



BIOMEDICAL SCIENCES

Determination of prognostic markers for COVID-19 disease severity using routine blood tests and machine learning

TAYNÁ E. LIMA, MATHEUS V.F. FERRAZ, CARLOS A.A. BRITO, PAMELLA B. XIMENES, CAROLLINE A. MARIZ, CYNTHIA BRAGA, GABRIEL L. WALLAU, ISABELLE F.T. VIANA & ROBERTO D. LINS

Abstract: The need for the identification of risk factors associated to COVID-19 disease severity remains urgent. Patients' care and resource allocation can be potentially different and are defined based on the current classification of disease severity. This classification is based on the analysis of clinical parameters and routine blood tests, which are not standardized across the globe. Some laboratory test alterations have been associated to COVID-19 severity, although these data are conflicting partly due to the different methodologies used across different studies. This study aimed to construct and validate a disease severity prediction model using machine learning (ML). Seventy-two patients admitted to a Brazilian hospital and diagnosed with COVID-19 through RT-PCR and/or ELISA, and with varying degrees of disease severity, were included in the study. Their electronic medical records and the results from daily blood tests were used to develop a ML model to predict disease severity. Using the above data set, a combination of five laboratorial biomarkers was identified as accurate predictors of COVID-19 severe disease with a ROC-AUC of 0.80 ± 0.13 . Those biomarkers included prothrombin activity, ferritin, serum iron, ATTP and monocytes. The application of the devised ML model may help rationalize clinical decision and care.

Key words: COVID-19, blood tests, machine learning, disease prognosis.

INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the coronavirus disease 2019 (COVID-19), has infected more than 750 million people worldwide (WHO 2023, Wajnberg et al. 2020). Despite of the several scientific advances and the intense vaccination programs currently ongoing in several countries, new infections continue to be daily registered (Moore et al. 2022) exposing the fragility of the healthcare systems across the globe.

The spectrum of the symptomatic disease varies from mild to severe, with mortality rates

ranging from 11 to 52% among hospitalized individuals (Abate et al. 2021, Calabrese et al. 2002). In most cases, the infection produces mild symptoms, with few respiratory signs, which may evolve to pneumonia and require hospitalization, or may evolve to severe acute respiratory syndrome and require admission to the intensive care unit (Rodriguez-Morales et al. 2020, Wang et al. 2020). The classification of disease severity is mostly based on the levels of oxygen saturation and clinical criteria (WHO 2022), although the latter may vary among sites. Multicentric studies have showed that the case definitions commonly used worldwide are not accurate in identifying those individuals who are

more susceptible to severe cases (Baruch et al. 2022). Therefore, one of the biggest challenges still consists of the identification of severity predictors that are sensitive and specific enough to aid the proper early identification of patients that may progress to a severe case and their proper clinical management. Some laboratorial parameters have been reported as an attempt to identify prognostic markers of COVID-19 disease severity, although the results are not uniform among studies partly due to the differences in sample characteristics and used methodologies (Galanter et al. 2021, Rodriguez-Morales et al. 2020, Kermali et al. 2020).

A tangible solution might be the combination of these tests, along with conventional clinical data and sophisticated analytical methods from the field of Artificial Intelligence (AI). Several studies have proposed to use Machine Learning (ML) techniques to help clinicians to better classify COVID-19 suspected cases and predict disease severity (Brinati et al. 2020, Kukar et al. 2021, Kistenev et al. 2022, Alves et al. 2021, Arpaci et al. 2021, Wan et al. 2021, Imran et al. 2020, Karthikeyan et al. 2021). These studies describe the development of methods that allow computers to learn tasks by examples based on training data sets composed by routine blood tests. These tests play an important role in the diagnosis and follow-up of infected individuals, with the deviation of different laboratory parameters already shown to correlate with the COVID-19 diagnosis and disease worsening progression (Brinati et al. 2020, Fernandes et al. 2021, Kermali et al. 2020).

While numerous studies have used ML methods to identify laboratory markers associated to COVID-19 severity, most of them are based on imaging data (such as chest X-rays, computed tomography and ultrasound) (Bottino et al. 2021, Kulkarni et al. 2021, Xiao et

al. 2020, Wang et al. 2021, Zhang et al. 2020, Feng et al. 2021). The use of this type of data may be challenging in clinical practice due to the need for specialized equipment and personnel (Frija et al. 2021). The high cost of the imaging apparatus, along with its limited availability in low- and middle-income countries, poses additional challenges to the development of computational prediction methods that can be easily overcome using routine blood tests. Accordingly, other studies have also attempted to develop predictive models of COVID-19 severity using laboratory markers only (Statsenko et al. 2021, Liu et al. 2021).

Recently, Wichmann et al. (2023) (Wichmann et al. 2023) conducted a study to assess the predictive performance of the death risk due to COVID-19 using routinely collected hospital variables in the five regions of Brazil using ML. The results showed that training ML models with data from the same hospital led to improved predictive performance. This highlights the significance of taking into consideration the unique context and characteristics of patients from individual hospitals when developing health outcome prediction models. Therefore, using a cohort from a hospital in Recife, Brazil, we applied supervised ML and ensemble learning techniques to identify laboratory markers that are correlated to the severity of COVID-19 in this population. We processed a data set composed by qRT-PCR, ELISA and laboratorial data daily collected to construct and validate a disease severity ML prediction model. Our findings indicate that parameters related to coagulation and iron metabolism can be used as predictors of COVID-19 severity. The identification of laboratory markers allows to infer the prognosis of COVID-19 patients and to guide clinical decision and care.

ABBREVIATIONS

AI = Artificial intelligence
 ANOVA = Analysis of variance
 ATTP = Activated partial thromboplastin time
 AUC = Area under the curve
 CI = Confidence interval
 COVID-19 = Coronavirus disease
 CV = Coefficient of variation
 ELISA = Enzyme-linked immunosorbent assay
 Fn = False negative
 Fp = False positive
 IgG = Immunoglobulin G
 INR = International normalized ratio
 LACEN-PE = Central Laboratory of Public Health of the State of Pernambuco
 LD = Linear discriminant
 LDA = Linear discriminant analysis
 LOR = Logistic regression
 ML = Machine learning
 qRT-PCR = Quantitative reverse transcription polymerase chain reaction
 RBD = Receptor-binding domain
 RNA = Ribonucleic acid
 ROC = Receiver Operating Characteristic
 RT-PCR = Reverse transcription polymerase chain reaction
 SARS-CoV-2 = Severe acute respiratory syndrome associated coronavirus
 SDS-PAGE = Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
 SPSS = Statistical Package for the Social Sciences
 SVC = Support vector classifier
 SVM = Support vector machine
 Tn = True negative
 Tp = True positive
 WHO = World Health Organization

