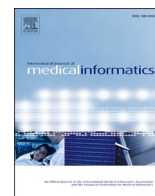




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Outcome prediction during an ICU surge using a purely data-driven approach: A supervised machine learning case-study in critically ill patients from COVID-19 Lombardy outbreak

Massimiliano Greco^{a,b,1}, Giovanni Angelotti^{b,1}, Pier Francesco Caruso^{a,b,*}, Alberto Zanella^{c,d}, Niccolò Stomeo^{a,b}, Elena Costantini^b, Alessandro Protti^{a,b}, Antonio Pesenti^{c,d}, Giacomo Grasselli^{c,d}, Maurizio Cecconi^{a,b}, on behalf of the Lombardy ICU Network

^a Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini 4, 20072 Pieve Emanuele, Milan, Italy

^b IRCCS Humanitas Research Hospital, Via Manzoni 56, 20089 Rozzano, Milan, Italy

^c Dipartimento di Anestesia, Rianimazione ed Emergenza-Urgenza, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

^d Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy

ARTICLE INFO

Keywords:

COVID-19
Machine learning
ICU management
Emergency organization
Outcomes
Epidemiology

ABSTRACT

Purpose: COVID-19 disease frequently affects the lungs leading to bilateral viral pneumonia, progressing in some cases to severe respiratory failure requiring ICU admission and mechanical ventilation. Risk stratification at ICU admission is fundamental for resource allocation and decision making. We assessed performances of three machine learning approaches to predict mortality in COVID-19 patients admitted to ICU using early operative data from the Lombardy ICU Network.

Methods: This is a secondary analysis of prospectively collected data from Lombardy ICU network. A logistic regression, balanced logistic regression and random forest were built to predict survival on two datasets: dataset A included patient demographics, medications before admission and comorbidities, and dataset B included respiratory data the first day in ICU.

Results: Models were trained on 1484 patients on four outcomes (7/14/21/28 days) and reached the greatest predictive performance at 28 days (F1-score: 0.75 and AUC: 0.80). Age, number of comorbidities and male gender were strongly associated with mortality. On dataset B, mode of ventilatory assistance at ICU admission and fraction of inspired oxygen were associated with an increase in prediction performances.

Conclusions: Machine learning techniques might be useful in emergency phases to reach good predictive performances maintaining interpretability to gain knowledge on complex situations and enhance patient management and resources.

1. Introduction

Towards the end of 2019, a novel strand of coronavirus, named Severe Acute Respiratory Syndrome coronavirus-2 (SARS-CoV-2), was identified as the causative agent of an outbreak of bilateral pneumonia in the city of Wuhan in China [1]. The clinical picture related with SARS-CoV-2 infection, was subsequently named COVID-19 disease, and is frequently characterized by severe bilateral pneumonia. The epidemic spread outside mainland China to an increasing number of countries, and on March 11th, 2020, it was declared a pandemic [2].

Lombardy region in Italy was the epicentre of the first outbreak of COVID-19 in the Western World. In Lombardy, the first cases were recognized at the end of February, and the number of Intensive Care Unit admissions rose substantially in the following weeks [3].

The outcomes of patients admitted to ICU for COVID-19 disease are severe, and comparable with those of patients with severe Acute Respiratory Distress Syndrome (ARDS), with mortality up to 50% in patients requiring mechanical ventilation [4-7]. Several factors have been associated with a negative outcome, including age, male gender, previous comorbidities, and level of respiratory support at ICU admission [4,8,9].

* Corresponding author at: Department of Biomedical Sciences, Humanitas University, Humanitas Clinical and Research Center-IRCCS.

E-mail address: pierfrancesco.caruso@humanitas.it (P.F. Caruso).

¹ Massimiliano Greco and Giovanni Angelotti joint co-first author.

Machine learning algorithms are increasingly employed in clinical medicine due to their potential of analysing large amount of information with reduced human supervision, resulting in high predictive performance [10,11]. This kind of models can similarly help to hasten data cleaning and finetuning of predictive models, a process which would normally require weeks of data cleaning and exploratory analysis, while time and human resources are scarce during an emergency.

On the contrary, a purely data-driven approach applied through artificial intelligence could yield good predictive performance while using less resources and in a lower time [12]. Better use of scarce resource through artificial intelligence could be useful both for the healthcare system, to enhance the allocation of resources, and for patients, to rapidly target the best therapeutic strategies with realistic goals. We tested the feasibility and performance of a purely data-driven machine learning model to predict mortality using emergency operative data.

2. Methods

This is a secondary analysis of data collected during the COVID-19 Lombardy outbreak from February 2020 to April 2020 using operational and clinical data from the Lombardy ICU network, as described in previous studies [3,4,13]. The aim of this study is to predict survival at 7, 14, 21 and 28 days from ICU admission, using several different supervised learning frameworks and comparing them to baseline models. Data on patient baseline characteristics, including medications, comorbidities and baseline ventilation parameters are included in the analysis. Data are described as mean (standard deviation) or frequency (percentage), as appropriate.

We first conducted univariate analysis of the data, testing associations with Chi-square test for categorical variables and Mann-Whitney U test for continuous variables. We chose non-parametric tests due to the nature of the data that do not follow a gaussian distribution according to the Shapiro-Wilk test for normality. Survival analysis was conducted plotting Kaplan-Meier curves. We built two different models' (model A and B), to predict survival at four different timepoints: 7-day, 14-day, 21-day and 28-day mortality from ICU admission.

Missing data were a small minority (<1.2%) in our categorical

variables and were imputed using the Simple Imputer; due to the categorical nature of the features, we opted for a median imputation.

Patients whose data from the ICU were missing were excluded as reported in the inclusion and exclusion criteria presented in Fig. 1. To assess the type of missingness of those data, we trained a multivariate model to test the hypothesis that patients with missing data belonged to a different cohort compared to those without missing data. The model, validated with 10-fold cross validation, and trained on baseline covariates, such as age, gender, and comorbidities, was not able to differentiate the two groups with average Area Under the Curve (AUC) of 0.5. Those results may indicate that missing data is not a result of clinical severity thus pointing to a missing at random (MAR) distribution.

Model A included only baseline patient data (age, gender, home medications and comorbidities); Model B included baseline data (the same data used to create model A) and ventilation parameters from the first 24 h in ICU: the list of training variables is presented in Table 1.

We created four subsequent models with increasing complexity. A baseline model was created using a uniform classifier (UNIF), a simple classifier commonly employed as baseline in ML models [14]. Our first prediction model was a logistic regression (LR). The second model was a composed logistic regression with under-over-sampling strategy to balance our classes known as a SMOTE-Tomek [15] (BAL-LR).

Lastly, we trained a random forest classifier with balanced-class weight (RF) [14].

Every technique created two different models (A and B) and every model was trained using nested cross-validation (10-fold each). Hyperparameters were optimized to maximize the out-of-fold F1-Score on a randomized grid space. The main scores to track models' performances were F1-score and area under the curve (AUC) and all the scores presented in the results section are on test sets.

