

# Early Biochemical Analysis of COVID-19 Patients Helps Severity Prediction

**Andres Roncancio-Clavijo**

Biodonostia Health Research Institute

**Miriam Gorostidi-Aicua**

Biodonostia Health Research Institute

**Ainhoa Alberro**

Biodonostia Health Research Institute

**Andrea Iribarren-Lopez**

Biodonostia Health Research Institute

**Ray Butler**

Butler Scientifics S.L., Barcelona

**Raul Lopez**

Butler Scientifics S.L., Barcelona

**Jose Antonio Iribarren**

Osakidetza Basque Health Service

**Diego Clemente**

Neuroimmune-repair Group

**Jose María Marimon**

Biodonostia Health Research Institute

**Javier Basterrechea**

Osakidetza Basque Health Service

**Bruno Martinez**

Osakidetza Basque Health Service

**Alvaro Prada**

Biodonostia Health Research Institute

**David Otaegui** (✉ [david.otaegui@biodonostia.org](mailto:david.otaegui@biodonostia.org))

Biodonostia Research Institute <https://orcid.org/0000-0002-6625-5976>

---

## Research Article

**Keywords:** COVID-19, Blood analysis, Severity prediction, Immature granulocytes

**Posted Date:** May 17th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-480470/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

COVID-19 pandemic has put the protocols and the capacity of our Hospitals to the test. The management of severe patients admitted to the Intensive Care Units has been a challenge for all health systems. To assist in this challenge, various models have been proposed to predict mortality and severity, however, there is no clear consensus for their use. In this work, we took advantage of data obtained from routine blood tests performed on all individuals on the first day of hospitalization. The test has been done by standardized cost-effective technique available in all the hospitals. We have analyzed the results of 1082 patients with COVID19 and using artificial intelligence we have generated a predictive model based on data from the first days of admission that predicts the risk of developing severe disease with an AUC = 0.78. Our results show the importance of immature granulocytes and their ratio with Lymphocytes in the disease and present an algorithm based on 5 parameters to identify a severe course.

This work highlights the importance of studying routine analytical variables in the early stages of hospital admission and the benefits of applying AI to identify patients who may develop severe disease.

## Introduction

COVID-19 is an infectious disease produced by the coronavirus SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2). The first outbreak was reported in a seafood market in Wuhan, (Hubei province, China) and the World Health Organization (WHO) declared the situation a public health emergency of international concern on January 30<sup>th</sup>, 2020 and as a pandemic on March 11<sup>th</sup>.

COVID-19 disease has a wide clinical spectrum. A relevant proportion of infected individuals is asymptomatic, while others develop from mild symptoms, such as low-grade inflammation, fever and cough, to severe symptoms, that include respiratory failure, metabolic acidosis, coagulopathy, septic shock and multiorgan failure [1]. Although it is a rapidly changing scenario, based on current data, most people infected by the SARS-CoV-2 virus present a mild disease. Besides, depending on the study, different proportions of severe and critical patients, as well as mortality rates have been reported. In the case of Spain, a work that included 4035 hospitalized COVID-19 patients from 127 centres, reported 18.5% intensive care unit (ICU) admission and one-month mortality of 28%, which raised to 42.4% among ICU patients [2]. Prediction of the clinical course of each individual becomes a priority to optimize both the personalized treatment of each patient and the use of the medical resources.

Several works have tried to identify risk and prognostic factors based on demographic, epidemiologic, genetic, and clinical variables [2–6], and several models have been proposed for mortality and severity prediction, but they present diverse outcomes and no consensus has been reached yet on their application[7].

With this in mind, in this work, we took advantage of the routine blood tests performed on all individuals on the first day of hospitalization and analyzed the results of 1,082 COVID-19 patients generating a predictive model, based on these first-day data, that allows us to efficiently predict the course of the

disease. Our results also highlight the importance of the immature granulocytes and their ratio with lymphocytes as parameters to be studied.

