





Article

Predicting in-Hospital Mortality of Patients with COVID-19 Using Machine Learning Techniques

Fabiana Tezza ¹, Giulia Lorenzoni ², Danila Azzolina ^{2,3}, Sofia Barbar ⁴, Lucia Anna Carmela Leone ⁵
and Dario Gregori ^{2,*}

¹ Geriatric Unit, Ospedali Riuniti di Padova Sud, AULSS 6 Euganea, 35043 Monselice, Italy; fabiana.tezza@aulss6.veneto.it

² Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padova, 35131 Padova, Italy; giulia.lorenzoni@unipd.it (G.L.); danila.azzolina@uniupo.it (D.A.)

³ Department of Translational Medicine, University of Eastern Piedmont, 28100 Novara, Italy

⁴ Internal Medicine Unit, Cittadella Hospital, AULSS 6 Euganea, 35013 Cittadella, Italy; sofia.barbar@aulss6.veneto.it

⁵ Internal Medicine Unit, Ospedali Riuniti Padova Sud, AULSS 6 Euganea, 35043 Monselice, Italy; lucia.leone@aulss6.veneto.it

* Correspondence: dario.gregori@unipd.it; Tel.: +39-049-8275-384; Fax: +39-02-700-445-089

Abstract: The present work aims to identify the predictors of COVID-19 in-hospital mortality testing a set of Machine Learning Techniques (MLTs), comparing their ability to predict the outcome of interest. The model with the best performance will be used to identify in-hospital mortality predictors and to build an in-hospital mortality prediction tool. The study involved patients with COVID-19, proved by PCR test, admitted to the “Ospedali Riuniti Padova Sud” COVID-19 referral center in the Veneto region, Italy. The algorithms considered were the Recursive Partition Tree (RPART), the Support Vector Machine (SVM), the Gradient Boosting Machine (GBM), and Random Forest. The resampled performances were reported for each MLT, considering the sensitivity, specificity, and the Receiving Operative Characteristic (ROC) curve measures. The study enrolled 341 patients. The median age was 74 years, and the male gender was the most prevalent. The Random Forest algorithm outperformed the other MLTs in predicting in-hospital mortality, with a ROC of 0.84 (95% C.I. 0.78–0.9). Age, together with vital signs (oxygen saturation and the quick SOFA) and lab parameters (creatinine, AST, lymphocytes, platelets, and hemoglobin), were found to be the strongest predictors of in-hospital mortality. The present work provides insights for the prediction of in-hospital mortality of COVID-19 patients using a machine-learning algorithm.

Keywords: machine learning techniques; COVID-19; Italy; in-hospital mortality; outcome prediction



Citation: Tezza, F.; Lorenzoni, G.; Azzolina, D.; Barbar, S.; Leone, L.A.C.; Gregori, D. Predicting in-Hospital Mortality of Patients with COVID-19 Using Machine Learning Techniques. *J. Pers. Med.* **2021**, *11*, 343. <https://doi.org/10.3390/jpm11050343>

Academic Editor: Moon-Soo Lee

Received: 9 March 2021

Accepted: 21 April 2021

Published: 24 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The prognosis prediction of patients infected with SARS-CoV-2 is one of the most relevant topics in COVID-19 research [1]. The pandemic poses a severe burden to the healthcare system of countries worldwide given the long symptoms' duration [2], the higher risk of severe complications requiring hospitalization/intensive care compared to seasonal influenza [3], and the documented excess of mortality associated with the virus spread [4]. Such conditions have led to a dramatic imbalance between healthcare resources and hospital/Intensive Care Unit (ICU) bed demands [5]. Italy, the first and most affected European country during the first COVID-19 wave [6], adopted emergency solutions to face such imbalance [5]. However, the COVID-19 wave is still ongoing in almost all European countries, including Italy. For these reasons, there is an urgent need for strategies to prevent such an imbalance occurs again. With the adoption of public health strategies to contain the virus spread [7] and the improvement of the health care resources available, the development of tools to predict patients' prognosis would be part

of the solution to prevent resource imbalance [8]. Such tools would be helpful to assist physicians' decision-making. Risk stratification is essential in this context to identify an adequate referral pathway for each COVID-19 patient, allowing for an appropriate resource allocation. Furthermore, such tools help the identification of the parameters that contribute the most to the outcome, allowing physicians to monitor/treat such parameters.

Several tools have been proposed from the beginning of the pandemic, especially for in-hospital mortality prediction [9]. A systematic review published at the beginning of the pandemic, i.e., April 2020, identified several mortality risk prediction tools for COVID-19 patients [9], and several others have been proposed after that. Several methods have been employed to develop such tools, from traditional statistical approaches, e.g., multivariable analysis [10], to more advanced machine learning techniques (MLTs) [11–14]. MLTs are increasingly used for outcome prediction in the clinical setting, since they present several advantages over traditional methods [15]. They help disentangle complex relationships between covariates and outcome of interest, even though a low number of events have occurred in front of many variables to be tested. Not least, the MLT predictive ability may improve as new data is provided.