3. Results

Table 1 [16] describes our cohort that included a total of 1484 patients. Overall mortality at 28 days was 49% ($n = 741$). Survivors were significantly younger and suffered fewer comorbidities at admission compared to non-survivors ($p < 0.001$). 44% of survivors had no pre-existing comorbidities, compared to 25% of non-survivors ($p <$

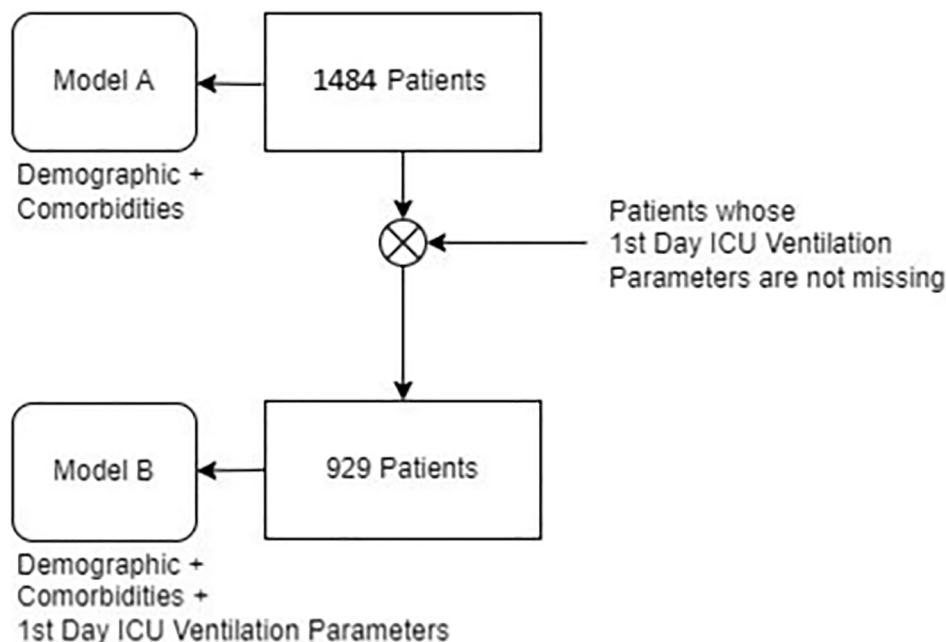


Fig. 1. Inclusion criteria diagram of our cohort population. Model A was built using data from 1484 patients, while Model B was trained on a reduced cohort (929 patients).

Table 1

Description of covariates characteristics according to survival analysis. IQR: 'InterQuartile Range', SD: 'Standard Deviation', COPD: 'Chronic Obstructive Pulmonary Disease', CKD: 'Chronic Kidney Disease', ACE: 'Angiotensin Converting Enzyme', ARBs: 'Angiotensin receptor blockers', PEEP: 'Positive End-Expiratory Pressure', FiO2: 'Fraction of inspired oxygen', PaO2: 'Partial pressure of oxygen', P/F: 'PaO2/FiO2', CPAP: 'Continuous positive airway pressure', IMV: 'Invasive Mechanical Ventilation', NIV: 'Non Invasive Ventilation', SB: 'Spontaneous Breathing'.

Variables	Categories	Overall	Survivors	Non-Survivors	p-value
Training variables present in both Model A and B					
Number of Patients		1484	743	741	
Age, median [IQR]		64 [56,70]	59 [52,65]	68 [62,73]	<0.001
Gender, n (%)	Females	273 (18.4)	160 (21.5)	113 (15.2)	0.002
	Males	1211 (81.6)	583 (78.5)	628 (84.8)	
Hypercholesterolemia, n (%)	No	1182 (80.8)	645 (87.5)	537 (74.1)	<0.001
	Yes	280 (19.2)	92 (12.5)	188 (25.9)	
Diabetes Mellitus, n (%)	No	1207 (82.6)	647 (87.8)	560 (77.2)	<0.001
	Yes	255 (17.4)	90 (12.2)	165 (22.8)	
Malignancy, n (%)	No	1297 (88.7)	675 (91.6)	622 (85.8)	0.001
	Yes	165 (11.3)	62 (8.4)	103 (14.2)	
Cardiac Disease, n (%)	No	1175 (80.4)	640 (86.8)	535 (73.8)	<0.001
	Yes	287 (19.6)	97 (13.2)	190 (26.2)	
COPD, n (%)	No	1406 (96.2)	723 (98.1)	683 (94.2)	<0.001
	Yes	56 (3.8)	14 (1.9)	42 (5.8)	
CKD, n (%)	No	1416 (96.9)	730 (99.1)	686 (94.6)	<0.001
	Yes	46 (3.1)	7 (0.9)	39 (5.4)	
Hepatic Disease, n (%)	No	1418 (97.0)	710 (96.3)	708 (97.7)	0.186
	Yes	44 (3.0)	27 (3.7)	17 (2.3)	
N° of comorbidities, mean (SD)		1.4 (1.4)	1.1 (1.2)	1.8 (1.5)	<0.001
>= 1 antihypertensives agents, n (%)	No	698 (47.4)	420 (56.6)	278 (38.1)	<0.001
	Yes	774 (52.6)	322 (43.4)	452 (61.9)	
ACE inhibitors, n (%)	No	1137 (77.2)	625 (84.2)	512 (70.1)	<0.001
	Yes	335 (22.8)	117 (15.8)	218 (29.9)	
ARBs, n (%)	No	1135 (77.1)	596 (80.3)	539 (73.8)	0.004
	Yes	337 (22.9)	146 (19.7)	191 (26.2)	
Beta Blockers, n (%)	No	1112 (75.5)	603 (81.3)	509 (69.7)	<0.001
	Yes	360 (24.5)	139 (18.7)	221 (30.3)	
Statins, n (%)	No	1104 (75.0)	619 (83.4)	485 (66.4)	<0.001
	Yes	368 (24.9)	123 (16.4)	245 (33.6)	
Immunosuppressors, n (%)	No	1318 (89.5)	676 (91.1)	642 (87.9)	0.058
	Yes	154 (10.5)	66 (8.9)	88 (12.1)	
Diuretics, n (%)	No	1354 (92.0)	706 (95.1)	648 (88.8)	<0.001
	Yes	118 (8.0)	36 (4.9)	82 (11.2)	
Training variables present in Model B only					
PEEP, median [IQR]		14.0 [12.0,16.0]	14.0 [12.0,15.0]	14.0 [12.0,16.0]	<0.001
FiO2, median [IQR]		70.0 [50.0,80.0]	60.0 [50.0,75.0]	75.0 [60.0,90.0]	<0.001
PaO2, median [IQR]		100.0 [80.0,136.0]	100.2 [80.0,141.0]	98.6 [80.0,129.2]	0.286
P/F, median [IQR]		160.0 [115.8,220.0]	177.5 [124.2,240.8]	146.0 [106.2,200.0]	<0.001
CPAP, n (%)	No	1414 (95.3)	694 (93.4)	720 (97.2)	0.001
	Yes	70 (4.7)	49 (6.6)	21 (2.8)	
IMV, n (%)	No	434 (29.2)	243 (32.7)	191 (25.8)	0.004
	Yes	1050 (70.8)	500 (67.3)	550 (74.2)	
NIV, n (%)	No	1429 (96.3)	713 (96.0)	716 (96.6)	0.590
	Yes	55 (3.7)	30 (4.0)	25 (3.4)	
SB, n (%)	No	1471 (99.1)	731 (98.4)	740 (99.9)	0.005
	Yes	13 (0.9)	12 (1.6)	1 (0.1)	

0.001). Male gender was associated with increased mortality ($p = 0.002$). All comorbidities were more common in non-survivors except for hepatic disease. Among home medications all drugs analyzed were more common in non-survivors ($p < 0.005$) except for immunosuppressors ($p = 0.058$). Concerning respiratory data, survivors were more often treated with Continuous Positive Airway Pressure (CPAP) and less often with invasive mechanical ventilation on the first day of their admissions. ($p < 0.005$).

Fig. 2 represents the Kaplan-Meier analysis stratified by age. Age was strongly associated with mortality, with 7-day survival ranging from 64% in the oldest age group to 93% in the 30–40 years group, with differences progressively increasing at 14 and 28-days survival. In the supplemental material, the overall Kaplan-Meier analysis shows a reduction in survival to 70% at 10 days and to 55% at 20 days from ICU admission, continuing as a plateau thereafter (Supplemental Figure 1). When considering gender in Kaplan-Meier analysis, females gender resulted in higher survival. (Supplemental Figure 2).

Comparison in F1-score and area under the curve for type A and B of the four models are reported in Fig. 3. Comparison between models A

and B of each model versus the baseline UNIF model are reported in Supplemental Figure 3. Comparison between models A and B of each model among each other are reported in Fig. 4.

All the three LR, BAL-LR and RF models performed better than baseline UNIF models according to the F1-score and AUC ($p < 0.01$) (Table 2). The only one exception was the F1-score of LR model at 7 days, that performed poorly due to low recall.

BAL-LR and RF had higher performance than LR when the outcome was set to 7 days ($p < 0.01$) and to 14 days for both model A ($p < 0.001$) and model B ($p < 0.03$). After that timeline the three models performed without any statistically significant difference (Fig. 4) and, as time intervals increased, F1-score and Area Under the Curve progressively increased as well for all the three models in both A and B.