## Methods

This study has been conducted at the Donostia University Hospital, which covers the region of Gipuzkoa (Basque Country, Spain), with a population of 720,000 inhabitants. Data of the first blood analysis after the COVID-19 diagnosis have been retrieved from the clinical records. The inclusion criteria were having a SARS-CoV-2 positive PCR between February and April 2020 and being admitted to the hospital. Demographic data and variables are presented in **Tables 1** and **2**. The data has been obtained anonymously and the follow-up has been done until 30 days after the discharge of the hospital. The project has been approved by the Ethical Committee from the hospital (PI2020076).

The severity categorization was done *a posteriori* according to [8]. “Severe cases” were defined as the ones meeting any of the following criteria: (1) Respiratory distress ( $\geq 30$  breaths/min); (2) Oxygen saturation  $\leq 93\%$  at rest; (3) Arterial partial pressure of oxygen (PaO<sub>2</sub>)/fraction of inspired oxygen (FiO<sub>2</sub>)  $\leq 300$ mmHg. Those patients that did not meet any of the conditions were considered “Non-severe cases”.

Continuous variables have been described by the mean, minimum and maximum data, while categorical variables have been described by the number of cases and percentages. For the continuous variables, the comparison of the means has been done using T-test, when the variables were normally distributed, and Mann-Whitney-Wilcoxon (2 categories) / Kruskal-Wallis (>2 categories) tests for the non-normally distributed. For the categorical variables, Chi-squared and Cramér’s V index test were applied. The fold change (FC) was calculated for the variables that had a mean value out of the normal range. For increased variables, the mean value was divided by the maximum value of the normal range, and in those below the normal range, by dividing the minimum value of the normal range by the mean value. FC values  $>|2|$  and a p-value lower than 0.05 were considered relevant. All statistical analyses have been done using R (R.3.6.3)[9]. Statistical significance was defined as follows:  $p < 0.05^*$ ,  $p < 0.01^{**}$  and  $p < 0.001^{***}$ .

We used AutoDiscovery (Butler Scientifics) for a stratified feature selection and to generate our final classification model. Briefly, a data set of 42 features with the data of 1,082 patients was constructed and 37 of these variables were chosen for the analysis [33 variables identified plus age group, sex, severity and the Lymphocyte to Immature Granulocyte ratio (L/IG)]. This L/IG ratio has been included based in clinical observation and in the analysis of our data. The next step was an automated data exploration, where the association and correlation between the variables were analyzed (AutoDiscovery Strength Score) to avoid model overfitting and to identify the patient’s features that were more associated with the predictive goal. In this specific case, the association of the variables with the severity variable was evaluated and one of the correlated pair of variables was removed, if any. To do that, AutoDiscovery selected the proper numerical method based on the data type and distribution of the variables assessed.

Numerical tests applied in each case are Spearman's Rank Correlation, Variance Analysis (ANOVA one-way, U Mann-Whitney, and Kruskal-Wallis; normality was tested in these cases with D'Agostino/Pearson methods), and Cramer's V Contingency Index.

This massive data exploration process was carried out systematically for each patient group (stratum), considering only those correlations evaluated in a data subset with  $n \geq 5$ , two-tailed  $p < 0.05$  and Spearman's rank correlation coefficient  $> 0.8$  to be relevant.

Given the nature of this multiple testing method, a high significance threshold was calculated based on Benjamini-Hochberg method (False-Discovery Rate) to classify the rest associations.

The Feature Selection process, where, based on Recursive Feature Elimination (RFE) and the AutoDiscovery Strength Score, age decade and the Lymphocytes to Immature Granulocyte ratio were added as variables for the model.

The model was trained and evaluated using three models: Top-5, Top-10 and Top-18( using the 5, 10 and 18 most significant variables), plus Age group and L/IG). The training and evaluation processes included the utilization of multi-class classifiers, such as Decision Trees, Logistic Regression, Naïve Bayes, K-Neighbors, SVM and Random Forest. The Model Training step is described in **Figure 1**, using the 60% of the samples for training, a 20% for validating and the final 20% for the testing, where the most optimized accuracy is obtained. Finally, a Cross-Validation method is applied to decrease overfitting impact.

The results of all these steps are the Confusion Matrix, the accuracy and the ROC curve of each of the three Top models.

## Results

The sex, age and clinical data of the 1082 participants are presented in **Table 1**. The incidence of severe cases, ICU admission and deceased patients was significantly higher among male individuals. A higher percentage of severe cases was reported with increasing age, while it is reduced in nonagenarians when compared to octogenarians (p-value = 0.04) (**Figure 2**).