MATERIALS AND METHODS

Cohort characterization

A set of 847 serum samples was obtained from 100 individuals admitted to the Hospital of

Public Employees of the State of Pernambuco from May to September 2020 exhibiting respiratory symptoms. Informed consent waivers were approved by the ethics committees from the Federal University of Pernambuco (protocol number: 4.016.659) and the Oswaldo Cruz Foundation (protocol number: 2.737.404). Demographic, clinical and comorbidity data were collected upon patients' admission, while laboratory parameters were measured daily throughout the hospitalization time. Serum samples and oropharyngeal swabs collected upon hospital admission were submitted to indirect ELISA and qRT-PCR, respectively, to confirm SARS-CoV-2 infection. RT-PCR tests were run following standard protocols by the Laboratório Central de Saúde Pública de Pernambuco (LACEN-PE). To rule out the possibility of co-infection, the detection of influenza RNA through qRT-PCR was also performed in those samples. ELISA tests were performed using an in-house procedure using serum samples collected on day 3 of hospitalization, as described in the following sections. Only individuals who presented a positive result in the qRT-PCR and/or ELISA tests were included in the COVID-19 positive group. COVID-19 severity classification was performed based on the World Health Organization (WHO) criteria (WHO 2022). A sample set composed of 100 serum samples from health donors collected prior the detection of the first COVID-19 case in Brazil was included in the COVID-19 negative group.

Production of recombinant SARS-CoV-2 RBD protein and detection of anti-RBD antibodies through ELISA

The plasmid DNA coding for the SARS-CoV-2 RBD protein (reference lineage HU-1) was kindly provided by Dr. Daniel Stadlbauer and his collaborators (Icahn School of Medicine at Mount Sinai, New York, NY, USA). SARS-CoV-2

RBD production was performed by transfection of HEK-293T cells according to protocols described elsewhere (Stadlbauer et al. 2020). Protein purification was performed by affinity chromatography through isocratic elution in 300 nM imidazole, 50 nM Tris-HCl, 300 nM NaCl₂ pH 8.0. Purified samples were run on SDS-PAGE to assess protein purity. Sample concentration was determined by spectrophotometry using Nanodrop™ One (Thermo scientific®).

The purified RBD protein was used to set up an ELISA assay to detect the presence of anti-RBD IgG antibodies in the sera from the enrolled patients according to a protocol previously described (Silva et al. 2021). Pools of serum samples from individuals with and without history of SARS-CoV-2 infection, as determined through qRT-PCR and ELISA, were used as positive and negative controls, respectively, and were run in quadruplicates to ensure assay reproducibility. All samples were tested in duplicates. Diagnostic performance of the assay was evaluated through a ROC curve analysis using 100 COVID-19 negative and 53 COVID-19 positive samples. Serum samples were considered positive when the sample absorbance/negative control ratio ≥ 4.139 (corresponding to 92.45% and 97% of sensitivity and specificity, respectively). Statistical analyses were performed using the GraphPad Prism v.7 software (San Diego CA, USA).

Data set building

The REDCap dataset used in this study consisted of clinical data daily collected from 100 patients during their medical treatment. Each day of monitoring was represented by a sample, with the measured parameters serving as features. The progression of the patients' clinical condition was classified as mild or severe. In order to address issues of missing data, a subset of the REDCap dataset was created. Specifically, certain biochemical markers were missing for

some patients in the original dataset, and these missing values were not evenly distributed among patients. We decided to exclude from the analysis any days on which a patient had one or more missing values for any parameter. Therefore, only parameters from patients recently admitted to the hospital (exhibiting mild symptoms) were used, regardless of being later discharged or evolved to a severe form of the disease (including death). This is an important aspect of our dataset, as it is absent of any bias arising from data that is unique/aggravated by disease severity. As a result, data from 41 different days for 39 patients were used for analysis in this study. The complete list of features used in this study is provided in the Supplementary Material - Table SI (in .CSV file).

Libraries and tools

The Python v.3.7.9 programming language was used along with the following libraries: Pandas (McKinney 2011), NumPy (Harris et al. 2020), and scikit-learn (Pedregosa et al. 2011). The standard implementation of the libraries was applied. The IBM® SPSS® Statistics version 25 (<https://www.ibm.com/products/spss-statistics>) (IBM Corp. Released 2017. IBM SPSS Statistics for Macintosh, Version 25.0. Armonk, NY: IBM Corp.) was utilized.

Data standardization

The input data was normalized by centering each variable at zero mean and scaling it to the unit variation.

Variables' selection

Aiming to reduce the complexity of the data set while leveraging the most important features, a subset of input variables considered as the most relevant to the target response was identified through two different approaches: analysis of variance (ANOVA) and mutual information statistics. The subset of relevant

features was constructed by combining the features predicted by both methods (Table S1). These techniques are commonly used in evaluating the importance of features for datasets consisting of numerical input and categorical output, such as in classification tasks. The ANOVA method is implemented in the *f_classif()* function of the Scikit-learn library and it was combined with the *SelectKBest* class with *K*, number of selected features, set to all. The mutual information approach selects features based on information gain (reduction of entropy), a topic from information theory. The mutual information measures the decrease in uncertainty for a feature for a previously known value of the other. In the Scikit-learn library, the mutual information is applied using the *mutual_info_regression()* function. For both methodologies, the data set was split into training and test sets using a 75% and 25% split percentage, respectively. To select the optimal number (*k*) of variables, we have systematically tested a range of different *k* of selected variables by a grid search. The performance of different configurations of variables was evaluated using a repeated stratified 10-fold cross validation.

Logistic regression

Logistic regression (LOR) is a linear model of parametric classification, which computes the probabilities of a discrete outcome given an input variable using a logistic function. One of the assumptions under LOR that must be satisfied is that there should be no multicollinearity in the independent variables, which occurs when the variables are highly correlated with one another Menard 2001. To verify this, a linear regression was performed using SPSS to diagnose multicollinearity in the selected feature data set. Starting with all variables, the ones presenting multicollinearity were individually removed and these sets were used to build LOR models. These

models were then estimated and fitted using the SPSS software. The models were compared to a null model that only contained the intercept and they were evaluated using the Nagelkerke R^2 , Hosmer and Lemeshow, and Omnibus tests (Hosmer et al. 2013). The statistical significance of each variable was assessed by its p-value (<0.05 for statistical significance). Outliers were also removed based on their standardized residuals, with a total of three instances removed.