BAL-LR model at 7-days model A and B yielded an F1-score of 0.35 and 0.45 and AUC of 0.68 and 0.76, respectively (Table 2). The performance increased at 14 days, with average F1-score 0.56 (model A) and 0.61 (model B) and AUC 0.72 (model A) and 0.75 (model B), and again at 21 days. The greatest predictive performance was reached at 28-days, with average F1-score of 0.71 (model A) and 0.75 (model B)

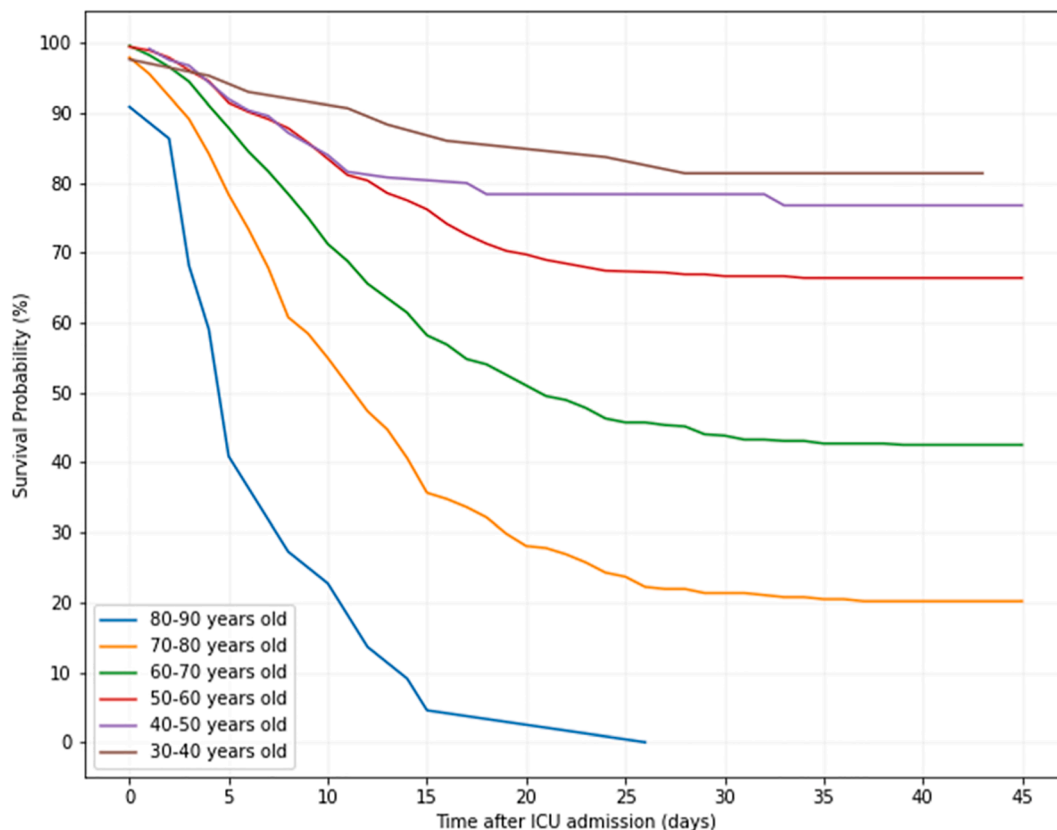


Fig. 2. Kaplan-Maier curve stratified by age.

AUC 0.77 for model A and 0.80 for model B. Precision and recall were balanced to maximize performances in all our models with only two exceptions with LR at 7 and 14 days where recall was very low compared to precision.

Fig. 5a reports the odds ratio of BAL-LR model: all the values higher than 1 were positively associated with death in ICU while those below 1 were considered protective by our models.

Fig. 5b reports the average features importance of RF model; in contrast with logistic regression models, it is not possible in RF model to reconstruct how the model considered each included variables, due to the nature of Random Forest. In RF model a numeric value was assigned to each variable to rank its importance in the decision process of the model.

4. Discussion

In this study we demonstrated the potential of a purely data-driven machine learning approach to predict relevant clinical outcomes, reaching good predictive performance.

Machine learning techniques have the advantage of automatic variable selection and model development with reduced human interaction. This can be applied not only in highly controlled settings with clean high-quality data -the best setting for performance of machine learning- but can also be considered in an emergency setting, based on operational data, with good results.

We tested this new approach with different models and techniques (LR, BAL-LR, RF) and compared it with a baseline model (UNIF) to evaluate the prediction capability of the models. We defined an outcome and set four different timepoints: 7-, 14-, 21- and 28- days. Lastly, we decided to evaluate two different data frameworks with every model: the first framework (model A) was trained on a dataset with baseline characteristics of admitted in ICU patients, while the second framework (model B) also included respiratory data gathered during the first day of

admission. We asked to every model 8 different questions that combines the 4 different timepoints (7–14–21–28 days) and the two different data frameworks (A and B).

When we compare the recall of the two models in the first two timepoints to our baseline model (which classify patients similarly to a coin toss), we can conclude that LR models have lower sensitivity compared to coin toss. This happens because imbalanced data classes can affect the predictive capability of methods like logistic regression (LR): this kind of models tend to optimize the overall accuracy without considering the relative distribution of the classes [17]. Accordingly, after 21 days LR seems to be reliable because the classes of the population considered happen to be balanced. On the contrary, techniques which allow balancing of data, such as BAL-LR (based on SMOTE-Tomek technique) or models that do not suffer from the class imbalance problem, like RF, can represent an advantage in prediction models, allowing increased performances regardless the distribution of the classes. Thus, we can conclude that BAL-LR and RF outperform UNIF in any single moment, while LR models are reliable only after 21-days (as confirmed by Fig. 4).

Models that were trained on baseline characteristics and respiratory data (model B) had higher overall performance than their baseline counterpart (model A), as shown in Fig. 3. The p-values calculated over the nested-cross validation were almost always statistically significant for the F1-score. Thus, the analysis of the respiratory condition at admission might be clinically relevant to establish survival predictions and correctly classify COVID-19 patients.

RF models were trained with very tight rules to avoid overfitting as the amount of data to train was limited. Since Random Forests often work better for the analysis of big data. we can suppose that higher amount of data would have probably allowed random forests to outperform the balanced logistic regression (BAL-LR).

In conclusion, as we are proposing a method that can be applied in any emergency setting with operational data, we would suggest training