**Table 1.** Demographic and clinical data of the study cohort. Percentages for the first column were calculated for the total number of patients (1082). The percentages shown for Females and Males were calculated with respect to the total number of patients in each of the conditions, and the p-value was calculated for these comparisons. ICU = Intensive care Unit.

	All	Female	Male	p-value
<b>Samples (%)</b>	1082	545 (50.37%)	537 (49.63%)	0.81
<b>Mean age (min-max)</b>	66.5 (10-99)	66.48 (10-99)	66.52 (17-95)	--
<b>Severe cases (%)</b>	449 (41.5%)	184 (40.98%)	265 (59.02%)	< 0.001 ***
<b>ICU admission (%)</b>	60 (5.55%)	16 (26.67%)	44 (73.33%)	< 0.001 ***
<b>Deceased (%)</b>	140 (12.94%)	54 (38.57%)	86 (61.43%)	0.004 **

The mean value of most of the variables was in the normal range, only 4 of them have a fold change > |2|. Patients have more D-dimer, Ferritin, IL-6, and remarkably, 10.03 times more C-reactive proteins (**Table 2**).

**Table 2.** The normal range of the study variables and the data from COVID-19 patients were compared. Fold change was calculated when the mean value of patients was above the maximum or below the minimum value of the normal range.

	Normal Range	COVID-19 (min-max)	Fold Change
Albumin (g/dl)	3.7 - 5.1	3.9 (2.07 - 4.94)	-
Creatine (mg/dl)	0.4 - 1	1.02 (0.27 - 13.5)	1.02
Urea (mg/dl)	10 - 65	43.97 (8 - 201)	-
Triglycerides (mg/dl)	35 - 135	129.04 (34 - 1265)	-
Creatine kinase (CK) (U/l)	0 - 188	150.92 (20 - 2662)	-
ALT/SGPT (%)	0 - 33	25.79 (2 - 395.5)	-
AST/SGOT (U/l)	0 - 31	30.33 (7 - 503)	-
LDH (U/l)	60 - 190	277 (110 - 2290)	1.46
Troponin T (TnT) (ng/l)	0 - 14	20.32 (2 - 393)	1.45
<b>C-reactive protein (mg/l)</b>	<b>0 - 5</b>	<b>50.16 (0 - 410)</b>	<b>10.03</b>
<b>IL-6 (pg/ml)</b>	<b>0 - 7</b>	<b>25.64 (1.72 - 166)</b>	<b>3.66</b>
Prothrombin Index (%)	70 - 140	93.58 (4 - 137)	-
Fibrinogen (mg/dl)	200 - 400	530.15 (266 - 780)	1.33
<b>D-dimer (ng/ml)</b>	<b>0 - 500</b>	<b>1351.25 (180 - 21000)</b>	<b>2.7</b>
<b>Ferritin (ng/ml)</b>	<b>15 - 150</b>	<b>426.71 (6.7 - 4557.3)</b>	<b>2.84</b>
Procalcitonin (ng/ml)	0 - 0.5	0.36 (0.01 - 18.51)	-
Lactate (mmol/l)	0.3 - 2	1.37 (0.5 - 6)	-
Platelets (10 <sup>3</sup> /μl)	140 - 400	215.69 (12 - 729)	-
Leukocytes (10 <sup>3</sup> /μl)	3.8 - 10	7.07 (0.03 - 77.44)	-
Neutrophils (10 <sup>3</sup> /μl)	1.6 - 7.5	4.5 (0 - 18.81)	-
Neutrophils (%)	40 - 75	63.16 (0 - 95.6)	-
Eosinophils (10 <sup>3</sup> /μl)	0 - 0.6	0.12 (0 - 1.85)	-
Eosinophils (%)	0.5 - 7	1.75 (0 - 16.9)	-
Basophils (10 <sup>3</sup> /μl)	0 - 0.2	0.03 (0 - 0.16)	-
Basophils (%)	0 - 1.5	0.45 (0 - 2.1)	-
Immature granulocytes (10 <sup>3</sup> /μl)	0 - 0.03	0.04 (0 - 0.97)	1.35
Immature granulocytes (%)	0 - 0.5	0.55 (0 - 23.2)	1.11
Lymphocytes (10 <sup>3</sup> /μl)	0.9 - 3.5	1.82 (0.01 - 65.51)	-
Lymphocytes (%)	19 - 48	25.33 (0.9 - 88.9)	-
Monocytes (10 <sup>3</sup> /μl)	0.2 - 0.9	0.58 (0.02 - 7.36)	-
Monocytes (%)	3.05 - 12	8.96 (0.5 - 66.7)	-
Partial pressure of oxygen (mmHg)	83 - 108	73.18 (14 - 271)	-1.13
Oxygen saturation (%)	95 - 99	92.94 (21.2 - 100.8)	-1.02