Our study aims to identify predictors of COVID-19 in-hospital mortality testing a set of MLTs, comparing their ability to predict the outcome of interest. The model estimation is based on patients treated in a COVID-19 referral center in the Veneto region, Italy, during the COVID-19 first wave. Together with Lombardia, Veneto was the first Italian region affected by the COVID-19 outbreak at the beginning of the year [16].

2. Materials and Methods

The study involved patients with COVID-19, proved by PCR test, admitted to the "Ospedali Riuniti Padova Sud" COVID-19 referral center in the Veneto region, Italy, in March and April 2020.

2.1. Data Collection

Information on sociodemographic characteristics; medical history (comorbidities, smoking habits, and weight status); clinical and instrumental parameters; and laboratory data were collected at hospital admission. Furthermore, drug therapy was registered. The outcome was represented by in-hospital mortality.

2.2. Statistical Analysis

Descriptive statistics were reported as I quartile/median/III quartile for continuous variables and percentages (absolute numbers) for categorical variables. Wilcoxon-type tests were performed for the continuous variables and the Pearson chi-square test, or Fisher-exact test, whichever is appropriate, for the categorical variables.

A set of MLTs was tuned to identify the predictors of in-hospital mortality. The predictive tools were tuned using a bootstrapping algorithm (50 runs). The algorithms considered for the classification task were the Recursive Partition Tree (RPART), the Support Vector Machine (SVM), the Gradient Boosting Machine (GBM), and Random Forest.

- The RPARTs are classification models based on a top-down methodology in which, starting from a root node, binary splits of data are generated until a certain criterion (i.e., the minimization of the node impurity) is encountered [17]. This method is prone to overfitting on training data. The cross-validation or bootstrapping procedure is a useful method to limit the overfitting, leading to defining the proper tuning of the Decision Tree (DT) parameters and optimizing the model accuracy [18].
- Random Forest is a tree-based algorithm, which involves the computation of hundreds to thousands of RPART trees; the method merges the DT output to increase the model's generalizability [19].
- The GBM is based on a sequential boosting improvement of weak RPART classifiers (high bias and low variance). The GBM idea is to add a classifier sequentially so that

the next classifier is trained to improve the already trained RPART. A random forest algorithm, instead, trains each classifier independently from the others [19].

- The SVM algorithm's main objective is to find an optimal hyperplane of the feature's N -dimensional space (N —the number of variables) that distinctly classify the data points into a binary partition [20]. Several hyperplanes may separate the classes of data points. The SVM algorithm considers the hyperplanes, which maximize the margin (the distance between data points of classes).

The resampled performances were reported for each MLT, considering the sensitivity, specificity, and the Receiving Operative Characteristic curve (ROC) measures, with 95% confidence intervals (C.I.s).

Among the proposed MLTs, the tool with the best performance was chosen for the classification task. *c*, pairwise differences were computed and tested to assess if the difference was equal to zero. The Bonferroni correction was used to calculate the *p*-values and adjust the confidence interval limits for the reported differences. For the most promising MLT, the variable importance plot was reported together with the ROC curve and the median balanced accuracy measure within the resampling.

The ROC and accuracy measures were also reported to evaluate the MLT tool's performance according to the increasing data availability during the study period. The database was ordered according to the hospitalization date. The MLT was built on the first 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, and 340 hospitalized patients. The ROC and accuracy measures were reported according to the different sample sizes. The performance was computed, together with the variable importance measures, by tuning and developing the tool for cumulative fractions of the sample.

A Shiny web application was developed based on the algorithm with the best performance. The tool calculates the in-hospital death probability, according to the patients' characteristics.

Analyses were performed using R 3.4.2 [21] with the rms [22] and caret [23] packages.

3. Results

The study enrolled 341 patients admitted to the "Ospedali Riuniti Padova Sud" COVID-19 referral center between March and April 2020. Table 1 presents the subjects' baseline characteristics according to the admission date (before–after 21 March 2020, the day on which the national lockdown was declared in Italy). Seventy-five patients out of 341 died while hospitalized.

Overall, the prevalence of the male gender was found to be higher than the female one (57% in the sample overall), even though it was found to be higher in the first period of admissions compared to the second one (65% vs. 51%; *p*-value 0.007). The median age of the subjects enrolled was 74 years, and it was lower in the first period compared to the second period (median of 70 vs. 78; *p*-value < 0.001). Subjects admitted to the hospital in the first period were significantly more likely to suffer from diabetes (*p*-value 0.031) and less likely to suffer from dementia (*p*-value < 0.001) and from cerebrovascular diseases (*p*-value 0.019). No significant differences were detected in the distribution of the other comorbidities. For what concerns the drug therapy, subjects admitted after March 21 were found to be significantly more likely to be treated with low molecular weight heparin (LMWH) (*p*-value 0.001) but less likely to be administered with antiviral drugs.