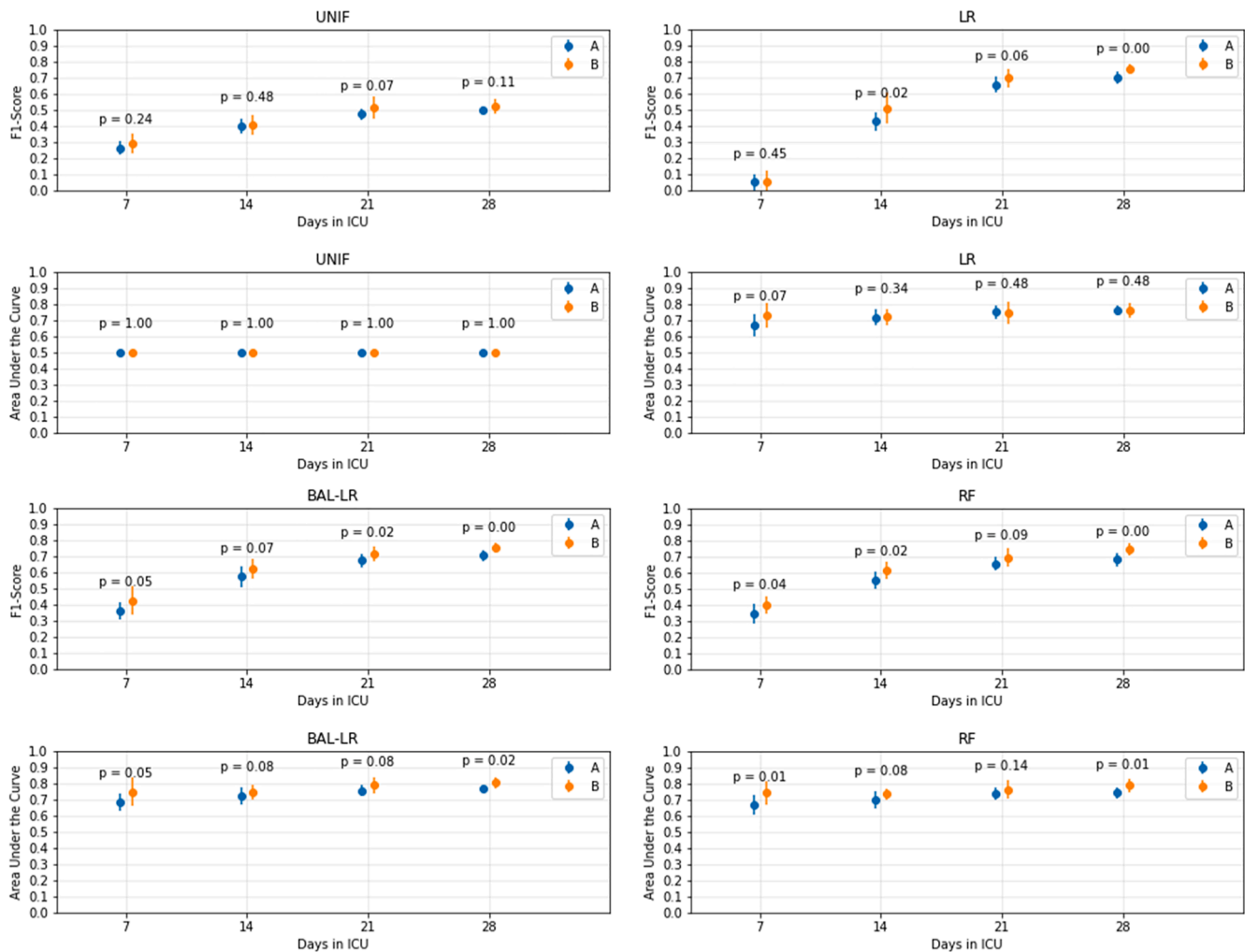


Fig. 3. Comparison of F1-score and Area Under the Curve between model A and model B for every model proposed. The point represents the mean, while the vertical line stands for the +/- 1 standard deviation among all cross-validation. UNIF: ‘Uniform Dummy Classifier’, LR: ‘Logistic Regression’, BAL-LR: ‘Balanced Logistic Regression’, RF: ‘Random Classifier’.

both models and modify the parameters of the forest according to the dimension of the dataset.

An analysis of how decisions were taken when predicting mortality at 28 days was performed (Fig. 5a and 5b): BAL-LR and RF models confirm that age is the strongest predictor of ICU survival in COVID-19 patients.

In both models B and A age was the most relevant feature and the most associated with mortality. The strong association between age and mortality is a constant finding in COVID-19 literature [4,18].

Chronic kidney disease (CKD) is highly correlated with mortality in our data (as shown in Table1), and this may be related to several factors. CKD affects older patients [19] that, as demonstrated widely [3] and confirmed by our models, are particularly fragile when hospitalized for COVID-19. Secondly, all stages of CKD are associated with an increased risk of premature mortality from all causes [20] and thirdly, CKD is associated in up to two-thirds of the cases with diabetes and hypertension [21], a proxy for older, multi-morbid patients [4]. COVID-19 disease is also associated with new onset acute kidney injury, that may further worsen previous kidney dysfunction, leading to organ failure [22,23].

Regarding type B ICU admission models, the level of oxygen therapy (FiO2) was highly correlated with the outcome, as it represents a proxy of severity. To a minor extent, there was positive association with end expiratory pressure (PEEP), while on the opposite side, an increased P/F ratio and an initial admission with a continuous positive airway pressure

(CPAP), non-invasive ventilation (NIV) or spontaneous breathing were associated with survival. All these features are a proxy of reduced severity because these patients needed an ICU hospitalization but were less critical than the rest of population in the ward.

Male gender is a negative predicting factor, a finding confirmed by previous studies [24]. An etiological justification might be linked to a difference between the sexes in cellular immunity as males present a poorer T-cell activation and an increase in proinflammatory cytokines, but further studies are required on the topic [25].

Chronic therapy with ACE inhibitors were associated with higher mortality. Initial reports linked the possible pharmacodynamics of this class of drugs to an up-regulation of ACE2 expression [26] and a consequent increase in the availability of target molecules for SARS-CoV-2 [27]. This association has been proven wrong by Mancia et al. [28], who performed a large population-based case-control study demonstrating that use of ACE inhibitors and ARBs was more frequent among COVID-19 patients due to their higher prevalence of cardiovascular disease, without evidence linking those drugs to a higher risk of infection by SARS-CoV-2.

According to BAL-LR, we found an association between chronic obstructive pulmonary disease (COPD) and mortality. COPD patients have both an increased risk of COVID-19 disease, and a poorer prognosis, with higher rates of hospitalization and mortality [29]. COPD is an independent predictor of mortality in patients admitted to ICU for COVID-19 pneumonia [4].

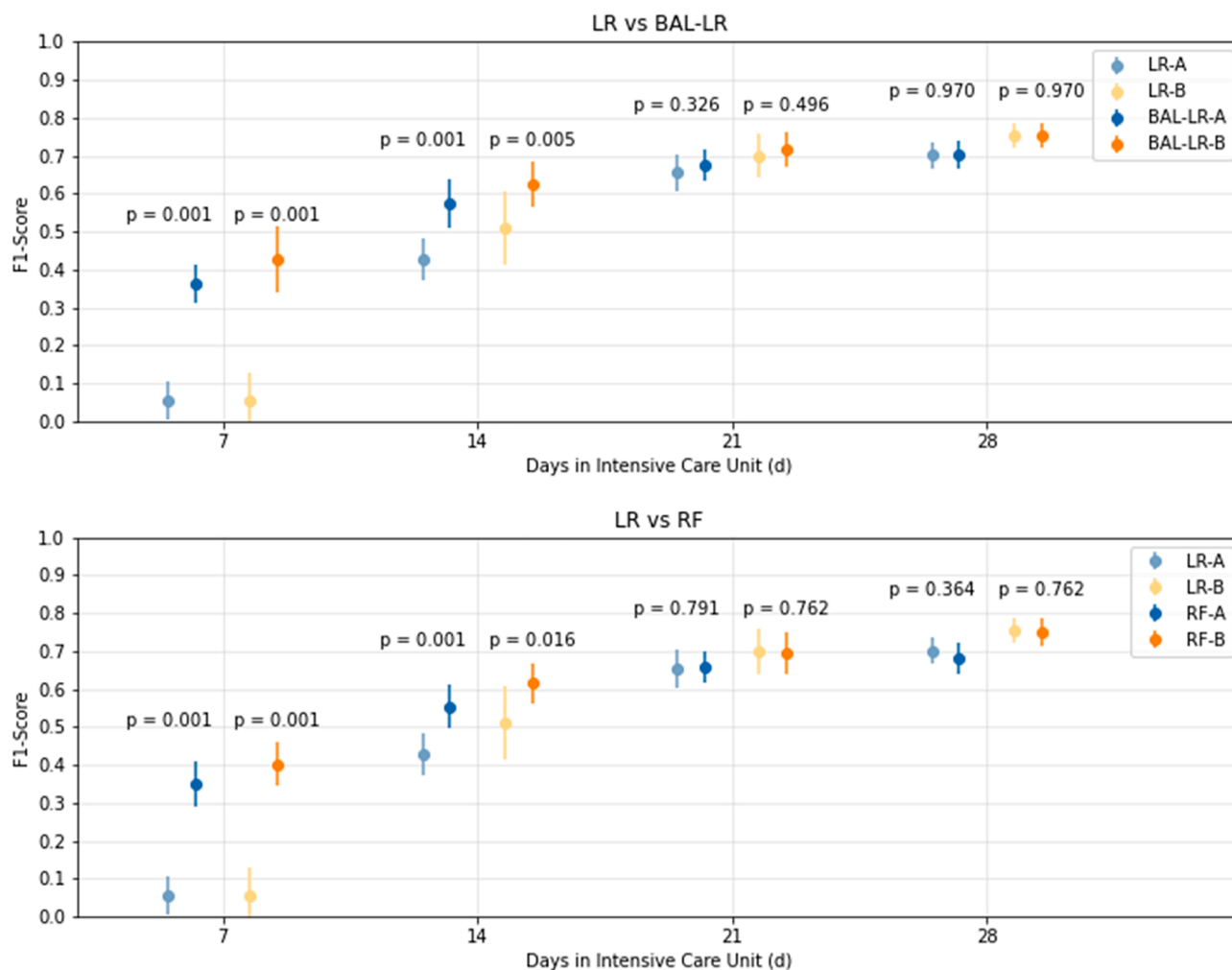


Fig. 4. Differences between Logistic Regression (LR) and the other two models (Balanced Logistic Regression (BAL-LR) and Random Forest classifier (RF)). There was no significance difference between BAL-LR, and RF. P-values refer to differences in results between LR and the second model (BAL-LR and RF, respectively). The point represents the mean, while the vertical line stands for the +/- 1 standard deviation among all cross-validation.