To determine if any of these variables could be good predictors of disease severity, patients were grouped based on their severity outcome. Notably, as shown in **Table 3**, we found that 32 out of the 33 variables tested were significantly different between severe and non-severe cases. Eleven variables show lower values in severe group, 21 were increased. There was a strong statistical significance for most of the variables, with 26 variables with a p-value < 0.001. We also calculated the L/IG for all the patients. The obtained results demonstrate that there is a strongly significant reduction of the L/IG in the first blood analysis of patients that then develop a severe disease.

**Table 3.** Comparison of the study variables between the COVID-19 patients that have a “Non-severe course” and the ones that develop a “Severe disease”. 32 of the 33 evaluated variables are significantly different. For an easier visual interpretation of the results, analytes with lower values in severe patients are shown in white and the analytes with higher values in gray. The “Sample n” column indicates the number of data available for each analyte.

	Normal Range	COVID-19 (min-max)	Non-Severe (min-max)	Severe (min-max)	p-value	Sample n
Albumin (g/dl)	3.7 - 5.1	3.9 (2.07 - 4.94)	4 (2.07 - 4.94)	3.76 (2.15 - 4.82)	< 0.001 ***	309
Creatine (mg/dl)	0.4 - 1	1.02 (0.27 - 13.5)	0.92 (0.27 - 13.5)	1.16 (0.29 - 10.44)	< 0.001 ***	869
Urea (mg/dl)	10 - 65	43.97 (8 - 201)	38.49 (8 - 193.2)	51.21 (11 - 201)	< 0.001 ***	685
Triglycerides (mg/dl)	35 - 135	129.04 (34 - 1265)	118.18 (34 - 492)	156.74 (54 - 1265)	< 0.001 ***	213
Creatine kinase (CK) (U/l)	0 - 188	150.92 (20 - 2662)	137.78 (25 - 2662)	168.97 (20 - 1322)	0.03 *	216
ALT/SGPT (%)	0 - 33	25.79 (2 - 395.5)	24.9 (5 - 395.5)	27.32 (2 - 228.5)	0.01 **	691
AST/SGOT (U/l)	0 - 31	30.33 (7 - 503)	29.79 (7 - 503)	31.26 (9 - 115)	< 0.001 ***	462
LDH (U/l)	60 - 190	277 (110 - 2290)	245.48 (117 - 727)	319.81 (110 - 2290)	< 0.001 ***	382
Troponin T (TnT) (ng/l)	0 - 14	20.32 (2 - 393)	11.55 (2 - 162)	32.79 (2 - 393)	< 0.001 ***	213
C-reactive protein (mg/l)	0 - 5	50.16 (0 - 410)	31.89 (0 - 206.1)	74.26 (0.2 - 410)	< 0.001 ***	603
IL-6 (pg/ml)	0 - 7	25.64 (1.72 - 166)	7.17 (1.72 - 25.6)	99.5 (33 - 166)	0.04 *	10
Prothrombin Index (%)	70 - 140	93.58 (4 - 137)	96.45 (6 - 136)	89.72 (4 - 137)	0.02 *	539
Fibrinogen (mg/dl)	200 - 400	530.15 (266 - 780)	473.67 (266 - 780)	600.75 (426 - 765)	0.01 **	27
