

Machine learning for prediction of in-hospital mortality in coronavirus disease 2019 patients: results from an Italian multicenter study

Marika Vezzoli^{a,*}, Riccardo Maria Inciardi^{b,*}, Chiara Oriecua^{a,b}, Sara Paris^b, Natalia Herrera Murillo^b, Piergiuseppe Agostoni^{c,u}, Pietro Ameri^d, Antonio Bellasi^e, Rita Camporotondo^f, Claudia Canale^d, Valentina Carubelli^b, Stefano Carugo^g, Francesco Catagnano^{f,h}, Giambattista Danzi^j, Laura Dalla Vecchiaⁱ, Stefano Giovinazzo^d, Massimiliano Gneccchi^f, Marco Guazzi^k, Anita Iorio^l, Maria Teresa La Rovere^m, Sergio Leonardi^f, Gloria Maccagni^k, Massimo Mapelli^{c,u}, Davide Margonato^{f,h}, Marco Merloⁿ, Luca Monzo^{o,t}, Andrea Mortara^h, Vincenzo Nuzziⁿ, Matteo Pagnesi^b, Massimo Piepoli^{b,q}, Italo Porto^d, Andrea Pozzi^l, Giovanni Provenzale^g, Filippo Sarullo^r, Michele Senni^l, Gianfranco Sinagraⁿ, Daniela Tomasoni^b, Marianna Adamo^b, Maurizio Volterrani^s, Roberto Maroldi^v, Marco Metra^b, Carlo Mario Lombardi^b and Claudia Specchia^a

Background Several risk factors have been identified to predict worse outcomes in patients affected by SARS-CoV-2 infection. Machine learning algorithms represent a novel approach to identifying a prediction model with a good discriminatory capacity to be easily used in clinical practice. The aim of this study was to obtain a risk score for in-hospital mortality in patients with coronavirus disease infection (COVID-19) based on a limited number of features collected at hospital admission.

Methods and results We studied an Italian cohort of consecutive adult Caucasian patients with laboratory-confirmed COVID-19 who were hospitalized in 13 cardiology units during Spring 2020. The Lasso procedure was used to select the most relevant covariates. The dataset was randomly divided into a training set containing 80% of the data, used for estimating the model, and a test set with the remaining 20%. A Random Forest modeled in-hospital mortality with the selected set of covariates: its accuracy was measured by means of the ROC curve, obtaining AUC, sensitivity, specificity and related 95% confidence interval (CI). This model was then compared with the one obtained by the Gradient Boosting Machine (GBM) and with logistic regression. Finally, to understand if each model has the same performance in the training and test set, the two AUCs were compared using the DeLong's test. Among 701 patients enrolled (mean age 67.2 ± 13.2 years, 69.5% male individuals), 165 (23.5%) died during a median hospitalization of 15 (IQR, 9–24) days. Variables selected by

the Lasso procedure were: age, oxygen saturation, PaO₂/FiO₂, creatinine clearance and elevated troponin. Compared with those who survived, deceased patients were older, had a lower blood oxygenation, lower creatinine clearance levels and higher prevalence of elevated troponin (all $P < 0.001$). The best performance out of the samples was provided by Random Forest with an AUC of 0.78 (95% CI: 0.68–0.88) and a sensitivity of 0.88 (95% CI: 0.58–1.00). Moreover, Random Forest was the unique model that provided similar performance in sample and out of sample (DeLong test $P = 0.78$).

Conclusion In a large COVID-19 population, we showed that a customizable machine learning-based score derived from clinical variables is feasible and effective for the prediction of in-hospital mortality.

Keywords: coronavirus disease 2019, inflammation, machine learning methods, mortality score

^aDepartment of Molecular and Translational Medicine, University of Brescia, Italy, ^bCardiology, ASST Spedali Civili di Brescia and Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, University of Brescia, Brescia, ^cCentro Cardiologico Monzino, IRCCS, Department of Clinical Sciences and Community Health, University of Milano, Milan, ^dIRCCS Ospedale Policlinico San Martino and Department of Internal Medicine, University of Genova, Genova, ^eInnovation and Brand Reputation Unit, Papa Giovanni XXIII Hospital, Bergamo, ^fFondazione IRCCS Policlinico S. Matteo and University of Pavia, Pavia, ^gDivision of Cardiology, Ospedale San Paolo, ASST Santi Paolo E Carlo, University of Milano, Milan, ^hCardiology Department, Policlinico Di Monza, Monza, ⁱDepartment of Cardiology, Istituti Clinici Scientifici Maugeri, IRCCS, Istituto Scientifico Di Milano, Milan, ^jDivision of Cardiology, Ospedale Maggiore Di Cremona, Cremona, ^kHeart Failure Unit, Cardiology Department, IRCCS San Donato Hospital, University of Milan, Milan, ^lCardiovascular Department and Cardiology Unit, Papa Giovanni XXIII Hospital-Bergamo, Bergamo, ^mDepartment of Cardiology, Istituti Clinici Scientifici Maugeri, IRCCS, Istituto Scientifico Di

