A Machine-learning Parsimonious Multivariable Predictive Model of Mortality Risk in Patients With Covid-19

Rita Murri (rita.murri@policlinicogemelli.it)
Fondazione Policlinico Universitario A. Gemelli IRCCS

Jacopo Lenkowicz
Fondazione Policlinico Universitario A. Gemelli IRCCS

Carlotta Masciocchi
Fondazione Policlinico Universitario A. Gemelli IRCCS

Chiara Iacomini
Fondazione Policlinico Universitario A. Gemelli IRCCS

Massimo Fantoni
Fondazione Policlinico Universitario A. Gemelli IRCCS

Andrea Damiani
Catholic University of the Sacred Heart

Antonio Marchetti
Datawarehouse, Fondazione Policlinico Universitario A. Gemelli IRCCS

Paolo Domenico Angelo Sergi
Datawarehouse, Fondazione Policlinico Universitario A. Gemelli IRCCS

Giovanni Arcuri
Unità Operativa Complessa Tecnologie Sanitarie, Fondazione Policlinico Universitario A. Gemelli IRCCS

Alfredo Cesario
Fondazione Policlinico Universitario A. Gemelli IRCCS

Stefano Patarnello
Fondazione Policlinico Universitario A. Gemelli IRCCS

Massimo Antonelli
Fondazione Policlinico Universitario A. Gemelli IRCCS

Rocco Bellantone
Fondazione Policlinico Universitario A. Gemelli IRCCS

Roberto Bernabei
Fondazione Policlinico Universitario A. Gemelli IRCCS

Stefania Boccia
Fondazione Policlinico Universitario A. Gemelli IRCCS

Paolo Calabresi
Fondazione Policlinico Universitario A. Gemelli IRCCS

Andrea Cambieri
Fondazione Policlinico Universitario A. Gemelli IRCCS

Roberto Cauda

Fondazione Policlinico Universitario A. Gemelli IRCCS
Cesare Colosimo
Fondazione Policlinico Universitario A. Gemelli IRCCS
Filippo Crea
Fondazione Policlinico Universitario A. Gemelli IRCCS
Ruggero De Maria
Catholic University of the Sacred Heart
Valerio De Stefano
Fondazione Policlinico Universitario A. Gemelli IRCCS
Francesco Franceschi
Fondazione Policlinico Universitario A. Gemelli IRCCS
Antonio Gasbarrini
Fondazione Policlinico Universitario A. Gemelli IRCCS
Ornella Parolini
Fondazione Policlinico Universitario A. Gemelli IRCCS
Luca Richeldi
Fondazione Policlinico Universitario A. Gemelli IRCCS
Maurizio Sanguinetti
Fondazione Policlinico Universitario A. Gemelli IRCCS
Andrea Urbani
Fondazione Policlinico Universitario A. Gemelli IRCCS
Maurizio Zega
Fondazione Policlinico Universitario A. Gemelli IRCCS
Giovanni Scambia
Fondazione Policlinico Universitario A. Gemelli IRCCS
Vincenzo Valentini
Fondazione Policlinico Universitario A. Gemelli IRCCS

Research Article

Keywords: COVID-19, SARS-CoV-2, AUROC, Fondazione Policlinico Gemelli, healthcare system, pandemic

Posted Date: June 10th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-544196/v1

Abstract

Background

The COVID-19 pandemic is impressively challenging the healthcare system. Several prognostic models have been validated but few of them are implemented in daily practice. The objective of the study was to validate a machine-learning risk prediction model using easy-to-obtain parameters, potentially available at home, to help identifying patients with COVID-19 who are at higher risk of death.

Methods

The training cohort included all patients admitted to Fondazione Policlinico Gemelli with COVID-19 from March 5, 2020 to November 5, 2020. Afterwards, the model was tested on all patients admitted to the same hospital with COVID-19 from November 6, 2020 to February 5 2021. The primary outcome was in-hospital mortality.

The out-of-sample performance of the model was estimated from the training set in terms of Area under the Receiving Operator Curve (AUROC) and classification matrix statistics by averaging the results of 5-fold cross validation repeated 3-times and comparing the results with those obtained on the test set. An explanation analysis of the model, based on the SHapley Additive exPlanations (SHAP), is also presented. To assess the subsequent time evolution, the change in paO2/FiO2 (P/F) at 48 hours after the baseline measurement was plotted against its baseline value.

Results

Among the 921 patients included in the training cohort, 120 died (13%). Variables selected for the model were age, platelet count, SpO2, blood urea nitrogen (BUN), hemoglobin, C-reactive protein, neutrophil count, and sodium. The results of the 5-fold cross-validation repeated 3-times gave AUROC of 0.87, and statistics of the classification matrix to the Youden index as follows: sensitivity 0.840, specificity 0.774, negative predictive value 0.971. Then, the model was tested on a new population (n=1463) in which the mortality rate was 22.6 %. The test model showed AUROC 0.818, sensitivity 0.813, specificity 0.650, negative predictive value 0.922. Considering the first quartile of the predicted risk score (low-risk score group), the mortality rate was 1.6%, 17.8% in the second and third quartile (high-risk score group) and 53.5% in the fourth quartile (very high-risk score group). The three risk score groups showed good discrimination for the P/F value at admission, and a positive correlation was found for the low-risk class to P/F at 48 hours after admission (adjusted R-squared= 0.48).

Conclusions

We developed a predictive model of death for people with SARS-CoV-2 infection by including only easy-to-obtain variables (abnormal blood count, BUN, C-reactive protein, sodium and lower SpO2). It demonstrated good accuracy and high power of discrimination. The simplicity of the model makes the risk prediction applicable for patients at home, in the Emergency Department, or during hospitalization.

Introduction

A rapid spread of SARS-CoV-2, the agent of coronavirus disease 2019 (COVID-19), has been observed first in China since early January 2020 and then in Italy since the last days of February 2020[1]. At this time, the number of COVID-19 cases and related deaths continue to increase. Patients hospitalized with COVID-19 had a relevant rate of clinical

deterioration. A first large study on more than 1000 patients with COVID-19 in China reported the need for transfer to an intensive care unit (ICU) in 5% of patients admitted with COVID-19, 2.3% mechanical ventilation; 1.4% died[2]. Other studies have reported a rate around 5% of people admitted as critically ill[3,4]. Mortality rates in persons with COVID-19 range from < 1–15%[5,6]. This implies a staggering challenge for the healthcare system. Unfortunately, treatment options are currently scarce, and as hospital resources are shrinking, systems to target respiratory support and other hospital resources to the highest-risk population, such as the ICU, is a priority. Several predictive models of adverse clinical outcomes in people with COVID-19[7–13] as well as a systematic review[14] have been published. Having a clinical algorithm to predict patients who can benefit most from available resources is a valuable aid for decision making and capacity allocation. However, few models have tested the predictive value of simple and readily available parameters. The objective of this study was to generate and validate a machine-learning risk prediction model using parameters that are also potentially available at home to help identify patients with COVID-19 who are at higher risk of death.