3.1. MLTs Performance

The Random Forest algorithm outperformed the other MLTs in predicting the in-hospital mortality (Figure 1), with a ROC of 0.84 (95% C.I. 0.78–0.9). In addition to that, looking at the values of sensitivity and specificity, Random Forest provided a better balance than the other techniques between two such measures. Such results were confirmed when the pairwise comparisons were made (Table S1 Supplementary Materials). Furthermore, no class imbalance issues were detected (Figure S1, Supplementary Materials).

Table 1. Subjects’ baseline characteristics according to the admission date (before–after 21 March 2020). Data are I quartile/Median/III quartile for continuous variables and percentages (absolute numbers) for categorical variables.

Baseline Characteristics	N	Before March 21 (n = 149)	After March 21 (n = 192)	Overall (N = 341)	p-Value
Gender: Female	341	35% (52)	49% (95)	43% (147)	0.007
Male		65% (97)	51% (97)	57% (194)	
Age	341	58.0/70.0/82.0	62.5/78.0/87.0	60.0/74.0/85.0	<0.001
Smoking: No	237	79% (94)	85% (100)	82% (194)	0.25
Yes		21% (25)	15% (18)	18% (43)	
Duration of Hospitalization (Days)	290	7.0/12.0/17.0	6.0/10.0/16.0	7.0/11.0/16.8	0.106
Comorbidities					
Diabetes Mellitus: No	341	73% (109)	83% (159)	79% (268)	0.031
Yes		27% (40)	17% (33)	21% (73)	
SHPT: No	341	48% (72)	53% (102)	51% (174)	0.379
Yes		52% (77)	47% (90)	49% (167)	
Dyslipidemia: No	341	82% (122)	86% (165)	84% (287)	0.309
Yes		18% (27)	14% (27)	16% (54)	
Ischemic heart disease: No	341	91% (136)	90% (172)	90% (308)	0.6
Yes		9% (13)	10% (20)	10% (33)	
Obesity: No	341	93% (138)	94% (181)	94% (319)	0.538
Yes		7% (11)	6% (11)	6% (22)	
Cerebrovascular disease: No	341	95% (142)	88% (169)	91% (311)	0.019
Yes		5% (7)	12% (23)	9% (30)	
Dementia: No	341	87% (129)	69% (132)	77% (261)	<0.001
Yes		13% (20)	31% (60)	23% (80)	
Chronic Kidney Disease: No	331	93% (136)	89% (164)	91% (300)	0.293
Yes		7% (11)	11% (20)	9% (31)	
Current cancer: No	341	96% (143)	91% (175)	93% (318)	0.078
Yes		4% (6)	9% (17)	7% (23)	
Previous Cancer: No	340	95% (142)	95% (182)	95% (324)	0.995
Yes		5% (7)	5% (9)	5% (16)	
COPD: No	340	95% (140)	90% (172)	92% (312)	0.096
Yes		5% (8)	10% (20)	8% (28)	
Lab parameters & Vital Signs					

Table 1. Cont.

