Journal of Information and Data Management, 2025, 16:1, doi: 10.5753/jidm.2025.4678

Enhancing COVID-19 Prognosis Prediction with Machine Learning and LIME Explanation

José Solenir Lima Figuerêdo [**University of Feira de Santana** | ***jslfigueredo@ecomp.uefs.br***]
Renata Freitas Araújo-Calumby [**Feira de Santana Higher Education Unit** | ***farm.renata@hotmail.com***]
Rodrigo Tripodi Calumby [**University of Feira de Santana** | ***rtcalumby@uefs.br***]

Postgraduate Program in Computer Science, University of Feira de Santana, Av. Transnordestina, s/n, Novo Horizonte, Feira de Santana, BA, 44036-900, Brazil.

Received: 7 July 2024 • **Published:** 20 June 2025

Abstract This study evaluates machine learning methods to predict the prognosis of patients in COVID-19 context. This study evaluates machine learning methods for predicting patient prognosis in the COVID-19 context. For the best-performing algorithm, we applied LIME to assess feature contributions to each decision, providing insights to assist experts in understanding the rationale behind the model's predictions. The results indicate that the developed model accurately predicted patient prognosis, achieving an ROC-AUC = 0.8524. The results also point out a higher risk of death among patients over 60 years of age, with comorbidities, and symptoms such as dyspnea and Oxygen saturation$< 95\%$, confirming results observed in other regions of the world. The results also indicated a higher percentage of deaths among those with little or no education. The prediction explanations allowed us to understand how each feature contributes to the decision made by the model, improving its transparency. For instance, in an illustrative case, LIME demonstrated that invasive ventilatory support and an age of 61 years positively contributed to the prediction of mortality, whereas hospitalization and the patient's race (being white) were not significant predictors for this particular patient.

Keywords: COVID-19, Machine Learning, Computer Aided Prognosis, Mortality Prediction, Explainable AI.

1 Introduction

Coronavirus disease (COVID-19) is an infection caused by Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-2). SARS-CoV-2 corresponds to a binuclear virus that has a broad clinical spectrum of infection [Lu *et al.*, 2020]. When infecting a host, this agent can trigger a series of symptoms, such as fever, cough, fatigue and mild to severe respiratory complications [Yan *et al.*, 2020]. Depending on the severity of the symptoms, the infection can lead the patient to death. With the large spread of SARS-CoV-2, the World Health Organization (WHO) declared a pandemic state on March 11, 2020 [WHO, 2021]. According to recent statistics (April 2023), more than 764 million people have been infected and more than 6.9 million have died from COVID-19 worldwide. Considering South America, Brazil appears as the country with the most deaths. Although there is currently a general understanding that the pandemic is under control, uncertainty regarding new pandemics is still a constant concern. Scenarios like these pose great challenges for health systems, especially regarding the clinical decision-making process [White and Lo, 2020]. Discussing and understanding strategies that support the decision-making process about care rationing is essential, especially in a pandemic context. Eventually, a right decision can contribute to reducing the number of deaths in scenarios analogous to COVID-19.

Care rationing demands a complex screening process, which can influence, the quality of care to the lethality and mortality rates. For this, biomarkers of effective prognosis could be applied. The purpose of this screening process is to help determine patients who require immediate medical attention, based on the estimation of the associated mortality risk in a data-driven approach. Although this stratification process is not ideal, in many situations it becomes necessary, due to the scarcity of hospital resources, whether human or technical. Estimating the risk of death would allow early intervention and potentially reduce mortality, since attention would be directed to patients similarly critical but with higher chances of death. To this end, current literature has identified different clinical characteristics associated with the severity of COVID-19 infection, especially of citizens from Wuhan [Xie *et al.*, 2020; Pan *et al.*, 2020]. Therefore, these characteristics could be used to make a prognosis, especially based on Artificial Intelligence (AI) methods [Kumar *et al.*, 2020].

AI systems have been applied to assist in the diagnosis and prognostication of many conditions [Yu *et al.*, 2018; Soares *et al.*, 2021]. In general, health systems are among the most promising fields for AI applications, mainly with sophisticated methods from the subarea called Machine Learning (ML). Nevertheless, the way AI systems make decisions may not be known, given the "black box" characteristic of some methods. For many algorithms, while achieving effective results, the recommendations are not easily interpretable or explainable. It is not always clear which information or specific reasoning was used to make the decision. Quite frequently, these constraints become barriers to the a broader adoption of ML solutions [Mittelstadt *et al.*, 2019]. This becomes even more critical when it comes to healthcare systems, especially for critical decision-making, which commonly affects the lives of patients. To attenuate this problem, techniques such as LIME [Ribeiro *et al.*, 2016] can be used to provide interpretation of the model's decisions.

LIME technique can significantly enhance decision-making related to COVID-19 mortality prediction by providing interpretable insights into the factors driving the model's outputs. It allows healthcare professionals to understand which features, such as age, race, education level, comorbidities, and the need for ventilatory support, contribute most to the prediction of a patient's mortality risk. By offering local explanations for individual predictions, LIME can help tailor clinical decisions to each patient's unique profile, in a personalized way. This transparency promotes trust in the model's predictions, enabling clinicians to assess the reliability of the outcomes and align them with their medical expertise. Furthermore, LIME can highlight actionable factors, guiding interventions that may improve patient outcomes. In this way, it can bridges the gap between complex ML models and practical, evidence-based clinical decision-making.

In the context of COVID-19, recent studies have proposed the application AI-based methods, for example, patient mortality prediction [Yan *et al.*, 2020]. Nevertheless, many of these works were carried out only in the Wuhan region using a restricted set of models, and in general they did not assess the interpretability of the decisions at the local/individual level. Thus, in this work, we propose and experimentally validate a pipeline of ML models to support computer-aided prognostication of patients in a pandemic context, like COVID-19. To achieve it, multiple predictive variables from individual information are exploited, such as sex, age, symptoms, among others. In addition, to better understand the relationship between the predictive variables and the model decisions, we enhance the ML decision support system with interpretability assets based on LIME technique. This study significantly extends our previous work Figuerêdo *et al.* [2023] by improving the problem formalization, presenting a more comprehensive analysis of the related work, including a descriptive analysis of the data, as well as the odds ratio analysis for risk factors.

The remainder of this article is organized as follows: Section 2 presents the related works and Section 3 describes the experimental process. The results and discussions are presented in Section 4. Finally, Section 5 brings the conclusions and future work.