Materials And Methods

Study population

The study cohort included all patients admitted to Fondazione Policlinico Gemelli with COVID-19 from March 5, 2020 to February 5, 2021. The diagnosis of SARS-CoV-2 infection was considered when the reverse transcription polymerase chain reaction (PCR) of the SARS-CoV-2 assay was detected from nasopharyngeal swab For each patient, time 0 was considered the date of hospitalization for SARS-CoV-2 infection.

Data collection

Patient data included demographics, comorbidities, vital signs, and laboratory characteristics, as well as exposure history, medical history, symptoms at onset, treatment, and outcome data on admission and during hospitalization. Pre-existing conditions collected were diabetes, hypertension, chronic heart disease, chronic respiratory disease, chronic kidney disease, mild to severe liver disease, pancreatitis, neurological impairment, connective tissue disease, transplantation, HIV infection, and malignancy. Vital signs included heart rate, respiratory rate, oxygen saturation by pulse oximetry (SpO2), temperature, body weight, and body mass index (BMI). Laboratory parameters included hematologic variables (white blood cells [WBC], neutrophils, lymphocytes, and eosinophils, platelet count, hematocrit), blood urea nitrogen (BUN); creatinine; total bilirubin; creatine kinase; glucose; sodium; potassium; C-reactive protein; procalcitonine, D-dimer; ferritin; lactate dehydrogenase (LDH); arterial blood oxygen partial pressure (paO2) and inspired oxygen fraction (FiO2), paO2/FiO2 ratio (P/F). SpO2 was grouped into three categories according to the interquartile range: SpO2 less than 94% (first quartile), SpO2 between 94% and 97.0% (second and third quartile), SpO2 greater than 97.0% (fourth quartile). All data were extracted from the electronic medical records of all patients. To obtain structural information from unstructured texts (such as clinical diary, radiology reports etc.), Natural Language Processing (NLP) algorithms were applied, based on text mining procedures such as: sentence/word tokenization; rule-based approach supported by annotations defined by the clinical SMEs, and using semantic / syntactic corrections where necessary.

Outcome

The primary outcome was in-hospital mortality.

Predictors

Candidate predictors were included when previously shown to be related to mortality in COVID-19 patients or other respiratory diseases (such as bacterial pneumonia) or possibly related because of clinical plausibility.

Statistical analysis

To capture the risk of death associated with early hospitalization, we developed a predictive model including only laboratory variables and oxygen saturation at the time of SARS-Cov2 infection. The rationale behind this choice was to provide a tool for early risk assessment. The variables for the model are routinely collected, available within a very short time after presentation, and the literature has reported their association with an increased likelihood of death; moreover, they could also be available at home through home services. In this way, an estimate of risk can be obtained at the time of hospital admission, and actions on the management of critical versus non-critical patients can be readily taken by hospital staff from the patient's initial clinical status as well as its evolution in a relatively short time frame. A binary logistic regression was applied to express the risk of death in analytical terms, and possibly use it in risk assessment tools based on model coefficients alone.

Candidate predictors were selected through a combination of prior domain knowledge and a data-driven approach: for example, cut-off values to classify SpO2 and sodium were heuristically defined by the interquartile range, confirmed by a-priori medical knowledge. Overall feature selection was conducted iteratively based on their added contribution to the model in terms of information criterion to minimize model redundancy. The model was trained on the first 8 months of data (March 5, 2020 – November 5, 2020), and tested on the next 3 months of data (November 6, 2020 – February 5 2021). The out-of-sample performance of the model was estimated from the training set in terms of area under the receiving operator curve (AUROC) and classification matrix statistics by averaging the results of the 5-fold cross validation repeated 3-times and comparing the results with those obtained on the test set. Finally, an analysis of lift and gain graphs is presented to identify segments of outcome probability where the model proves particularly useful compared to having no model at all. A model explanation analysis, based on the SHapley Additive exPlanations (SHAP) framework, is also presented to derive information about the contribution of individual variables to the model beyond that obtained from simple logistic regression coefficients.

Baseline laboratory variables for each patient were included by taking the first value after the date-time of hospital admission; only variables with less than 20% missing values were retained for further analysis. This set of variables, along with age and sex, and study outcome, were given as input to a routine of 100-iteration of AIC-based stepwise selection on 80% subsets of the randomly partitioned training data, and characteristics selected at least 50 times were considered to train the final logistic regression model. A level of 0.05 was considered significant for statistical testing. Statistical analysis was done with R version 3.6. Data were stored in SAS Viya V.03.05 and accessed through R with SWAT library version 1.5.0.

According to TRIPOD guidelines[15], the study should be considered a TRIPOD 2b because it involves a chronological division between training and testing data from a single institution. All methods were performed in accordance with the relevant guidelines and regulations.

Ethical aspects

This study was approved by Ethics Committee of the Fondazione Policlinico Gemelli (IRB 3447). A waiver of consent was approved by Ethics Committee of the Fondazione Policlinico Gemelli.

Results

The eligible training cohort included a total of 1126 patients with confirmed COVID-19 admitted from 5 March, 2020, to 5 November, 2020. In this cohort, the in-hospital mortality rate was 13.0%. Characteristics of the study population are shown in Table 1.

Table 1. Characteristics of patients included in the training and testing subsets.