Linear discriminant analysis

The linear discriminant analysis (LDA) is a technique for classifying data by assuming that the class-conditional densities are multivariate Gaussian and generalizes the Fisher's linear discriminant (Fisher 1936). The LDA uses a homoskedastic model, where the covariance matrices are assumed to be equal, and generates linear optimal decision boundaries. For a binary classification problem, LDA aims to maximize the difference in class means while minimizing the within-class scatter in the feature space. For a sample (x) with covariance matrix Σ and class-means μ_0 and μ_1 , the LDA discriminant is calculated as:

$$D_{L,n}(x) = (\mu_1 - \mu_0)^T \Sigma^{-1} \left(x - \frac{\mu_0 + \mu_1}{2} \right) \quad (\text{Braga - Neto 2020})$$

In this study, the LDA method was implemented in Python using singular value decomposition without shrinkage. The method involved two steps: fitting and transforming the data according to the class proportions inferred from the training data. Since the classification is binary, only the first linear discriminant was calculated. The LDA values were calculated as a linear combination of the variables, where each weight is referred as a linear discriminant loading (or coefficient) and measures the influence of the variable for the classification. The values of the first linear discriminant were also used as

inputs for the support vector classifier (SVC), as described in the next section.

Support vector machine

Support vector machine (SVM) is a type of non-probabilistic and non-linear classifier and regressor that utilizes a non-linear transformation of data into high-dimensional space to separate different classes using hyperplanes. The transformed data is then retransformed back into the original space, providing non-linear models. SVM aims to find the optimal separation line between classes by identifying the hyperplanes that are maximally oriented by support vectors. For a given training data set of n points of the form $(x_1, y_1), \dots, (x_n, y_n)$, where y_i either indicates the class $\{A, B\}$ of x_i . Any hyperplane can be written as the set of points x that satisfies the following equation:

$$w^T x - b = 0$$

Here, the scikit-learn library was used to conduct support vector classification (SVC) with a linear kernel and $C=1$ and $\gamma=0.01$ as the chosen hyperparameters. These values were determined by fine-tuning the hyperparameters through a grid search across a predefined set of values. The C parameter is used to balance correct classification and maximize the decision boundary by acting as a penalty for errors, while γ controls the complexity of the SVM. The SVM models were built using the original dataset of selected features as input and using linear discriminant as the output value.

Models' performance evaluation

The accuracy of the models was determined using k -fold cross validation, where $k = 3$ and 5 . To assess the performance of the models, the accuracy was computed. In addition, the Receiver Operating Characteristic (ROC) curve was evaluated based on precision and recall.

In binary classification tasks, precision, recall, and f1-score are defined using true positive (tp), true negative (tn), false positive (fp), and false negative (fn) rates according to the following equations:

$$Precision = \frac{tp}{tp + fp}$$

$$Recall = \frac{tp}{tp + fn}$$

A detailed description of each evaluation metric can be found in Powers (2011) (Powers 2011).

RESULTS AND DISCUSSION

Serological, epidemiological and clinical cohort characterization

Determining the appropriate biomarkers for COVID-19 disease severity requires the adequate definition of the infection status of all subjects included in the study. Oropharyngeal swabs collected on the admission day from all subjects enrolled in the study were tested for the presence of SARS-CoV-2 viral RNA. Out of the 100 patients admitted to the local hospital, 64 individuals presented a positive COVID-19 qRT-PCR result. Due to the possibility of a false negative result in the molecular diagnosis due to the sample collection time, all samples were tested for the presence of anti-SARS-CoV-2 antibodies through ELISA. Among the hospitalized individuals, 72 presented a positive COVID-19 qRT-PCR and/or ELISA result and therefore met the criteria for inclusion in the study. Viral RNA and RBD-specific IgG antibodies were detected in 88.9% (64/72) and 90.3% (65/72) of those samples, respectively (Table 1). Variables comprising laboratory tests, demographics and comorbidities for each of these patients are also depicted in the table.

Although other studies have demonstrated that varying degrees of disease severity might be associated with different IgG profiles against SARS-CoV-2 (Wellinghausen et al. 2020, Fill

Table I. Clinical and epidemiological characterization of the COVID-19 positive study population.

Characteristics	N (%)
Diagnostic test	
Viral RNA detection (qRT-PCR)	64 (88.9%)
Anti-RBD IgG detection (ELISA)	65 (90.3%)
Age (years)	69 (13.3) ^a
Gender	
Female	49 (68.1%)
Male	23 (31.9%)
Comorbidities	
Hypertension	48 (66.7%)
Diabetes	25 (34.7%)
Cancer under therapy	1 (1.4%)
Asthma	9 (12.5%)
Vascular disease	3 (4.2%)
Heart disease	15 (20.8%)
COPD ^b	4 (5.6%)
None	14 (19.4%)
Occupation	
Health workers	9 (12.5%)
Education professionals	8 (11.1%)
Retirees	25 (34.7%)
Other / not informed	30 (41.7%)
COVID-19 severity	
Mild	39 (54.2%)
Severe	33 (45.8%)
Clinical outcome	
Hospital Discharge	60 (83.3%)
Death	7 (9.7%)
Transferred to another facility	1 (1.4%)
Not informed	3 (4.2%)
Hospitalization days	
2 - 5	12 (16.7%)
6 - 10	32 (44.4%)
11 - 15	12 (16.7%)
16 - 20	4 (5.6%)
> 20	10 (13.9%)
Not informed	2 (2.8%)

N = sample number; ^aStandard deviation; ^bCOPD = chronic obstructive pulmonary disease.

Malfertheiner et al. 2020, Rijkers et al. 2020, Ng et al. 2020, Guthmiller et al. 2021, Miller et al. 2020, Petersen et al. 2021, Wajnberg et al. 2020), conventional statistical analysis did not show differences in terms of seroconversion among the non-severe, mild and severe groups (Supplementary Material - Figure S1).

Variables' selection suggests the relevance of coagulation and iron-related markers

Our first attempt to identify the appropriate biomarkers for disease severity relied on descriptive statistical analysis of laboratory data, although no significant differences could be observed at first (Figure S2). Therefore, a predictive model of disease severity combining computational data analysis and human knowledge was proposed (Figure 1).

Aiming to select the adequate number of variables to build a model to accurately determine COVID-19 disease severity, two techniques were employed: ANOVA and mutual information. To identify the optimal number of variables for each method, the accuracy of different combinations of variables against a baseline model (LOR) was assessed. The two methods identified different optimal numbers of laboratory markers. While ANOVA identified two variables with the highest importance (international normalized ratio (INR) and prothrombin activity), the mutual information method selected four variables (monocytes, activated partial thromboplastin time (ATTP), serum iron, and ferritin). INR reflects the time required for blood coagulation, while prothrombin activity indicates the clotting tendency of blood. Monocytes are markers associated with chronic or sub-acute infections. ATTP is a measure of the time it takes for blood to clot using a different method other than INR. Serum iron assesses the quantity of iron present in the blood, and ferritin measures the amount of ferritin stored in the blood.