D-dimer (ng/ml)	0 - 500	1351.25 (180 - 21000)	861.93 (180 - 12650)	2010.07 (270 - 21000)	< 0.001 ***	359
Ferritin (ng/ml)	15 - 150	426.71 (6.7 - 4557.3)	317.18 (6.7 - 2100)	629.86 (20.4 - 4557.3)	< 0.001 ***	354
Procalcitonin (ng/ml)	0 - 0.5	0.36 (0.01 - 18.51)	0.17 (0.01 - 16.8)	0.59 (0.02 - 18.51)	< 0.001 ***	349
Lactate (mmol/l)	0.3 - 2	1.37 (0.5 - 6)	1.19 (0.5 - 3.9)	1.5 (0.6 - 6)	< 0.001 ***	268
Platelets (10 <sup>3</sup> /μl)	140 - 400	215.69 (12 - 729)	221.79 (44 - 729)	206.29 (12 - 560)	< 0.001 ***	874
Leukocytes (10 <sup>3</sup> /μl)	3.8 - 10	7.07 (0.03 - 77.44)	6.56 (1.41 - 54.32)	7.86 (0.03 - 77.44)	< 0.001 ***	875
Neutrophils (10 <sup>3</sup> /μl)	1.6 - 7.5	4.5 (0 - 18.81)	4.01 (0.61 - 14.21)	5.25 (0 - 18.81)	< 0.001 ***	875
Neutrophils (%)	40 - 75	63.16 (0 - 95.6)	60.28 (7.3 - 90.5)	67.59 (0 - 95.6)	< 0.001 ***	875
Eosinophils (10 <sup>3</sup> /μl)	0 - 0.6	0.12 (0 - 1.85)	0.13 (0 - 1.85)	0.1 (0 - 1.36)	< 0.001 ***	875
Eosinophils (%)	0.5 - 7	1.75 (0 - 16.9)	2.01 (0 - 16.9)	1.35 (0 - 14.6)	< 0.001 ***	875
Basophils (10 <sup>3</sup> /μl)	0 - 0.2	0.03 (0 - 0.16)	0.03 (0 - 0.13)	0.03 (0 - 0.16)	0.02 *	875
Basophils (%)	0 - 1.5	0.45 (0 - 2.1)	0.48 (0 - 2.1)	0.4 (0 - 1.8)	< 0.001 ***	875
Immature granulocytes (10 <sup>3</sup> /μl)	0 - 0.03	0.04 (0 - 0.97)	0.03 (0 - 0.29)	0.06 (0 - 0.97)	< 0.001 ***	875
Immature granulocytes (%)	0 - 0.5	0.55 (0 - 23.2)	0.42 (0 - 8.5)	0.75 (0 - 23.2)	< 0.001 ***	875
Lymphocytes (10 <sup>3</sup> /μl)	0.9 - 3.5	1.82 (0.01 - 65.51)	1.79 (0.26 - 48.29)	1.86 (0.01 - 65.51)	< 0.001 ***	875
Lymphocytes (%)	19 - 48	25.33 (0.9 - 88.9)	27.77 (3.5 - 88.9)	21.58 (0.9 - 86)	< 0.001 ***	875
Monocytes (10 <sup>3</sup> /μl)	0.2 - 0.9	0.58 (0.02 - 7.36)	0.57 (0.12 - 1.9)	0.6 (0.02 - 7.36)	-	875
Monocytes (%)	3.05 - 12	8.96 (0.5 - 66.7)	9.18 (1.7 - 38.6)	8.63 (0.5 - 66.7)	< 0.001 ***	875
Partial pressure of oxygen (mmHg)	83 - 108	73.18 (14 - 271)	81.73 (27 - 154)	68.53 (14 - 271)	< 0.001 ***	321
Oxygen saturation (%)	95 - 99	92.94 (21.2 - 100.8)	96.66 (46.5 - 99.9)	90.92 (21.2 - 100.8)	< 0.001 ***	318
Lymphocyte to Immature Granulocyte ratio		84.97 (0.75 - 889)	101.26 (2.75 - 889)	59.94 (0.75 - 430)	< 0.001 ***	875

Since almost all variables appeared significantly different between severity groups, even if the mean value of most of them was in the normal range, it was decided to create a classification model to find the most relevant variables that would be able to predict the severity outcome of the patients at an early stage of the disease.