Characteristics		Training			Test		
		All patients (N=921)	Alive (n=801)	Died (n=120)	All patients (N=1463)	Alive (n=1131)	Died (n=332)
Demographics	Age, median(SD)	68.0 (15.9)	64.0 (15.4)	84.0 (10.1)	70.0 (18.2)	65.0 (18.5)	80.0 (11.3)
	Male	566 (61.4 %)	501 (62.5 %)	65 (54.2 %)	798 (54.5 %)	596 (52.7 %)	202 (60.8 %)
	BMI, median (IQR)	26.0 (24.2 ; 29.1)	26.1 (24.2 ; 28.7)	26.0 (23.5 ; 29.3)	26.1 (24.2 ; 28.2)	26.1 (24.2 ; 28.2)	26.1 (23.9 ; 28.1)
Coexisting Conditions	Any	685 (74.4 %)	574 (71.7 %)	111 (92.5 %)	1127 (77.0 %)	827 (73.1 %)	300 (90.4 %)
	Current or Former Smoker	24 (2.6 %)	23 (2.9%)	1 (0.8 %)	23 (1.6 %)	20 (1.8%)	3 (0.9 %)
	Arteriopathy	8 (0.9 %)	4 (0.5 %)	4 (3.3 %)	13 (0.9 %)	5 (0.4 %)	8 (2.4 %)
	Chronic Liver Disease	11 (1.2 %)	10 (1.2 %)	1 (0.8 %)	14 (1.0 %)	9 (0.8 %)	5 (1.5 %)
	Cirrhosis	6 (0.7 %)	3 (0.4 %)	3 (2.5 %)	12 (0.8 %)	8 (0.7 %)	4 (1.2 %)
	Diabetes	149 (16.2 %)	120 (15.0 %)	29 (24.2 %)	279 (19.1 %)	206 (18.2 %)	73 (22.0 %)
	Dyslipidemia	78 (8.5 %)	70 (8.7 %)	8 (6.7 %)	106 (7.2 %)	80 (7.1 %)	26 (7.8 %)
	Hiv	27 (2.9%)	26 (3.2 %)	1 (0.8 %)	22 (1.5%)	18 (1.6 %)	4 (1.2 %)
	Myocardial Infarction	116 (12.6 %)	87 (10.9 %)	29 (24.2 %)	227 (15.5 %)	148 (13.1 %)	79 (23.8 %)
	Kidney Failure	44 (4.8 %)	32 (4.0 %)	12 (10.0 %)	107 (7.3 %)	53 (4.7 %)	54 (16.3 %)
	Hypertension	374 (40.6 %)	315 (39.3 %)	59 (49.2 %)	636 (43.5 %)	480 (42.4 %)	156 (47.0 %)
	Autoimmune Disease	41 (4.5 %)	38 (4.7 %)	3 (2.5 %)	62 (4.2 %)	48 (4.2 %)	14 (4.2 %)
	Hematologic Neoplasm	6 (0.7 %)	4 (0.5 %)	2 (1.7 %)	29 (2.0 %)	16 (1.4 %)	13 (3.9 %)
	Neurologic Impairment	102 (11.1 %)	59 (7.4 %)	43 (35.8 %)	171 (11.7 %)	101 (8.9 %)	70 (21.1 %)
	Pancreatitis	5 (0.5 %)	5 (0.6 %)	0 (0.0 %)	13 (0.9 %)	8 (0.7 %)	5 (1.5 %)
	Cardiovascular Pathology	155 (16.8 %)	118 (14.7%)	37 (30.8 %)	295 (20.2 %)	182 (16.1%)	113 (34.0 %)
	Lung Pathology	108 (11.7 %)	81 (10.1 %)	27 (22.5 %)	162 (11.1 %)	97 (8.6 %)	65 (19.6 %)
	Radiotherapy	15 (1.6 %)	13 (1.6 %)	2 (1.7 %)	43 (2.9 %)	28 (2.5 %)	15 (4.5 %)
	Heart Failure	44 (4.8 %)	27 (3.4 %)	17 (14.2 %)	84 (5.7 %)	39 (3.4 %)	45 (13.6 %)
	Transplantation	6 (0.7 %)	5 (0.6 %)	1 (0.8 %)	22 (1.5 %)	13 (1.1 %)	9 (2.7 %)
	Tumor	236 (25.6 %)	188 (23.5 %)	48 (40.0 %)	544 (37.2 %)	406 (35.9 %)	138 (41.6 %)
	Hepatic Ulcer	15 (1.6 %)	11 (1.4 %)	4 (3.3 %)	30 (2.1 %)	16 (1.4 %)	14 (4.2 %)
Symptoms At Admission	Any	807 (87.6 %)	706 (88.1 %)	101 (84.2 %)	1162 (79.4 %)	887 (78.4 %)	275 (82.8 %)
	Cough	344 (37.4 %)	327 (40.8 %)	17 (14.2 %)	366 (25.0 %)	309 (27.3 %)	57 (17.2 %)
	Dyspnea	503 (54.6 %)	429 (53.6 %)	74 (61.7 %)	755 (51.6 %)	542 (47.9 %)	213 (64.2 %)
	Fever	712 (77.3 %)	632 (78.9 %)	80 (66.7 %)	949 (64.9 %)	744 (65.8 %)	205 (61.7 %)
	Nausea or Vomiting	47 (5.1 %)	43 (5.4 %)	4 (3.3 %)	65 (4.4 %)	53 (4.7 %)	12 (3.6 %)
	Diarrhea	99 (10.7 %)	97 (12.1 %)	2 (1.7 %)	93 (6.4 %)	80 (7.1 %)	13 (3.9 %)
Time From Symptom Onset to Admission, median (IQR)		7 (3 ; 10)	7 (3 ; 10)	3 (2 ; 7)	7 (3 ; 10)	7 (3 ; 11)	5 (2 ; 9.5)
Vital Signs on the Day of Admission, median (IQR)	Temperature, °C	37.0 (36.0 ; 38.3)	37.1 (36.0 ; 38.4)	36.8 (36.0 ; 37.9)	36.5 (36.0 ; 37.9)	36.8 (36 ; 37.6)	36.3 (36.0 ; 37.9)
	Systolic Blood Pressure, mm Hg	130.0 (118.0 ; 143.0)	130 (120 ; 142)	122 (108 ; 140)	130.0 (118.0 ; 145.0)	131.0 (120.0 ; 145.0)	123.5 (110.0 ; 145.0)
	Heart rate, /min	80.0 (71.0 ; 88.0)	78.0 (70.0 ; 88.0)	83.0 (72.0 ; 90.0)	80.0 (73.0 ; 90.0)	80.0 (74.0 ; 90.0)	80.0 (72.0 ; 90.0)
Laboratory Findings on the Day of Admission, median (IQR)	White Blood Cell Count, /μL	7.0 (5.1 ; 9.5)	6.9 (5.0 ; 9.3)	8.1 (5.3 ; 10.5)	8.1 (5.9 ; 11.5)	7.9 (6.0 ; 11.5)	8.5 (5.6 ; 11.5)
	Lymphocyte Count, /μL	1.1 (0.8 ; 1.5)	1.1 (0.8 ; 1.5)	1.1 (0.7 ; 1.5)	1.0 (0.7 ; 1.5)	1.1 (0.8 ; 1.5)	0.9 (0.6 ; 1.3)
	Hemoglobin Level, g/dL	14.3 (13.1 ; 15.3)	14.5 (13.4 ; 15.3)	12.9 (11.6 ; 14.4)	14.1 (12.5 ; 15.2)	14.3 (12.9 ; 15.3)	13.3 (11.2 ; 14.6)
	Platelets, μL	198.0 (158.0 ; 257.0)	202.0 (160.0 ; 263.0)	176.0 (143.0 ; 214.0)	204.0 (154.0 ; 273.0)	207.0 (162.0 ; 278.0)	196.5 (140.0 ; 253.5)
	Creatinine Level, mg/dL	0.9 (0.8 ; 1.1)	0.9 (0.8 ; 1.1)	1.1 (0.9 ; 1.6)	1.0 (0.8 ; 1.4)	0.9 (0.8 ; 1.2)	1.3 (0.9 ; 2.0)
	D-dimer level, ng/mL	740.5 (400.0 ; 1396.0)	695.0 (380.0 ; 1314.0)	1158.0 (863.0 ; 2872.0)	853.0 (468.0 ; 2031.0)	718.0 (396.0 ; 1488.0)	1715.5 (811.5 ; 3667.0)
	C-reactive protein level, mg/L	60.4 (23.5 ; 130.0)	58.0 (22.1 ; 130.0)	80.4 (38.7 ; 129.9)	77.4 (32.6 ; 143.1)	66.4 (25.5 ; 132.7)	99.9 (61.7 ; 164.3)
	Urea Nitrogen, mg/dL	18.0 (15.0 ; 24.0)	17.0 (14.0 ; 22.0)	27.0 (20.0 ; 38.0)	22.0 (17.0 ; 32.0)	20.0 (16.0 ; 27.0)	34.0 (23.0 ; 50.0)
	Albumin, g/L	33.0 (30.0; 37.0)	33.0 (30.0 ; 37.0)	31.0 (26.0 ; 35.0)	31.0 (28.0; 35.0)	33.0 (29.0 ; 36.0)	29.0 (26.0 ; 32.0)
	Vitamin D, ng/mL	15.7 (10.7 ; 20.1)	15.8 (10.7 ; 20.1)	12.8 (12.8 ; 12.8)	16.4 (13.2 ; 28.4)	19.3 (13.2 ; 28.4)	15.6 (14.3 ; 22.8)
	P/F	290.5 (201.4 ; 361.9)	297.4 (208.2 ; 366.7)	248.3 (164.5 ; 351.7)	228.8 (159.5 ; 323.0)	249.8 (182.0 ; 332.8)	166.4 (104.0 ; 250.0)