Pavia, Pavia, ¹Cardiovascular Department, Azienda Sanitaria Universitaria Integrata, Trieste, ²Istituto Clinico Casal Palocco, Rome, ³Heart Failure Unit, G da Saliceto Hospital, AUSL Piacenza, Piacenza, ⁴Institute of Life Sciences, Sant'Anna School of Advanced Studies, Pisa, ⁵Cardiovascular Rehabilitation Unit, Buccheri La Ferla Fatebenefratelli Hospital, Palermo, ⁶Department of Medical Sciences, Istituto Di Ricovero E Cura a Carattere Scientifico (IRCCS) San Raffaele Pisana, ⁷Policlinico Casilino, Rome, ⁸Department of Clinical Sciences and Community Health, University of Milano, Milan and ⁹Radiology ASST Spedali Civili di Brescia and Department of Medical and Surgical, Specialties,

Radiological Sciences and Public Health, University of Brescia, Brescia, Italy

Correspondence to Carlo Mario Lombardi, Cardiology, Spedali Civili di Brescia, Piazza, Spedali Civili 1, 25123, Brescia, Italy
E-mail: carlo.lombardi@unibs.it

Introduction

Regardless of the measures put in place to contain the spread of coronavirus disease 2019 (COVID-19), Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has affected more than 190 million people, resulting in over 4 million deaths as of July 2021, with a significant impact on the healthcare system worldwide, overwhelming the local medical resources. SARS-CoV-2 infection ranges from mild or asymptomatic status to severe viral pneumonia with respiratory insufficiency. In some cases, there is a rapid and often unpredictable clinical deterioration with the appearance of multiorgan failure and even death. As previously reported, patients with risk factors including concomitant cardiovascular disease on admission, history of heart failure, older age, and immunosuppression have an extremely poor prognosis, with higher mortality and elevated risk of ARDS (acute respiratory distress syndrome), thrombo-embolic events, and septic shock.¹⁻³ Identifying patients at a higher risk of events or those that already have severity markers of infection enables the adoption of more targeted therapeutic strategies, potentially improving the prognosis and allowing a better use of available resources.

To date, clinical models have been developed to predict worse outcomes among COVID-19 patients. Nevertheless, despite the predominant prognostic role of cardiovascular disease, few data derive from a COVID-19 population with a relevant burden of cardiovascular comorbidities.

By the application of machine learning algorithms, the aim of our study is to identify the most important variables collected at hospital admission in COVID-19 patients that have a strong impact on the death prediction. Moreover, a risk score was developed to predict the in-hospital mortality of COVID-19 patients when they arrive in the cardiology units (independently from the time to death) based on a limited number of features.

Methods

Population and outcome

This multicenter observational study involved a cohort of consecutive adult Caucasian patients with laboratory-confirmed COVID-19 who were hospitalized in 13 Italian cardiology units from 1 March to 9 April 2020.⁴⁻⁶ Diagnosis of COVID-19 was made by real-time reverse transcriptase-PCR (RT-PCR) assay of nasal and pharyngeal swabs. RT-PCR of lower respiratory tract aspirates was also performed when clinically indicated. Acute cardiovascular diagnosis (i.e. acute heart failure, acute coronary

syndrome and new-onset arrhythmias) upon admission were exclusion criteria. Patients were followed up after the COVID-19 diagnosis and all causes of in-hospital mortality or discharge were ascertained until 23 April 2020.

This study complied with the Declaration of Helsinki and was approved by the ethical committee of Spedali Civili di Brescia, Brescia, Italy (no. NP 4105), and each recruiting center.

Data collection

Patients' data at admission were extracted from the electronic medical records of each designated hospital. Detailed demographics information, medical history (particularly cardiovascular diseases), and in-hospital clinical course including treatments were recorded. Laboratory examinations including routine blood tests; lymphocyte subsets; inflammatory or infection-related biomarkers; and cardiac, renal, liver, and coagulation function tests were obtained at initial diagnosis. Renal function was measured as estimated glomerular filtration rate (eGFR) and was calculated by the chronic kidney disease epidemiology collaboration equation⁷; chronic kidney disease was defined when eGFR was less than 60 ml/kg/1.73 m². Cardiac injury was defined by plasma levels of high-sensitivity troponin, either troponin T or troponin I, greater than the 99th percentile of normal values, as per manufacturer's indications.

Statistical analysis

Variables with more than 20% of missing values were excluded from the analyses.

To obtain a risk score of in-hospital mortality for COVID-19 based on a limited number of covariates, a variable selection method was used, namely the Lasso procedure, where the λ parameter was set to retain only $\cong 12\%$ of covariates. For estimating Lasso, missing values must be imputed, and to do this the MissForest algorithm⁸ was chosen; it is able to deal with multivariate data consisting of continuous and categorical variables simultaneously. This imputation method, well suited when data are then modeled by means of ensemble tree procedures, considers complex interactions and nonlinear relationships and is robust to noisy data and multicollinearity.

