of the patients were females, and 226 were males. There was no statistically significant difference in age between the African-American and white patients. However, patients who identified as Hispanic or Latino were significantly younger than other patients ($p < 0.001$).

2.3 Preprocessing

The original dataset consisted of data collected at multiple time points during the patients’ treatment. We used only the data from the first time point (i.e., the initial evaluation and the subsequent measures up to 24 h), consisting of the earliest available diagnostic characteristics. We retained the minimal and maximal values if multiple measurements and/or laboratory

Table 1. General demographic characteristics of the investigated cohort.

Race	Female					Male				
	Alive	Alive %	Died	Died %	Total	Alive	Alive %	Died	Died %	Total
Asian	3	100.0%	0	0.0%	3	5	83.3%	1	16.7%	6
Black or Afr. Amer.	123	87.9%	17	12.1%	140	85	80.2%	21	19.8%	106
Refused to identify	1	33.3%	2	66.7%	3	2	100.0%	0	0.0%	2
Unknown	10	100.0%	0	0.0%	10	5	100.0%	0	0.0%	5
White	77	86.5%	12	13.5%	89	88	82.2%	19	17.8%	107
Total	214	87.3%	31	12.70%	245	185	81.9%	41	18.1%	226

<https://doi.org/10.1371/journal.pdig.0000327.t001>

results were provided. Because many features describe the patient's status as very detailed and granular, some binary factors were observed for only a few (one or two) patients. These variance-deficient features were removed to prevent the model from overfitting, even though they might have been informative. Another step of feature engineering (not pursued in this study) is likely needed to construct virtual features that summarize these descriptors.

The presence of multiple correlated features related to delirium is an example of the described problem. We found that the large number of delirium-related descriptors would lead to the emergence of 29 distinct categories. However, because over 300 cases showed no evidence of delirium, an alternative approach would be to combine all of the positive categories into one (e.g., "some evidence of delirium"). In our LR-NET and XGBoost models, this engineered feature was not included. Nonetheless, a distinct analysis was carried out to assess the predictive value of the "delirium" secondary feature, with its findings detailed in the "Results" section.

2.4 Logistic regression with elastic-net regularization model

We established a set of relevant diagnostic features by implementing an *ante-hoc* explainable, predictable statistical model with embedded feature selection capability. We utilized a logistic regression model regularized with a ridge (ℓ_2), LASSO (ℓ_1), or a combination of both penalties (elastic net) [9, 23, 24]. This approach allowed us to (1) create a simple model capturing all the significant sources of variability, incorporating all of the diverse clinical descriptors/features, and (2) perform simultaneous feature selection and feature ranking, allowing identification of the major drivers of correct prediction. It is worth highlighting that the inclusion of ℓ_1 regularization may generate instability within the feature selection process [25]. However, this challenge can be addressed by re-executing the feature selection multiple times. Please refer to the [S1 Appendix](#) for details.

2.5 XGBoost model

We utilize the extreme gradient boosting decision tree (XGBoost). This method developed by Chen and Guestrin, is a highly effective, portable, and scalable machine learning system for tree boosting that is optimized under the Gradient Boosting framework [10]. It combines a series of low-accuracy weak classifiers using the gradient descent architecture to produce a strong classifier with higher classification performance.

2.6 SHapley Additive exPlanations (SHAP) values

The subsequent analysis of features identified by LR-ENET and XGBoost is performed using approach based on cooperative game theory concept known as SHAP (or Shapley) values. The

subsequent analysis of the features identified by LR-ENET and XGBoost was performed using an approach based on the cooperative game theory concept known as SHAP (or Shapley) values [11–13, 26]. The SHAP values can determine the importance of a feature and its directionality influence by comparing what a model predicts with and without that feature for each observation in the training data and calculating the marginal contribution [12].

In simple terms, the Shapley analysis is a method that allocates “credit” or “blame” fairly among a set of contributing factors or variables in a model. This is particularly useful in clinical research where numerous variables can influence an outcome, and it can be challenging to determine the importance or contribution of each variable, especially when they interact in complex ways.

The Shapley factor analysis addresses this challenge by considering all possible combinations of variables and their contributions to the accuracy of a model or multiple complementary models. Essentially, it answers the question: “What is the average contribution of each variable to the outcome when all possible combinations of variables are taken into account?” Therefore, each variable’s SHAP can be interpreted as its average marginal contribution across all possible combinations of variables. Even with non-linear and complex interactions between features, SHAP values remain meaningful and interpretable.

3 Results

3.1 Patients characteristics

Patients who died of COVID-19 were significantly older than patients who survived ($p < 0.001$) in both the female and male groups. The average Braden score was slightly lower among patients who died (females: $p = 0.035$, males: $p = 0.021$). There was no observable difference between the mean Glasgow Coma scores of patients who died or survived (females $p = 0.14$, males: $p = 0.15$). Patients who presented with clear speech and, therefore, presumably had intact cognitive ability were significantly over-represented among those who survived (odds ratio of being admitted with clear speech for surviving patients: 5.65 ($p < 0.001$) for females, 3.6 ($p < 0.001$) for males). See Table 2 for a summary of the findings.

Delirium-related symptoms were noted across all patient age groups, but were particularly prevalent among the elderly. Interestingly, when the cohort was stratified by gender, older female patients were found to be at a significantly higher risk of developing delirium, whereas no statistically significant increase in risk was observed among their male counterparts. See Table 3 and Table A in the S1 Appendix. When adjusting for age, there was no difference in delirium occurrence between patients in the *white* and *African American* subcohorts ($p = 0.52$).

The composite feature that describes the presence of any delirium-related symptoms has been identified as an excellent univariate predictor of patient outcomes. Patients presenting delirium symptoms have a significantly higher probability of death, a relationship that has been observed in both female and male cohorts. See Table 4.

To compare the formulated feature selection models to a benchmark, we created a simple GLM (non-regularized regression) employing the composite delirium feature, sex, race,

Table 2. Mean ages, Braden scores, Glasgow coma scores, and probabilities of demonstrating a clear speech pattern for patients in the investigated cohort.

Sex	Status	Age, mean	Age, IQR	Braden score, mean	Coma score, mean	\mathcal{P} (Clear speech)	(95% CI)
Female	alive	55.08	25.61	16.92	13.80	0.84	(0.79, 0.88)
	died	80.27	19.35	13.90	12.84	0.48	(0.32, 0.65)
Male	alive	56.70	20.18	17.11	14.14	0.82	(0.76, 0.87)
	died	75.31	12.85	14.15	13.29	0.56	(0.41, 0.7)

<https://doi.org/10.1371/journal.pdig.0000327.t002>

Table 3. Occurrence of delirium in different age groups of male and female patients in the study.

Age	Female					Male				
	No delirium	No delirium %	Delirium	Delirium %	Total	No delirium	No delirium %	Delirium	Delirium %	Total
[17.1, 51.8)	84	92.30%	7	7.70%	91	59	89.40%	7	10.60%	66
[51.8, 66.9)	66	91.70%	6	8.30%	72	72	84.70%	13	15.30%	85
[66.9,101.5]	57	69.50%	25	30.50%	82	60	80.00%	15	20.00%	75
Total	207	84.50%	38	15.50%	245	191	84.50%	35	15.50%	226

<https://doi.org/10.1371/journal.pdig.0000327.t003>

Table 4. Probability of death given signs of delirium at the hospital admittance and accompanying lower (LCL) and upper (UCL) 95-percentile confidence limits.