Baseline Characteristics	N	Before March 21 (n = 149)	After March 21 (n = 192)	Overall (N = 341)	p-Value
Neutrophils ($\times 10^9$ /L)	339	3.16/4.38/6.56	3.24/4.57/7.23	3.17/4.50/6.72	0.235
Lymphocytes ($\times 10^9$ /L)	339	0.640/0.915/1.263	0.670/0.890/1.345	0.650/0.900/1.320	0.784
Neutrophils/Lymphocytes ($\times 10^9$ /L)	338	2.69/4.42/7.96	3.23/5.09/9.36	2.97/4.93/8.93	0.297
Hemoglobin (g/L)	339	116/130/142	116/128/140	116/128/140	0.586
D-dimer (μ g/L)	124	134/192/416	148/301/675	142/227/636	0.089
Creatinine (μ mol/L)	337	67.8/84.0/104.2	63.0/78.0/96.0	67.0/81.0/102.0	0.075
Platelets ($\times 10^9$ /L)	338	148/189/234	150/198/246	148/196/244	0.445
AST (U/L)	292	25.2/35.5/49.0	26.0/37.0/60.0	26.0/36.5/55.2	0.265
ALT (U/L)	298	20.0/27.0/40.0	15.8/25.0/43.2	17.0/26.0/42.0	0.2
Procalcitonin (PCT) (μ g/L)	208	0.0400/0.0800/0.1400	0.0500/0.1200/0.3950	0.0475/0.1000/0.2600	0.03
Troponine (ng/L)	31	6.0/25.0/90.0	10.5/26.5/64.5	9.5/25.0/68.0	0.848
Oxygen saturation (%)	244	93.8/96.0/98.0	93.0/95.0/97.0	93.0/95.5/98.0	0.541
Quick SOFA	341	0/0/1	0/0/1	0/0/1	0.048
pO ₂ (mmHg)	213	54.0/64.0/71.0	56.0/64.0/73.2	55.0/64.0/72.0	0.567
Glasgow Coma Scale \geq 15 : No	341	84% (125)	70% (135)	76% (260)	0.003
Yes		16% (24)	30% (57)	24% (81)	
Drug therapy					
ARIXTRA: No	111	59% (37)	35% (17)	49% (54)	0.015
Yes		41% (26)	65% (31)	51% (57)	
LMWH: No	286	28% (35)	12% (20)	19% (55)	0.001
Yes		72% (88)	88% (143)	81% (231)	
Antiviral: No	340	51% (76)	97% (186)	77% (262)	<0.001
Yes		49% (73)	3% (5)	23% (78)	
Plaquenil: No	340	24% (36)	18% (34)	21% (70)	0.15
Yes		76% (113)	82% (157)	79% (270)	
Antibiotic: No	340	5% (7)	4% (8)	4% (15)	0.82
Yes		95% (142)	96% (183)	96% (325)	
Tocilizumab: No	340	95% (141)	97% (186)	96% (327)	0.189
Yes		5% (8)	3% (5)	4% (13)	