Diabetes Mellitus is associated with mortality in our results. Type-2 diabetes mellitus is more frequent in older patients, male gender, and is part of the metabolic syndrome with hypertension and obesity, which was previously demonstrated to have strong association with COVID-19 outcomes [30]. The association of diabetes and survival has been questioned by other studies, where the association was lost after controlling for other factors [4,31].

A higher number of comorbidities was associated with increased risk of death. As most of the comorbidities were associated with an increased risk, it does not surprise that their sum leads to an increased overall risk. In concordance with another work performed on the COVID-19 Lombardy population 4, most of the comorbidities we analyzed are associated with an increased risk of death with very few exceptions (e.g., hepatic diseases).

5. Limitations

The study presents several limitations. First, it is an observational study based on operative data collected during an emergency crisis by a regional coordination center, hindering the quality of data assured by a research targeted database. Despite being a limiting factor, our aim was to assess the ability of machine learning models on operational/emergency data collected during the escalation phase of the spread of SARS-CoV2, where a hold-out validation could not be retrieved. Some variables that could be useful to increase the predictive performance of the

model were not collected, including more specific data about comorbidities (i.e., CKD Stage, hypertension severity stage) and other physiological parameters (weight, body mass index, more complete ventilatory data, patient frailty). Availability of more data could have improved predictive performances in this population.

The number of patients and data included in this study is not comparable with *big data* analysis, where machine learning techniques really shine. However, with this study we were able to demonstrate that machine learning approach may be used even with smaller dataset in an emergency setting and reach high predictive performance.

6. Conclusions

Supervised machine learning models with a completely data-driven approach may be employed in emergency setting to assess the major risk factors of critical COVID-19 patients, despite sub-optimal numerosity and clean-up of the datasets. We propose a four-step machine learning approach which can be used in similar setting to gain knowledge on complex situations and enhance patient management and resources, sparing resources and time compared to classic statistical techniques.

7. Summary Table

What was already know on the topic?

Table 2

Nested cross-validated performances of the models on the test set. UNIF is the baseline models, LR is the first logistic regression, BAL-LR is the composed balanced logistic regression and RF is a random forest classifier. F1-score is a combined harmonization of precision and recall. SD: ‘Standard Deviation’, AUC: ‘Area Under the Curve’, precision is also known as Positive Predictive Value (PPV), recall also known as sensitivity.

Model Name	UNIF		LR		BAL-LR		RF	
Model	A	B	A	B	A	B	A	B
Time Outcome	7 days		7 days		7 days		7 days	
F1-score mean (SD)	0.27 (0.04)	0.29 (0.06)	0.05 (0.05)	0.06 (0.07)	0.35 (0.06)	0.45 (0.08)	0.35 (0.05)	0.40 (0.06)
AUC mean (SD)	0.5 (0)	0.5 (0)	0.67 (0.07)	0.72 (0.07)	0.68 (0.06)	0.76 (0.08)	0.66 (0.06)	0.73 (0.07)
Precision mean (SD)	0.18 (0.03)	0.19 (0.04)	0.44 (0.45)	0.35 (0.47)	0.25 (0.04)	0.34 (0.07)	0.23 (0.04)	0.30 (0.05)
Recall mean (SD)	0.54 (0.09)	0.58 (0.12)	0.03 (0.03)	0.03 (0.04)	0.63 (0.11)	0.68 (0.13)	0.58 (0.09)	0.65 (0.07)
Time Outcome	14 days		14 days		14 days		14 days	
F1-score mean (SD)	0.40 (0.05)	0.41 (0.06)	0.42 (0.07)	0.52 (0.09)	0.56 (0.06)	0.61 (0.05)	0.55 (0.08)	0.61 (0.06)
AUC mean (SD)	0.5 (0)	0.5 (0)	0.71 (0.04)	0.72 (0.05)	0.72 (0.06)	0.75 (0.04)	0.70 (0.06)	0.72 (0.04)
Precision mean (SD)	0.33 (0.04)	0.36 (0.05)	0.61 (0.10)	0.59 (0.12)	0.49 (0.05)	0.56 (0.04)	0.48 (0.06)	0.54 (0.05)
Recall mean (SD)	0.51 (0.06)	0.47 (0.09)	0.32 (0.06)	0.47 (0.09)	0.66 (0.10)	0.68 (0.09)	0.65 (0.12)	0.79 (0.09)
Time Outcome	21 days		21 days		21 days		21 days	
F1-score mean (SD)	0.47 (0.03)	0.52 (0.07)	0.65 (0.04)	0.70 (0.06)	0.67 (0.04)	0.73 (0.04)	0.66 (0.04)	0.69 (0.07)
AUC mean (SD)	0.5 (0)	0.5 (0)	0.75 (0.04)	0.74 (0.07)	0.75 (0.04)	0.79 (0.05)	0.73 (0.04)	0.75 (0.07)
Precision mean (SD)	0.45 (0.02)	0.51 (0.05)	0.66 (0.05)	0.65 (0.06)	0.65 (0.05)	0.72 (0.05)	0.65 (0.04)	0.68 (0.06)
Recall mean (SD)	0.50 (0.05)	0.53 (0.09)	0.64 (0.05)	0.75 (0.08)	0.70 (0.05)	0.74 (0.07)	0.67 (0.05)	0.69 (0.10)
Time Outcome	28 days		28 days		28 days		28 days	
F1-score mean (SD)	0.50 (0.03)	0.52 (0.05)	0.70 (0.04)	0.75 (0.03)	0.71 (0.04)	0.75 (0.03)	0.68 (0.04)	0.74 (0.05)
AUC mean (SD)	0.5 (0)	0.5 (0)	0.76 (0.03)	0.76 (0.05)	0.77 (0.03)	0.80 (0.03)	0.74 (0.04)	0.78 (0.04)
Precision mean (SD)	0.49 (0.02)	0.55 (0.03)	0.69 (0.03)	0.68 (0.03)	0.70 (0.03)	0.76 (0.06)	0.68 (0.04)	0.74 (0.05)
Recall mean (SD)	0.51 (0.05)	0.50 (0.06)	0.72 (0.06)	0.84 (0.05)	0.72 (0.06)	0.75 (0.07)	0.69 (0.06)	0.71 (0.08)

- COVID-19 patients admitted to ICU have very high mortality.
- Age, gender, and previous comorbidities are associated to negative outcome.
- Resource allocation during surges would be helpful for both patients and healthcare systems.

What this study added to our knowledge?

- Supervised machine learning predictive models perform robustly in predicting mortality at ICU admission for COVID-19 patients.
- Operational/emergency datasets collected during escalation phases can create robust machine learning predictive models even in smaller datasets.
- Easy-to-deploy machine learning pipelines should be created in advance, so that, during emergencies phases, faster insights on the patients admitted in ICU could be retrieved in a completely data-driven manner.

8. Ethics approval and consent to participate

The institutional ethics board of Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico, Milan, approved this study and waived the need for informed consent from individual patients owing to the retrospective nature of the study. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

9. Authors’ contributions

On behalf of the Lombardy ICU Network:
 Concept and design: MG, GA, PFC.
 Acquisition, analysis, or interpretation of data: MG, GA, PFC, AZ.
 Drafting of the manuscript: MG, GA, PFC, NS.
 Critical revision of the manuscript for important intellectual content: AZ, EC, Alessandro Protti, Antonio Pesenti, GG, MC.
 Supervision: Antonio Pesenti, GG, MC.