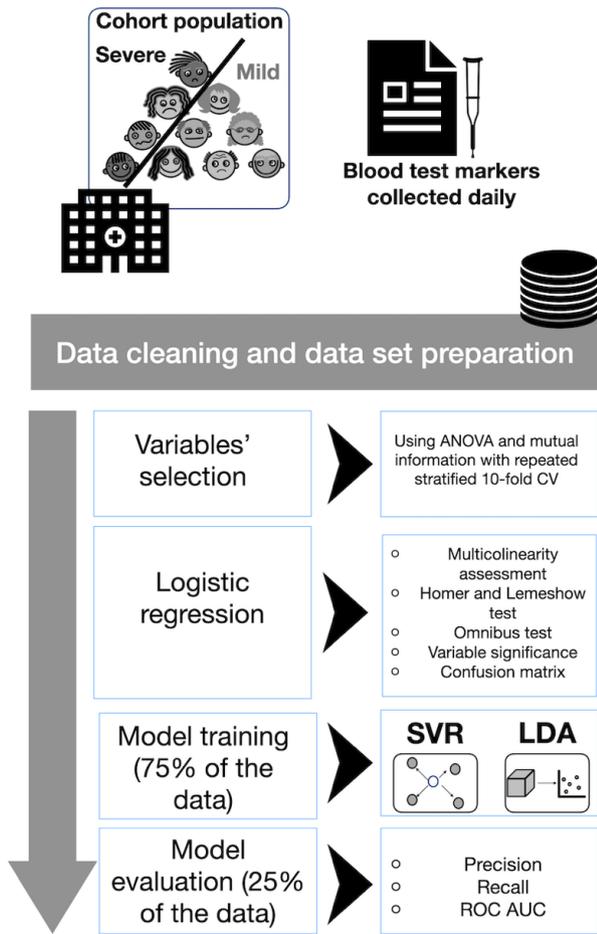


Figure 1. Workflow used for model training, selection, and testing/validation. Oropharyngeal swabs and blood samples were collected from individuals presenting clinical signs compatible with SARS-CoV-2 infection and hospitalized in a Brazilian local hospital. Infection status was confirmed through qRT-PCR and ELISA. Data from daily blood tests were used as input to train and validate the ML methods presented in this manuscript. The data was curated to select most relevant set of variables (see method section for details) and subsequently split for ML model training and validation. Data from 75% of the patients was used to train a support vector regression (SVR) and a linear discriminant analysis (LDA) models, and assessment was performed by loading the remaining 25% of patients' data onto the trained models and their accuracy measured. Precision, recall and area under the curve from receiver operating characteristic curve (ROC AUC) methods were used as quality metrics.

An increase in the ATTP and serum iron values represents an increase in the likelihood of infected patients evolve to severe cases of COVID-19

The selected variables were used as inputs for the LOR model. However, the variables INR and prothrombin activity were shown to have a high degree of correlation ($R_{\text{Pearson}} = 0.98$), resulting in multicollinearity within the input set as diagnosed by variable inflation factors greater than 10 and tolerance values smaller than 0.1 (as detailed in Table SII) (Hair et al. 1995). Therefore, two separate models were created: one including prothrombin activity, and another including INR. The LOR model with prothrombin activity had a goodness-of-fit with a Nagelkerke R^2 of 0.68, while the model with INR had a Nagelkerke R^2 of 0.67. It is important to note that for logistic regression, the explained variance (R^2) is less demanding than for a linear model. Another metric for comparison was the percentage of correct predictions in the classification table, which will be further discussed. The model with prothrombin activity had a correct prediction rate of 80%, while the model with INR had a correct prediction rate of 73%. As a result, no further analyses were conducted with the variable INR.

Two different tests were employed to evaluate the overall fit of the model. Firstly, the Hosmer and Lemeshow test was applied since this test has been shown to provide reliable results for small sample sizes (Garson 2014). Using this test, a non-significant p-value, *i.e.*, $p > 0.05$ indicates that the estimated model exhibits a superior performance when compared to the null model. The estimated model yielded a chi-square (χ^2) of 3.65 and p-value of 0.89, indicating an appropriate fit. In addition, the Omnibus test of model coefficients was also utilized. Unlike the Hosmer and Lemeshow test, a significant result indicates suitable fit for the model. The

estimated model had a χ^2 of 27.44 and p-value of 0.00. Therefore, according to both tests, the estimated model is superior to the null model, suggesting that the independent variables have an impact on the dependent variable. Analyzing the significance of each of them for the model, the variables that presented p-value < 0.05 were serum iron (p-value = 0.01), ATTP (p-value = 0.02), and prothrombin activity (p-value = 0.01). It is worth noting that the statistical insignificance of the other variables may be related to the small size of the dataset. Furthermore, the odds ratio (Table SIII) indicates that an increase of one unit in the ATTP values represents an increase of 10.4% in the likelihood of infected patients to progress to COVID-19 severe cases, while one unit increase in the serum iron marker values represents 7% higher probability of progressing to severe cases. For the remaining metrics, the odds ratios fall within the confidence intervals and have not been considered.

Lastly, the classification table, also known as the confusion matrix, was analyzed. The classification table provides a measure of the model's predictive power. The classification table uses the standard cut-off of 50% to allocate cases as severe (if the predicted probability is greater than 0.5) or mild (if it is lower than 0.5) disease. For our model, the accuracy, *i.e.*, the proportion of true positive and true negatives, is 0.8 (Table SIV). These results suggest that the logistic regression can be used as an efficient predictor of the severity outcome of SARS-CoV-2 infected patients.

Combining LDA and SVC improves the accuracy of the severity outcome prediction

To create a predictive model in addition to logistic regression, we used machine learning-based techniques. Linear discriminant analysis (LDA) classifier was applied to the selected features, excluding INR, to evaluate their ability to linearly

separate the severity of COVID-19 as severe or mild outcomes based on laboratory parameters. LDA is also commonly used for dimensionality reduction of the data. One advantage of LDA is that it allows to assess the contribution of each feature to the binary separation through LD loading (or weight). The variation in the data was concentrated in the first linear discriminant (LD) component, which constituted 100% of the data's variability in discriminating between mild and severe COVID-19 cases using the subset of the five selected variables. The weights of the LDs indicated which features mostly contributed to the separation between the classes (Figure 2a). Among the assessed variables, discrimination between mild and severe COVID-19 disease was achieved using ferritin and prothrombin activity as disease severity markers. As shown, positive LD values are predominantly associated to mild cases, while negative LD values are associated to severe cases (Figure 2b). The LD can consistently discriminate between severe and mild cases.