2 Related Works

Since the emergence of SARS-CoV-2 in December 2019 in Wuhan, the global research community has engaged in ongoing investigations to identify mechanisms and insights that may help mitigate the effects of COVID-19. All that effort has enabled significant advances in multiple fields in a short time. Research has been developed with multiple purposes, such as to aid in patient diagnosis, in the prediction of pandemic progress, to improve the care of critical patients, radiology-based diagnosis, among others [Islam *et al.*, 2020; Kumar *et al.*, 2020].

Specifically, in such circumstances, accurate prognostication of patients is a crucial task. Therefore, systems designed to address this need can serve as valuable allies in managing COVID-19. In fact, these modern systems can assist in decision making in healthcare units. A direct application of such systems regards supporting the process of patient prioritization according to their clinical characteristics and overall condition. With this purpose, some studies have already been conducted and are described in the following.

Yan *et al.* [2020] developed predictive models to perform COVID-19 prognosis prediction according to predictive biomarkers. To support the methods, the authors used epidemiological, demographic, clinical and laboratory data. The model discovery relied on data from 375 patients from the city of Wuhan. The predictive model was built using the XGBoost algorithm. The experimental results indicated that the model managed to select three biomarkers which were enough to predict the mortality of individual patients with an accuracy of 90%. Specifically, the most predictive biomarkers were: lactic dehydrogenase (LDH), highly sensitive C-reactive protein and lymphocytes. Using also data from patients in the city of Wuhan, Xie *et al.* [2020] conducted a retrospective study to assess the association between hypoxemia and mortality in patients with COVID-19 in a survival analysis. Numerous relevant results have been found, among them is the fact that hypoxemia is independently associated with in-hospital mortality. In addition, the researchers found that oxygen saturation values (SpO_2) greater than 90% with oxygen supplementation indicate a high probability of survival.

In order to establish a reliable nomogram to predict mortality in patients with COVID-19, Pan *et al.* [2020] developed a model using critical patient data of the Optical Valley Branch of Tongji Hospital from the Huazhong University of Science and Technology. The researchers collected data from 21 patients who died of COVID-19 between February 9 and March 10, 2020. In addition, they also selected data from 99 patients recovered in the same period. A predictive model was developed using data from these 120 patients. For validation, the researchers used an independent cohort of 84 patients. The predictive model relied on multivariate logistic regression based on: reactive protein, PaO_2/FiO_2[1], and cardiac troponin I (cTnI)[2]. The model achieved an ROC-AUC of 0.956 in the validation set.

In Souza *et al.* [2020], the authors conducted a study similar to those previously described, but using data from the state of Espírito Santo – Brazil. In addition to the geographical difference between the data used in these works, there were also differences regarding the sources of the data, such as the absence of data regarding laboratory tests, factors that are known to increase the model's effectiveness. To determine the prognosis in patients with COVID-19, the authors used numerous machine learning algorithms. For model construction purposes, data from clinical records of 13,690 patients (cases closed due to cure or death) were used. The experiments performed by the authors revealed that the out-

[1] PaO_2/FiO_2 is the ratio of arterial oxygen partial pressure (PaO_2 in mmHg) to fractional inspired oxygen (FiO_2 expressed as a fraction, not a percentage).

[2] cTnI is a cardiac regulatory protein that control the calcium mediated interaction between actin and myosin [Sharma *et al.*, 2004]

come by COVID-19 could be predicted with an a ROC-AUC of 0.92. Likewise, using data from patients in Brazil (national scale), Mattos *et al.* [2020] assessed the correlation between the manifestation of symptoms/comorbidities and the patients' survival response through Kaplan-Meier survival estimates. The authors identified that the observed comorbidities and symptoms are in accordance with the main clinical markers of the disease already reported in the literature. In addition, the authors also identified that such clinical aspects may present different distributions of comorbidities and present symptoms differently from the results reported in patients from other countries.

Although our work has a similar objective to the works already mentioned, this work presents some significant contributions and innovations. For example, the works in Yan *et al.* [2020]; Xie *et al.* [2020] were done mostly in Wuhan, China, while ours was done with data from patients in Brazil. On the other hand, the work carried out by Souza *et al.* [2020] was also conducted in Brazil, but used a limited database, containing only data from a single Brazilian state. Despite Mattos *et al.* [2020] using a more comprehensive dataset than Souza *et al.* [2020], the main objective was to perform an initial analysis of clinical factors related to admission in ICU or death of SARS-CoV-2, and not the development of predictive models from ML. In addition to these aforementioned remarks, the previous works included no explicit resources and analysis for explaining why the model made a particular decision. Differently, our work explicitly introduce an Explainable Artificial Intelligence (XAI) step to the predictions, with the objective of helping in the understanding of the decisions made by the model. Consequently, it helps the experts to comprehend the context and reliability for each prediction.

3 Methodology

The experimental process followed in this work is illustrated in Figure 1. There are four stages: data collection, preprocessing, model training (including optimization and validation) and model assessment and analysis, which includes an explainability stage of the learned models.

3.1 Dataset and preprocessing

The experiments relied on the same database used in a previous work [Figuerêdo *et al.*, 2021], i.e., the *Database for Severe Acute Respiratory Syndrome 2020* (SARS2020), available in the OpenDATASUS[3] portal. Such repository is maintained by the Ministry of Health of Brazil, through the Secretariat of Health Surveillance, which conducts the surveillance of Severe Acute Respiratory Syndrome in Brazil, since 2009. Previous to the publication in the portal, the database is submitted to preparation procedures that include anonymization procedures in compliance with current regulations. With the new Coronavirus pandemic, multiple data on COVID-19 cases were incorporated into the surveillance network and are updated on a weekly basis.

In this study, only completed cases (death or cure) were considered. Thus, data were discarded for the patients whose outcome was reported as unknown. After removing these cases, the database remained with 274,493 patient records, 164,535 of cure (59.94%) and 109,958 of death (40.06%). In addition to information regarding the evolution of the infection, the dataset also includes basic individual information, such as gender and age group, symptoms, comorbidities, among others.