For the variables selected by Lasso, data were presented stratified by death status. Continuous variables were shown as means and standard deviations, skewed variables as medians and interquartile ranges (IQR),

dichotomous variables as counts and percentages. Comparisons were made, respectively, using *t* test for means, Wilcoxon test for medians and chi-squared test for proportions.

Spearman correlations (ρ_s) between couples of variables were visualized by means of a correlation plot where blue and red circles correspond to positive and negative correlations, respectively.^{9–11} The circle diameters and color intensity are proportional to the magnitude of the Spearman indexes (bigger circles with more intense colors correspond to higher correlations) and black crosses on them identify correlation not significantly different from zero (P -values >0.05). The correlation matrix was reordered according to the hierarchical cluster analysis on the quantitative variables. To overcome potential multicollinearity problems and modeling nonlinear relationships,^{12–15} in-hospital mortality of COVID-19 (outcome) was estimated by means of a Random Forest,¹⁶ an ensemble tree method^{17,18} belonging to the machine learning approach,¹⁹ where the covariates were the variables selected by Lasso. The dataset was randomly divided in two subsamples with the same percentage of dead/alive people of the entire sample: the training set contained 80% of the data and the test set the remaining 20%. The training set was used for estimating Random Forest, which modeled in-hospital mortality with the covariates selected by Lasso. Its accuracy was measured by means of the ROC curve, extracting AUC, the optimal threshold obtained with the Youden index, accuracy, sensitivity, specificity, negative predictive values (NPV), positive predictive values (PPV) and related 95% CI computed with 10 000 stratified bootstrap replicates. Moreover, the Brier score is used for model comparison. In order to identify which variables have the best predictive power, the relative Variable Importance Measure (relVIM, known as Total Decrease in Node Impurity) was estimated. relVIM computes the importance of each feature as the sum over the number of splits across all trees containing the feature, proportionally to the number of samples it splits. This measure provides a ranking from the most (relVIM = 100) to the less important variable and it is visualized by means of a bar plot. Finally, the marginal effect that one variable has on the predicted outcome of Random Forest was visualized by means of a 2D plot called a Partial Dependence Plot (PDP). This graph reports in the *x*-axis one covariate and in the *y*-axis the Random Forest predictions, showing if their relationship is linear, monotonic or more complex. Results show a PDP for each feature selected by the Lasso procedure and used as a covariate in Random Forest. The model obtained was compared with another ensemble tree method, which follows the idea of perturbing and combining nonaccurate trees, namely the Gradient Boosting Machine²⁰ (GBM), and with the logistic regression where the continuous variables were treated as Restricted Cubic Splines with five nodes. In the case of logistic regression,

predictions were cross validated (10-fold) to make logistic regression performance comparable with Out-Of-Bag (OOB) predictions of Random Forest and GBM. Finally, to understand if each model had the same performance in sample (training) and out of sample (test), the two AUCs were compared by means of the DeLong's test. Details on parameters of the machine learning algorithms used in this article are reported in Table S2, Supplementary Materials, <http://links.lww.com/JCM/A463>.

Statistical analyses were performed using R version 4.1.0 (R Core Team 2019, Vienna, Austria) and SAS statistical software version 9.4 (SAS Institute, Inc., Cary, North Carolina, USA).

Results

Between 1 March and 9 April 2020, 701 patients were enrolled (mean age 67.2 ± 13.2 years, 69.5% male individuals) of whom 165 (23.5%) died during a median hospitalization of 15 (IQR, 9–24) days.

Original dataset contained 53 variables: 12 of these were excluded because of the percentage of missing values ($>20\%$). The remaining were imputed by the MissForest algorithm, selected by the Lasso procedure where λ was set as equal to 0.07 to retain only five variables ($\cong 12\%$) resulting in a risk score based on a limited number of covariates.

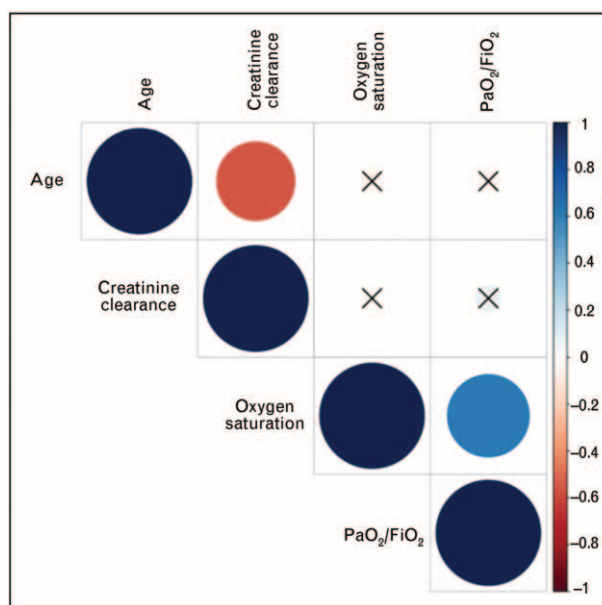
Variables selected by the Lasso procedure, stratified by death status, are shown in Table 1. Compared with those who survived, deceased patients were older (mean age 74.6 ± 10.2 vs. 65.0 ± 13.2 years; P -value <0.001), with a lower oxygen saturation and PaO₂/FiO₂ ratio [89 (IQR, 82–94) vs. 93 (IQR, 89–96) and 158 mmHg/% (IQR, 101–249) vs. 257 (IQR, 147–329) respectively; both P -values <0.001] and lower creatinine clearance levels [57 ml/min (IQR, 33–81) vs. 80 (IQR, 59–93); P -value <0.001] (Table 1). A higher prevalence of patients with elevated troponin values was found among those who died (70.3 vs. 37.3%; P -value <0.001). Table S1, <http://links.lww.com/JCM/A463> in Supplementary Materials