To develop a classificatory model between severe and mild COVID-19 cases, SVC was used since it is one of the most suitable machine learning algorithms for small data sets (Kramer et al. 2009). The model training was conducted using the standardized selected variables, as well as the value of the first LD parsed as input. A 3-fold CV (coefficient of variation) was used to diagnose the performance of the classifier and the calculated accuracy for the SVC model using the 5-dimensional input data is of 0.69 ± 0.07 , while for the model based on the LD value as input, the accuracy is of 0.76 ± 0.07 . To the latter we refer to it as LDA-SVC protocol, while the former we refer to as SVC. In addition, the ROC curves for the classification models using either SVC or LDA-SVC were plotted (Figure 3a, b). To this end, we have used different settings of the train/test split by randomly varying the random state (using 10, 15, 20, and 42) yielding

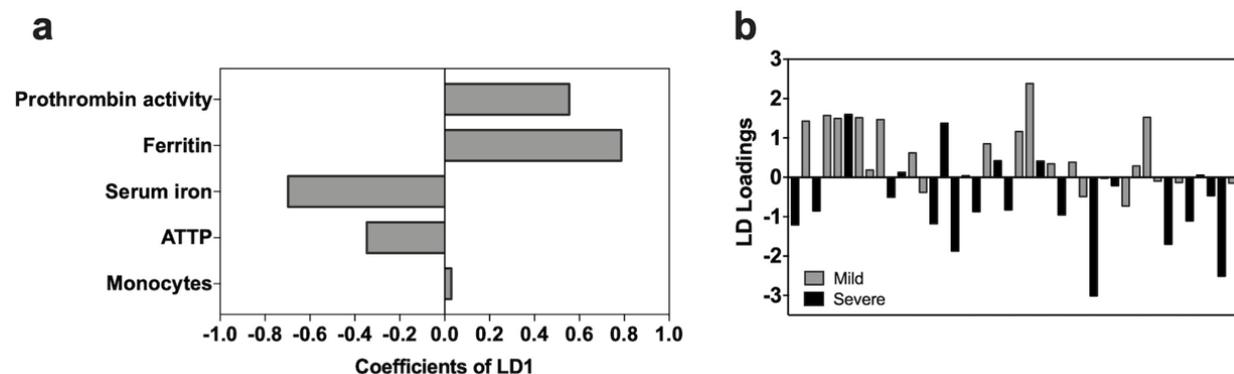


Figure 2. Linear discriminant analysis. a) Loading of each feature used to calculate de LDs. b) One dimensional LDA. The bars in black are the LD values for severe COVID-19 cases whereas the bars in gray consist of the mild ones.

areas under curve between 0.25 and 0.60 for SVC and 0.79 and 0.85 for LDA-SVC (Figure 3b). As a further evaluation step of the LDA-SVC model, we have calculated the ROC curve considering the confidence interval for 3 and 5 cross validations, resulting in an AUC of 0.77 ± 0.06 and 0.76 ± 0.08 (Figure S3), respectively, which is in line with what was observed for the ROC curve within the train-split method. Therefore, the LDA-SVC model showed to be a consistent predictor for severity outcome.

A recent study using a set of laboratory markers for 10 937 patients from 9 laboratories across the US and Spain has evaluated the prediction of severity and mortality of COVID-19 patients using deep learning (Singh et al. 2021). In this work, the developed model presented an AUC of 0.78 for the need to use a ventilator (*i.e.*, severe cases). Their AUC was similar to what our model exhibited, even though our built model presented a limited number of patients compared to the study by Singh *et al.* Interestingly, their final model presented nine blood biomarkers associated to early independent predictors, which contains ferritin and coagulation-related parameters (in their case, d-dimer and INR, in which the latter was also predicted to be relevant from our variable selection analysis), in accordance with our results. It is well established that some infected patients with COVID-19

develop a unique coagulopathy characterized by systemic hypercoagulability, leading to altered coagulation-related parameters (Wool & Miller 2021). Coagulation dysfunction is more common in patients with severe COVID-19 and has been used as a predictor of mortality (Long et al. 2020).

Our findings are also supported by several studies that have found that higher levels of serum ferritin are associated with hospital mortality and disease severity in COVID-19 patients (Alroomi et al. 2021, Dahan et al. 2020, Zhou et al. 2020, Taneri et al. 2020, Gandini et al. 2020, Kaushal et al. 2022). Ferritin is produced following the cytokine storm and the release of IL-6 and TNF- α (Guan et al. 2020, Goyal et al. 2020, Mahroum et al. 2022), and has been associated to an increase of the iron sequestration to the inside of the cell, low serum levels, decrease in hemoglobin and consequent hypoxia. Besides, intracellular iron leads to the formation and release of oxygen reactive species that may lead to cellular injury (Wenzhong & Hualan 2021, Taneri et al. 2020, Engin et al. 2022). In a retrospective cohort involving 50 COVID-19 patients, ferritin levels above 162 ng/mL were associated to the development of severe cases with 86.9% of sensitivity and 70.3% of specificity (Zhou et al. 2020). In a meta-analysis study of 29 grouped studies involving 13,620 individuals, serum ferritin was higher in individuals with

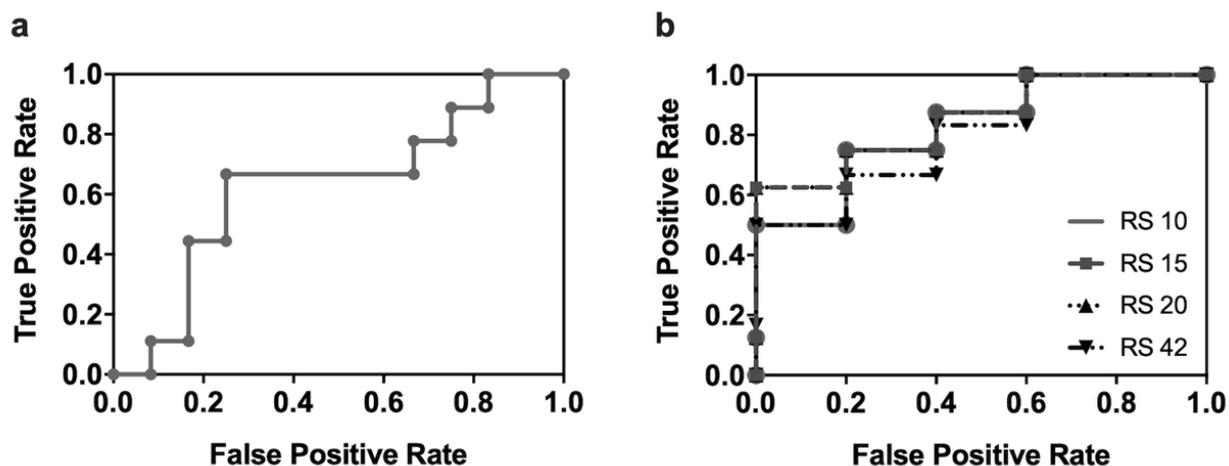


Figure 3. Assessment of the classification models' performance through their characteristic ROC curves. ROC curves for SVC (a) and LDA-SVC (b) models. Curves using four different random states (RS) are shown for the LDA-SVC model. RS values of 10, 15, 20 and 42 provide AUC values of 0.82, 0.85, 0.82 and 0.79, respectively.