The original dataset has 156 attributes. However, a preliminary analysis showed that some of these variables do not add relevant predictive content, e.g., patient name. For this reason, some variables considered non-relevant for the task were removed, resulting in a final set of 39 variables (including the outcome attribute). In addition to employing a conceptual attribute selection process that identified irrelevant features (e.g., patient name, notification date, region identifier, hospital unit identifier and so on), we also engaged the collaboration of an expert who selected the features deemed most relevant to our context. After both processes, we retained the following variables: age, gender, race, education, geographic area, dyspnea, fever, cough, Oxygen saturation $< 95\%$ (Yes, No), sore throat, respiratory discomfort, diarrhea, vomiting, other symptoms, heart disease, diabetes, neuropathy, pneumopathy, kidney disease, asthma, immunodepression, hemopathy, liver failure, down syndrome, postpartum, obesity, other comorbidities, hospitalization (Yes, No, Unknown), intensive care unit (ICU), antiviral treatment, antiviral type, severe syndrome outbreak, nosocomial[4], bird or swine contact, pregnancy, risk factor (Yes, No), Chest X-ray (Normal, Interstitial infiltrate, Consolidation, Mixed, Other, Unrealized, Unknown), ventilatory support (Yes, invasive; Yes, non-invasive; No; Unknown), and evolution (outcome).

Moreover, in the preprocessing phase, it was detected that the database had a large amount of missing data. Some attributes such as "Obesity" and "Kidney disease", for instance, had more than 60% absence of data. Thus, in order to avoid possible inconsistencies in the experiments, data standardization was performed considering missing data as non-occurrence of the event in particular (e.g., for the cough attribute, if the information was missing, it was indicated as the patient not having this symptom). With the exception of the "Age" attribute, all other variables were categorical. Thus, the "Age" attribute was discretized into the following categories: child (0-10), teenager (11-17), young adult (18-29), average adult (30-40), adult (41-59) and elderly (60 or more).

3.2 Experiments, assessment, and explanations

The database was partitioned into training and test sets. Before partitioning, a sub-sampling was applied to balance the dataset (cures and deaths). Absolute partitioning was performed through a stratified random procedure. For the test set, 30,000 records were randomly hold-out (15,000 cures and 15,000 deaths) and the remainder (202,164 samples) was retained only as the training set. The training set was used to

[3] `https://opendatasus.saude.gov.br/dataset/srag-2020`

[4] Refers to infection acquired in the hospital.

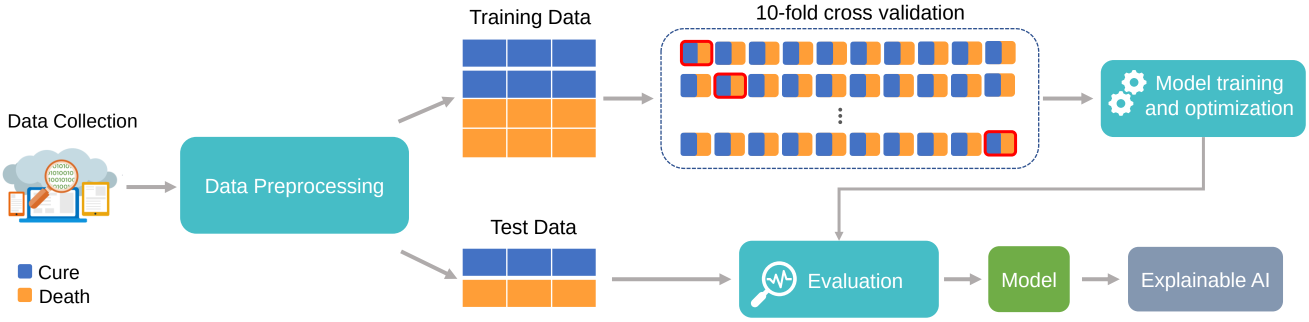

Figure 1. Experimental process used in this study

Table 1. Hyperparameters tested for each of the algorithms (except naive bayes) used in this work

Algorithm/Classifier	Hyperparameters	Tested values	Best parameter
Decision Tree	Quality Measure	Gini index, Gain Ratio	Gain Ratio
	Pruning method	Without Pruning, MDL	Without Pruning
	Minimum number of records per node	(10, 20, 30, 50, 100)	50
Gradient Boosted Trees	Learning rate	(0.5, 0.1, 0.01, 0.001)	0.1
	Number of models	(500, 700, 1000)	1000
	Tree depth	(5, 10, 20)	5
Logistic Regression	Epochs	(100, 300, 500)	300
	Learning rate	(1, 0.5, 0.1, 0.01, 0.001)	0.1
Random Forest	Minimum number of records per node	(10, 20, 30, 50, 100)	20
	Number of models	(500, 700, 1000)	1000
	Tree depth	(5, 10, 20)	20

discover the best predictive model supported by a hyperparameter optimization (Table 1) with k-fold cross-validation and stratified sampling with $k = 10$. In turn, the test set was used to verify the effectiveness and perform the explanation analysis of the models. Five algorithms were used in this study, namely: Decision Tree, Logistic Regression, Naive Bayes, Random forest and Gradient Boosted Trees. The best model was selected considering the F_1 measure.

The predictive models were assessed using the Area Under the ROC Curve (AUC). The ROC curve corresponds to a graphic technique widely used to evaluate the effectiveness of binary classifiers, based on multiple confidence thresholds. This technique relates the false-positive rate to the true-positive rate. On the other hand, the AUC provides a numerical summary for the two-dimensional area below the ROC curve. The AUC varies between 0 and 1, with 0 representing a model that provides all predictions erroneously, while 1 represents a models that provides 100% of correct predictions. The effectiveness of the developed models was also evaluated based on classical ML measures, such as Precision, Recall and F_1. While Precision quantifies the portion of samples correctly predicted as belonging to the class of interest (death), Recall quantifies the portion of samples of the class of interest the were correctly predicted as belonging to that class. Finally, the F_1 measure is taken as the harmonic mean of Precision and Recall.

Understanding how features affect the decision-making is considerably important for the confidence on the model. This may be decisive to select which model to be deployed or to support further actions based on model predictions [Ribeiro *et al.*, 2016]. When using ML, especially in the healthcare, actions may not be taken based only on predictions from a black box oracle, as the consequences can be catastrophic. Thus, in addition to the traditional effectiveness assessment, we also evaluate the models using a ML explanation technique, named LIME. LIME is a technique that aims at explaining the individual predictions of a black box model by training a local surrogate model that is easier to understand (e.g., a linear model) [Ribeiro *et al.*, 2016]. The rationale behind this approach is that a globally nonlinear model might actually be linear within a small local region of the feature space. To provide this, LIME creates a dataset of perturbed samples for a single sample of interest, predicts it with the black box model and then learns a local surrogate, which approximates the predictions of the black box model.