AutoDiscovery software was used for the creation of a predictive model. In the Automated Data Exploration step an association of almost all variables with severity variable was found. The ranking of the variables obtained in the Feature Selection step is shown in **Fig 3**, where the specific variables of each model are described.

Among all the multi-class classifiers, Random Forest showed the best performance. After the Cross-Validation process and the application of RFE, the results have shown an accuracy of around 60% for the three models, being the Top-5 the best model, with an accuracy of 67.33%. The Top-5 model also achieved the best confusion matrix and ROC curve (**Figure 4**).

## Discussion

Our study showed higher severity, ICU admission and death rates among male patients (**Table 1**), in accordance with previously reports [10]. The age distribution of severe cases followed a very similar pattern to that reported in the Italian population, with an increasing severity with age and a maximum peak of severe cases in octogenarians, followed by a decrease of this percentage in nonagenarians. Besides, the ICU occupancy (critical patients) was also similar, resulting in a 5.5% in our analysis and a 5.3% in the Italian study. The mortality rate was also analogous to previous reports, which was 12.94% in the present study, close to the 11% estimated for the Spanish population in the literature[2].

We analyzed classical parameters recovered from blood analysis. In total, 33 variables were considered for the analysis, among which, only the absolute number of monocytes did not show significant differences between groups (**Table 3**). The analytical alterations observed in the biochemistry of COVID-19 patients are the reflection of the degree and the severity of multi-organ involvement [11], similar to what is observed in septic conditions of bacterial origin. These conditions commonly appear due to endothelial dysfunction and abnormal activation of the coagulation cascade, whose origins and mechanism are not completely understood yet [12].

It should be noted that the mean values of most of the analyzed variables were within normal ranges (**Table 2**), and their power as biomarker comes from the conjunction of the data more than from each data itself. This was indeed confirmed by AI algorithms that offer the opportunity to identify new variables that could be used as good predictors even when they are within normal ranges.

It is worth mentioning that sepsis – a complex and potentially lethal syndrome, in which an infection produces physiopathological changes that cannot be resolved to get back to homeostasis – is commonly associated with a bacterial infection. However, the trigger of the symptoms and analytical alterations characteristic of sepsis can have other origins, including viral infections. Sepsis cases originated by viral infections account for 1-33% of cases in adults, and it rises to 76% in pediatric patients [13]. Among the viral etiology of sepsis, the herpes simplex viruses, the enteroviruses, the human parechoviruses, the influenza viruses, the adenoviruses and the dengue virus have been well characterized, recently, cases of septic shocks have also been reported in patients infected with the SARS-CoV-2 virus[14]. Independently of the etiology, septic shocks are characterized by hyperinflammatory processes, with augmented IL-6 and other proinflammatory cytokines like IL-1beta, IL-2, IL-8 and IFN-alpha[15,16]. As reported in COVID-19, the cytokine storm is largely responsible for disease severity and it can have lethal outcomes[17]. Regarding the study of lymphocyte populations, a marked T-cell lymphopenia was reported in the literature in patients with severe disease. There was a decrease in the absolute number of CD4 lymphocytes, an increase in the percentage of naïve CD4+ lymphocytes and a decrease in the percentage of memory CD4 T lymphocytes. Besides, a decrease in regulatory T lymphocytes and an increase in subpopulation CD8+ and CD28+ T lymphocytes (suppressor cytotoxic lymphocytes) were found [13]. Likewise, a decrease in CD8+ lymphocytes [18] and an increase in exhausted CD8+ T lymphocytes have been shown, indicating that the proportions of the lymphocyte populations are different between severe and non-severe COVID-19 patients [11]. Accordingly, in our cohort, the percentage of lymphocytes also showed statistically significant lower levels in severe compared to non-severe patients. Other parameters

such as platelets, leukocytes, neutrophils, eosinophils, basophils and the percentage of monocytes showed statistically significant differences between severe and non-severe patients. These results are in accordance with what is reported in the literature [19]