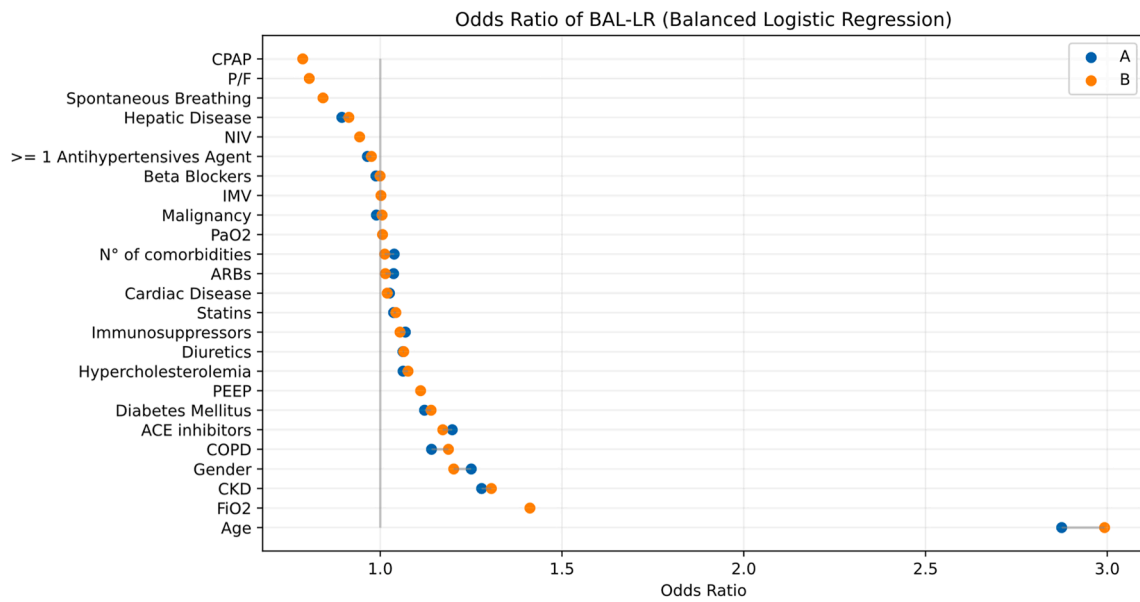


Fig. 5a. Odds Ratio of BAL-LR models for 28 days predictions. CPAP: ‘Continuous positive airway pressure’, P/F: ‘PaO2/FiO2’, NIV: ‘Non Invasive Ventilation’, IMV: ‘Invasive Mechanical Ventilation’, ARBs: ‘Angiotensin receptor blockers’, PaO2: ‘Partial pressure of oxygen’, PEEP: ‘Positive End-Expiratory Pressure’, COPD: ‘Chronic Obstructive Pulmonary Disease’, ACE: ‘Angiotensin Converting Enzyme’, CKD: ‘Chronic Kidney Disease’, FiO2: ‘Fraction of inspired oxygen’.

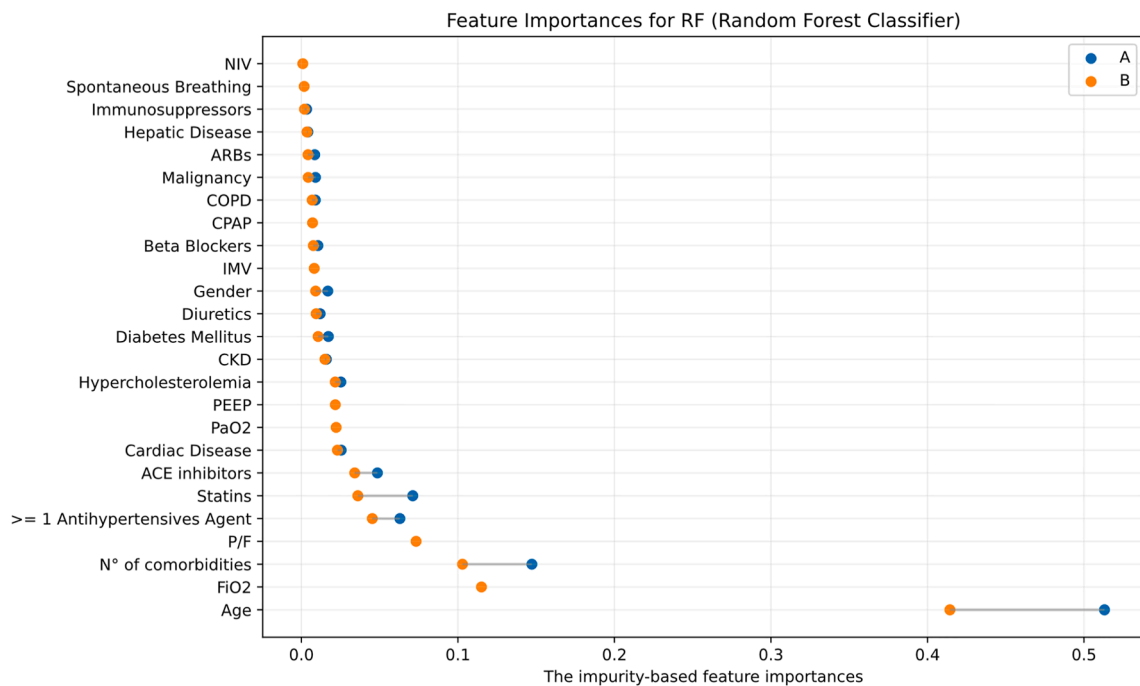


Fig. 5b. Mean feature importance of RF models for 28 days predictions. Differently from odds ratio, there is no association between the importance and the outcome, but it represents how relevant the features were for the model.

CRedit authorship contribution statement

Massimiliano Greco: . Giovanni Angelotti: . Pier Francesco Caruso: . Alberto Zanella: . Niccolò Stomeo: . Elena Costantini: . Alessandro Protti: . Antonio Pesenti: Supervision. Giacomo Grasselli: Supervision. Maurizio Cecconi: Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Acknowledgements

We would like to thank all the collaborators of the Lombardy ICU network.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijmedinf.2022.104807>.