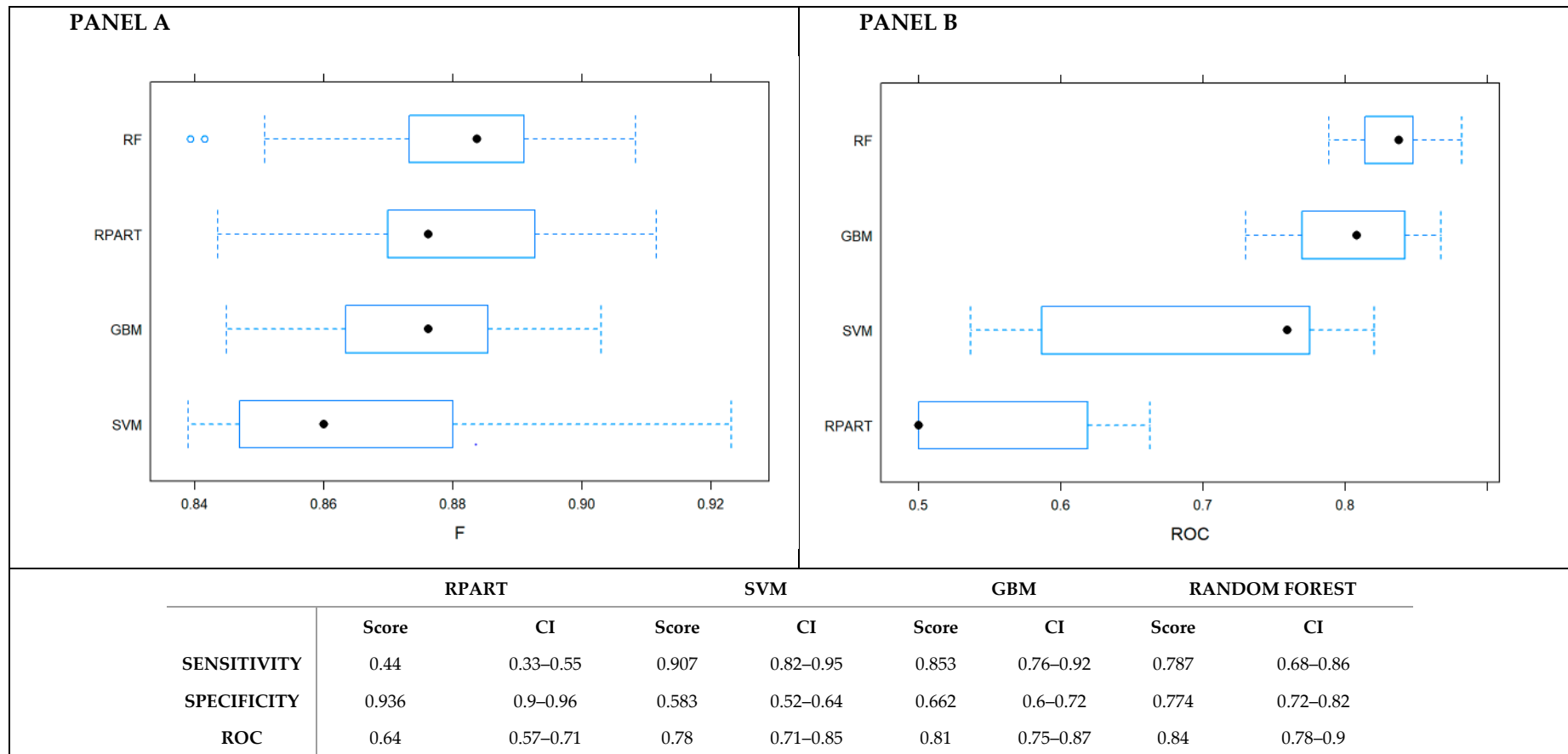


Figure 1. Comparison of the performance of the MLT algorithms. **Panel A** presents the F-score measure. **Panel B** presents the ROC.

3.2. Variable Importance in Predicting in-Hospital Mortality According to the Random Forest

Since Random Forest was found to be the MLT with the best performance, it was chosen to identify the predictors of in-hospital mortality. The algorithm was tuned to achieve the optimal performance in the correspondence of 500 trees and 6 mtry (number of variables available for splitting each tree node).

Figure 2 reports the plots of the variable importance measures according to the mean decrease accuracy. Age, together with vital signs (oxygen saturation and the quick SOFA) and lab parameters (creatinine, AST, lymphocytes, platelets, and hemoglobin), were found to be the strongest predictors of in-hospital mortality (Figure 2). Conversely, comorbidities were found to provide only a small contribution in predicting the in-hospital mortality in such patients.

Furthermore, the algorithm stability investigation as the availability of data increases was performed to inform clinicians about the minimum number of cases required to obtain a reliable algorithm. Interestingly, the ROC parameter underwent stabilization after the first 100 patients (Figure 3). The variable importance plots calculated on cumulative fractions of the sample are provided in Table S2 (Supplementary Materials).

The web application based on the algorithm developed is available at <https://r-ubesp.dctv.unipd.it/shiny/Schiavonia/>.