Table 1 Distribution of the selected variables in the study sample stratified by death status

Variables	Alive (N = 536)		Dead (N = 165)		P-value
	N		N		
Age (years)	536	65.0 \pm 13.2	165	74.6 \pm 10.2	$<0.001^a$
Oxygen saturation (ambient air,%)	524	93 (89–96)	163	89 (82–94)	$<0.001^b$
PaO ₂ /FiO ₂ (mmHg/%)	469	257 (147–329)	139	158 (101–249)	$<0.001^b$
Creatinine clearance (ml/min)	525	80 (59–93)	160	57 (33–81)	$<0.001^b$
Troponin (elevated, %)	466	174 (37.3)	148	104 (70.3)	$<0.001^c$

Data shown as mean \pm standard deviation, median (IQR), or count (%). ^a *t*-test. ^b Wilcoxon test. ^c Chi-squared test.

Fig. 1



Correlation plot on variables selected by the Lasso procedure (ρ_S). The circle diameter and the color intensity are proportional to the strength of the Spearman correlation between the variables. Grey and light greys represent a positive or negative relationship, respectively. Correlations without the black symbol X are statistically significant (P -values < 0.05).

reports the descriptive statistics for the remaining 35 variables in the dataset, stratified by death status.

Spearman correlation coefficients (ρ_S) computed between couples of the quantitative variables selected by the Lasso procedure (troponin was excluded from this analysis as it is a dichotomic variable) are visualized by means of a correlation plot (Fig. 1). The graph highlights a strong negative correlation between age and creatinine ($\rho_S = -0.56$; P -value < 0.001) and a positive correlation between oxygen saturation and $\text{PaO}_2/\text{FiO}_2$ ($\rho_S = 0.61$; P -value < 0.001).

The dataset was randomly divided into two subgroups: training (80% of patients, $N_{\text{train}} = 561$) and test set (remaining 20%, $N_{\text{test}} = 140$), used for estimating and validating the model, respectively. The random partition of patients in these two sets was performed stratifying

with respect to the outcome. In this way, the two subgroups contained the same percentage of dead people of the entire sample (23.54%).

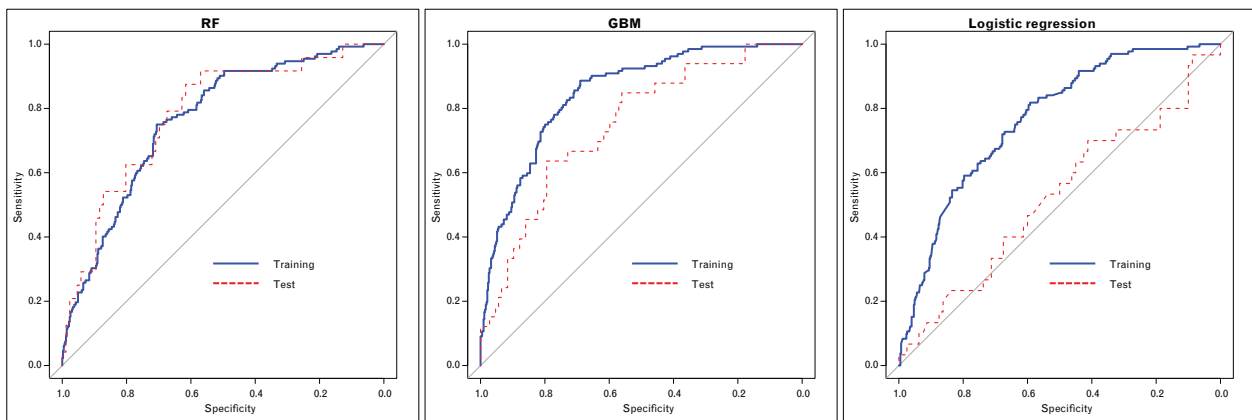
The risk score for predicting mortality was estimated on the training set by means of Random Forest where the covariates were the five selected variables by the Lasso procedure: age, oxygen saturation, $\text{PaO}_2/\text{FiO}_2$, creatinine clearance, troponin.

Accuracy in sample (training) and out of sample (test) was measured by the ROC curve: AUC, the optimal threshold obtained with the Youden index, accuracy, sensitivity, specificity, NPV, PPV (with corresponding 95% CI) are reported in Table 2, separately for the Random Forest, the GBM and the logistic regression model. Figure 2 visualizes the ROC curves of the three models computed in training (blue lines) and in test (red lines).