References

- [1] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L.I. Zhang, novel coronavirus pneumonia in Wuhan, China: a descriptive study, *Lancet* 395 (10223) (2020) 507–513.
- [2] WHO announces COVID-19 outbreak a pandemic. Published online March 12, 2020.
- [3] G. Grasselli, A. Zangrillo, A. Zanella, M. Antonelli, L. Cabrini, A. Castelli, D. Cereda, A. Coluccello, G. Foti, R. Fumagalli, G. Iotti, N. Latronico, L. Lorini, S. Merler, G. Natalini, A. Piatti, M.V. Ranieri, A.M. Scandroglio, E. Storti, M. Cecconi, A. Pesenti, E. Agosteo, V. Alaimo, G. Albano, A. Albertin, A. Alborghetti, G. Aldegheri, B. Antonini, E. Barbara, N. Belgioirno, M. Belliato, A. Benini, E. Beretta, L. Bianciardi, S. Bonazzi, M. Borelli, E. Boselli, N. Bronzini, C. Capra, L. Carnevale, G. Casella, G. Castelli, E. Catena, S. Cattaneo, D. Chiumello, S. Cirri, G. Citerio, S. Colombo, D. Coppini, A. Corona, P. Cortellazzi, E. Costantini, R. D. Covello, G. De Filippi, M. Dei Poli, F. Della Mura, G. Evasi, R. Fernandez-Olmos, A. Forastieri Molinari, M. Galletti, G. Gallioli, M. Gemma, P. Gnesin, L. Grazioli, S. Greco, P. Gritti, P. Grosso, L. Guatterri, D. Guzzon, F. Harizay, R. Keim, G. Landoni, T. Langer, A. Lombardo, A. Malara, E. Malpetti, F. Marino, G. Marino, M. Mazzoni, G. Merli, A. Micucci, F. Mojoli, S. Muttini, A. Naulescu, M. Panigada, P. Perazzo, G.B. Perego, N. Petrucci, A. Pezzi, A. Protti, D. Radrizzani, M. Raimondi, M. Ranucci, F. Rasulo, M. Riccio, R. Rona, C. Roscitant, P. Ruggeri, A. Sala, G. Sala, L. Salvi, P. Sebastiano, P. Severgnini, I. Sforzini, F.D. Sigurtà, M. Subert, P. Tagliabue, C. Troiano, R. Valsecchi, U. Viola, G. Vitale, M. Zambon, E. Zoia, Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy, *JAMA* 323 (16) (2020) 1574.
- [4] G. Grasselli, M. Greco, A. Zanella, G. Albano, M. Antonelli, G. Bellani, E. Bonanomi, L. Cabrini, E. Carlesso, G. Castelli, S. Cattaneo, D. Cereda, S. Colombo, A. Coluccello, G. Crescini, A. Forastieri Molinari, G. Foti, R. Fumagalli, G.A. Iotti, T. Langer, N. Latronico, F.L. Lorini, F. Mojoli, G. Natalini, C.M. Pessina, V. M. Ranieri, R. Rech, L. Scudeller, A. Rosano, E. Storti, B.T. Thompson, M. Tirani, P. G. Villani, A. Pesenti, M. Cecconi, E. Agosteo, G. Albano, A. Albertin, A. Alborghetti, G. Aldegheri, B. Antonini, E. Barbara, G. Bardelloni, S. Basilio, N. Belgioirno, G. Bellani, E. Beretta, A. Berselli, L. Bianciardi, E. Bonanomi, S. Bonazzi, M. Borelli, N. Bottino, N. Bronzini, S. Brusatori, L. Cabrini, C. Capra, L. Carnevale, G. Castelli, E. Catena, S. Cattaneo, M. Cecconi, S. Celotti, S. Cerutti, D. Chiumello, S. Cirri, G. Citerio, S. Colombo, A. Coluccello, D. Coppini, A. Corona, P. Cortellazzi, E. Costantini, R.D. Covello, G. Crescini, G. De Filippi, M. Dei Poli, P. Dughi, F. Fieni, G. Florio, A. Forastieri Molinari, G. Foti, R. Fumagalli, M. Galletti, G.A. Gallioli, H. Gay, M. Gemma, P. Gnesin, G. Grasselli, S. Greco, M. Greco, P. Grosso, L. Guatterri, D. Guzzon, G.A. Iotti, R. Keim, T. Langer, N. Latronico, A. Lombardo, F.L. Lorini, F. Mamprin, G. Marino, F. Marino, G. Merli, A. Micucci, C.R. Militano, F. Mojoli, G. Monti, S. Muttini, S. Nadalin, G. Natalini, P. Perazzo, G.B. Perego, L. Perotti, A. Pesenti, C.M. Pessina, N. Petrucci, A. Pezzi, S. Piva, G. Portella, A. Protti, M. Raccagni, D. Radrizzani, M. Raimondi, M. Ranucci, R. Rech, M. Riccio, A. Rosano, P. Ruggeri, G. Sala, L. Salvi, P. Sebastiano, P. Severgnini, D. Sigurtà, N. Stocchetti, E. Storti, M. Subert, M. Tavola, S. Todaro, F. Torigliola, D. Tubiolo, R. Valsecchi, P.G. Villani, U. Viola, G. Vitale, M. Zambon, A. Zanella, E. Zoia, Risk Factors Associated with Mortality among Patients with COVID-19 in Intensive Care Units in Lombardy, Italy, *JAMA Internal Med.* 180 (10) (2020) 1345.
- [5] C. Karagiannidis, C. Mostert, C. Hentschker, T. Voshaar, J. Malzahn, G. Schillinger, J. Klauber, U. Janssens, G. Marx, S. Weber-Carstens, S. Kluge, M. Pfeifer, L. Grabenhenrich, T. Welte, R. Busse, Case characteristics, resource use, and outcomes of 10 021 patients with COVID-19 admitted to 920 German hospitals: an observational study, *Lancet Respir Med.* 8 (9) (2020) 853–862.
- [6] S. Gupta, S.S. Hayek, W. Wang, L. Chan, K.S. Mathews, M.L. Melamed, S. K. Brenner, A. Leonberg-Yoo, E.J. Schenck, J. Radbel, J. Reiser, A. Bansal, A. Srivastava, Y. Zhou, A. Sutherland, A. Green, A.M. Shehata, N. Goyal, A. Vijayan, J.C.Q. Velez, S. Shaefi, C.R. Parikh, J. Arunthamakun, A.M. Athavale, A.N. Friedman, S.A.P. Short, Z.A. Kibbelaar, S. Abu Omar, A.J. Admon, J. P. Donnelly, H.B. Gershengorn, M.A. Hernán, M.W. Semler, D.E. Leaf, C.P. Walther, S.J. Anumudu, K.F. Kopecky, G.P. Milligan, P.A. McCullough, T.-D. Nguyen, M. L. Krajewski, S. Shankar, A. Pannu, J.D. Valencia, S.S. Waikar, P. Hart, O. Ajiboye, M. Itteera, J.-S. Rachoïn, C.A. Schorr, L. Shea, D.L. Edmonston, C.L. Mosher, A. Karp, Z. Cohen, V. Allusson, G. Bambrick-Santoyo, N.u.A. Bhatti, B. Mehta, A. Williams, P. Walters, R.C. Go, K.M. Rose, A.M. Zhou, E.C. Kim, R. Lisk, S. G. Coca, D.R. Altman, A. Saha, H. Soh, H.H. Wen, S. Bose, E.A. Leven, J.G. Wang, G. Mosoyan, G.N. Nadkarni, J. Guirguis, R. Kapoor, C. Meshberger, B.T. Garibaldi, C.P. Corona-Villalobos, Y. Wen, S. Menez, R.F. Malik, C.E. Cervantes, S.C. Gautam, H.B. Nguyen, A. Ahoum, L.F. Thomas, D.R. Sirganagari, P.K. Guru, P.A. Bergl, J. Rodriguez, J.A. Shah, M.S. Gupta, P.N. Kumar, D.G. Lazarous, S.G. Kassaye, T. S. Johns, R. Mocerino, K. Prudhvi, D. Zhu, R.V. Levy, Y. Azzi, M. Fisher, M. Yunes, K. Sedaliu, L. Golestaneh, M. Brogan, R. Raichoudhury, S.J. Cho, M. Plataki, S. L. Alvarez-Mulet, L.G. Gomez-Escobar, D.i. Pan, S. Lee, J. Kirshnan, W. Whalen, D. Charytan, A. Macina, D.W. Ross, A.S. Leidner, C. Martinez, J.M. Kruser, R. G. Wunderink, A.J. Hodakowski, E.G. Price-Haywood, L.A. Matute-Trochez, A. E. Hasty, M.B. Mohamed, R.S. Avasare, D. Zonies, R.M. Baron, M.E. Sise, E. T. Newman, K.K. Pokharel, S. Sharma, H. Singh, S. Correa, T. Shaikat, O. Kamal, H. Yang, J.O. Boateng, M. Lee, I.A. Strohbehn, J. Li, S.A. Muhsin, E.I. Mandel, A. L. Mueller, N.S. Cairl, C. Rowan, F. Madhai-Lovely, V. Peev, J.J. Byun, A. Vissing, E.M. Kapania, Z. Post, N.P. Patel, J.-M. Hermes, A. Patrawalla, D.G. Finkel, B. A. Danek, S. Arikapudi, J.M. Paer, S. Puri, J. Sunderram, M.T. Scharf, A. Ahmed, I. Berim, S. Hussain, S. Anand, J.E. Levitt, P. Garcia, S.M. Boyle, R. Song, J. Zhang, M.A. Sharshir, V.V. Rusnak, A.S. Podoll, M. Chonchol, S. Sharma, E.L. Burnham, A. Rashidi, R. Hejal, E.T. Judd, L. Latta, A. Tolwani, T.E. Albertson, J.Y. Adams, S. Y. Chang, R.M. Beutler, C.E. Schulze, E. Macedo, H. Rhee, K.D. Liu, V.K. Jotwani, J. L. Koyner, C.V. Shah, V. Jaikaransingh, S.M. Toth-Manikowski, M.J. Joo, J.P. Lash, J.A. Neyra, N. Chaaban, A. Iardino, E.H. Au, J.H. Sharma, M.A. Sosa, S. Taldone, G. Contreras, D.D.L. Zerda, P. Blakely, H. Berlin, T.U. Azam, H. Shadid, M. Pan, P. O'Hayer, C. Meloche, R. Feroze, K.J. Padalia, A. Bitar, J.E. Flythe, M.J. Tugman, B.R. Brown, R.C. Spiardi, T.A. Miano, M.S. Roche, C.R. Vasquez, A.D. Bansal, N. C. Ernecoff, C.P. Kovesdy, M.Z. Molnar, A. Azhar, S.S. Hedayati, M.V. Nadamuni, S. S. Khan, D.L. Willett, A.D. Renaghan, P.K. Bhatraju, B.A. Malik, C.M. Joy, T. Li, S. Goldberg, P.F. Kao, G.L. Schumaker, A.J. Faugno, C.M. Hsu, A. Tariq, L. Meyer, D.E. Weiner, M. Christov, F.P. Wilson, T. Arora, U. Ugwuowo, in the US, *JAMA Internal Med.* 180 (11) (2020) 1436.
- [7] P.F. Caruso, G. Angelotti, M. Greco, et al., The effect of COVID-19 epidemic on vital signs in hospitalized patients: a pre-post heat-map study from a large teaching hospital, *J. Clin. Monitor. Computing.* (2021), <https://doi.org/10.1007/s10877-021-00715-y>.
- [8] M. Arentz, E. Yim, L. Klaff, S. Lokhandwala, F.X. Riedo, M. Chong, M. Lee, Characteristics and Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State, *JAMA* 323 (16) (2020) 1612.
- [9] Dan T, Li Y, Zhu Z, et al. Sable but Lost Lives When ICU Is Overloaded: A Model from 733 Patients in Epicenter Wuhan, China; 2021. <https://www.who.int/>.
- [10] M. Greco, P.F. Caruso, M. Cecconi, Artificial Intelligence in the Intensive Care Unit, *Semin. Respir. Crit. Care Med.* 42 (01) (2021) 002–009.
- [11] J. Deasy, P. Liò, A. Ercole, Dynamic survival prediction in intensive care units from heterogeneous time series without the need for variable selection or curation, *Sci. Rep.* 10 (1) (2020), <https://doi.org/10.1038/s41598-020-79142-z>.
- [12] T. Dan, Y. Li, Z. Zhu, et al. Machine Learning to Predict ICU Admission, ICU Mortality and Survivors' Length of Stay among COVID-19 Patients: Toward Optimal Allocation of ICU Resources. In: Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020. Institute of Electrical and Electronics Engineers Inc.; 2020:555-561. doi:10.1109/BIBM49941.2020.9313292.
- [13] G. Grasselli, A. Pesenti, M. Cecconi, Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response, *JAMA* 323 (16) (2020) 1545.
- [14] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [15] Z. Wang, C. Wu, K. Zheng, X. Niu, X. Wang, SMOTETomek-based resampling for personality recognition, *IEEE Access* 7 (2019) 129678–129689, <https://doi.org/10.1109/ACCESS.2019.2940061>.
- [16] T.J. Pollard, A.E.W. Johnson, J.D. Raffa, R.G. Mark, Tableone: an open source Python package for producing summary statistics for research papers, *JAMIA Open* 1 (1) (2018) 26–31, <https://doi.org/10.1093/jamiaopen/ooy012>.
- [17] M.M. Rahman, D.N. Davis, Addressing the Class Imbalance Problem in Medical Datasets, *IJMLC* (2013) 224–228.
- [18] J. O'Brien, K.Y. Du, C. Peng, Incidence, clinical features, and outcomes of COVID-19 in Canada: impact of sex and age, *J. Ovarian Res.* 13 (1) (2020), <https://doi.org/10.1186/s13048-020-00734-4>.
- [19] F. Bill, Foundation MG. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Published online 2017. doi:10.1016/S0140-6736(20)30045-3.
- [20] N.R. Hill, S.T. Fatoba, J.L. Oke, et al. Global prevalence of chronic kidney disease – a systematic review and meta-analysis. Published online 2016:1-18. doi:10.5061/dryad.3s7rd.Funding.
- [21] KDIGO. Official Journal of the international supplements Society of nephrology KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int. Supplements. Published online 2013.
- [22] G. Louis, T. Belyve, C. Goetz, R. Gaci, V. Dinot, Acute kidney injury in severe SARS-CoV-2 infection: An experience report in Eastern France, *Anaesthesia Crit. Care Pain Med.* 40 (1) (2021), 100788, <https://doi.org/10.1016/j.accpm.2020.100788>.
- [23] P. Gabarre, G. Dumas, T. Dupont, M. Darmon, E. Azoulay, L. Zafrani, Acute kidney injury in critically ill patients with COVID-19, *Intensive Care Med.* 46 (7) (2020) 1339–1348, <https://doi.org/10.1007/s00134-020-06153-9>.
- [24] R. Pastor-Barriuso, B. Pérez-Gómez, M.A. Hernán, et al. SARS-CoV-2 infection fatality risk in a nationwide seroepidemiological study. *medRxiv*. Published online 2020:1-13. doi:10.1101/2020.08.06.20169722.
- [25] T. Takahashi, P. Wong, M. Ellingson, et al., Sex differences in immune responses to SARS-CoV-2 that underlie disease outcomes, *medRxiv* (2020), <https://doi.org/10.1101/2020.06.06.20123414>.
- [26] M.J. Soler, C. Barrios, R. Oliva, D. Batlle, Pharmacologic modulation of ACE2 expression, *Curr. Hypertension Rep.* 10 (5) (2008) 410–414, <https://doi.org/10.1007/s11906-008-0076-0>.
- [27] M. Esler, D. Esler, Can angiotensin receptor-blocking drugs perhaps be harmful in the COVID-19 pandemic? *J. Hypertension* 38 (5) (2020) 781–782, <https://doi.org/10.1097/HJH.0000000000002450>.
- [28] G. Mancía, F. Rea, M. Ludergnani, G. Apolone, G. Corrao, Renin-angiotensin-aldosterone system blockers and the risk of COVID-19, *New Engl. J. Med.* 382 (25) (2020) 2431–2440, <https://doi.org/10.1056/NEJMoa2006923>.