Sex	Status	$\mathcal{P}(\text{Death})$	LCL	UCL
Female	No delirium	0.08	0.05	0.12
	Delirium	0.39	0.25	0.56
Male	No delirium	0.13	0.09	0.19
	Delirium	0.46	0.30	0.62

<https://doi.org/10.1371/journal.pdig.0000327.t004>

Braden score, and age category discretized into three tertiles: *young* [<55.7], *middle* [51.7, 66.8] and *older* [>66.9]. The mathematical specification of the model is provided in the [S1 Appendix](#) (Equation G).

Inspection of the model demonstrates that age is the most important predictive factor, followed by delirium symptoms and the Braden score. There is no significant difference between patients of different races. Table B in the [S1 Appendix](#) shows the detailed results of the statistical analysis, and Table C presents the average marginal effects (AME) computed from the model. After fitting, the model has an AUC of 0.87, a sensitivity of 0.94, and a specificity of 0.3. Post-hoc analysis demonstrated that after controlling for race, sex, age group, and the Braden score, the odds of death for patients exhibiting delirium symptoms increased by 5.23 ($p < 0.001$).

3.2 Regularized logistic regression model results

LR-ENET model training was performed using a grid search through the space of parameters λ and α . An example of a training grid is illustrated in [Fig 1](#).

The performance of the trained LR models is demonstrated in [Fig 2](#). The classifier performed well and was consistently able to reach an AUC of approximately 0.9.

Each regularized logistic regression model in the ensemble was trained with the objective of limiting the feature count to around twenty. To address data imbalance, the training and feature selection process included an augmentation step where synthetic data points were generated. This was achieved by applying the ROSE algorithm [27] directly to each set of bootstrapped data during the 0.632 bootstrap cross-validation training. This approach not only addressed the imbalance in the data but also introduced additional variability, presenting a robust challenge to the classifier.

After each run, absolute values of the different LR-ENET models were collected and scaled to [0, 1] intervals. These measures were considered reflective of the predictors' importance. Due to instability, several lower-performing predictors were observed with non-zero coefficients, though only in a few runs. On the other hand, consistent predictors emerged in most or all of the runs. All of the scaled predictive importance values were subsequently analyzed, and the top 20 were picked for the final selection.

The results illustrating the identified predictive EHR features are summarized in [Fig 3](#). As mentioned previously, due to the inherent instability of sparse classifiers, each run may result

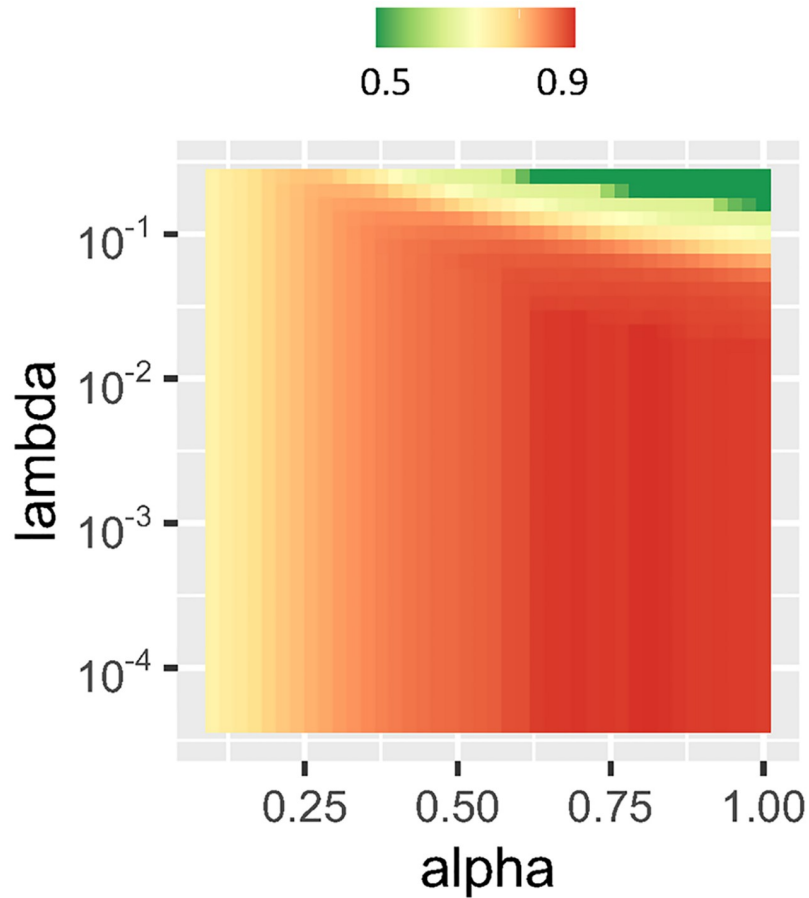


Fig 1. Changes in the performance of the elastic-net regularized model in the classification of COVID-19 patients, expressed as AUC, given different values of α and λ parameters.

<https://doi.org/10.1371/journal.pdig.0000327.g001>

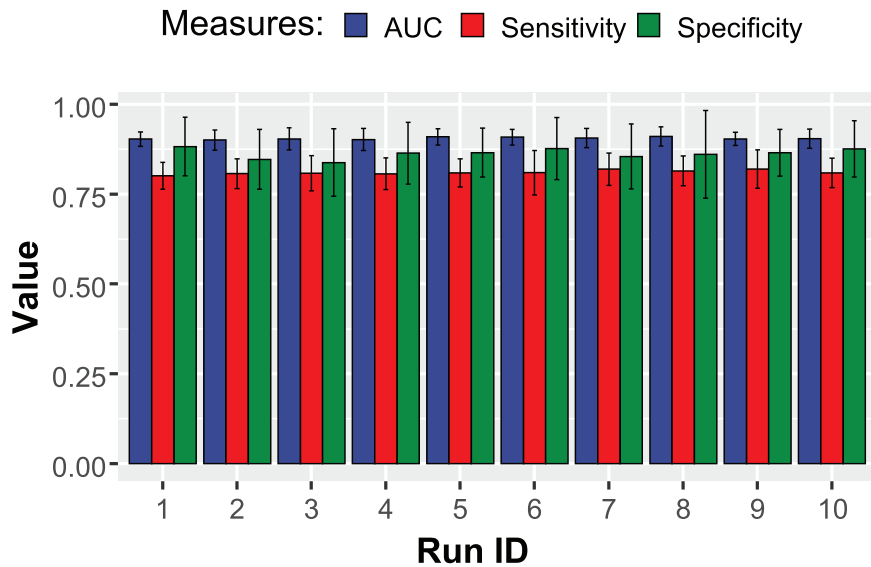


Fig 2. Performance of the elastic-net regularized regression model in the classification of COVID-19 patients. Ten independent rounds of model training are shown.