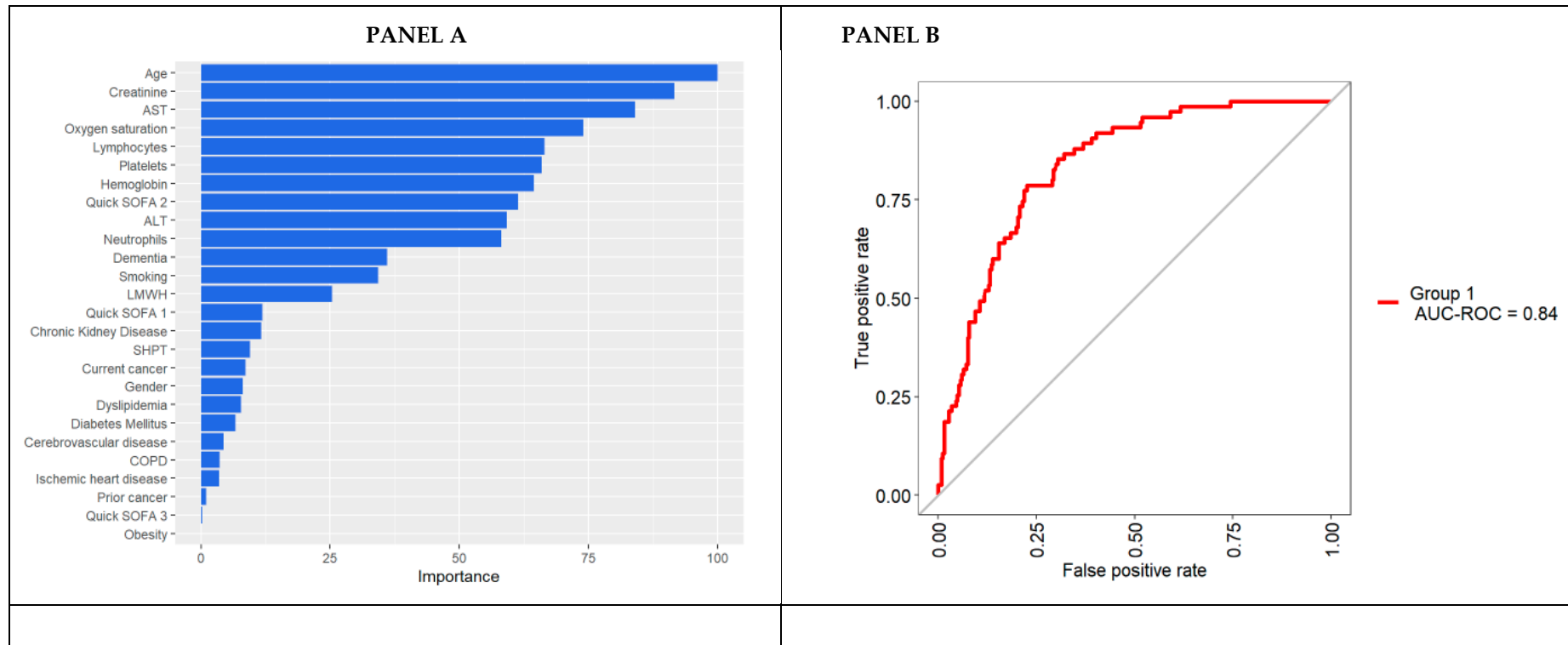


Figure 2. Random Forest’s variable importance plot (**Panel A**) and ROC curve on the Random Forest predictions (**Panel B**). The median resampled accuracy is 0.86.

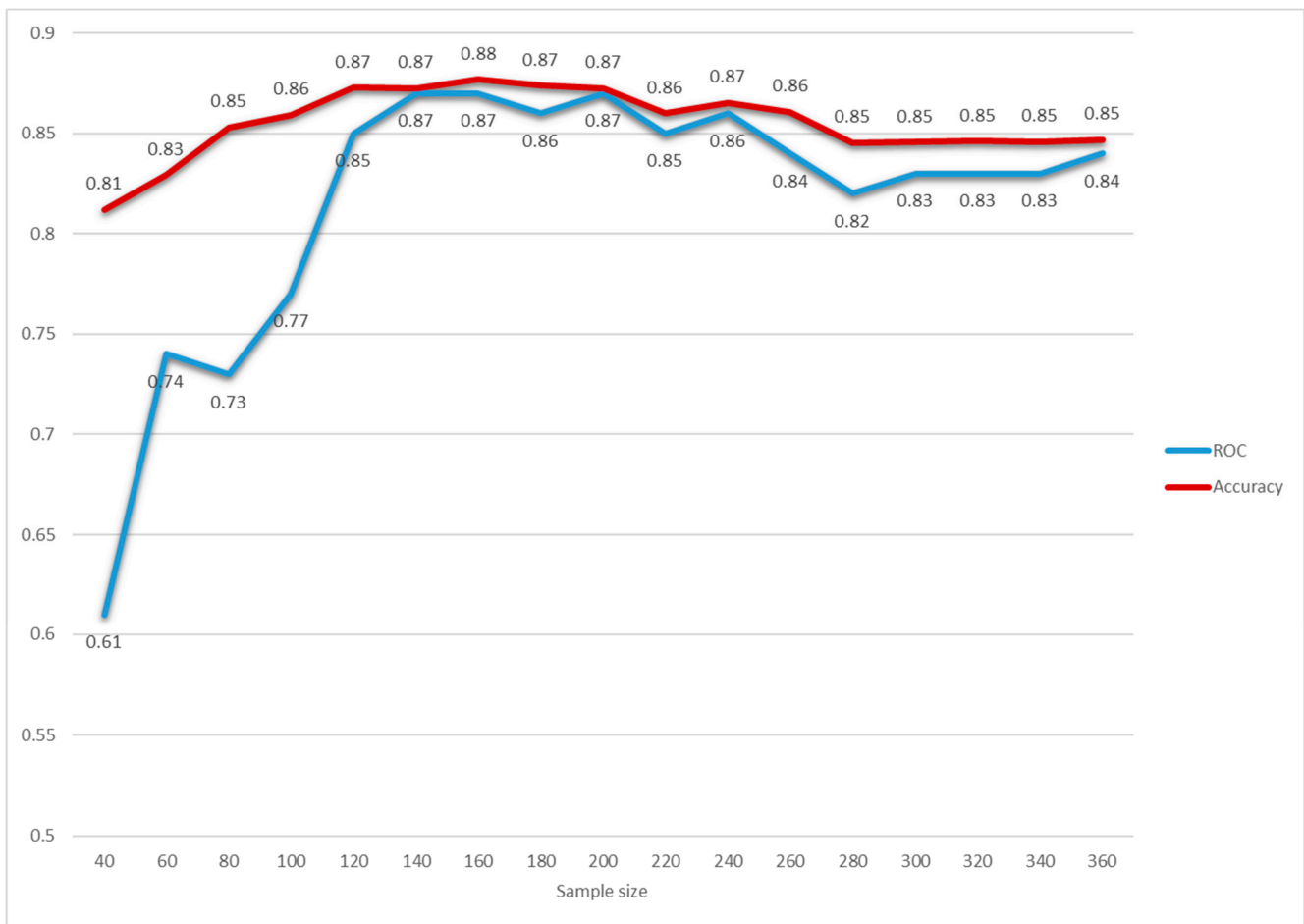


Figure 3. Random Forest’s ROC and accuracy as the availability of the data increased. The database was ordered according to the hospitalization date. Random Forests were built on the first 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, and 340 hospitalized patients. The ROC (blue line) and accuracy (red line) measures were reported according to the different sample sizes. The performance was computed by tuning and developing the Random Forest tool for the cumulative fractions of the sample.