Table 2 Performance metrics of Random Forest, Gradient Boosting Machine and Logistic regression

Metrics	Random Forest (RF)		Gradient Boosting Machine (GBM)		Logistic regression	
	Training set (in sample)	Test set (out of sample)	Training set (in sample)	Test set (out of sample)	Training set (in sample)	Test set (out of sample)
AUC (95% CI)	0.77 (0.72–0.81)	0.78 (0.68–0.89)	0.85 (0.81–0.88)	0.75 (0.65–0.84)	0.77 (0.72–0.81)	0.52 (0.40–0.65)
Threshold ^a (95% CI)	0.23 (0.11–0.29)	0.21 (0.17–0.43)	0.24 (0.18–0.29)	0.25 (0.17–0.33)	0.20 (0.15–0.37)	0.15 (0.01–0.55)
Accuracy (95% CI)	0.70 (0.59–0.75)	0.71 (0.60–0.85)	0.77 (0.70–0.82)	0.72 (0.59–0.83)	0.68 (0.60–0.79)	0.58 (0.34–0.75)
Sensitivity (95% CI)	0.79 (0.69–0.94)	0.88 (0.58–1.00)	0.87 (0.76–0.94)	0.73 (0.55–0.94)	0.79 (0.55–0.89)	0.63 (0.09–1.00)
Specificity (95% CI)	0.68 (0.49–0.76)	0.65 (0.52–0.92)	0.72 (0.64–0.82)	0.76 (0.47–0.87)	0.64 (0.53–0.85)	0.46 (0.20–0.95)
NPV (95% CI)	0.91 (0.88–0.97)	0.95 (0.88–1.00)	0.94 (0.91–0.97)	0.91 (0.85–0.98)	0.91 (0.86–0.95)	0.80 (0.74–1.00)
PPV (95% CI)	0.42 (0.35–0.48)	0.42 (0.34–0.67)	0.50 (0.43–0.58)	0.45 (0.35–0.62)	0.41 (0.35–0.55)	0.35 (0.29–1.00)

AUC, area under the curve; CI, confidence interval; NPV, negative predictive values; PPV, positive predictive values. ^a The threshold was computed by means of the Youden index.

Fig. 2

ROC curves of Random Forest, Gradient Boosting Machine and Logistic regression.

Given that an accurate model should have good predictions also on observations not used in the training set, Random Forest shows the best performance [from Table 2, the most interesting metrics are: AUC 0.78 (95% CI: 0.68–0.89), sensitivity 0.88 (95% CI: 0.58–1.00) and specificity 0.65 (95% CI: 0.52–0.92)]. Moreover, Random Forest is the unique model that shows same performance in sample and out of sample (DeLong’s test P -value = 0.782); on the contrary, for GBM and logistic regression the DeLong’s test P -values are 0.047 and less than 0.001, respectively. Finally, the smallest Brier Score (computed out of sample) is in correspondence with Random Forest (BS = 0.14) confirming that this model is more accurate than GBM (BS = 0.15) and logistic regression (BS = 0.22). For better understanding of Random Forest, two additional pieces of information were extracted: the relVIM and the PDPs. The relVIM (Fig. 3a) shows a ranking from the most (creatinine clearance, relVIM = 100) to less important variables (troponin, relVIM = 30.37) in predicting the mortality in patients hospitalized for COVID-19.

The five PDPs (one for each covariate in the model), from Fig. 3b–f, show the nonlinear relationships between a variable (x -axis) and the risk score (y -axis, called the mortality risk score) of in-hospital death for COVID-19 patients obtained from Random Forest. Note that when creatinine clearance, $\text{PaO}_2/\text{FiO}_2$, and oxygen saturation increase, the risk score decreases. Otherwise, when age and troponin increase, the risk score increases.