<https://doi.org/10.1371/journal.pdig.0000327.g002>

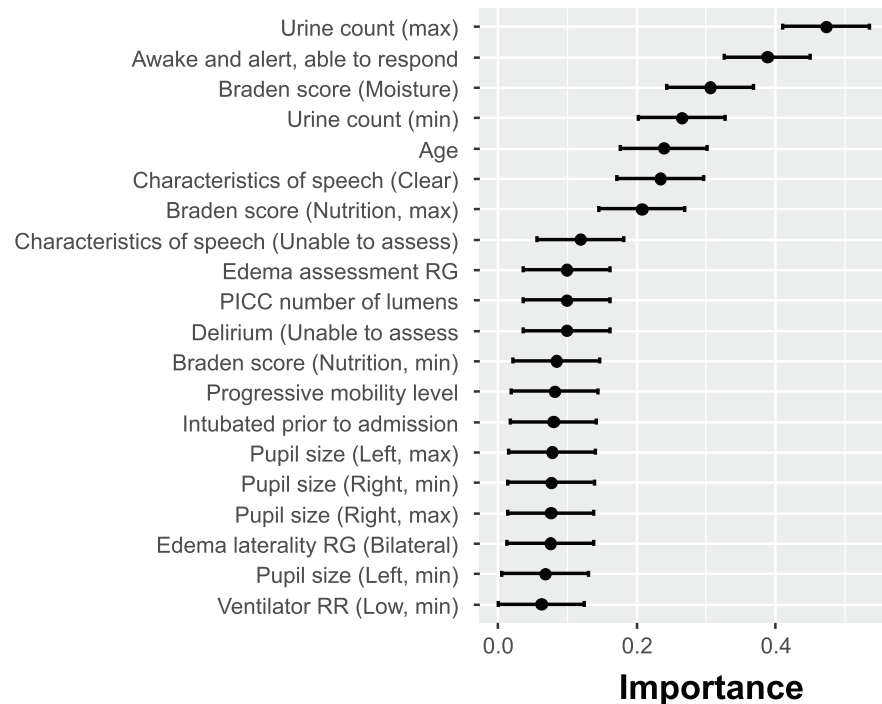


Fig 3. Importance of elastic-net selected features shown as the normalized absolute value of the regression coefficients. Multiple runs of the feature selection were executed to alleviate the inherent instability of the sparse models.

<https://doi.org/10.1371/journal.pdig.0000327.g003>

in a slightly different order of selected features. This is expected since informative feature sets may have multiple subsets that lead to the same classification success. Therefore, formally, the SHAP analysis should be performed for every separate feature selection run. However, to obtain some preliminary insight into the explanatory power of the selected features, we conducted the SHAP analysis for only one of the ten LR-ENET-trained classifiers, which well represented the mean feature importance. The results are demonstrated in Fig 4.

3.3 XGBoost model results

XGBoost performed comparably to the LR-ENET algorithm, although it is noted that the acquired specificities were often lower. We optimized the XGBoost hyperparameters to maximize the AUC as opposed to explicitly minimizing the number of utilized features. Therefore, the algorithm was allowed to employ as few or as many features as were required to optimize its performance.

The XGBoost model can be used as a feature selection wrapper. In the process of training the features are selected in ignored in the created trees. On the basis of that, the XGBoost method ranks the most significant characteristics according to “Gain,” “Cover,” and “Frequency.” The gain reflects how crucial a characteristic is for making a branch of a decision tree pure. Coverage measures the proportion of observations affected by a feature. A feature’s frequency is the number of times it is used in all created trees (See Fig 5).

The XGBoost algorithm was not directly constrained by the number of used features. The performance of the XGBoost classifier is shown in Fig 6. The XGBoost-discovered features were exposed to the SHAP analysis, the results of which are demonstrated in Fig 7.

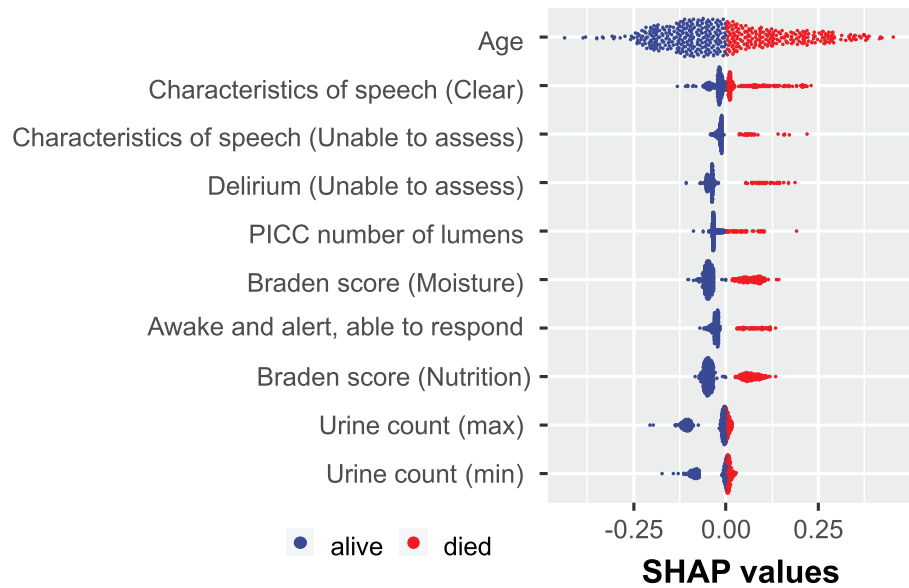


Fig 4. Example of the SHAP value distribution for all of the tested patients and the augmented dataset. The visualization was created based on one of the trained LR-ENET classifiers.

<https://doi.org/10.1371/journal.pdig.0000327.g004>

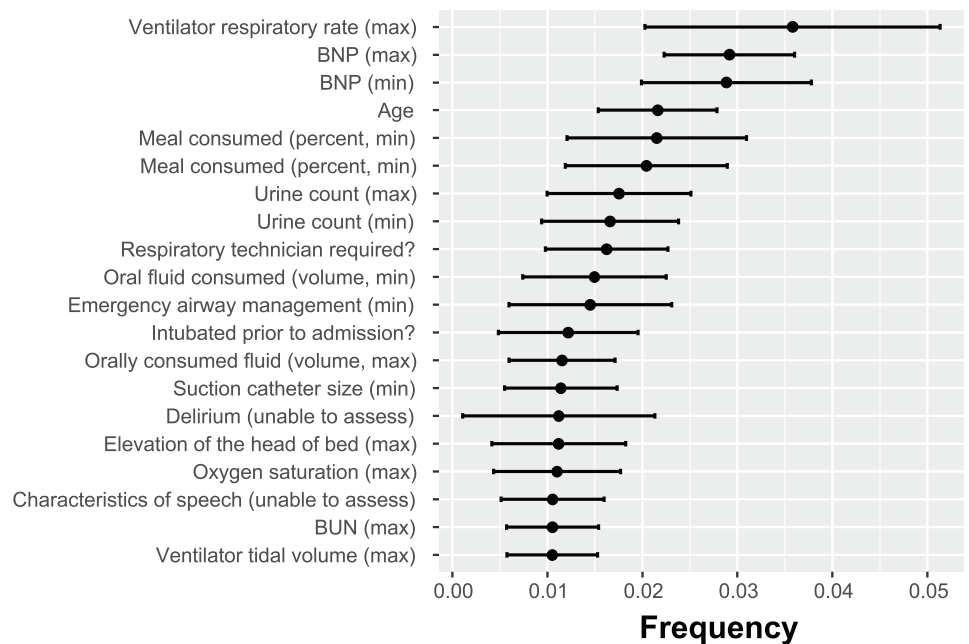


Fig 5. Importance of features employed by XGBoost algorithm expressed as the frequency at which each feature was used in the created classification trees.