severe COVID-19 disease, when compared to mild cases [weighted median deviation, 473.25 ng/mL (95% CI 382.52 – 563.98); I₂ = 91.8%, p value for heterogeneity < 0.001]. High levels of ferritin were also observed in the non-survival group, compared to the survival group with a difference of medium levels of ferritin of 606.37 ng/mL (CI 95% 461.86 to 750.88) (Taneri et al. 2020). Another study evaluating 158 COVID-19 patients found an association between iron homeostasis disturb and severe cases. Those included increased ferritin levels and low serum levels of iron, transferrin and iron-binding capacity. Higher levels of ferritin have also been associated to lesions in multiple organs, including SRAG, coagulopathy, cardiac injury, acute hepatic lesion, sepsis, UCI admission, use of mechanical ventilator and death ($p < 0.005$) (Lv et al. 2021). However, a different study by Carubbi et al. (2021) (Carubbi et al. 2021) found that while ferritin is associated with the severity of lung involvement in COVID-19 patients, it is not associated with disease outcomes. These findings suggest that, while serum ferritin level is an important descriptor to be considered when predicting the severity of COVID-19 and

the likelihood of hospitalization or death, it cannot be used *per se* as predictor of the COVID-19 severity. This is supported by our results, as standard statistical analyses could not associate a single descriptor to prognosis of disease severity. Disease outcome could only be predicted by a combination of five descriptors as depicted by the ML model.

In summary, our results show that routinely used laboratory results have a synergistic effect in predicting COVID-19 severity that is captured by machine learning approaches. As demonstrated by the previous metanalysis approach, the use of a large data set leads to an increased model accuracy. Nevertheless, we show that when dealing with reasonably small datasets, adequate prediction accuracy can be obtained by combining LDA and SVC. Even though COVID-19 is no longer considered a global emergence, the present work provides fundamental knowledge regarding the development of a machine learning model that could potentially be used to predict not only the clinical outcome of COVID-19 patients, but that could also be applied to virtually any other disease settings.

Acknowledgments

The authors thank the following research financial support agencies: Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco - FACEPE (grant numbers: APQ-0346-2-09-19 to RDL, BFP-0010-2.11/22 to IFTV and IBPG-0549-2-10/22 to TEL); Conselho Nacional de Desenvolvimento Científico e Tecnológico of Brazil - CNPq (grant numbers: 400738/2019-8, 303001/2018-6, 425997/2018-9 and INCT-FCx to RDL; grant number 142297/2019-4 to MVFF); CAPES-DAAD (MVFF, grant number: 91819540); Programa de Inovação da Fundação Oswaldo Cruz (INOVA, grant numbers: VPPCB-005-FIO-20-2-87 to RDL and VPPCB-007-18-2-134 to IFTV and RDL); and Rede de Monitoramento Genômico da Fiocruz (GLW and IFTV, grant number: VPGDI-050-FIO-20-2-13-36). Computer allocation was partly granted by the Brazilian National Scientific Computing Center (LNCC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- ABATE SM, CHECKOL YA & MANTEFARDO B. 2021. Global prevalence and determinants of mortality among patients with COVID-19: A systematic review and meta-analysis. *Ann Med Surg (Lond)* 64: 102204.
- ALROOMI M ET AL. 2021. Ferritin level: A predictor of severity and mortality in hospitalized COVID-19 patients. *Immun Inflamm Dis* 9: 1648-1655.
- ALVES MA, CASTRO GZ, OLIVEIRA BAS, FERREIRA LA, RAMIREZ JA, SILVA R & GUIMARAES FG. 2021. Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. *Comput Biol Med* 132: 104335.
- ARPACI I, HUANG S, AL-EMRAN M, AL-KABI MN & PENG M. 2021. Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms. *Multimed Tools Appl* 80: 11943-11957.
- BARUCH J ET AL. 2022. Symptom-based case definitions for COVID-19: Time and geographical variations for detection at hospital admission among 260,000 patients. *Influenza Other Respir Viruses* 16: 1040-1050.
- BOTTINO F, TAGLIENTE E, PASQUINI L, NAPOLI AD, LUCIGNANI M, FIGA-TALAMANCA L & NAPOLITANO A. 2021. COVID Mortality Prediction with Machine Learning Methods: A Systematic Review and Critical Appraisal. *J Pers Med* 11: 893.
- BRAGA-NETO U. 2020. Fundamentals of Pattern Recognition and Machine Learning. Springer, Cham, 351 p.
- BRINATI D, CAMPAGNER A, FERRARI D, LOCATELLI M, BANFI G & CABITZA F. 2020. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *J Med Syst* 44: 135.
- CALABRESE LH ET AL. 2002. Placebo-controlled trial of cyclosporin-A in HIV-1 disease: implications for solid organ transplantation. *J Acquir Immune Defic Syndr* 29: 356-362.
- CARUBBI F ET AL. 2021. Ferritin is associated with the severity of lung involvement but not with worse prognosis in patients with COVID-19: data from two Italian COVID-19 units. *Sci Rep* 11: 4863.
- DAHAN S, SEGAL G, KATZ I, HELLOU T, TIETEL M, BRYK G, AMITAL H, SHOENFELD Y & DAGAN A. 2020. Ferritin as a Marker of Severity in COVID-19 Patients: A Fatal Correlation. *Isr Med Assoc J* 22: 494-500.
- ENGIN AB, ENGIN ED & ENGIN A. 2022. Can iron, zinc, copper and selenium status be a prognostic determinant in COVID-19 patients? *Environ Toxicol Pharmacol* 95: 103937.
- FENG Y-Z ET AL. 2021. Severity Assessment and Progression Prediction of COVID-19 Patients Based on the LesionEncoder Framework and Chest CT. *Information* 12: 471.
- FERNANDES FT, DE OLIVEIRA TA, TEIXEIRA CE, BATISTA AFM, DALLA COSTA G & CHIAVEGATTO FILHO ADP. 2021. A multipurpose machine learning approach to predict COVID-19 negative prognosis in Sao Paulo, Brazil. *Sci Rep* 11: 3343.
- FILL MALFERTHEINER S ET AL. 2020. Immune response to SARS-CoV-2 in health care workers following a COVID-19 outbreak: A prospective longitudinal study. *J Clin Virol* 130: 104575.
- FISHER RA. 1936. The use of multiple measurements in taxonomic problems. *Annals of eugenics* 7: 179-188.
- FRIJA G, BLAŽIĆ I, FRUSH DP, HIERATH M, KAWOYOA M, DONOSOBACH L & BRKLJAČIĆ B. 2021. How to improve access to medical imaging in low- and middle-income countries? *eClinicalMedicine* 38: 101034.
- GALANTER W, RODRIGUEZ-FERNANDEZ JM, CHOW K, HARFORD S, KOCHENDORFER KM, PISHGAR M, THEIS J, ZULUETA J & DARABI H. 2021. Predicting clinical outcomes among hospitalized COVID-19 patients using both local and published models. *BMC Med Inform Decis Mak* 21: 224.
- GANDINI O, CRINITI A, BALLELIO L, GIGLIO S, GALARDO G, GIANNI W, SANTORO L, ANGELONI A & LUBRANO C. 2020. Serum Ferritin is an independent risk factor for Acute Respiratory Distress Syndrome in COVID-19. *J Infect* 81: 979-997.