A parameter that was initially little explored in COVID-19 disease and that has shown to have prognostic implications in bacterial sepsis is the number of immature granulocytes (IG) [20,21], suggesting that the dysregulation of the myeloid compartment plays a role in the physiopathology of the COVID-19 disease. In our study, IG was the variable that showed the greatest discriminatory weight, between those chosen by the AI algorithm. Automated hematology analyzers enable a rapid and cost-effective blood cell count can be performed in fresh blood avoiding artifacts in the populations studies due to congelation. The percentage of immature granulocytes obtained by this technique give us an approach to the presence of a higher production of myeloid-derived immature cells with suppression capacity (MDSCs). Interestingly, augmented MDSCs have been found in COVID-19 patients with a severe disease course [22–24]. It has been proposed that MDSCs expand their numbers under inflammatory stimuli and that they can explain, at least in part, processes such as L-arginine mediated lymphopenia [25].

Therefore, the observed inverse relation between lymphocytes and immature myeloid cells, the positive relation between absolute numbers of IG and IL-6 (previously reported in COVID-19 [23], and the physiopathological basis that explains, in part, the lymphopenia, led us to calculate the Lymphocyte to Immature Granulocyte ratio. Importantly, we found that this indirect measure of systemic inflammation was significantly decreased in patients that developed a severe COVID-19 disease when compared to the ones that only suffered a non-severe disease (Table 3). In line with our results, other authors found the immature neutrophil-to-VD2 (or CD8) T-cell ratio as an early marker for severe COVID-19 disease, predicting pneumonia and hypoxia onset. Moreover, it has been also reported, among other variables, a differential immature granulocyte to lymphocyte ratio between critical and non-critical COVID-19 patients, reinforcing the relevance of this measure [26].

The alterations in the studied variables turn out to be useful for the prediction of unfavorable clinical outcomes, with the help and the utilization of IA and Machine learning tools. The algorithm chose ferritin, troponin, D-dimer and triglycerides among the most powerful variables to predict severity, the model with 5 variables shows an AUC=0.78. This tool needs to be tested in the real world but shows promising utility as a clinical tool for predicting disease severity.

It is interesting how the algorithm, without receiving normality ranges data, nor data of clinical relevance, choose these parameters among best predictors, which are, precisely, related to myocardial injury[27], thrombotic phenomena[28] and features that resemble hemophagocytic syndrome, all potentially fatal in a short period and related to the hyperinflammatory state[29].

In conclusion, the present work highlights the importance of the study of analytic variables in the early phases of hospitalization by an easy and standard cost-effective technique present in all the labs. Our results also underline the benefits of applying AI methods to identify the best variables, even although they present values in normal ranges, to identify the best classifier. The prompt identification of the risk

group is essential to thoroughly follow them and implement the necessary analytical and clinical interventions.

## Declarations

### *Ethical Approval*

The data has been obtained anonymously and the follow-up has been done until 30 days after the discharge of the hospital. This retrospective chart review study involving human participants was in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The Human Investigation Committee (IRB) of Hospital Universitario Donostia (PI2020076) approved this study.

### *Authors Contributions*

Roncancio-Clavijo A <sup>1-2\*</sup>, Gorostidi-Aicua M <sup>1\*</sup>, Alberro A <sup>1</sup>, Iribarren-Lopez A <sup>1</sup>, Butler R<sup>3</sup>, Lopez R<sup>3</sup>, Iribarren JA <sup>4</sup>, Clemente D<sup>5</sup>, Marimon JM <sup>6</sup>, Basterrechea J <sup>7</sup>, Martinez B <sup>7</sup>, Prada A <sup>1-2 ‡</sup>, Otaegui D <sup>1</sup>

Conception of the study: ARC, JAI, DO

Design of the study: ARC, MGA, AP, DO

Draft of the paper: ARC, MGA, AA, AIL, RB, JAI, JMM, DO

Acquisition of the data: ARC, JAI, JMM, JB, BM

Analysis of the data: MGA, RB, RL, DO

Interpretation of the data: ARC, AA, AIL, JAI, DC, AP, DO

Supervision and funding: AP, DO

Critical reading and writing of the final version: ARC, AA, JAI, DC, DO

### *Funding*

This project has been funded by Instituto de Salud Carlos III (COV20/00314) that include part of FEDER funds. AA is funded by a post-doctoral Grant from the Basque Government.

### *Competing interests.*

The authors declare that they have no conflict of interest

### *Availability of data and materials*

Biochemical raw Data will be available under request.