<https://doi.org/10.1371/journal.pdig.0000327.g005>

3.4 Discovered predictive features

Both feature-selection strategies identified a number of clinical characteristics that reflect overall patient health, cognitive status, and hospitalization risks associated with the patient’s condition. Here, we list some of the identified features, with a particular focus on those discovered

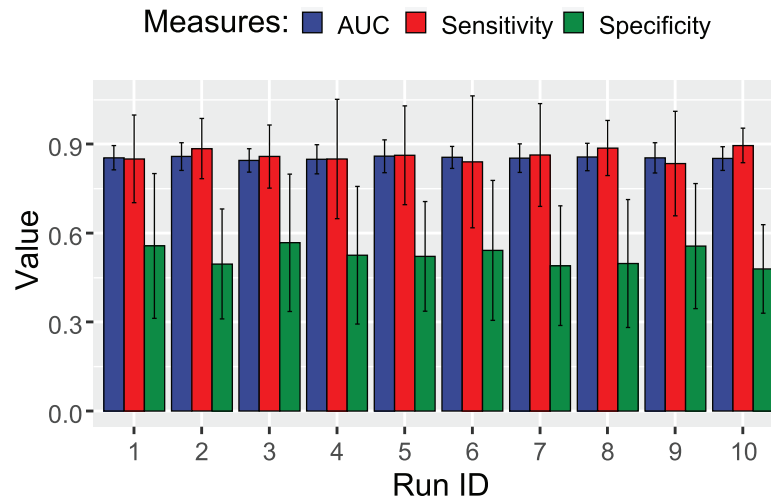


Fig 6. Performance of the XGBoost model in repeated independent cross-validation rounds classifying the COVID-19 patients.

<https://doi.org/10.1371/journal.pdig.0000327.g006>

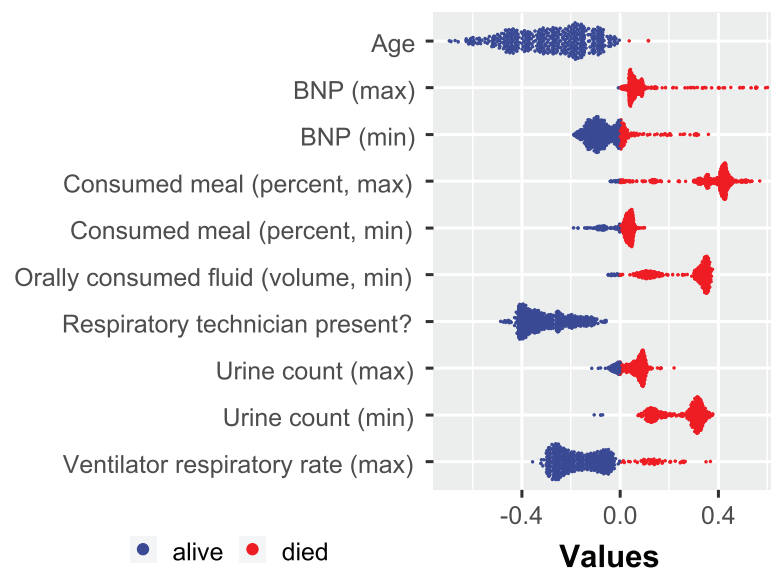


Fig 7. Example of SHAP value distribution for all of the tested patients. The visualization was created based on one of the trained XGBoost classifiers.

<https://doi.org/10.1371/journal.pdig.0000327.g007>

by the LR-ENET model, which demonstrated superior AUC, sensitivity, and specificity compared to the XGBoost approach.

CNS and cognition-related features. A set of features related to the patient’s neurologic status was selected. The LR-ENET approach generated a model richer in these features than the XGBoost model.

- **Inability to assess the patient’s speech** may be caused by neurologic symptoms (such as loss of consciousness or the presence of delirium), but it may also be related to intubation. This

EHR observation represents a surrogate measurement of the patient state in both instances, and its existence significantly suggests an elevated mortality risk.

- **Unclear and slurred speech** observed by nurses or physicians could be a key indicator of delirium [28]. Delirium is related to adverse outcomes during hospitalization (e.g., increased risk of complications) in post-acute care settings, and long-term follow-up (e.g., prolonged cognitive and functional impairment). The lack of speech clarity is only a surrogate measure for delirium that might be considered low value given that the presence of delirium is another explicitly defined feature in the analyzed dataset. However, it should be noted that there is a significant rate of under-recognition and lack of documentation of delirium in admitted patients, with less than 3% of cases documented by International Classification of Diseases-9 (ICD-9) codes in patients' medical records [28, 29]
- **Any evidence of delirium.** Among the delirium-related characteristics included in the dataset is the expressly stated "Presence of delirium." However, the feature selected as predictive was "Inability to assess delirium," which may appear odd and counterintuitive. Examining the explicit delirium feature reveals a biserial correlation between the feature and the patient's registered death of only 0.13, whereas the inability to assess delirium scores -0.25. This curious selection may be partially explained by the procedure required to assess delirium, which also explains the under-recognition of delirium in general. The Confusion Evaluation Method, the most used delirium assessment tool, requires an in-person, bedside discussion with the patient [30]. Because delirium fluctuates, interview-based approaches may overlook delirium that occurs beyond bedside interviews. It has been reported that manual searches of all records (e.g., nursing and physician notes, discharge summaries) may allow the determination of signs of delirium despite the lack of explicit notes in the records [28]
- **"Awake and able to respond"** is another important neurologic assessment feature. A patient is scored positively if considered "awake," "able to respond" (i.e., responding appropriately), and "oriented" (aware of self, place, and time) [31]. However, the patient's situation may change on the first day after admission. Therefore, patients who were sedated and unable to be subjected to the neurologic examination may demonstrate full-strength cognitive abilities later. Thus, yet again, this feature provides value in combination with other features and cannot be considered in isolation. Interestingly, the Glasgow coma scale (GCS), which is a measure used to determine the level of consciousness in trauma or critically ill individuals with impaired consciousness and which was also available in the dataset, has not been utilized by the models.

Patient frailty-related features. Several identified features described the patient's overall condition and frailty upon hospital admission.

- **Braden score** describes the frailty of the patients. The Braden scale was created to identify early pressure, sore-prone patients. Six sub-scales of the score measure sensory perception, skin wetness, activity, mobility, friction and shear, and nutrition [32]. Although ample evidence exists for the usefulness and applicability of the Braden scale in predicting patients' conditions during hospitalization [33], the scale has been criticized for lacking explainability and detail from the machine learning perspective [34]. On the other hand, a retrospective study of 146 COVID-19 patients demonstrated that the Braden score is indeed helpful for risk stratification at hospital admission, as the mortality among patients with $BS \leq 15$ was significantly higher than in patients with $BS > 15$ [35].

In our research context, Braden score features also appear to identify particularly vulnerable patients. Interestingly, only two sub-scales (moisture and nutrition) have been included by the LR-ENET classifier in the final model. On the other hand, the XGBoost model did not rely on Braden scores.