Figure 3 presents an example of how the explanations are generated. In short, LIME produces the explanations through five main steps:

- **Instance Selection:** First, LIME focuses on a single instance from the dataset (e.g., a patient record) for which the model has made a prediction and needs to explain the decision taken.
- **Data Perturbation:** Next, LIME generates a series of perturbations of the instance under analysis, producing synthetic data points similar to the original. These data can be generated using different approaches. In our context, this is done by slightly modifying the variable values.
- **Model Predictions:** For each new synthetic instance,

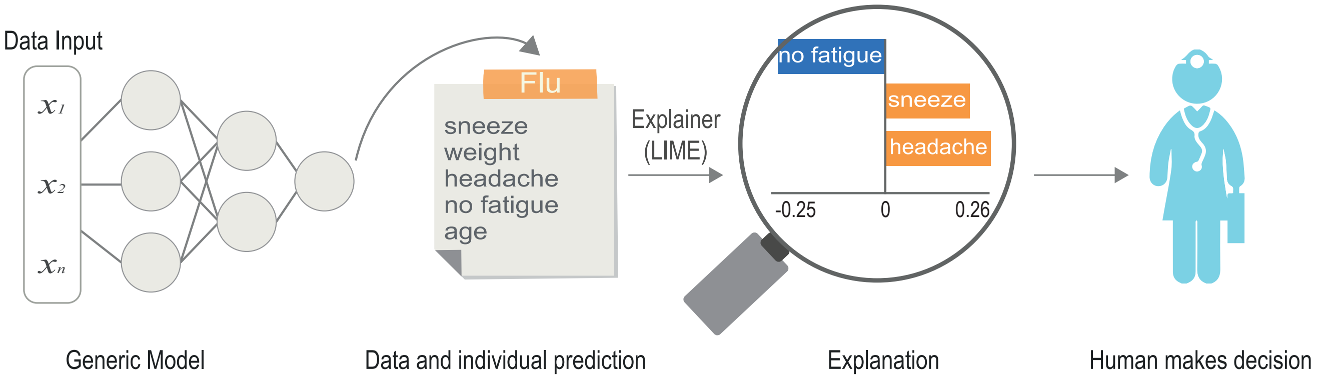

Figure 2. Explaining individual predictions.

the original model (f) is used to generate new predictions. This allows to identify how small changes in the input data affect the model's output.

- **Building a Simple Model (surrogate):** After collecting the model's predictions on the synthetic data, LIME fits a simple interpretable model (such as linear regression) around the local region of the instance. This simple model is easily interpretable and approximates the behavior of the complex model only within that neighborhood.
- **Explanations:** Based on the simple model, LIME identifies which features of the original data were most important for the prediction. It assigns weights to each feature, indicating their relevance to the model's decision.

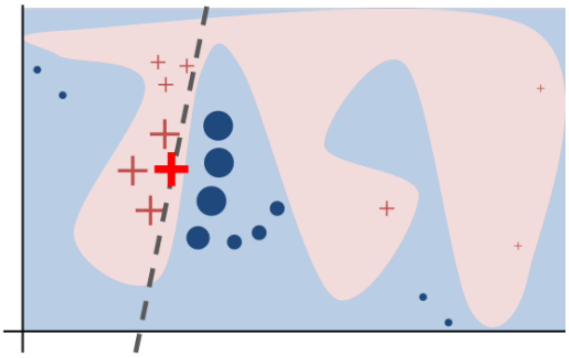

Figure 3. Didactic example to present intuition for LIME. The black-box model's complex decision function f (unknown to LIME) is represented by the blue/pink background, which cannot be approximated well by a linear model. The bold red cross is the instance being explained. LIME samples instances, gets predictions using f, and weighs them by the proximity to the instance being explained (represented here by size). The dashed line is the learned explanation that is locally (but not globally) faithful.

Figure 2 (adapted from Ribeiro *et al.* [2016].) illustrates the process of explaining individual predictions. This example shows a situation in which the model predicts that a patient has flu, and LIME highlights the symptoms in the patient's history that led to the prediction. Sneeze and headache are stated as contributing to the "flu" prediction, while "no fatigue" is evidence against it. Hence, a doctor can decide whether to trust the model's prediction [Ribeiro *et al.*, 2016]. Therefore, an explanation, as the case illustrated in Figure 5, corresponds to a small list of symptoms with relative weights that either contribute to the prediction (in orange) or are evidence against it (in blue). In this work, we considered a sample set of 8 records to evaluate the individual predictions, thus simulating the practical application of the predictive models. It is worth noting that this explanation of the model not necessarily means causality, and such investigations are out of the scope of this work.

4 Results and Discussions

The results are organized in four sections. In section 4.1 we present a descriptive analysis of the data. In turn, section 4.2 addresses an odds ratio analysis. In the section 4.3 we discuss the results achieved by the machine learning models developed. Finally, in the section 4.4, the explainability perspective is presented.

4.1 Clinical and Demographic Context

Table 2 presents the descriptive analysis for selected attributes. In general, the results found from this analysis corroborate what has been described in the literature about older age, male gender, and the presence of comorbidities as factors associated with hospitalization from COVID-19 and that can be used as potential risk indicators [Niquini *et al.*, 2020]. Additionally, the descriptive analysis revealed that symptoms related to breathing, such as dyspnea and respiratory distress, cause a higher percentage of deaths among patients who develop them.

The results of descriptive analysis from Table 2 resemble the ones in previous studies carried out in different regions of the world [Zhou *et al.*, 2020; Grasselli *et al.*, 2020; Richardson *et al.*, 2020; Onder *et al.*, 2020], which report that older age, male gender and the presence comorbidities are associated with hospitalization by COVID-19 and, therefore, can be used as potential risk factors. These findings are particularly significant because, despite Brazil's distinct geographic region, climate, and sociodemographic characteristics from those of the initial studies, the identified risk factors identified were similar.

4.2 Mortality Risk Analysis

Complementing the descriptive analysis, an estimate of the odds ratio of death was performed considering the binary at-

Table 2. Descriptive analysis. Selected variables.