4. Discussion

The present study showed that the Random Forest is a feasible machine-learning algorithm to stratify the mortality risk in a sample of patients admitted to a COVID-19 referral center in the Veneto region, Italy, during the first COVID-19 wave.

The sample characteristics, in terms of gender and comorbidities distribution, are consistent with those reported by previous studies on subjects admitted to the Italian hospitals during the first wave of the pandemic [24], even though the median age of the patients enrolled was about ten years higher than that previously reported [24]. For what concerns drug therapy, the increasing use of LMWH documented between the first and second study periods reflects the publication of the preliminary evidence about the beneficial effects of LMWH in patients with COVID-19 [25].

Risk stratification is a hot topic in COVID-19 research, given the shortage of healthcare resources for the management of COVID-19 patients. For this reason, an impressive number of risk stratification tools for in-hospital mortality in COVID-19 patients has been developed. Both traditional, i.e., logistic regression, and nontraditional, i.e., MLTs, statistical techniques have been employed to create such tools [9], but the traditional ones have been most frequently used, probably for the ease of use and interpretability in the clinical setting. However, MLTs have been shown to be feasible to develop risk stratification

tools in the COVID-19 research setting [11,13], with promising results underlying such algorithms' potentials to assist clinicians' decision-making in everyday clinical practice.

The present work showed an outperformance of Random Forest compared to the other algorithms tested. Interestingly, the variables found to contribute to in-hospital mortality were age and those related to hospital presentation (vital signs and lab parameters). At the same time, the comorbidities played a minor role in the mortality risk prediction. Such findings are only partially in-line with those of international studies, showing that, together with clinical presentation, comorbidities also play a relevant role in risk prediction [11]. However, it is worth pointing out that it is difficult to compare results from the literature, since each tool has been developed on different patient populations, e.g., hospitalized patients, mechanically ventilated patients, elderly patients hospitalized, and the general population. Furthermore, each tool has been developed using a different set of baseline variables, including sociodemographic ones, laboratory data, instrumental parameters, and clinical measures.

5. Conclusions

The present work provides a useful tool able to assist physicians' decision-making in facing the COVID-19 emergency. Furthermore, the feasibility of MLTs in this research context has been shown, which represents an added value of the present work, since MLTs are able to overcome the limitations of more traditional statistical techniques.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/jpm11050343/s1>: Table S1. Differences in ROC (Panel A) and F statistic (Panel B) among the MLTs employed to predict in-hospital death. Table S2: Variable importance plots for each Random Forest developed. Figure S1. ROC, sensitivity, and specificity measures of the null model and the random forest.

Author Contributions: Conceptualization, D.G.; Formal analysis, D.A.; Investigation, F.T., S.B., and L.A.C.L.; Writing—original draft, G.L.; and Writing—review and editing, F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Review and approval were waived for this study, due to the retrospective nature of the study based on anonymized data.

Informed Consent Statement: Patient consent was waived, since only anonymized data was collected.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sperrin, M.; McMillan, B. Prediction models for covid-19 outcomes. *BMJ* **2020**, *371*, m3777. [[CrossRef](#)]
2. Tenforde, M.W.; Kim, S.S.; Lindsell, C.J.; Rose, E.B.; Shapiro, N.I.; Files, D.C.; Gibbs, K.W.; Erickson, H.L.; Steingrub, J.S.; Smithline, H.A.; et al. Symptom Duration and Risk Factors for Delayed Return to Usual Health Among Outpatients with COVID-19 in a Multistate Health Care Systems Network—United States, March–June 2020. *Morb. Mortal. Wkly. Rep.* **2020**, *69*, 993. [[CrossRef](#)]
3. Cates, J.; Lucero-Obusan, C.; Dahl, R.M.; Schirmer, P.; Garg, S.; Oda, G.; Hall, A.J.; Langley, G.; Havers, F.P.; Holodniy, M.; et al. Risk for In-Hospital Complications Associated with COVID-19 and Influenza—Veterans Health Administration, United States, October 1, 2018–May 31, 2020. *Morb. Mortal. Wkly. Rep.* **2020**, *69*, 1528. [[CrossRef](#)]
4. Michelozzi, P.; De'Donato, F.; Scortichini, M.; Pezzotti, P.; Stafoggia, M.; De Sario, M.; Costa, G.; Noccioli, F.; Riccardo, F.; Bella, A.; et al. Temporal dynamics in total excess mortality and COVID-19 deaths in Italian cities. *BMC Public Health* **2020**, *20*, 1–8. [[CrossRef](#)]
5. Grasselli, G.; Pesenti, A.; Cecconi, M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. *JAMA* **2020**, *23*, 1545–1546. [[CrossRef](#)] [[PubMed](#)]
6. Remuzzi, A.; Remuzzi, G. COVID-19 and Italy: What next? *Lancet* **2020**, *395*, 1225–1228. [[CrossRef](#)]
7. Gregori, D.; Azzolina, D.; Lanera, C.; Prosepe, I.; Destro, N.; Lorenzoni, G.; Berchiolla, P. A first estimation of the impact of public health actions against COVID-19 in Veneto (Italy). *J. Epidemiol. Community Health* **2020**, *74*, 858–860. [[CrossRef](#)] [[PubMed](#)]