- **Number of lumens** Multiple studies have demonstrated a substantial correlation between the number of PICC lumens and the risk of complications, including central-line associated bloodstream infection (CLABSI), venous thrombosis, and catheter occlusion [36, 37].
- **The requirement for a respiratory technician** to be present during the transportation of the patient is yet another predictive EHR feature that communicates the severity of the patient's condition.
- **The urine voiding count feature** is connected with the lower urinary tract symptoms. The medical literature describes an association between LUTS and COVID [38–40]. According to these reports, there was a high prevalence of abnormal urinary storage symptoms, urine frequency, urgency, and urinary incontinence among the SARS-CoV-2-infected patients. The data indicate that the majority of COVID-19 patients may experience increased urination frequency, nocturia, and urgency during the infection. Also, patients with urine storage symptoms were found to have considerably higher COVID-19 severity levels than those without urine storage symptoms [40]. The urinary symptoms might be caused directly by inflammation or indirectly by COVID-19-related general dysfunction in the autonomic nervous system [41].

Clinical laboratory test results. Report on the patient's physiological status. Specifically, two test results were identified by our feature selection methods.

- **The measurement of urea nitrogen in serum or plasma** (BUN SerPl test) assesses the kidney's function. High urea nitrogen levels in the BUN test indicate problems with renal function or a reduction in blood supply to the kidneys. A reduction in urea nitrogen, as measured by the BUN test, indicates serious liver illness or malnutrition. The outcomes of the BUN test were put into the XGBoost model's identified collection of features. However, they were not among the top 20 features identified by LR-ENET. In the analyzed cohort, there was a significant difference in test findings between patients who died and those who survived ($p < 0.001$). These observations are consistent with literature reports [42, 43].
- **Brain natriuretic peptide (BNP)** is an active fragment (1–32) of the cardiac cell-produced ProBNP. It is elevated in right-sided and left-sided heart failure, as well as systolic and diastolic heart failure. It is, therefore, used to identify and treat heart failure. The BNP test was recognized as an important feature by the XGBoost-feature selection, but not the LR-ENET. In the tested cohort, the values of the BNP test were substantially higher for male patients who died ($p < 0.001$) but not for females ($p = 0.23$). Others have recently postulated that BNP should be considered an early predictor of clinical severity in patients with COVID-19 pneumonia [44].

4 Discussion

4.1 Clinical feature selection

When analyzing disease mortality causes, the search for predictive factors typically begins with the formulation of a hypothesis based on domain knowledge of the underlying diseases and initial preliminary evidence, such as case studies and anecdotal reports. This hypothesis-driven process is philosophically well-established and operationally widely accepted. Despite the fact

that this conventional path of hypothesis-driven research has been challenged numerous times in recent years, particularly by the rise of genomics, for many researchers it is virtually synonymous with the scientific method itself [45]. The alternative paradigm, often referred to as data-driven research, begins with an agnostic stance that does not involve a preconceived hypothesis and instead employs either a data reduction process that results in the emergence of a model or a supervised model-building process that reveals predictive features that explain observed outcomes.

Here, we examined three methods for identifying factors predictive of mortality among COVID-19 patients with severe disease necessitating hospitalization. As a baseline, we utilized a standard statistical technique for formulating a hypothesis and generating or selecting hypothesized predictive factors through a non-regularized GLM framework. In carrying out this method, we used a composite delirium factor, which is a combination of multiple EHR characteristics/symptoms associated with delirium occurrence. We also accounted for age, race, and sex, which we recognized as notable confounding variables plausibly related to the outcome. The developed model revealed a significant increase in the likelihood of death for hospitalized patients exhibiting any symptom of delirium.

Subsequently, we developed two feature discovery models using two well-established machine learning techniques. The first model utilized regularized elastic-net logistic regression. A sparse collection of predictors was generated using the model's inbuilt feature selection capability. However, as previously established, the sparse models are inherently unstable, and the selection of reliably predictive features necessitated many model runs and a compilation of the findings [25]. The most consistently predicted variables across numerous runs were evaluated using the SHAP method to obtain insight into their local significance for patient classification.

The second machine learning model utilized the XGBoost tree learning technique. Unlike the LR-ENET approach which enforced sparsity, this method freely utilized all available clinical features. After multiple rounds of independent training and cross-validation, the algorithm demonstrated high stability, ultimately producing very similar results. To represent feature importance, data from the internals of trained classifiers (gain and frequency) were extracted. A secondary XGBoost model was then trained using the top features, and its SHAP values were assessed.

Our findings demonstrate that the two distinct classifiers relied on very different sets of predictive features during optimization and training. Age, surrogate measures for the patients' cognitive status (neurologic observations), features broadly describing the patients' overall condition upon admission (such as the Braden score), and features associated with a risk of serious complications requiring hospitalization (such as the number of catheter lumens) were the descriptors that the LR-ENET model selected. It is interesting to note that common clinical laboratory test results were not chosen during the feature selection process by LR-ENET.

XGBoost, on the other hand, placed considerable emphasis on laboratory test results, including values obtained from the BUN and BNP tests. XGBoost captured the patient's overall condition by analyzing variables such as oral fluid consumption, oxygen saturation, ventilator use, etc. Despite the fact that the features chosen by XGBoost are frequently more quantitative and objective, the overall performance of the XGBoost model was marginally inferior in terms of specificity (such as in the case of the laboratory test results).

4.2 Limitations

It is important to acknowledge that the study presented here has certain critical limitations that require a thorough understanding.

The study lacks consideration of the patients' comorbidity status, which is a complex and multifaceted area. It is widely acknowledged that patients with severe and potentially fatal cases of SARS-CoV-2 infection typically have multiple comorbidities [46–52]. This is also true for our specific cohort, and has been demonstrated by the data continuously collected by the Regenstrief Institute on patients from Indiana [53, 54]. However, due to the limited sample size and diverse range of comorbidities present, we would not be able to draw any conclusive inferences about which specific comorbidities may impact the likelihood of mortality.

The other limitation stems directly from the question of the extent to which the deaths among the patients in the study cohort can be directly associated with COVID-19, rather than other unrelated factors. The patients were admitted to the hospital's ICU due to the onset of symptoms, not because of results of a screening test. As the admission occurred before the availability of rapid testing, it was assumed that the severity of symptoms was at least correlated with their later confirmed COVID status. It is important to bear in mind that in the early phase of the pandemic, the admission of patients without severe respiratory distress symptoms was minimized across hospital systems in the US and elsewhere [55, 56]. It would be impossible to ascertain whether every individual patient in the cohort who died indeed died “from” COVID or rather “with” COVID unless an autopsy followed by a detailed inquest were performed in every case. However, the studies on overall mortality in the US indicate an underestimation, rather than an overestimation, of COVID mortality [57, 58].

Fillmore et al. showed that the number of hospitalized patients testing positive for COVID-19 without exhibiting related symptoms rose in tandem with greater virus exposure and advancements in vaccination programs [59]. Consequently, it is now more crucial to differentiate between patients who died “with” COVID-19 and those who died “due to” COVID-19, especially when analyzing recent data, as opposed to the historical data from the early and mid-2020s, which we used.

The described limitation is related to the wider issue of clinical coding, as possible errors in coding for mortality and morbidity significantly impact any use of machine learning in clinical decision support. We recognize that potential inaccuracies in ICD code assignment, and resulting data inconsistencies, may lead to flawed machine learning models, yielding unreliable predictions.

Lastly, it is important to note that our research utilizes a specifically narrow understanding of causal links. While feature importance scores as well as Shapley factors are valuable tools for attributing potentially causal effects to individual variables or factors within a presupposed causal model—particularly in cases where multiple variables are involved—we do not claim that a high Shapley value (or another metric of statistical association) is, on its own, proof of a causal relationship [60].