SD: standard deviation; IQR = interquartile range; BMI: Body mass index; P/F = paO2/FiO2 ratio

Survivors differed from nonsurvivors for being younger, having few preexisting medical conditions (specifically, lower rates of diabetes, hypertension, cardiovascular diseases, chronic respiratory diseases, renal failure, solid tumors, and arteriopathy), more cough and diarrhea at onset but less dyspnea, a longer time from symptoms onset to hospitalization, a higher P/F, albumin and hemoglobin value, a higher platelet count, lower WBC and lymphocyte count, a lower creatinine, BUN, C-reactive protein, and D-dimer.

From an initial dataset of 1126 patient records, a total of 921 complete records were included. After the feature selection phase, the selected variables were age (relative selection frequency [RSF] 100%), platelet count (RSF 97%), SpO2 (RSF 80%), BUN (RSF 72%), hemoglobin (RSF 71%), C-reactive protein (RSF 68%), neutrophil count (RSF 60%), and sodium (RSF 58%). These variables were used to fit the logistic regression model. The estimated coefficients of the logistic model are shown in Table 2, along with p-values.

Table 2
Logistic regression model at the start of hospitalization for SARS-CoV-2 infection

Variable	Coefficient	P-value
Intercept	-8.022163	3.34e-09 ****
Age (continuous)	0.090299	9.32e-15 ****
Hemoglobin (continous)	-0.124580	0.03666 *
Blood urea nitrogen (continuous)	0.016342	0.00956 **
Platelet count (continuous)	-0.004924	0.00057 ***
C-reactive protein (continuous)	0.003086	0.04838 *
Neutrophils (continuous)	0.092127	0.00203 **
Sodium < = 136 mmol/l	0.015663	0.95494
Sodium > = 141 mmol/l	0.720771	0.01388 *
SpO2 94.4–97.0%	0.501530	0.15757
SpO2 < = 94.3%	1.060584	0.00521 **

Each variable in the model is associated with a distribution of importance values among all instances of the dataset (patients), ordered by the value of the variable from low to high. It emerges, for example, that a lower value of platelet count is associated with a higher risk of death, whereas higher values of BUN, C-reactive protein, neutrophils and age are associated with a higher risk of death. The sodium variable was subdivided according to the interquartile range: in this three-category version of the variable (low, normal, high), it can be seen that the ”low sodium” group (< = 136 mmol/l) does not impact death for this cohort of patients, whereas the ”high sodium” class (> = 141 mmol/l) does. Similarly, SpO2 < 94% has a greater impact in the model than the variable representing SpO2 values between 94 and 97. Figure 1 is a representation of the importance of the variables in the model based on the SHAP framework.

Figure 1. SHAP (SHapley Additive exPlanations) framework for the features in the logistic model.

The overall statistical significance of the model according to chi-squared residual deviance test was confirmed with a p-value zero. The 5-fold cross-validation repeated 3-times resulted in an AUROC of 0.87, and the statistics of the classification matrix at the Youden index as follows: sensitivity 0.840, specificity 0.774, negative predictive value 0.971. The model was then tested on the cohort of patients admitted between November 6, 2020, and February 5, 2021, (n = 1463), recording the model variable of interest and the clinical outcome. In this cohort of patients, the mortality rate was 22.6 %. The model test results in terms of AUROC statistics and confounding matrix are AUROC 0.818, sensitivity 0.813, specificity 0.650, negative predictive value 0.922 (Table 3 and Fig. 2).

Table 3
Classification matrix and statistics at training set Youden classification threshold on training (cross-validation) and test data.

Dataset	AUROC	Sensitivity	Specificity	PPV	NPV	TN	FN	TP	FP
Training set (cross validation)	0.870	0.840	0.766	0.341	0.971	639	19	100	193
Testing set	0.818	0.813	0.650	0.405	0.922	734	62	270	397

AUROC: Area under the Receiver Operating Characteristics; PPV: positive predictive value; NPV: negative predictive value; TN: true negative; FN: false negative; TP: true positive; FP: false positive

To get a quantification of how the model performs in different segments of probability outputs compared to a random classifier, a gain and lift curve analysis is shown (Fig. 3).