Discussion

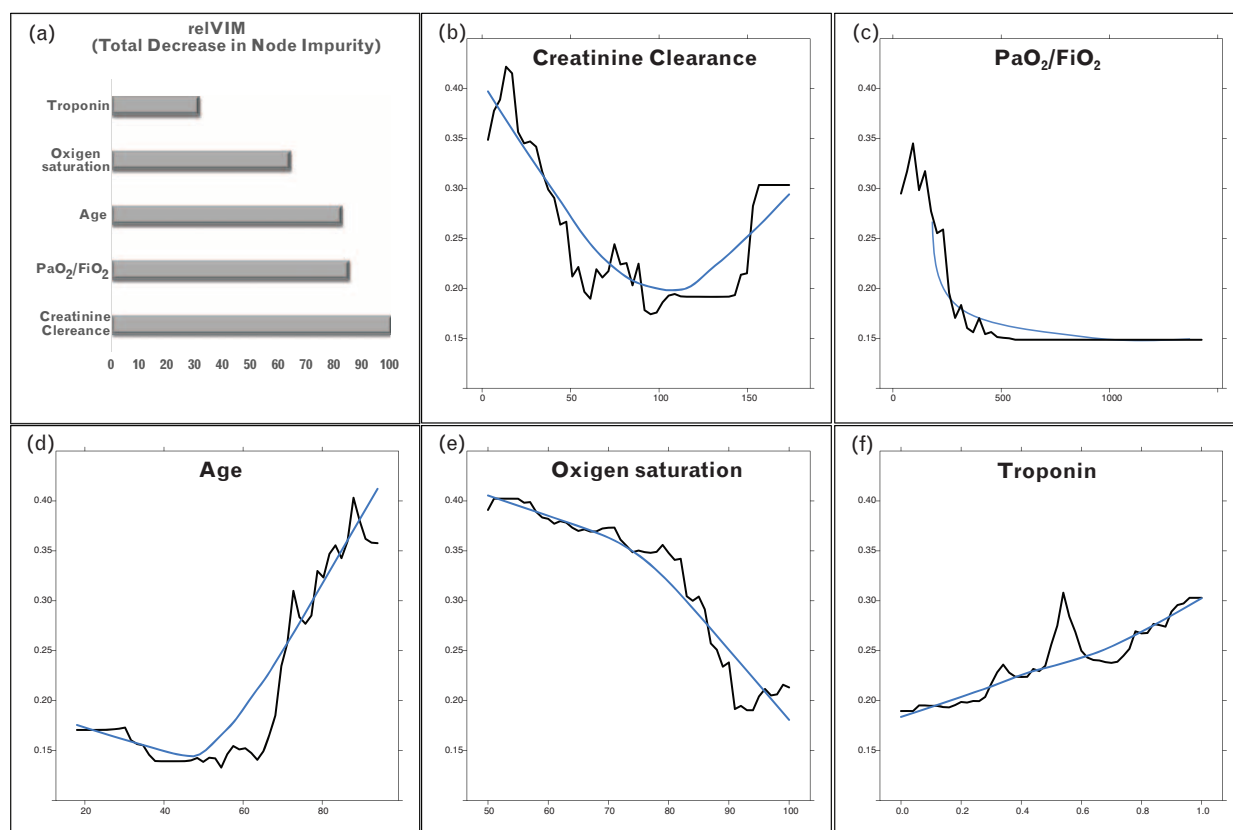
In our study, we developed a machine learning-based risk score to predict mortality among COVID-19 patients hospitalized in several Italian cardiology units. Machine learning methods have the advantage of overcoming problems related to multicollinearity, missing values,

and mixed-type variables (qualitative and quantitative). Moreover, they model nonlinear relationships between outcome and covariates and the predictions obtained are more accurate with respect to classical models (e.g. logistic regression). Combining machine learning algorithms with the Lasso procedure that selects only the strongest predictors of the outcome produces an in-hospital death risk score easily interpretable and based on data that could simply be acquired on admission to the cardiology units.

The proposed score is based on readily available clinical characteristics to be screened at hospital admission. The model achieved a good statistical performance out of a sample with an AUC of 0.78 (95% CI: 0.68–0.89) and a sensitivity of 0.88 (95% CI: 0.58–1.00), proving a not inferior performance to existing models in literature. In fact, during the COVID-19 pandemic, several risk stratification scores using clinical parameters have been elaborated to predict severity and in-hospital mortality, allowing a better management of patients and optimization of resource allocation.

A robust and validated risk stratification model has been proposed by the ISARIC 4C investigators, the 4C Mortality Score, which is based on eight variables readily available at initial hospital assessment: age, sex, number of comorbidities, respiratory rate, peripheral oxygen saturation, level of consciousness, urea level and C-reactive protein.²¹ The model showed high discrimination for in-hospital death with an AUC of 0.79 (95% CI: 0.78–0.79) within the derivation cohort. Similar discriminatory capacity has been achieved by the Piacenza score, consisting of six variables (age, mean corpuscular hemoglobin concentration, $\text{PaO}_2/\text{FiO}_2$ ratio, temperature, previous stroke and gender).²² The score has been created to predict mortality in 852 patients hospitalized for COVID-19 pneumonia and achieved comparable

Fig. 3



Relative Variable Importance (relVIM) and Partial Dependence Plots (PDPs), one for each covariate, extracted from Random Forest. The five PDPs (one for each covariate in the model), (b–f) show the nonlinear relationships between a variable (on x -axis) and the risk score (called mortality risk score and reported on y -axis) of in-hospital death for COVID-19 patients obtained from Random Forest (RF) algorithm. The nonlinear functions are represented by the black lines whereas loess smooth functions (grey lines) are added to each graph.

predictive power, despite the smaller number of patients considered, with an AUC of 0.78 (95% CI: 0.74–0.84, Brier score = 0.19), sensitivity of 94% and specificity of 37%. Yuan *et al.*²³ developed an easier prognostic risk score that included uniquely laboratory markers: lactate dehydrogenase, high-sensitivity C-reactive protein, and lymphocyte percentage. The accuracy in predicting the risk of mortality was more than 95%, but did not take into account clinical conditions and comorbidities that could significantly affect outcomes. Moreover, the cohort under investigation was predominantly represented by patients with severe or critical COVID-19 disease, and it therefore may not be as accurate in the case of asymptomatic or mild forms. Some differences in variables identified may be related to the population enrolled. In our study, we investigated a large cohort of patients hospitalized for COVID-19 infection in several Italian cardiology units, with variable degrees of disease severity and a higher burden of cardiovascular disease. Our results are consistent with available evidence and show that the main drivers related to COVID-19 mortality are older age, lower oxygen saturation and PaO₂/FiO₂ ratio, lower

creatinine clearance values and elevated serum troponin levels. The higher proportion of cardiovascular comorbidities in our population may have influenced the observed results.