However, the measures of feature importance can indicate to which extent changing the variable (intervention) may contribute to changes in the outcome. This aligns with the causal inference goal of understanding the effect of interventions. They also help to decompose these interactions between variables. This is crucial in causal inference, as it aids in understanding how combined interventions might affect an outcome.

The measures of clinical feature importance do inspire counterfactual reasoning, as they allow us to consider the contributions of variables under different intervention scenarios. For instance, the significant impact of delirium suggests that preventing its onset could be a compelling strategy for intervention. Such variables can then be further investigated for causal relationships using a randomized clinical trial setting.

In essence, the clinical features (surrogate measures) identified as important indicators for certain underlying conditions, as detailed in the “Results” section, highlight noteworthy statistical associations. It's important to note that the presence of these associations alone does not

establish causality. For that, interventions are necessary, and our model offers recommended intervention strategies that show the greatest potential.

5 Conclusions

Risk stratification of hospitalized COVID-19 patients is crucial for informing individual treatment decisions that also account for resource allocation. Multiple risk models proposed so far have been based on mechanistic hypotheses regarding the SARS-CoV-2 mode of action, the association between comorbidities and observed outcomes, as well as hypothesized models of disease progression. In this report, we demonstrated the use of a machine learning-based, hypothesis-agnostic methodology for the discovery of predictive risk factors, which produces an easily understandable, observable, explainable, and actionable set of clinical features that may cause or be closely associated with in-hospital COVID-19 mortality.

This work sets the stage for future intervention-based studies aimed at addressing hypotheses that the clinical features identified here are causally linked to COVID-19 mortality. Even in the absence of established causal relationships, our findings suggest that analyses of these features should be prioritized to identify patients with COVID-19 (and, potentially, other forms of acute respiratory distress syndrome) having an elevated risk of mortality.

Supporting information

S1 Appendix. Table A. Summary of differences in the occurrence of delirium among the three age groups of patients. **Table B.** Statistical summary of the regression model shown in S1 Appendix. Equation G. **Table C.** Average marginal effect (AME), *p*-values, and confidence intervals associated with the factors incorporated in the benchmark model introduced in S1 Appendix. Equation G. The AME represents the average effect of a unit change in a predictor variable on the predicted outcome, averaged across all observations in the dataset. It provides a straightforward measure of the influence of each independent variable, revealing the expected average change in the dependent variable with a one-unit increase in that independent variable, while keeping all other variables constant. (PDF)

Author Contributions

Conceptualization: Bartek Rajwa, Babar A. Khan, M. Murat Dunder, Jean-Christophe Rochet.

Data curation: Bartek Rajwa, Md Mobasshir Arshed Naved, Mohammad Adibuzzaman, M. Murat Dunder.

Formal analysis: Bartek Rajwa, Md Mobasshir Arshed Naved, M. Murat Dunder.

Funding acquisition: Bartek Rajwa, Babar A. Khan, M. Murat Dunder, Jean-Christophe Rochet.

Investigation: Bartek Rajwa, Mohammad Adibuzzaman, Babar A. Khan, M. Murat Dunder, Jean-Christophe Rochet.

Methodology: Bartek Rajwa, Md Mobasshir Arshed Naved, Mohammad Adibuzzaman, Ananth Y. Grama, Babar A. Khan, M. Murat Dunder, Jean-Christophe Rochet.

Project administration: Babar A. Khan, Jean-Christophe Rochet.

Resources: Ananth Y. Grama, Babar A. Khan.

Software: Bartek Rajwa, Md Mobasshir Arshed Naved, Ananth Y. Grama.

Supervision: Bartek Rajwa, Jean-Christophe Rochet.

Visualization: Bartek Rajwa.

Writing – original draft: Bartek Rajwa, Md Mobasshir Arshed Naved, Babar A. Khan, Jean-Christophe Rochet.

Writing – review & editing: Bartek Rajwa, Jean-Christophe Rochet.

References

1. Zhou Y, Yang Q, Chi J, Dong B, Lv W, Shen L, et al. Comorbidities and the Risk of Severe or Fatal Outcomes Associated with Coronavirus Disease 2019: A Systematic Review and Meta-Analysis. *International Journal of Infectious Diseases*. 2020; 99:47–56. <https://doi.org/10.1016/j.ijid.2020.07.029> PMID: 32721533
2. Mesas AE, Cavero-Redondo I, Álvarez-Bueno C, Cabrera MAS, de Andrade SM, Sequí-Dominguez I, et al. Predictors of In-Hospital COVID-19 Mortality: A Comprehensive Systematic Review and Meta-Analysis Exploring Differences by Age, Sex and Health Conditions. *PLOS ONE*. 2020; 15(11): e0241742. <https://doi.org/10.1371/journal.pone.0241742> PMID: 33141836
3. Belsky JA, Tullius BP, Lamb MG, Sayegh R, Stanek JR, Auletta JJ. COVID-19 in Immunocompromised Patients: A Systematic Review of Cancer, Hematopoietic Cell and Solid Organ Transplant Patients. *Journal of Infection*. 2021; 82(3):329–338. <https://doi.org/10.1016/j.jinf.2021.01.022> PMID: 33549624
4. Li C, Liu J, Lin J, Shang H. COVID-19 and Risk of Neurodegenerative Disorders: A Mendelian Randomization Study. *Translational Psychiatry*. 2022; 12(1):1–6. <https://doi.org/10.1038/s41398-022-02052-3> PMID: 35835752
5. Frontera JA, Boutajangout A, Masurkar AV, Betensky RA, Ge Y, Vedvyas A, et al. Comparison of Serum Neurodegenerative Biomarkers among Hospitalized COVID-19 Patients versus Non-COVID Subjects with Normal Cognition, Mild Cognitive Impairment, or Alzheimer's Dementia. *Alzheimer's & Dementia*. 2022; 18(5):899–910. <https://doi.org/10.1002/alz.12556> PMID: 35023610
6. Taquet M, Geddes JR, Husain M, Luciano S, Harrison PJ. 6-Month Neurological and Psychiatric Outcomes in 236 379 Survivors of COVID-19: A Retrospective Cohort Study Using Electronic Health Records. *The Lancet Psychiatry*. 2021; 8(5):416–427. [https://doi.org/10.1016/S2215-0366\(21\)00084-5](https://doi.org/10.1016/S2215-0366(21)00084-5) PMID: 33836148
7. Farhadian SF, Seilhean D, Spudich S. Neuropathogenesis of Acute Coronavirus Disease 2019. *Current Opinion in Neurology*. 2021; 34(3):417–422. <https://doi.org/10.1097/WCO.0000000000000944> PMID: 33852526
8. Thepmankorn P, Bach J, Lasfar A, Zhao X, Souayah S, Chong ZZ, et al. Cytokine Storm Induced by SARS-CoV-2 Infection: The Spectrum of Its Neurological Manifestations. *Cytokine*. 2021; 138:155404. <https://doi.org/10.1016/j.cyto.2020.155404> PMID: 33360025
9. Zou H, Hastie T. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005; 67(2):301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
10. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD'16; 2016. p. 785–794. <https://doi.org/10.1145/2939672.2939785>
11. Shapley LS. A Value for N-Person Games. In: Roth AE, editor. *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. Cambridge: Cambridge University Press; 1988. p. 31–40. <https://doi.org/10.1017/CBO9780511528446.003>
12. Song E, Nelson BL, Staum J. Shapley Effects for Global Sensitivity Analysis: Theory and Computation. *SIAM/ASA Journal on Uncertainty Quantification*. 2016; 4(1):1060–1083. <https://doi.org/10.1137/15M1048070>
13. Owen AB, Prieur C. On Shapley Value for Measuring Importance of Dependent Inputs. *SIAM/ASA Journal on Uncertainty Quantification*. 2017; 5(1):986–1002. <https://doi.org/10.1137/16M1097717>
14. Smith M, Alvarez F. Identifying Mortality Factors from Machine Learning Using Shapley Values—a Case of COVID19. *Expert Systems with Applications*. 2021; 176:114832. <https://doi.org/10.1016/j.eswa.2021.114832> PMID: 33723478