	Category	All n (%)	Cure n (%)	Death n (%)
All		274493 (100)	164535 (59.9)	109958 (40.1)
Age	0-35	31186 (11.7)	27607 (88.5)	3579 (11.5)
	36-59	99811 (36.7)	75543 (75.7)	24268 (24.3)
	60 or more	**143496 (52.3)**	**61385 (42.8)**	**82111 (57.2)**
Gender	Female	120023 (43.7)	73998 (61.7)	46025 (38.4)
	Male	**154410 (56.3)**	**90499 (58.6)**	**63911 (41.4)**
	Unknown	60 (0.02)	38 (63.3)	22 (36.7)
Dyspnea	**Yes**	**189309 (69.0)**	**106204 (56.1)**	**83105 (43.9)**
	No	51998 (19.0)	38924 (74.9)	13074 (25.1)
	Unknown	3405 (1.2)	1681 (49.4)	1724 (50.6)
	Missing	29781 (10.9)	17726 (59.5)	12055 (40.5)
Fever	**Yes**	**182306 (66.4)**	**115290 (63.2)**	**67016 (36.**8)
	No	58699 (21.4)	33958 (57.9)	24741 (42.2)
	Unknown	4227 (1.5)	1654 (39.1)	2573 (60.9)
	Missing	29261 (10.7)	13633 (46.6)	15628 (53.4)
Cough	**Yes**	**198174 (72.2)**	**125367 (63.3)**	**72807 (36.7)**
	No	45881 (16.7)	25784 (56.2)	20097 (43.8)
	Unknown	3718 (1.4)	1347 (36.2)	2371 (63.8)
	Missing	26720 (9.7)	12037 (45.1)	14683 (55.0)
Saturation	**Yes**	**150698 (54.9)**	**78931 (52.4)**	**71767 (47.6)**
	No	73057 (26.6)	55368 (75.8)	17689 (24.2)
	Unknown	5817 (2.1)	2905 (49.9)	2912 (50.1)
	Missing	44921 (16.4)	27331 (60.8)	17590 (39.1)
Respiratory Discomfort	**Yes**	**151419 (55.2)**	82588 (54.5)	**68831 (45.5)**
	No	71790 (26.2)	51870 (72.3)	19920 (27.8)
	Unknown	4871 (1.8)	2549 (52.3)	2322 (47.7)
	Missing	46413 (16.9)	27528 (59.3)	18885 (40.7)
Heart disease	**Yes**	**88770 (32.3)**	**44716 (50.4)**	**44054 (49.6)**
	No	48011 (17.5)	27768 (57.8)	20243 (42.2)
	Unknown	1871 (0.7)	841 (45.0)	1030 (55.1)
	Missing	135841 (49.5)	91210 (67.1)	44631 (32.9)
Neuropathy	**Yes**	**10641 (3.9)**	**4002 (37.6)**	**6639 (62.4)**
	No	93384 (34.0)	51964 (55.7)	41420 (44.4)
	Unknown	3448 (1.3)	1509 (43.8)	1939 (56.2)
	Missing	167020 (60.9)	107060 (64.1)	59960 (35.9)
Pneumopathy	**Yes**	**9991 (3.6)**	**3974** (39.8)	**6017 (60.2)**
	No	93526 (34.1)	51819 (55.4)	41707 (44.6)
	Unknown	3519 (1.3)	1552 (44.1)	1967 (55.9)
	Missing	167457 (61.0)	107190 (64.0)	60267 (36.0)
Homeopathy	**Yes**	**2210 (0.8)**	**1072 (48.5)**	**1138 (51.5)**
	No	98698 (36.0)	53785 (54.5)	44913 (45.5)
	Unknown	3672 (1.3)	1572 (42.8)	2100 (57.2)
	Missing	169913 (61.9)	108106 (63.6)	61807 (36.4)
Obesity	**Yes**	**11385 (4.2)**	**6385 (56.1)**	**5000 (43.9)**
	No	89595 (32.6)	48652 (54.3)	40943 (45.7)
	Unknown	5588 (2.0)	2553 (45.7)	3035 (54.3)
	Missing	167925 (61.18)	106945 (63.69)	60980(36.3)

tributes (yes or no to the presence of a specific characteristic). Although age does not fall into this category, after binary discretization, it was also included in the analysis. The odds ratio was calculated using the 95% confidence interval. The results are presented in Table 3 and the attributes that indicate greater chances of death are highlighted in bold. That is, these attributes are potential risk factors and, therefore, deserve special attention. According to the data, the chance of death is greater among patients admitted to the ICU, aged 60 years or older or who used invasive ventilation support. Signs and symptoms such as saturation, dyspnea and respiratory distress were also indicated as factors that increase the chances of death. In addition, comorbidities also indicate an increased chance of death. On the other hand, symptoms such as fever, cough, sore throat and diarrhea were identified as having the least contribution to the patient's death. Peculiarly, diabetes and asthma, usually indicated as risk factors, did not indicate greater chances of death. Nevertheless, it is worth mentioning that the large amount of missing and ignored data may have influenced these findings.

In particular, considering the ICU and the ventilation support attributes, it is worth mentioning that the fact that they are among the characteristics that increase the chances of the patient dying does not necessarily mean that these factors were the cause of death. Their presence is possibly justified because the patients who require the use of these resources are typically more severe cases, for whom other risk factors may exist.

Table 3. Evaluation of odds ratio. The attributes that indicate greater chances of death are highlighted in bold.

Condition	OR	95% CI	*P*	z-score
ICU	**5.04**	(4.94 - 5.13)	< 0.0001	170.96
Age >=60	**4.76**	(4.68 - 4.84)	< 0.0001	184.75
Ventilatory Support	**4.18**	(4.09 – 4.27)	< 0.0001	126.32
Oxygen saturation	**2.85**	(2.79 - 2.90)	< 0.0001	103.97
Dyspnea	**2.33**	(2.28 - 2.38)	< 0.0001	76.06
Risk factor	**2.25**	(2.21 - 2.28)	< 0.0001	96.16
Kidney disease	**2.19**	(2.11- 2.28)	< 0.0001	38.24
Heart Disease	**2.18**	(2.13 - 2.22)	< 0.0001	71.94
Respiratory Discomfort	**2.17**	(2.13 - 2.21)	< 0.0001	79.03
Neuropathy	**2.08**	(2.00 - 2.17)	< 0.0001	34.79
Pneumopathy	**1.88**	(1.80 - 1.96)	< 0.0001	29.42
Liver Failure	**1.81**	(1.67 - 1.96)	< 0.0001	14.21
Immunodepression	**1.41**	(1.34 - 1.48)	< 0.0001	14.22
Diabetes	**1.35**	(1.32 - 1.38)	< 0.0001	26.91
Hemopathy	**1.27**	(1.17 - 1.38)	< 0.0001	5.58
Down Syndrome	**1.20**	(1.03 - 1.39)	= 0.0185	2.35
Antiviral treatment	**1.07**	(1.05 - 1.10)	< 0.0001	7.25
Obesity	0.93	(0.90 - 0.97)	< 0.0001	3.59
Fever	0.80	(0.78 - 0.81)	< 0.0001	23.36
Cough	0.75	(0.73 - 0.76)	< 0.0001	28.02
Sore throat	0.74	(0.72 - 0.76)	< 0.0001	27.63
Diarrhea	0.74	(0.72 - 0.76)	< 0.0001	24.17
Asthma	0.56	(0.52 - 0.59)	< 0.0001	22.11
Postpartum	0.30	(0.26 - 0.36)	< 0.0001	14.37
Vomiting	0.00	(0.003 - 0.004)	< 0.0001	266.54