Age and respiratory parameters are common in different models and are crucial in predicting risk of death. In previous reports, hospital mortality ranged from less than 5% among people under the age of 40 years to 35% for patients aged 70–79 years and higher than 60% for those aged over 80 years.²⁴ Older patients are at a higher risk of worse outcome because of the higher burden of comorbidities and frailty. From the PDPs (Fig. 3d), the risk appears higher for those above 60 years old with a linear trend above this cut-off.

After adjustment for age-related risk factors, a 2.7% risk increase for disease severity was observed, without any additional risk for death per year of age.²⁵ The higher susceptibility to COVID-19 may be linked to an age-related defect in immune function and control of viral replication, with prolonged proinflammatory responses and a more pronounced procoagulant state.²⁶

As found in our results, assessment of respiratory parameters is critical for early detection of respiratory failure. SpO₂ 93% or less and PaO₂/FiO₂ less than 300 mmHg are two parameters used to identify severe COVID-19 in adults, in addition to the presence of dyspnea, tachypnea with greater than 30 breaths per minute and evidence of infiltrates in more than 50% of the lung field.²⁷ This has been confirmed from our analysis that showed a similar threshold of severity. From PDPs (Fig. 3c and e), lower values of both PaO₂/FiO₂ and oxygen saturation at admission identify patients with a higher lung involvement and disease severity. The thresholds are consistent from the physiological and clinical perspective with a higher risk below 300 and 85%, respectively. These parameters are essential for early recognition of respiratory insufficiency and to optimize the breathing support, allowing a better prognosis.^{28,29}

In our study, impaired renal function has also proved to be a relevant factor. This may in part be related to the high percentage of renal insufficiency encountered in our population and to a greater susceptibility to the development of kidney damage in patients with cardiovascular disease.

It has been shown that patients with concomitant kidney disease have a significantly higher risk of in-hospital death. In fact, elevated serum creatinine, elevated blood urea nitrogen, acute kidney injury at baseline, as well as the presence of proteinuria and/or hematuria are independent risk factors for in-hospital mortality after adjusting for confounders.³⁰ Lower levels of creatinine clearance on admission could also be related to direct kidney involvement by SARS-CoV-2 infection and secondary systemic effects.³¹

Finally, cardiac involvement and detection of myocardial injury are common findings in COVID-19, reported in about 7.2% of all patients and, in particular, in 22.2% of those admitted to the ICU.^{32,33} Increased troponin concentration on admission represents a marker of disease severity and may predict a worse outcome, as it is associated with mortality and elevated risk of cardiovascular and noncardiovascular complications (i.e. heart failure, sepsis, acute kidney failure, multiorgan failure, pulmonary embolism, delirium, major bleeding), irrespective of concomitant cardiac disease.^{4,34}

Limitations

Our population was characterized by a significant burden of cardiovascular comorbidities justifying the relatively high rate of death. Data derive from the first pandemic wave when no effective treatment strategies were available. This should be considered for the interpretation of the results. Our analysis lacks postdischarge follow-up data, thus we could not assess long-term mortality.

Conclusion

In a large COVID-19 population, we showed that a customizable machine learning-based score derived from

clinical variables is feasible and effective for the prediction of in-hospital mortality.

This score showed good performances in terms of sensitivity and involves only five clinical parameters that may be obtained quickly at the emergency department, supporting clinicians in identifying patients with a higher risk of developing complications, which might need a more aggressive treatment.

Conflicts of interest

The authors have no conflicts of interest regarding this paper.