15. Booth AL, Abels E, McCaffrey P. Development of a Prognostic Model for Mortality in COVID-19 Infection Using Machine Learning. *Modern Pathology*. 2021; 34(3):522–531. <https://doi.org/10.1038/s41379-020-00700-x> PMID: 33067522
16. Ramón A, Torres AM, Milara J, Cascón J, Blasco P, Mateo J. eXtreme Gradient Boosting-based Method to Classify Patients with COVID-19. *Journal of Investigative Medicine*. 2022; 70(7):1472–1480. <https://doi.org/10.1136/jim-2021-002278> PMID: 35850970
17. Vaid A, Somani S, Russak AJ, Freitas JKD, Chaudhry FF, Paranjpe I, et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients with COVID-19 in New York City: Model Development and Validation. *Journal of Medical Internet Research*. 2020; 22(11):e24018. <https://doi.org/10.2196/24018> PMID: 33027032
18. Bertsimas D, Lukin G, Mingardi L, Nohadani O, Orfanoudaki A, Stellato B, et al. COVID-19 Mortality Risk Assessment: An International Multi-Center Study. *PLOS ONE*. 2020; 15(12):e0243262. <https://doi.org/10.1371/journal.pone.0243262> PMID: 33296405
19. Cavallaro M, Moiz H, Keeling MJ, McCarthy ND. Contrasting Factors Associated with COVID-19-related ICU Admission and Death Outcomes in Hospitalised Patients by Means of Shapley Values. *PLOS Computational Biology*. 2021; 17(6):e1009121. <https://doi.org/10.1371/journal.pcbi.1009121> PMID: 34161326
20. Banerjee T, Paul A, Srikanth V, Strümke I. Causal Connections between Socioeconomic Disparities and COVID-19 in the USA. *Scientific Reports*. 2022; 12(1):15827. <https://doi.org/10.1038/s41598-022-18725-4> PMID: 36138106
21. Ikram M, Shaikh NF, Vishwanatha JK, Sambamoorthi U. Leading Predictors of COVID-19-related Poor Mental Health in Adult Asian Indians: An Application of Extreme Gradient Boosting and Shapley Additive Explanations. *International Journal of Environmental Research and Public Health*. 2023; 20(1):775. <https://doi.org/10.3390/ijerph20010775>
22. Bottino F, Tagliente E, Pasquini L, Napoli AD, Lucignani M, Figà-Talamanca L, et al. COVID Mortality Prediction with Machine Learning Methods: A Systematic Review and Critical Appraisal. *Journal of Personalized Medicine*. 2021; 11(9):893. <https://doi.org/10.3390/jpm11090893> PMID: 34575670
23. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*. 1970; 12(1):55–67. <https://doi.org/10.1080/00401706.1970.10488634>
24. Tibshirani R. Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996; 58(1):267–288. <http://www.jstor.org/stable/2346178>
25. Xu H, Caramanis C, Mannor S. Sparse Algorithms Are Not Stable: A No-Free-Lunch Theorem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2012; 34(1):187–193. <https://doi.org/10.1109/TPAMI.2011.177>
26. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017; 30.
27. Menardi G, Torelli N. Training and Assessing Classification Rules with Imbalanced Data. *Data Mining and Knowledge Discovery*. 2014; 28(1):92–122. <https://doi.org/10.1007/s10618-012-0295-5>
28. Puelle MR, Kosar CM, Xu G, Schmitt E, Jones RN, Marcantonio ER, et al. The Language of Delirium: Keywords for Identifying Delirium from Medical Records. *Journal of Gerontological Nursing*. 2015; 41(8):34–42. <https://doi.org/10.3928/00989134-20150723-01> PMID: 26248142
29. Inouye SK, Leo-Summers L, Zhang Y, Bogardus ST Jr, Leslie DL, Agostini JV. A Chart-Based Method for Identification of Delirium: Validation Compared with Interviewer Ratings Using the Confusion Assessment Method. *Journal of the American Geriatrics Society*. 2005; 53(2):312–318. <https://doi.org/10.1111/j.1532-5415.2005.53120.x> PMID: 15673358
30. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegel AP, Horwitz RI. Clarifying Confusion: The Confusion Assessment Method. *Annals of Internal Medicine*. 1990; 113(12):941–948. <https://doi.org/10.7326/0003-4819-113-12-941> PMID: 2240918
31. Shahrokhi M, Asuncion RMD. Neurologic Exam. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2022. PMID: 32491521
32. Bergstrom N, Braden BJ, Laguzza A, Holman V. The Braden Scale for Predicting Pressure Sore Risk. *Nursing Research*. 1987 Jul-Aug; 36(4):205–210. <https://doi.org/10.1097/00006199-198707000-00002> PMID: 3299278
33. Brown SJ. The Braden Scale: A Review of the Research Evidence. *Orthopaedic Nursing*. 2004; 23(1):30–38. <https://doi.org/10.1097/00006416-200401000-00010> PMID: 14999950
34. Alderden J, Kennerly SM, Wilson A, Dimas J, McFarland C, Yap DY, et al. Explainable Artificial Intelligence for Predicting Hospital-Acquired Pressure Injuries in COVID-19-Positive Critical Care Patients. *Computers, Informatics, Nursing*. 2022; 40(10):659–665. <https://doi.org/10.1097/CIN.0000000000000943> PMID: 36206146