4.3 Effectiveness of Mortality Prediction

In Figure 4, we present the contingency matrices for the developed models. Considering specifically True Positive (TP) and True Negative (FN) (in blue), which are strongly related to the main class of interest (i.e., death cases), it was observed that the models obtained from GB and RF were the

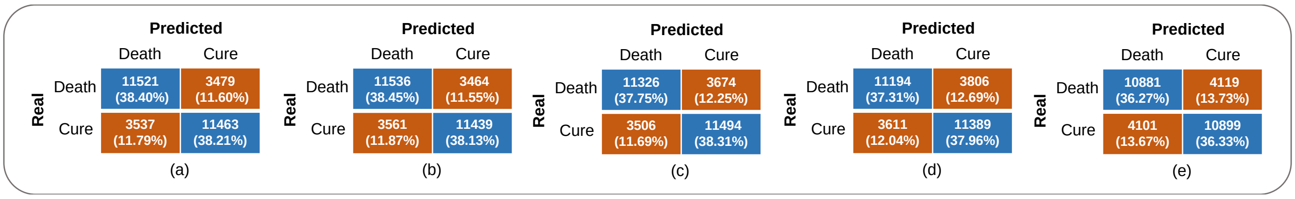

Figure 4. Contingency matrix for the models developed. (a) Gradient Boosted Trees; (b) Random Forest; (c) Logistic Regression; (d) Decision Tree; and (e) Naive Bayes.

ones that showed the best effectiveness. Considering a scenario in which a large number of patients would be evaluated by predictive models, the GB-based and RF-based would be the most suitable, as the system would make more assertive decisions regarding the identification of patients with greater chances of death, those who would therefore need intervention as soon as possible.

Still considering Figure 4, we can notice a balance between the number of False Positive (FP) and False Negative (FN) (in orange). Considering a scenario of practical use, the ideal would be the absence of errors. However, depending on the purpose of the application, some errors can have a greater negative impact than others. In our context, a FN is especially critical given it represents a patient that would not be identified as at risk and possibly not properly treated. For this scenario, the GB-based and RF-based models would be the most suitable for use, as they reached a lower FN percentage than the others. Ultimately, disregarding financial and other indirect burdens, when compared to FN, the occurrence of FP would be less harmful, considering that the patient would be directed to immediate care and submitted to further examination. However, it is important to highlight that this decision could overburden the hospital more quickly, which could deteriorate the care of patients that in fact need assistance.

Table 4 shows the effectiveness of the models developed, considering the Recall, Precision, F_1 and AUC measures. In general, the models achieved promising effectiveness, especially the GB and RF. Among the models developed, the NB achieved the worst effectiveness. It is worth mentioning that the classification process used a standard probability threshold of 0.5. The AUC values achieved (above 80% for all cases), show that the models developed were able to obtain high and promising predictive effectiveness. In summary, these results point to the possible effective use of machine learning models to face current and similar problems as those imposed by the COVID-19 pandemic.

Table 4. Prediction effectiveness of the developed models in terms of Recall, Precision and F_1.

Algorithm/Classifier	Recall	Precision	F_1	AUC
Gradient Boosted Trees	0.7661	0.7661	0.7661	0.8524
Random Forest	0.7691	0.7641	0.7666	0.8471
Logistic Regression	0.7607	0.7607	0.7607	0.8438
Decision Tree	0.7528	0.7528	0.7528	0.8382
Naíve Bayes	0.7260	0.7260	0.7260	0.8114

These results described in this paper become even more relevant, considering a high percentage of missing data had to be handled. In addition, it is worth mentioning that among the attributes used to characterize users, there was no data from clinical tests and laboratory tests. Such kind o data would possibly contribute to the process of class separation, thus improving the process of identifying cases with greater chances of death, consequently increasing the effectiveness of the models.

4.4 Underlying Mortality Factors

For model explanation, LIME was applied over the GB models, since it achieved the greatest global effectiveness in terms of AUC. In addition, among the algorithms used, the GB is also the one that most represent a “black box” algorithm. In the application of LIME, we limit the number of features to 10 in the process of explaining predictions. Although this value represents less than 50% of the total variables of the dataset, it is highly recommended to use a reduced number of features, otherwise, the explanations could be difficult to understand. Explanations of the individual predictions are illustrated in Figure 5. The explanation illustration regards 8 records selected at random, including 2 cases from each prediction assessment category, i.e., TP, TN, FN, and FP. For these cases, the explanations correspond to a list of 10 features with relative weights - features that contribute to the prediction (in orange) or are evidence against it (in blue). These explanations can be used to help an expert in the decision-making process. The specialist, with knowledge of the domain, can use the provided explanations to accept (trust) or reject a prediction by more clearly understanding the reasoning behind it.

With LIME, an individualized analysis is performed for each patient. For instance, considering TP-Case 1 in Figure 5 (a), we observe that the use of invasive ventilatory support contributed positively to the prediction of death. Likewise, the age of 61 was also positively correlated. On the other hand, the fact that the patient is hospitalized and is white is presented as not contributing to the prediction of death. It also suggests that a greater number of positive correlation features led the model to predict the case as in risk of death. Thus, when observing these data, a specialist in the field of application could perceive, for example, that although this patient is hospitalized, a number of features commonly related to the patient’s death are shown as contributing to the prediction of death. With these decision-support resources, the specialist may take the decision made by the model as reliable and more probably correct.