Moreover, the lift plot on the testing data in Fig. 3 shows that for the first decile of predictions, the model performs more than 3 times better than random guessing based on prevalence only. Specifically, when considering the first quartile of the predicted risk score on the test set, it contains 6 death events out of 366 total predictions in that risk group. Similarly, the highest 25% of risk scores on the test set contain 196 actual death events, which is more than 50% of the population classified in that risk group (Table 4).

Table 4
Risk groups as defined from gains and lift chart analysis on the test data by applying the thresholds defined on the trained data.

Risk Group	Number of patients	Lower Threshold	Higher Threshold	Death Prevalence
Mild Risk (< 25th percentile)	366	0.000	0.019	1.6%
High Risk (25–75 th percentile)	731	0.019	0.270	17.8%
Very High Risk (> 75th percentile)	366	0.270	1.000	53.5%

Evolution of respiratory condition by initial risk group

In addition to having an instrument capable of distinguish between low-risk, high-risk and very high-risk cases with a fair degree of accuracy, we evaluated the evolution of the different groups of patients in the first few hours after hospital admission. Considering the cohort of patients used for model training and taking the first available value of P/F within 24 hours of hospital admission, the three model-defined risk groups had a mean value of P/F of 301, 273, 273 for low-risk, high-risk and very high-risk, respectively. A t-test between the low-risk group versus the other two categories showed a statistically significant difference. To assess the subsequent time course, the change in P/F at 48 hours after the baseline measurement can be plotted against its baseline value (Fig. 4).

A positive correlation, with an adjusted R-squared of 0.48, was found for the low-risk class whereas for the high-risk and very high-risk risk classes the R-squared for this simple linear model is 0.14 and zero, respectively.

Adoption in clinical practice

The risk of death score for each patient with SARS-CoV-2 infection was made available to clinicians along with real-time predictions directly on the Electronic Health Record (Fig. 5).

Discussion

Given the high rate of patients with complications of SARS-CoV-2 infection, prioritization of patients who need higher levels of care or immediate medical attention is critical. In the present study on a total of 2384 patients hospitalized with COVID-19, of whom 18.9% died, we presented an artificial intelligence-driven clinical algorithm to predict risk of death. The algorithm showed that abnormal blood counts (hemoglobin, platelets, neutrophils), high levels of BUN, C-reactive protein, sodium and lower SpO2 were associated with an increased risk of death. From the model, we were able to identify three risk level groups: *low-risk*, with a prevalence of death of 1.6%, *high-risk*, with a prevalence of death of 17.8%, and *very high-risk* with a prevalence of death of 53.5%. Our model includes only easy-to-obtain variables: its simplicity makes the risk prediction applicable for different purposes for patients at home, in the Emergency Department, or during the hospitalization. For example, when the calculated individual risk of death is low, the physician may choose to monitor the patient at home, whereas high risk estimates suggest more aggressive monitoring or resource allocation or may be useful in anticipating organizational needs in terms of intensive, sub-intensive, and rehabilitation rooms and staff allocation. Safely discharging patients from the Emergency Department is of a great benefit in saving beds for other critically ill patients. Such a parsimonious model is exploitable even in medically resource-limited settings.

The discriminatory performance of the model is very high and testing of the model on a new cohort of the very newly diagnosed patients confirmed its validation. The model also demonstrated good accuracy in predicting respiratory evolution when P/F at baseline and at 48 hours were considered.

The two major strengths of the present study are the parsimonious inclusion of simple and easy-to-obtain variables, also available in primary care settings, and the immediate translation of a mathematical model into an comprehensible and implementable number in EHR for clinical decision making in daily practice. Some published studies provide a computational tool for easy use in a variety of settings[10,11]. Unfortunately, such a calculator requires data entry that is cumbersome in a busy clinical practice. Real-time processing of the model directly from the EHR provides an immediate and seamless calculation, a score that can be used to support clinical decision making and support prioritization, especially when the healthcare system is overloaded.

Other predictive models have been published previously, many of which report age, hematologic measures, C-reactive protein and spO2 as the main variables explaining the predictive model[7,8]. Our results confirm and extend those of other large cohort studies[7–13] demonstrating the predictive value of renal function and, in particular, of blood urea nitrogen for mortality[14,16]. In addition, we share 4 of 9 variables from a machine-learning-based study with the largest included population[14].

The model of the present study shares some variables among those included in CURB-65, a well-validated and widely used score for predicting mortality in persons with community-acquired pneumonia[17], with an AUROC of 0.72 (0.71–0.73) in patients with COVID-19[14]. Age and BUN are included in both CURB-65 and our predictive model.

whereas respiratory function was described by respiratory rate in CURB-65 and SpO2 in our model. The variables in the present model also share many parameters with other risk scores used to predict mortality in patients with sepsis, such as the widely used SOFA score[18], probably reflecting a clinical presentation of COVID-19 very close to sepsis. These findings may help highlight the complex pathogenesis of the SARS-CoV-2 infection.

Only a few published models implement machine learning techniques for statistical analysis. Machine learning methods can synthesize data from thousands of patients to generate tailored predictions for each new patient in real time. In addition, model explanations were made available to physicians along with real-time predictions.

The present study includes several limitations: the scalability and the interoperability of the entire data architecture must be demonstrated in other centers and clinical settings. Moreover, the impact of clinical implementation of this predictive model in daily clinical life has not yet been demonstrated. Studies demonstrating changes in clinical management based on model prediction are strongly warranted.

Currently, containing the COVID-19 epidemic is an urgent global priority. Dealing with a severe pandemic disease such as COVID-19 is also very challenging because rapidly changing variables (vaccination, new SARS-CoV-2 variants, saturation of hospital capacity) alter the risk of death over time[19]. Our predictive model, which includes parsimonious and easy-to-obtain variables, is pragmatic and effective in identifying individuals at particularly high risk for a poorer hospital course. Computational infrastructure could enhance this process, and data repository, updated in real time, can continuously inform the planning of diagnostic and treatment strategies. Finally, an integrated system that gives a measure of the risk of death for any patient admitted at any time could be of particular value. Predictive models can help provide appropriate care and optimize the use of limited resources, such as during a pandemic.

Finally, sharing large amounts of data among centers around the world can be a formidable response to the tremendous challenge of the COVID-19 pandemic.

Declarations

The authors declare no competing interests.

AUTHOR CONTRIBUTION

R.M., J.L., S.P. and V.V conceived of the presented idea and drafted the manuscript.