- GARSON GD. 2014. Logistic Regression: Binary & Multinomial. Statistical Associates Publishing, Asheboro: 224 p.
- GOYAL P ET AL. 2020. Clinical Characteristics of Covid-19 in New York City. *The New England journal of medicine* 382: 2372-2374.
- GUAN WJ ET AL. 2020. Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England journal of medicine* 382: 1708-1720.
- GUTHMILLER JJ ET AL. 2021. SARS-CoV-2 Infection Severity Is Linked to Superior Humoral Immunity against the Spike. *mBio* 12(1): e02940-20.
- HAIR JFA, ANDERSON RE, TATHAM RL & BLACK WC. 1995. *Multivariate data analysis*. 7th ed, New York, 761 p.
- HARRIS CR, MILLMAN KJ, VAN DER WALT SJ, GOMMERS R, VIRTANEN P, COURNAPEAU D, WIESER E, TAYLOR J, BERG S & SMITH NJ. 2020. *Array programming with NumPy*. *Nature* 585: 357-362.
- HOSMER DW, LEMESHOW S & STURDIVANT RX. 2013. *Applied Logistic Regression*. 3rd ed., Wiley, New York: 528 p.
- IMRAN A, POSOKHOVA I, QURESHI HN, MASOOD U, RIAZ MS, ALI K, JOHN CN, HUSSAIN MI & NABEEL M. 2020. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Inform Med Unlocked* 20: 100378.
- KARTHIKEYAN A, GARG A, VINOD PK & PRIYAKUMAR UD. 2021. Machine Learning Based Clinical Decision Support System for Early COVID-19 Mortality Prediction. *Front Public Health* 9: 626697.
- KAUSHAL K ET AL. 2022. Serum ferritin as a predictive biomarker in COVID-19. A systematic review, meta-analysis and meta-regression analysis. *J Crit Care* 67: 172-181.
- KERMALI M, KHALSA RK, PILLAI K, ISMAIL Z & HARKY A. 2020. The role of biomarkers in diagnosis of COVID-19 - A systematic review. *Life Sci* 254: 117788.
- KISTENEV YV, VRAZHNOV DA, SHNAIDER EE & ZUHAYRI H. 2022. Predictive models for COVID-19 detection using routine blood tests and machine learning. *Heliyon* 8: e11185.
- KRAMER KA, HALL LO, GOLDFGOF DB, REMSEN A & LUO T. 2009. Fast support vector machines for continuous data. *IEEE Trans Syst Man Cybern B Cybern* 39: 989-1001.
- KUKAR M, GUNCAR G, VOVKO T, PODNAR S, CERNELC P, BRVAR M, ZALAZNIK M, NOTAR M, MOSKON S & NOTAR M. 2021. COVID-19 diagnosis by routine blood tests using machine learning. *Sci Rep* 11: 10738.
- KULKARNI AR ET AL. 2021. Deep learning model to predict the need for mechanical ventilation using chest X-ray images in hospitalised patients with COVID-19. *BMJ Innov* 7: 261-270.
- LIU C ET AL. 2021. Laboratory Testing Implications of Risk-Stratification and Management of COVID-19 Patients. *Front Med (Lausanne)* 8: 699706.
- LONG H, NIE L, XIANG X, LI H, ZHANG X, FU X, REN H, LIU W, WANG Q & WU Q. 2020. D-Dimer and Prothrombin Time Are the Significant Indicators of Severe COVID-19 and Poor Prognosis. *Biomed Res Int* 2020: 6159720.
- LV Y, CHEN L, LIANG X, LIU X, GAO M, WANG Q, WEI Q & LIU L. 2021. Association between iron status and the risk of adverse outcomes in COVID-19. *Clin Nutr* 40: 3462-3469.
- MAHROUM N, ALGHORY A, KIYAK Z, ALWANI A, SEIDA R, ALRAIS M & SHOENFELD Y. 2022. Ferritin - from iron, through inflammation and autoimmunity, to COVID-19. *J Autoimmun* 126: 102778.
- MCKINNEY W. 2011. *pandas: a foundational Python library for data analysis and statistics*. *Python for High Performance and Scientific Computing* 14: 1-9.
- MENARD S. 2001. *Applied logistic regression analysis*. 2nd edition ed: SAGE Publications, Inc.
- MILLERTE ET AL. 2020. Clinical sensitivity and interpretation of PCR and serological COVID-19 diagnostics for patients presenting to the hospital. *FASEB J* 34: 13877-13884.
- MOORE S, HILL EM, DYSON L, TILDESLEY MJ & KEELING MJ. 2022. Retrospectively modeling the effects of increased global vaccine sharing on the COVID-19 pandemic. *Nature Medicine* 28: 2416-2423.
- NG DL ET AL. 2020. SARS-CoV-2 seroprevalence and neutralizing activity in donor and patient blood. *Nat Commun* 11: 4698.
- PEDREGOSA F, VAROQUAUX G, GRAMFORT A, MICHEL V, THIRION B, GRISEL O, BLONDEL M, PRETTENHOFER P, WEISS R & DUBOURG V. 2011. *Scikit-learn: Machine learning in Python*. *J Mach Learn Res* 12: 2825-2830.
- PETERSEN LR ET AL. 2021. Lack of Antibodies to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in a Large Cohort of Previously Infected Persons. *Clin Infect Dis* 73: e3066-e3073.
- POWERS DM. 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol* 1: 37-63.
- RIJKERS G, MURK JL, WINTERMANS B, VAN LOOY B, VAN DEN BERGE M, VEENEMANS J, STOHR J, REUSKEN C, VAN DER POL P & REIMERINK J. 2020. Differences in Antibody Kinetics and Functionality Between Severe and Mild Severe Acute

Respiratory Syndrome Coronavirus 2 Infections. *J Infect Dis* 222: 1265-1269.