35. Lovicu E, Faraone A, Fortini A. Admission Braden Scale Score as an Early Independent Predictor of In-Hospital Mortality among Inpatients with COVID-19: A Retrospective Cohort Study. *Worldviews on Evidence-Based Nursing*. 2021; 18(5):247–253. <https://doi.org/10.1111/wvn.12526> PMID: 34275200
36. Chopra V, Anand S, Hickner A, Buist M, Rogers MA, Saint S, et al. Risk of Venous Thromboembolism Associated with Peripherally Inserted Central Catheters: A Systematic Review and Meta-Analysis. *The Lancet*. 2013; 382(9889):311–325. [https://doi.org/10.1016/S0140-6736\(13\)60592-9](https://doi.org/10.1016/S0140-6736(13)60592-9) PMID: 23697825
37. Bozaan D, Skicki D, Brancaccio A, Snyder A, Friebe S, Tupps M, et al. Less Lumens-less Risk: A Pilot Intervention to Increase the Use of Single-lumen Peripherally Inserted Central Catheters. *Journal of Hospital Medicine*. 2019; 14(1):42–46. <https://doi.org/10.12788/jhm.3097> PMID: 30379146
38. Mumm JN, Osterman A, Ruzicka M, Stihl C, Vilsmaier T, Munker D, et al. Urinary Frequency as a Possibly Overlooked Symptom in COVID-19 Patients: Does SARS-CoV-2 Cause Viral Cystitis? *European Urology*. 2020; 78(4):624–628. <https://doi.org/10.1016/j.eururo.2020.05.013> PMID: 32475747
39. Swatesutipun V, Tangpaitoon T. Lower Urinary Tract Symptoms (LUTS) Related to COVID-19: Review Article. *Journal of the Medical Association of Thailand*. 2021; 104(6):1045–9. <https://doi.org/10.35755/jmedassocthai.2021.06.12671>
40. Swatesutipun V, Tangpaitoon T. The Prevalence and Risk Factors of Storage Urinary Symptoms in Symptomatic COVID-19 Patients Who Were Treated in Cohort Ward and Field Hospital. *Siriraj Medical Journal*. 2022; 74(2):134–141. <https://doi.org/10.33192/Smj.2022.17>
41. Buoite Stella A, Furlanis G, Frezza NA, Valentinotti R, Ajcevic M, Manganotti P. Autonomic Dysfunction in Post-COVID Patients with and Without Neurological Symptoms: A Prospective Multidomain Observational Study. *Journal of Neurology*. 2022; 269(2):587–596. <https://doi.org/10.1007/s00415-021-10735-y> PMID: 34386903
42. Ok F, Erdogan O, Durmus E, Carkci S, Canik A. Predictive Values of Blood Urea Nitrogen/Creatinine Ratio and Other Routine Blood Parameters on Disease Severity and Survival of COVID-19 Patients. *Journal of Medical Virology*. 2021; 93(2):786–793. <https://doi.org/10.1002/jmv.26300> PMID: 32662893
43. Cheng A, Hu L, Wang Y, Huang L, Zhao L, Zhang C, et al. Diagnostic Performance of Initial Blood Urea Nitrogen Combined with D-dimer Levels for Predicting in-Hospital Mortality in COVID-19 Patients. *International Journal of Antimicrobial Agents*. 2020; 56(3):106110. <https://doi.org/10.1016/j.ijantimicag.2020.106110> PMID: 32712332
44. Cilingir BM, Askar S, Meral A, Askar M. Can B-type Natriuretic Peptide (BNP) Levels Serve as an Early Predictor of Clinical Severity in Patients with COVID-19 Pneumonia. *Clinical laboratory*. 2022; 68(1). PMID: 35023694
45. Defining the Scientific Method. *Nature Methods*. 2009; 6(4):237–237. <https://doi.org/10.1038/nmeth0409-237>
46. Booth A, Reed AB, Ponzo S, Yassaee A, Aral M, Plans D, et al. Population Risk Factors for Severe Disease and Mortality in COVID-19: A Global Systematic Review and Meta-Analysis. *PLOS ONE*. 2021; 16(3):e0247461. <https://doi.org/10.1371/journal.pone.0247461> PMID: 33661992
47. Gold MS, Sehayek D, Gabrielli S, Zhang X, McCusker C, Ben-Shoshan M. COVID-19 and Comorbidities: A Systematic Review and Meta-Analysis. *Postgraduate Medicine*. 2020; 132(8):749–755. <https://doi.org/10.1080/00325481.2020.1786964> PMID: 32573311
48. Guan Wj, Liang Wh, Zhao Y, Liang Hr, Chen Zs, Li Ym, et al. Comorbidity and Its Impact on 1590 Patients with COVID-19 in China: A Nationwide Analysis. *European Respiratory Journal*. 2020; 55(5). <https://doi.org/10.1183/13993003.00547-2020>
49. Ng WH, Tipih T, Makoah NA, Vermeulen JG, Goedhals D, Sempa JB, et al. Comorbidities in SARS-CoV-2 Patients: A Systematic Review and Meta-Analysis. *mBio*. 2021; 12(1). <https://doi.org/10.1128/mBio.03647-20> PMID: 33563817
50. Russell CD, Lone NI, Baillie JK. Comorbidities, Multimorbidity and COVID-19. *Nature Medicine*. 2023; 29(2):334–343. <https://doi.org/10.1038/s41591-022-02156-9> PMID: 36797482
51. Sanyaolu A, Okorie C, Marinkovic A, Patidar R, Younis K, Desai P, et al. Comorbidity and Its Impact on Patients with COVID-19. *SN Comprehensive Clinical Medicine*. 2020; 2(8):1069–1076. <https://doi.org/10.1007/s42399-020-00363-4> PMID: 32838147
52. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors Associated with COVID-19-related Death Using OpenSAFELY. *Nature*. 2020; 584(7821):430–436. <https://doi.org/10.1038/s41586-020-2521-4> PMID: 32640463
53. Regenstrief Institute. COVID-19 Dashboard—Regenstrief Institute; 2023. <https://www.regenstrief.org/covid-dashboard/>.
54. Dixon BE, Grannis SJ, McAndrews C, Broyles AA, Mikels-Carrasco W, Wiensch A, et al. Leveraging Data Visualization and a Statewide Health Information Exchange to Support COVID-19 Surveillance

- and Response: Application of Public Health Informatics. *Journal of the American Medical Informatics Association*. 2021; 28(7):1363–1373. <https://doi.org/10.1093/jamia/ocab004> PMID: 33480419
55. Blumenthal D, Fowler EJ, Abrams M, Collins SR. Covid-19—Implications for the Health Care System. *New England Journal of Medicine*. 2020; 383(15):1483–1488. <https://doi.org/10.1056/NEJMs2021088> PMID: 32706956
 56. Riley B, Packer M, Gallier S, Sapey E, Atkin C. Acute, Non-COVID Related Medical Admissions during the First Wave of COVID-19: A Retrospective Comparison of Changing Patterns of Disease. *Acute medicine*. 2020; 19(4):176–182. <https://doi.org/10.52964/AMJA.0825> PMID: 33215170
 57. Stokes AC, Lundberg DJ, Elo IT, Hempstead K, Bor J, Preston SH. COVID-19 and Excess Mortality in the United States: A County-Level Analysis. *PLOS Medicine*. 2021; 18(5):e1003571. <https://doi.org/10.1371/journal.pmed.1003571> PMID: 34014945
 58. Wang H, Paulson KR, Pease SA, Watson S, Comfort H, Zheng P, et al. Estimating Excess Mortality Due to the COVID-19 Pandemic: A Systematic Analysis of COVID-19-related Mortality, 2020–21. *The Lancet*. 2022; 399(10334):1513–1536. [https://doi.org/10.1016/S0140-6736\(21\)02796-3](https://doi.org/10.1016/S0140-6736(21)02796-3)
 59. Fillmore NR, La J, Zheng C, Doron S, Do NV, Monach PA, et al. The COVID-19 Hospitalization Metric in the Pre- and Postvaccination Eras as a Measure of Pandemic Severity: A Retrospective, Nationwide Cohort Study. *Infection Control & Hospital Epidemiology*. 2022; 43(12):1767–1772. <https://doi.org/10.1017/ice.2022.13> PMID: 35012694
 60. Ma S, Tourani R. Predictive and Causal Implications of Using Shapley Value for Model Interpretation. In: *Proceedings of the 2020 KDD Workshop on Causal Discovery*. PMLR; 2020. p. 23–38. <https://proceedings.mlr.press/v127/ma20a.html>