J.L., C.M., C.I., S.P., A.D., A.M., P.D.A.S. extracted and analysed the data.

All other authors contributed equally, discussed the results and concurred to the final manuscript

References

1. https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6
2. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Zhong NS; China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020

Apr 30;382(18):1708-1720. doi: 10.1056/NEJMoa2002032. Epub 2020 Feb 28. PMID: 32109013; PMCID: PMC7092819

3. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA. 2020 Apr 7;323(13):1239-1242. doi: 10.1001/jama.2020.2648. PMID: 32091533.
4. Coronavirus disease 2019 (COVID-19). https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html
5. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, Ma K, Xu D, Yu H, Wang H, Wang T, Guo W, Chen J, Ding C, Zhang X, Huang J, Han M, Li S, Luo X, Zhao J, Ning Q. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. BMJ. 2020 Mar 26;368:m1091. doi: 10.1136/bmj.m1091.
6. Zhou F, Yu T, Du R et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020 Mar 28;395(10229):1054-1062. doi: 10.1016/S0140-6736(20)30566-3. Epub 2020 Mar 11.
7. Castro VM, McCoy TH, Perlis RH. Laboratory Findings Associated With Severe Illness and Mortality Among Hospitalized Individuals With Coronavirus Disease 2019 in Eastern Massachusetts. JAMA Netw Open. 2020 Oct 1;3(10):e2023934. doi: 10.1001/jamanetworkopen.2020.23934. PMID: 33125498; PMCID: PMC7599467.
8. Berenguer J, Ryan P, Rodríguez-Baño J, et al. Characteristics and predictors of death among 4035 consecutively hospitalized patients with COVID-19 in Spain. Clin Microbiol Infect. 2020 Nov;26(11):1525-1536. doi: 10.1016/j.cmi.2020.07.024. Epub 2020 Aug 4. PMID: 32758659; PMCID: PMC7399713.
9. Gupta S, Hayek SS, Wang W, et al. Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. JAMA Intern Med. 2020 Jul 15;180(11):1–12. doi: 10.1001/jamainternmed.2020.3596.
10. Liang W, Liang H, Ou L, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med. 2020 Aug 1;180(8):1081-1089. doi: 10.1001/jamainternmed.2020.2033. PMID: 32396163; PMCID: PMC7218676.
11. Razavian N, Major VJ, Sudarshan M, et al. A validated, real-time prediction model for favorable outcomes in hospitalized COVID-19 patients. NPJ Digit Med. 2020 Oct 6;3:130. doi: 10.1038/s41746-020-00343-x. PMID: 33083565; PMCID: PMC7538971
12. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ. 2020 Apr 7;369:m1328. doi: 10.1136/bmj.m1328. Update in: BMJ. 2021 Feb 3;372:n236.
13. Bennett TD, Moffitt RA, Hajagos JG et al. The National COVID Cohort Cllaborative: Clinical characterization and early severity prediction. medRxiv 2021.01.12.21249511; doi: https://doi.org/10.1101/2021.01.12.21249511
14. Knight SR, Ho A, Pius R, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ. 2020 Sep 9;370:m3339. doi: 10.1136/bmj.m3339CURB-65
15. https://www.equator-network.org/reporting-guidelines/tripod-statement/
16. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med. 2020 Jun 25;58(7):1021-1028. doi: 10.1515/cclm-2020-0369. PMID: 32286245.
17. Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003 May;58(5):377-82. doi: 10.1136/thorax.58.5.377. PMID: 12728155; PMCID: PMC1746657

18. Raith EP, Udy AA, Bailey M, et al. Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit. JAMA. 2017 Jan 17;317(3):290-300. doi: 10.1001/jama.2016.20328. PMID: 28114553.
19. Asch DA, Sheils NE, Islam MN, et al. Variation in US Hospital Mortality Rates for Patients Admitted With COVID-19 During the First 6 Months of the Pandemic. JAMA Intern Med. 2020 Dec 22:e208193. doi: 10.1001/jamainternmed.2020.8193. Epub ahead of print. PMID: 33351068; PMCID: PMC7756246.

Figures

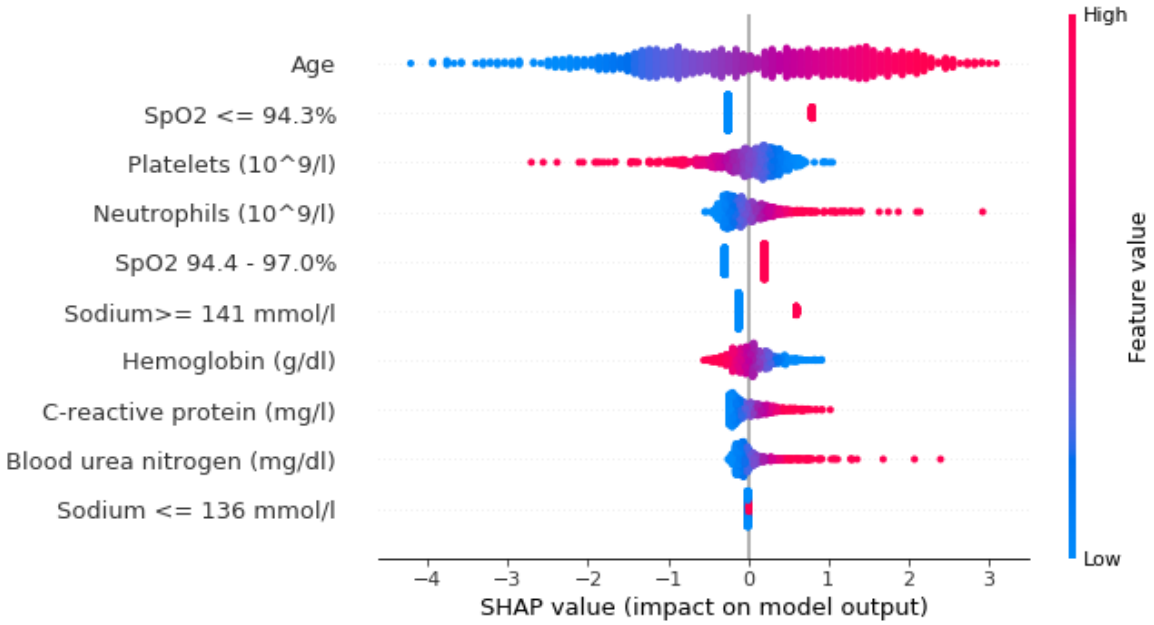

Figure 1

SHAP (SHapley Additive exPlanations) framework for the features in the logistic model.