Figure 5 (b) illustrates the cases of true negatives. Analyzing TN-Case 2, we found that the fact the patient feels respiratory discomfort and is brown contributes positively to the prediction of death. On the other hand, the facts that the patient is 10 years old, is not in an ICU, and uses non-invasive

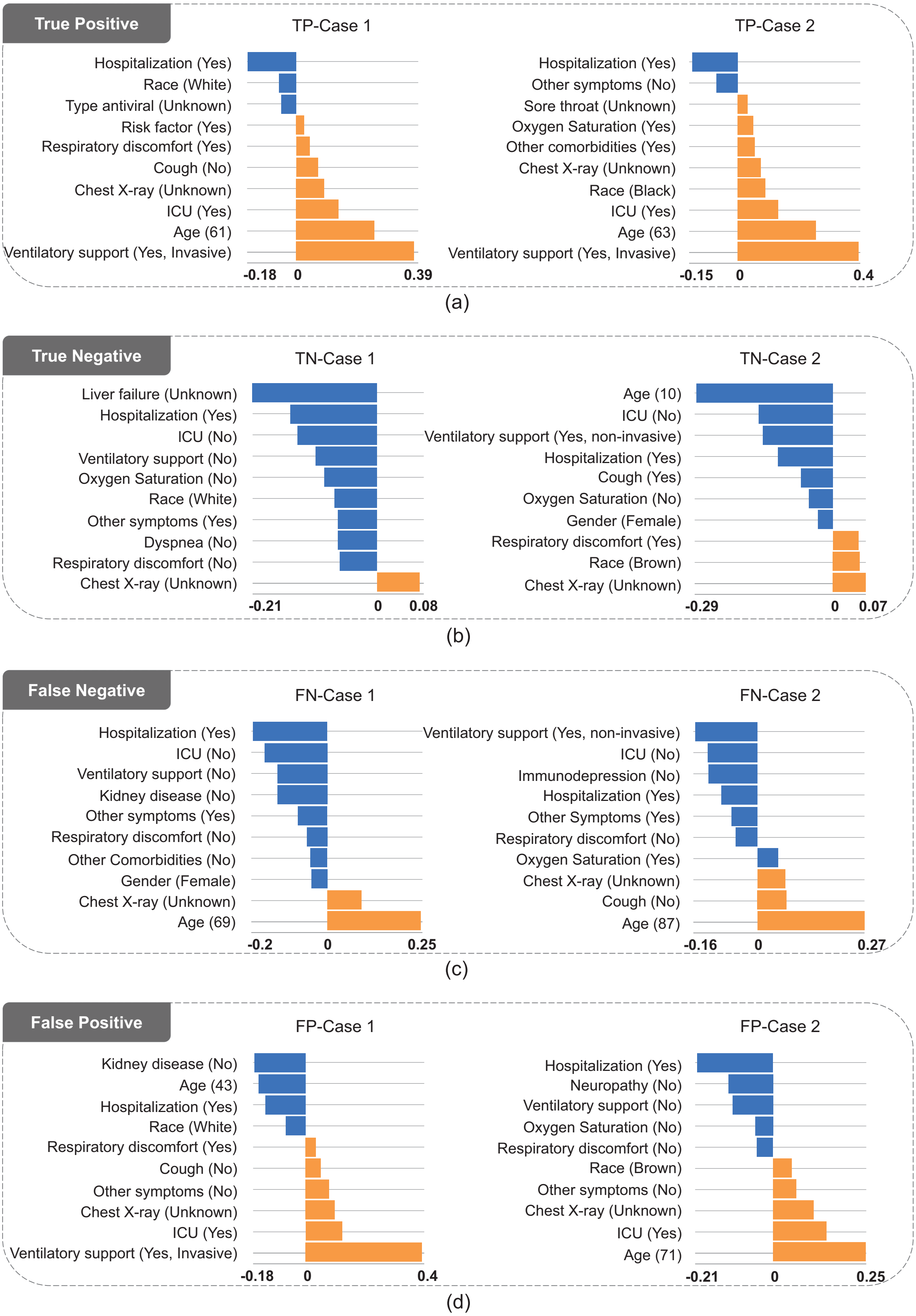

Figure 5. Explaining individual predictions. (a) TP cases; (b) TN cases; (c) FN cases; (d) FP cases.

ventilatory support are pointed out as not contributing to the prediction of death. The contributions against the prediction of death outweigh the correlated contributions. Therefore, it suggests why the system indicates the patient's cure. When confronted with these explanations, a professional with domain knowledge could verify that the decision made by the system is consistent.

The cases described above represented two situations in which the system correctly predicted the evolution of the patient's health status. Although this situation is ideal, the system is susceptible to errors, generating false negatives or false positives. Figure 5 (c) shows cases of false negatives. Many features are presented as not contributing to death, which outweighs the features that contribute to death, leading the system to mistakenly predict the patient's cure. However, it is important to carefully analyze this decision. Although in both cases the contributing features against death outweigh the features that contribute in favor, these are elderly patients. That is, these are patients belonging to the known risk group. In addition, this feature presented a reasonably high death contribution weight. Hence, the professional's knowledge is still crucial in the decision-making process. The health professional can link their knowledge to the explanations provided, to verify that the prediction is possibly inaccurate.

Figure 5 (d) presents explanations for predictions that generated false positives. Taking the prediction on FP-Case 1, notice that the contributions towards a death prediction are higher than the opposite, especially the ventilatory support attribute. However, it is worth noting that this particular patient has some features that do not contribute to the death and must be better analyzed. For example, the patient has no kidney disease, is hospitalized, and is not an elderly patient. Hence, a professional with domain knowledge would more carefully consider their decision from these data. After all, even though there are features indicating a strong relationship with death, some factors, such as age, could lead the professional to identify that prediction is possibly misleading. Ultimately, for this particular case, with hospital resources available, healthcare professionals could rely on the system's response and continue with the care for this patient.

The main risk factors associated with COVID-19 mortality identified in our results, align with the international literature [Rod *et al.*, 2020; Pijls *et al.*, 2021]. Some of these factors were also highlighted by the LIME technique. For example, LIME indicated that factors such as advanced age, the use of invasive ventilatory support, and oxygen saturation had a greater contribution to the prediction of patient mortality. While these results align with previous international studies, additional research is required to thoroughly validate these results. Despite this need for validation, our findings suggest a promising direction for applying and further investigating this approach in similar context. LIME is one of several methods to provide explanations for model decisions, thereby enhancing users' confidence in model-based predictions.

5 Conclusions

The descriptive analysis performed revealed that being of older age, male gender and the presence of comorbidities are factors that contribute to death, suggesting that these factors can be used to support decision making. In addition, considering the education characteristics of the population, it was found that the number of deaths was higher among patients with low education levels. In fact, from the patients with COVID-19 who declared themselves as illiterate, roughly 63% died. These findings highlight the impact of direct and indirect education-related factors on COVID-19 outcomes. The calculation of odds ratio confirms that the chance of death is greater among patients aged 60 years or older, and among those who have comorbidities such as kidney, neuropathy, and cardiac diseases. Some symptoms were also pointed out as factors that increase the chance of death, such as dyspnea and saturation. Some of these findings had already been pointed out as factors present in patients from other regions. Our results demonstrate that this behavior was also observed in Brazil.