- [29] D. Graziani, J.B. Soriano, C. Del Rio-Bermudez, D. Morena, T. Díaz, M. Castillo, M. Alonso, J. Ancochea, S. Lumbreras, J.L. Izquierdo, Characteristics and Prognosis of COVID-19 in Patients with COPD, *J. Clin. Med.* 9 (10) (2020) 3259.
- [30] A. Simonnet, M. Chetboun, J. Poissy, V. Raverdy, J. Noulette, A. Duhamel, J. Labreuche, D. Mathieu, F. Pattou, M. Jourdain, R. Caizzo, M. Caplan, N. Cousin, T. Duburcq, A. Durand, A. El kalioubie, R. Favory, B. Garcia, P. Girardie, J. Goutay, M. Houard, E. Jaillette, N. Kostuj, G. Ledoux, D. Mathieu, A.S. Moreau, C. Niles, S. Nseir, T. Onimus, E. Parmentier, S. Préau, L. Robriquet, A. Rouze, S. Six, H. Verkindt, High Prevalence of Obesity in Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) Requiring Invasive Mechanical Ventilation, Obesity. 28 (7) (2020) 1195–1199.
- [31] C. Izzi-Engbeaya, W. Distaso, A. Amin, W. Yang, O. Idowu, J.S. Kenkre, R.J. Shah, E. Woin, C. Shi, N. Alavi, H. Bedri, N. Brady, S. Blackburn, M. Leczycka, S. Patel, E. Sokol, E. Toke-Bjølgerud, A. Qayum, M. Abdel-Malek, D.C.D. Hope, N.S. Oliver, V. Bravis, S. Misra, T.M. Tan, N.E. Hill, V. Salem, Adverse outcomes in COVID-19 and diabetes: a retrospective cohort study from three London teaching hospitals, *BMJ Open Diabetes Research Care* 9 (1) (2021) e001858.