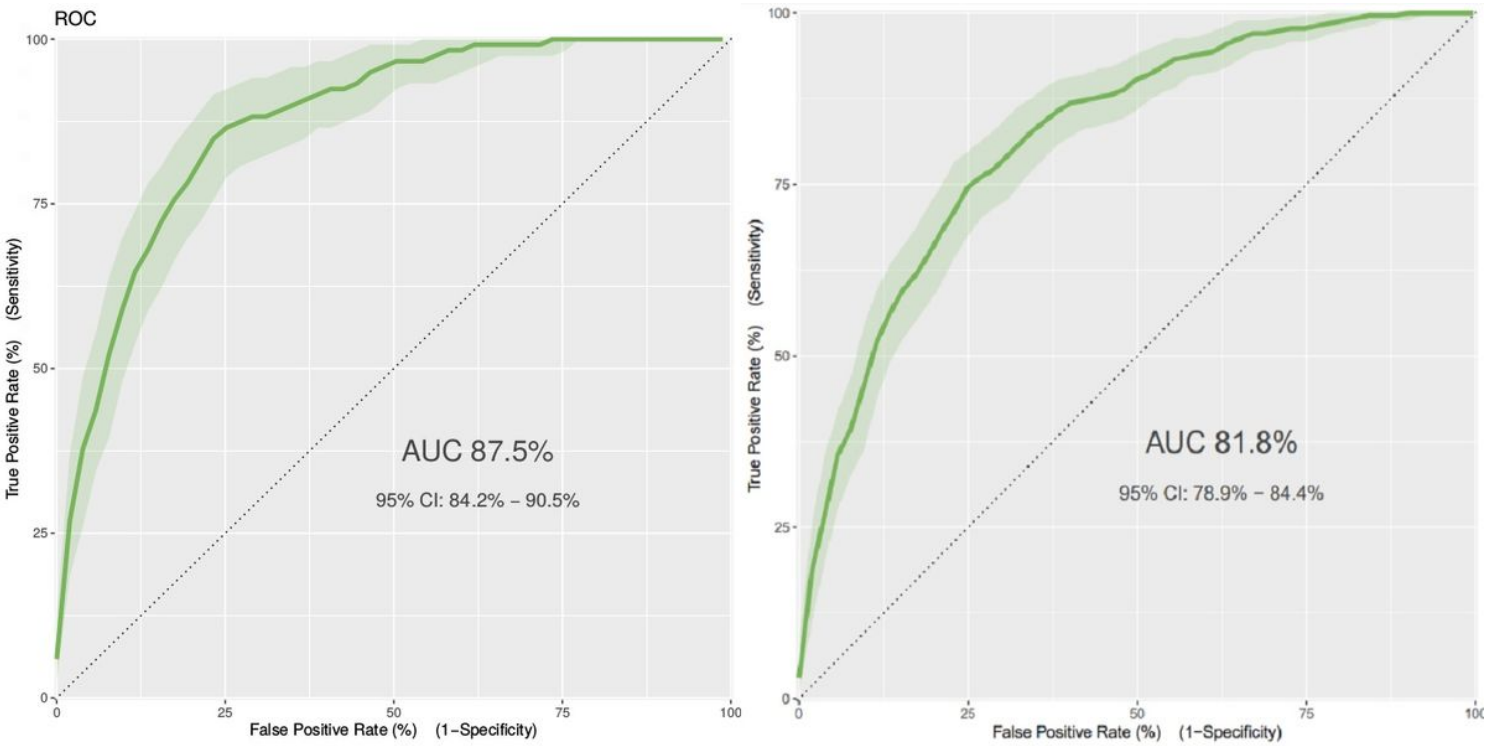

Figure 2

Receiver operator characteristics (ROC) on training set (left) and testing set (right)

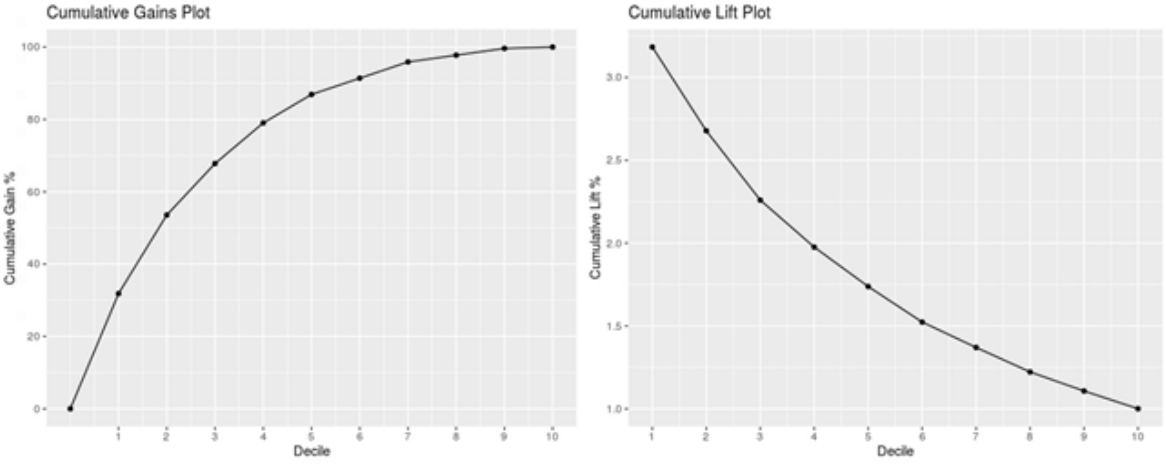

Figure 3

Cumulative gain and lift charts on testing data.