RODRIGUEZ-MORALES AJ ET AL. 2020. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Med Infect Dis* 34: 101623.

SILVA LTD ET AL. 2021. SARS-CoV-2 recombinant proteins stimulate distinct cellular and humoral immune response profiles in samples from COVID-19 convalescent patients. *Clinics (São Paulo)* 76: e3548.

SINGHV ET AL. 2021. A deep learning approach for predicting severity of COVID-19 patients using a parsimonious set of laboratory markers. *iScience* 24: 103523.

STADLBAUER D ET AL. 2020. SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup. *Curr Protoc Microbiol* 57: e100.

STATSENKO Y, AL ZAHMI F, HABUZA T, GORKOM KN & ZAKI N. 2021. Prediction of COVID-19 severity using laboratory findings on admission: informative values, thresholds, ML model performance. *BMJ Open* 11: e044500.

TANERI PE ET AL. 2020. Anemia and iron metabolism in COVID-19: a systematic review and meta-analysis. *Eur J Epidemiol* 35: 763-773.

WAJNBERG A ET AL. 2020. Humoral response and PCR positivity in patients with COVID-19 in the New York City region, USA: an observational study. *Lancet Microbe* 1: e283-e289.

WAN Y, ZHOU H & ZHANG X. 2021. An Interpretation Architecture for Deep Learning Models with the Application of COVID-19 Diagnosis. *Entropy (Basel)* 23: 204.

WANG D ET AL. 2020. [Clinical analysis of 31 cases of 2019 novel coronavirus infection in children from six provinces (autonomous region) of northern China]. *Zhonghua Er Ke Za Zhi* 58: 269-274.

WANG S ET AL. 2021. A Deep Learning Radiomics Model to Identify Poor Outcome in COVID-19 Patients With Underlying Health Conditions: A Multicenter Study. *IEEE J Biomed Health Inform* 25: 2353-2362.

WELLINGHAUSEN N, VOSS M, IVANOVA R & DEININGER S. 2020. Evaluation of the SARS-CoV-2-IgG response in outpatients by five commercial immunoassays. *GMS Infect Dis* 8: Doc22.

WENZHONG L & HUALAN L. 2021. COVID-19: captures iron and generates reactive oxygen species to damage the human immune system. *Autoimmunity* 54: 213-224.

WICHMANN RM ET AL. 2023. Improving the performance of machine learning algorithms for health outcomes predictions in multicentric cohorts. *Sci Rep* 13: 1022.

WOOL GD & MILLER JL. 2021. The Impact of COVID-19 Disease on Platelets and Coagulation. *Pathobiology* 88: 15-27.

WORLD HEALTH ORGANIZATION – WHO. 2022. Clinical management of COVID-19: Living guideline. Available: <https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2023.2>.

WORLD HEALTH ORGANIZATION – WHO. 2023. WHO Coronavirus (COVID-19) Dashboard [Online]. Available: <https://covid19.who.int> [2023].

XIAO LS. ET AL. 2020. Development and Validation of a Deep Learning-Based Model Using Computed Tomography Imaging for Predicting Disease Severity of Coronavirus Disease 2019. *Front Bioeng Biotechnol* 8: 898.

ZHANG K ET AL. 2020. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* 182: 1360.

ZHOU C, CHEN Y, JI Y, HE X & XUE D. 2020. Increased Serum Levels of Hcpidin and Ferritin Are Associated with Severity of COVID-19. *Med Sci Monit* 26: e926178.

SUPPLEMENTARY MATERIAL

Figures S1-S3.
Tables SI-SIV.

How to cite

LIMA TE, FERRAZ MVF, BRITO CAA, XIMENES PB, MARIZ CA, BRAGA C, WALLAU GL, VIANA IFT & LINS RD. 2024. Determination of prognostic markers for COVID-19 disease severity using routine blood tests and machine learning. *An Acad Bras Cienc* 96: e20230894. DOI 10.1590/0001-3765202420230894.

Manuscript received on August 09, 2023;
accepted for publication on February 22, 2024

TAYNÁ E. LIMA¹

<https://orcid.org/0000-0002-4609-1433>

MATHEUS V.F. FERRAZ^{1,2}

<https://orcid.org/0000-0002-6958-3115>

CARLOS A.A. BRITO³

<https://orcid.org/0000-0002-5963-8178>

PAMELLA B. XIMENES⁴

<https://orcid.org/0009-0007-3181-0566>

CAROLLINE A. MARIZ⁵

<https://orcid.org/0000-0002-7651-9429>

CYNTHIA BRAGA⁵

<https://orcid.org/0000-0002-7862-6455>

GABRIEL L. WALLAU⁶

<https://orcid.org/0000-0002-1419-5713>

ISABELLE F.T. VIANA¹

<https://orcid.org/0000-0003-4648-6635>

ROBERTO D. LINS¹

<https://orcid.org/0000-0002-3983-8025>

¹Fundação Oswaldo Cruz, Instituto Aggeu Magalhães, Departamento de Virologia, Av. Professor Moraes Rego, s/n, Cidade Universitária, 50740-465 Recife, PE, Brazil

²Universidade Federal de Pernambuco, Departamento de Química Fundamental, Av. Professor Moraes Rego, s/n, Cidade Universitária, 50740-560 Recife, PE, Brazil

³Universidade Federal de Pernambuco, Hospital das Clínicas, Av. Professor Moraes Rego, 1235, Cidade Universitária, 50670-901 Recife, PE, Brazil

⁴Hospital dos Servidores Públicos do Estado de Pernambuco, Av. Conselheiro Rosa e Silva, s/n, Espinheiro, 52020-020 Recife, PE, Brazil

⁵Fundação Oswaldo Cruz, Instituto Aggeu Magalhães, Departamento de Parasitologia, Av. Professor Moraes Rego, s/n, Cidade Universitária, 50740-465 Recife, PE, Brazil

⁶Fundação Oswaldo Cruz, Instituto Aggeu Magalhães, Departamento de Entomologia, Av. Professor Moraes Rego, s/n, Cidade Universitária, 50740-465 Recife, PE, Brazil

Correspondence to: **Roberto Dias Lins**

E-mail: roberto.neto@fiocruz.br

Author contributions

IFTV, RDL and CAAB conceived the study. IFTV, RDL, CAAB, TEL and MVFF wrote the manuscript. CAAB and PBX provided the clinical samples. TEL conducted the experiments. MVFF conducted the computational analysis. IFTV, RDL, CAAB, CAM, CB and GLW reviewed the paper. RDL, CAAB and CB acquired funding. All authors read and approved the final manuscript.