8. Castro, V.M.; McCoy, T.H.; Perlis, R.H. Laboratory Findings Associated With Severe Illness and Mortality Among Hospitalized Individuals With Coronavirus Disease 2019 in Eastern Massachusetts. *JAMA Netw. Open* **2020**, *3*, e2023934. [CrossRef] [PubMed]
9. Wynants, L.; Van Calster, B.; Collins, G.S.; Riley, R.D.; Heinze, G.; Schuit, E.; Bonten, M.M.J.; Dahly, D.L.; A Damen, J.; A Debray, T.P.; et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* **2020**, *369*, m1328. [CrossRef]
10. Hajifathalian, K.; Sharaiha, R.Z.; Kumar, S.; Krisko, T.; Skaf, D.; Ang, B.; Redd, W.D.; Zhou, J.C.; Hathorn, K.E.; Mccarty, T.R.; et al. Development and external validation of a prediction risk model for short-term mortality among hospitalized U.S. COVID-19 patients: A proposal for the COVID-AID risk tool. *PLoS ONE* **2020**, *15*, e0239536. [CrossRef] [PubMed]
11. An, C.; Lim, H.; Kim, D.-W.; Chang, J.H.; Choi, Y.J.; Kim, S.W. Machine learning prediction for mortality of patients diagnosed with COVID-19: A nationwide Korean cohort study. *Sci. Rep.* **2020**, *10*, 1–11. [CrossRef] [PubMed]
12. Yadaw, A.S.; Li, Y.-C.; Bose, S.; Iyengar, R.; Bunyavanich, S.; Pandey, G. Clinical features of COVID-19 mortality: Development and validation of a clinical prediction model. *Lancet Digit. Health* **2020**, *2*, e516–e525. [CrossRef]
13. Gao, Y.; Cai, G.-Y.; Fang, W.; Li, H.-Y.; Wang, S.-Y.; Chen, L.; Yu, Y.; Liu, D.; Xu, S.; Cui, P.-F.; et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat. Commun.* **2020**, *11*, 1–10. [CrossRef]
14. Vaid, A.; Somani, S.; Russak, A.J.; De Freitas, J.K.; Chaudhry, F.F.; Paranjpe, I.; Johnson, K.W.; Lee, S.J.; Miotto, R.; Richter, F.; et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. *J. Med. Internet Res.* **2020**, *22*, e24018. [CrossRef]
15. Ngiam, K.Y.; Khor, W. Big Data and Machine Learning Algorithms for Health-Care Delivery. *Lancet Oncol.* **2019**, *20*, e262–e273. [CrossRef]
16. Lorenzoni, G.; Lanera, C.; Azzolina, D.; Berchiolla, P.; Gregori, D.; Gregori, D.; Lanera, C.; Berchiolla, P.; Catelan, D.; Azzolina, D.; et al. Is a More Aggressive COVID-19 Case Detection Approach Mitigating the Burden on ICUs? Some Reflections from Italy. *Crit. Care* **2020**, *24*, 175. [CrossRef]
17. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. Classification and Regression Trees. Available online: <https://cds.cern.ch/record/2253780> (accessed on 12 October 2019).
18. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2009; ISBN 978-0-387-84857-0.
19. Breiman, L. Machine Learning. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
20. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
21. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
22. Harrell, F.E.J. *Rms: Regression Modeling Strategies*, R Package Version 4.1-3. 2014.
23. Kuhn, M. Building Predictive Models in R Using the caretPackage. *J. Stat. Softw.* **2008**, *28*. [CrossRef]
24. Grasselli, G.; Zangrillo, A.; Zanella, A.; Antonelli, M.; Cabrini, L.; Castelli, A.; Cereda, D.; Coluccello, A.; Foti, G.; Fumagalli, R. Baseline Characteristics and Outcomes of 1591 Patients Infected with SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA* **2020**, *323*, 1574–1581. [CrossRef]
25. Thachil, J. The Versatile Heparin in COVID-19. *J. Thromb. Haemost.* **2020**, *18*, 1020–1022. [CrossRef] [PubMed]