References

- Berlin DA, Gulick RM, Martinez FJ. Severe Covid-19. *N Engl J Med* 2020; **383**:2451–2460.
- Inciardi RM, Adamo M, Lupi L, et al. Characteristics and outcomes of patients hospitalized for COVID-19 and cardiac disease in Northern Italy. *Eur Heart J* 2020; **41**:1821–1829.
- Inciardi RM, Lupi L, Zaccone G, et al. Cardiac involvement in a patient with coronavirus disease 2019 (COVID-19). *JAMA Cardiol* 2020; **5**:819–824.
- Lombardi CM, Carubelli V, Iorio A, et al. Association of troponin levels with mortality in Italian patients hospitalized with coronavirus disease 2019: results of a multicenter study. *JAMA Cardiol* 2020; **5**:1274–1280.
- Tomasoni D, Inciardi RM, Lombardi CM, et al. Impact of heart failure on the clinical course and outcomes of patients hospitalized for COVID-19: results of the Cardio-COVID-Italy multicentre study. *Eur J Heart Fail* 2020; **22**:2238–2247.
- Nuzzi V, Merlo M, Specchia C, et al. The prognostic value of serial troponin measurements in patients admitted for COVID-19. *ESC Heart Fail* 2021; **8**:3504–3511.
- Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009; **150**:604–612.
- Stekhoven DJ, Bühlmann P. MissForest—nonparametric missing value imputation for mixed-type data. *Bioinformatics* 2012; **28**:112–118.
- Dancelli L, Manisera M, Vezzoli M. On two classes of weighted rank correlation measures deriving from the Spearman's ρ . *Studies in classification, data analysis, and knowledge organization*. New York: Springer; 2013.
- Salvi A, Vezzoli M, Busatto S, et al. Analysis of a nanoparticle-enriched fraction of plasma reveals miRNA candidates for Down syndrome pathogenesis. *Int J Mol Med* 2019; **43**:2303–2318.
- Codenotti S, Vezzoli M, Poliani PL, et al. Caveolin-1, Caveolin-2 and Cavin-1 are strong predictors of adipogenic differentiation in human tumors and cell lines of liposarcoma. *Eur J Cell Biol* 2016; **95**:252–264.
- Vezzoli M, Ravaggi A, Zanotti L, et al. RERT: a novel regression tree approach to predict extrauterine disease in endometrial carcinoma patients. *Sci Rep* 2017; **7**:10528.
- Carpita M, Vezzoli M. Statistical evidence of the subjective work quality: the fairness drivers of the job satisfaction. *Electron J Appl Stat Anal* 2012; **5**:89–107.
- Abate G, Vezzoli M, Polito L, et al. A conformation variant of p53 combined with machine learning identifies alzheimer disease in preclinical and prodromal stages. *J Pers Med* 2021; **11**:1–16.
- Garrafa E, Vezzoli M, Ravanelli M, et al. Early prediction of in-hospital death of COVID-19 patients: a machine-learning model based on age, blood analyses, and chest x-ray score. *eLife* 2021; **10**:e70640.
- Breiman L. Random Forests. *Mach Learn* 2001; **45**:5–32.
- Vezzoli M. Exploring the facets of overall job satisfaction through a novel ensemble learning. *Electron J Appl Stat Anal* 2011; **4**:23–38.
- Savona R, Vezzoli M. Fitting and forecasting sovereign defaults using multiple risk signals. *Oxf Bull Econ Stat* 2015; **77**:66–92.
- Azzolina D, Ileana B, Giulia B, et al. Machine learning in clinical and epidemiological research: isn't it time for biostatisticians to work on it? *Epidemiol Biostat Public Health* 2019; **16**:e13245-1–e13245-3.
- Friedman JH. Greedy Function Approximation: a gradient boosting machine. *Ann Stat* 2001; **29**:1189–1232.
- 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; **370**:m3339.

- 22 Halasz G, Sperti M, Villani M, *et al.* A machine learning approach for mortality prediction in COVID-19 pneumonia: development and evaluation of the Piacenza score. *J Med Internet Res* 2021; **23**:e29058.
- 23 Yuan Y, Sun C, Tang X, *et al.* Development and validation of a prognostic risk score system for COVID-19 inpatients: a multi-center retrospective study in China. *Eng Beijing China* 2020; **8**:116–121.
- 24 Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. *JAMA* 2020; **324**:782–793.
- 25 Romero Starke K, Petereit-Haack G, Schubert M, *et al.* The age-related risk of severe outcomes due to COVID-19 infection: a rapid review, meta-analysis, and meta-regression. *Int J Environ Res Public Health* 2020; **17**: E5974.
- 26 Opal SM, Girard TD, Ely EW. The immunopathogenesis of sepsis in elderly patients. *Clin Infect Dis* 2005; **41** (Suppl 7):S504–512.
- 27 Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 2020; **323**:1239–1242.
- 28 Gibson PG, Qin L, Pua SH. COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS. *Med J Aust* 2020; **213**:54.e1–56.e1.
- 29 Grasselli G, Greco M, Zanella A, *et al.* Risk factors associated with mortality among patients with COVID-19 in intensive care units in Lombardy, Italy. *JAMA Intern Med* 2020; **180**:1345–1355.
- 30 Cheng Y, Luo R, Wang K, *et al.* Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* 2020; **97**:829–838.
- 31 Wang M, Xiong H, Chen H, Li Q, Ruan XZ. Renal injury by SARS-CoV-2 infection: a systematic review. *Kidney Dis* 2021; **7**:100–110.
- 32 Wang D, Hu B, Hu C, *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 2020; **323**:1061–1069.
- 33 Inciardi RM, Solomon SD, Ridker PM, Metra M. Coronavirus 2019 disease (COVID-19), systemic inflammation, and cardiovascular disease. *J Am Heart Assoc* 2020; **9**:e017756.
- 34 Nie S-F, Yu M, Xie T, *et al.* Cardiac troponin I is an independent predictor for mortality in hospitalized patients with COVID-19. *Circulation* 2020; **142**:608–610.