The experiments carried out indicate that the model developed is capable of predicting patients' prognosis, with the model obtained with GB as the most effective. The GB model reached $ROC - AUC = 0.8524$. Using the LIME ML model explainability technique, we illustrate for a sample of patients, how each feature influences decision-making, showing whether the feature correlated negatively or positively with the prediction provided by the model. In summary, the results showed the potential of using this technique as a strategy to increase users' confidence in the models, refine the decision-making process, and increase its transparency, and, therefore, enable wider adoption. Although this work was developed considering the context of COVID-19, the procedures performed could be replicated in similar health-related contexts. In future work we intend to evaluate deep learning methods, also with a larger amount of data, as it is continuously updated. With this, it would be possible to verify, among other aspects, whether the risk factors have remained the same over time and if more accurate predictions emerge from more data or more complex models. Additionally, for a better understanding of the predictions, complementary explainability techniques may be integrated.

Acknowledgements

We thank to the University of Feira de Santana for providing the necessary infrastructure to conduct the experiments.

Funding

This work was partially supported by UEFS AUXPPG 2023 and CAPES PROAP 2023 grants.

Authors' Contributions

José Soenir Lima Figuerêdo contributed to the conceptualization, methodology, performed the experiments, results analysis, writing (original draft), data curation, investigation and writing (review & editing). **Renata Freitas Araujo-Calumby** contributed to writing (review & editing) and supervision. Finally, **Rodrigo Tripodi Calumby** contributed to the conceptualization, writing (review & editing), results analysis, investigation and supervision.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials

The datasets used in this study are available in `https://opendatasus.saude.gov.br/dataset/srag-2020`.

References

Figuerêdo, J., Araujo-Calumby, R., and Calumby, R. (2023). Towards effective and reliable data-driven prognostication: An application to covid-19. In *Anais do XI Symposium on Knowledge Discovery, Mining and Learning*, pages 81–88, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/kdmile.2023.232894.

Figuerêdo, J. *et al.* (2021). Machine learning for prognosis of patients with covid-19: An early days analysis. In *Anais do XVIII ENIAC*, pages 59–70, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/eniac.2021.18241.

Grasselli, G. *et al.* (2020). Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA*, 323(16):1574–1581. DOI: 10.1001/jama.2020.5394.

Islam, M. N. *et al.* (2020). A survey on the use of AI and ML for fighting the COVID-19 pandemic. *CoRR*, abs/2008.07449.

Kumar, A. *et al.* (2020). A review of modern technologies for tackling covid-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4):569 – 573. DOI: https://doi.org/10.1016/j.dsx.2020.05.008.

Lu, R. *et al.* (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224):565–574. DOI: 10.1016/S0140-6736(20)30251-8.

Mattos, J. *et al.* (2020). Clinical risk factors of icu & fatal covid-19 cases in brazil. In *Anais do VIII KDMiLe*, pages 33–40, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/kdmile.2020.11956.

Mittelstadt, B. *et al.* (2019). Explaining explanations in ai. In *Proceedings of the Conference on FAT*, page 279–288, New York, NY, USA. ACM. DOI: 10.1145/3287560.3287574.

Niquini, R. P. *et al.* (2020). Description and comparison of demographic characteristics and comorbidities in sari from covid-19, sari from influenza, and the brazilian general population. *CSP*, 36. DOI: 10.1590/0102-311x00149420.

Onder, G., Rezza, G., and Brusaferro, S. (2020). Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA*, 323(18):1775–1776. DOI: 10.1001/jama.2020.4683.

Pan, D. *et al.* (2020). A predicting nomogram for mortality in patients with covid-19. *Frontiers in Public Health*, 8:461. DOI: 10.3389/fpubh.2020.00461.

Pijls, B. G. *et al.* (2021). Demographic risk factors for covid-19 infection, severity, icu admission and death: a meta-analysis of 59 studies. *BMJ Open*, 11(1). DOI: 10.1136/bmjopen-2020-044640.

Ribeiro, M. T. *et al.* (2016). "why should i trust you": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD*, page 1135–1144, New York, USA. ACM. DOI: 10.1145/2939672.2939778.

Richardson, S. *et al.* (2020). Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*, 323(20):2052–2059. DOI: 10.1001/jama.2020.6775.

Rod, J., Oviedo-Trespalacios, O., and Cortes-Ramirez, J. (2020). A brief-review of the risk factors for covid-19 severity. *Revista de Saúde Pública*, 54:60. DOI: 10.11606/s1518-8787.2020054002481.

Sharma, S., Jackson, P., and Makan, J. (2004). Cardiac troponins. *JCP*, 57:1025–1026. DOI: 10.1136/jcp.2003.015420.

Soares, F. *et al.* (2021). Analysis and prediction of childhood pneumonia deaths using machine learning algorithms. In *Anais do IX KDMiLe*, pages 16–23, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/kdmile.2021.17456.

Souza, F. S. H. *et al.* (2020). Predicting the disease outcome in covid-19 positive patients through machine learning: a retrospective cohort study with brazilian data. *medRxiv*. DOI: 10.1101/2020.06.26.20140764.

White, D. B. and Lo, B. (2020). A Framework for Rationing Ventilators and Critical Care Beds During the COVID-19 Pandemic. *JAMA*, 323(18):1773–1774. DOI: 10.1001/jama.2020.5046.

WHO (2021). Coronavirus disease 2019 Situation Report. `https://covid19.who.int/`. Accessed 01 April 2021.

Xie, J. *et al.* (2020). Association Between Hypoxemia and Mortality in Patients With COVID-19. *Mayo Clinic Proceedings*, 95(6):1138–1147. DOI: 10.1016/j.mayocp.2020.04.006.

Yan, L. *et al.* (2020). An interpretable mortality prediction model for COVID-19 patients. *Nat. Mach. Intell.*, 2(5):283–288. DOI: 10.1038/s42256-020-0180-7.

Yu, K.-H. *et al.* (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10):719–731. DOI: 10.1038/s41551-018-0305-z.

Zhou, F. *et al.* (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229):1054–1062. DOI: 10.1016/S0140-6736(20)30566-3.