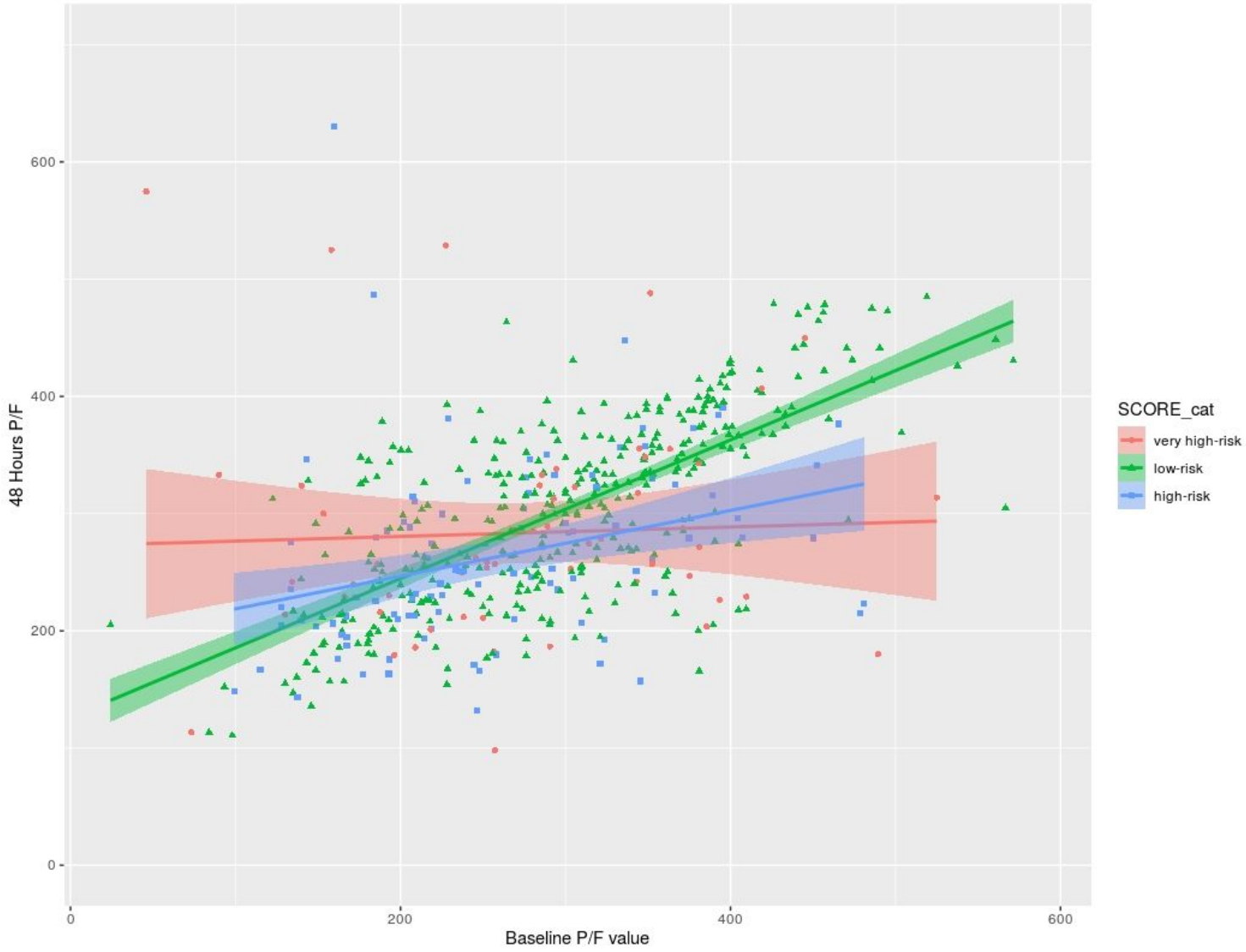

Figure 4

Scatter plot of baseline P/F value and its variation at 48 hours for the three groups of risk class patients according to the logistic regression model. The line of best linear fit is reported for ease of visualization.

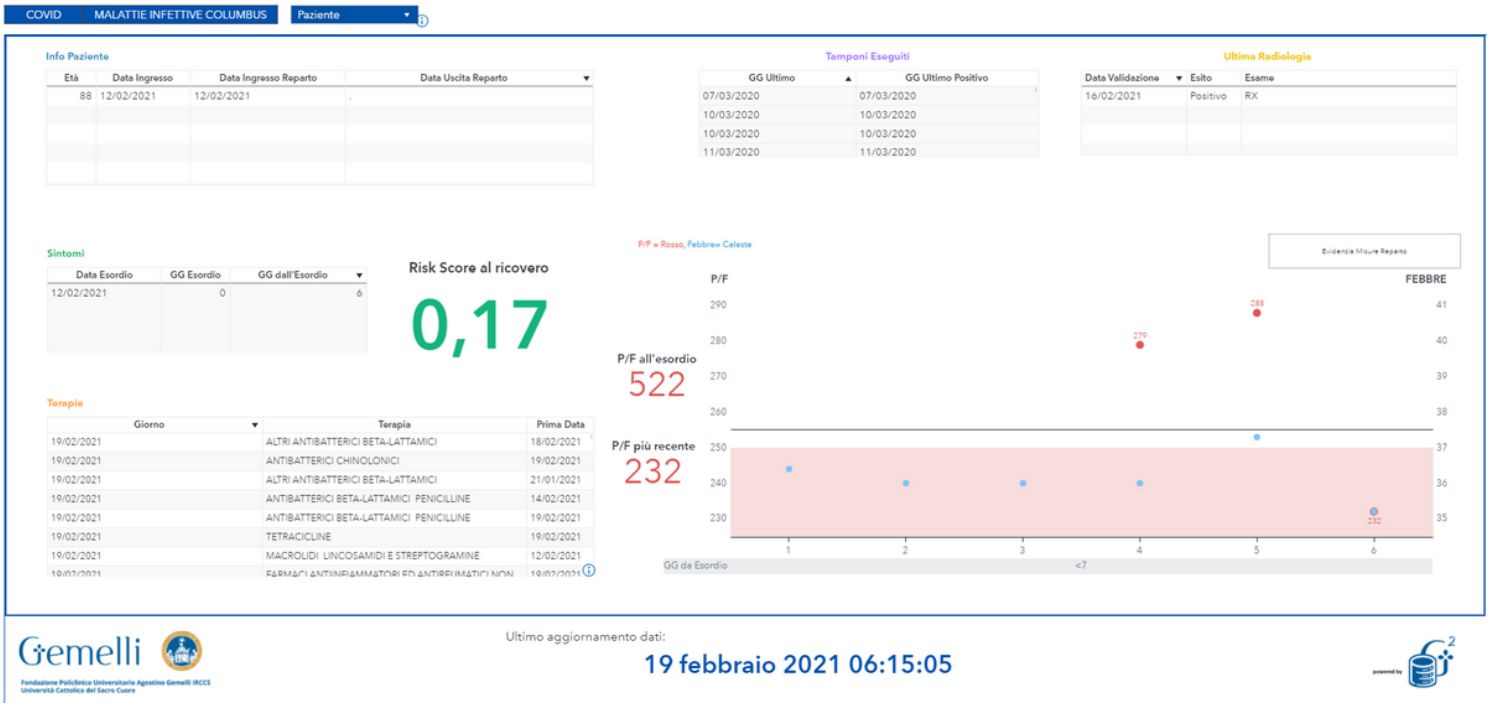

Figure 5

Example of the availability on HER of the risk score of death at admission for a patient with SARS-CoV-2 infection