## ***Consents:***

All the authors have given their explicit consent to publish this work and have obtained the consent from the entities involved.

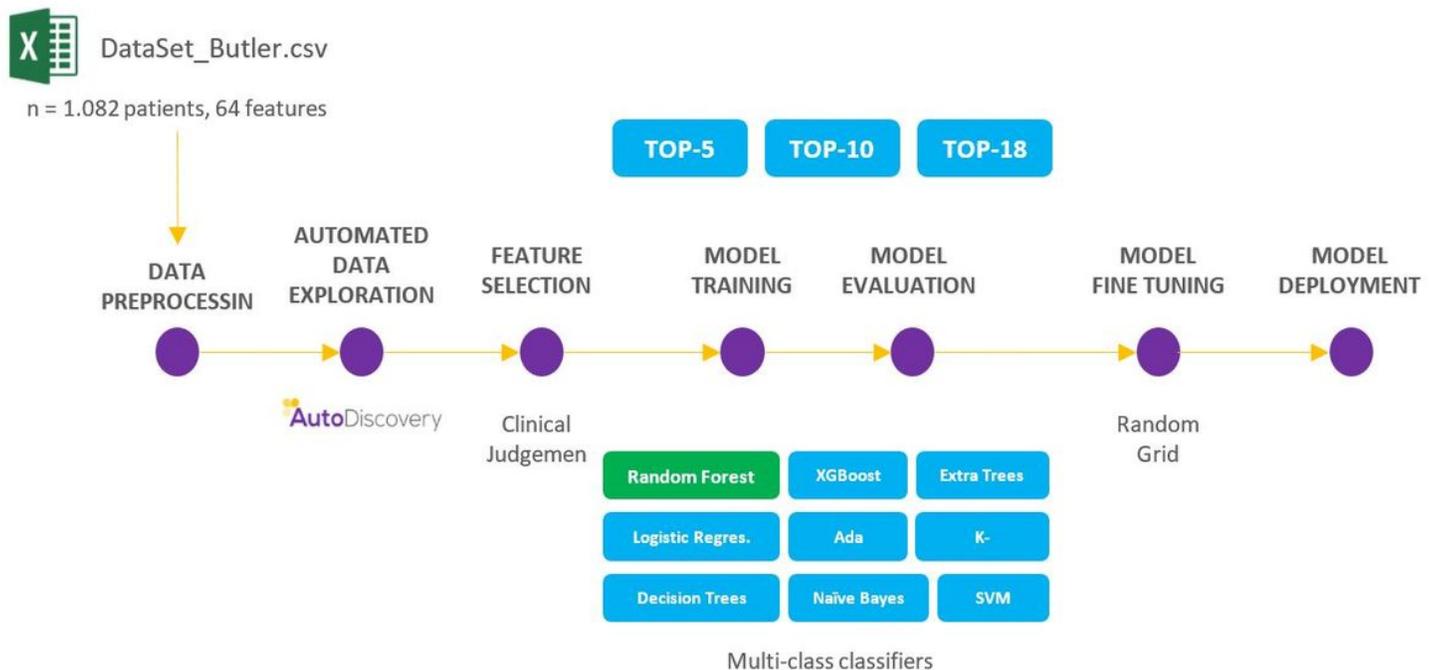
## **References**

- [1] Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. Features, Evaluation, and Treatment of Coronavirus. 2020.
- [2] Berenguer J, Ryan P, Rodríguez-Baño J, Jarrín I, Carratalà J, Pachón J, et al. Characteristics and predictors of death among 4035 consecutively hospitalized patients with COVID-19 in Spain. *Clin Microbiol Infect* 2020;26:1525–36. <https://doi.org/10.1016/j.cmi.2020.07.024>.
- [3] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054–62. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3).
- [4] Bonetti G, Manelli F, Patroni A, Bettinardi A, Borrelli G, Fiordalisi G, et al. Laboratory predictors of death from coronavirus disease 2019 (COVID-19) in the area of Valcamonica, Italy. *Clin Chem Lab Med* 2020;58:1100–5. <https://doi.org/10.1515/cclm-2020-0459>.
- [5] Van Dam PMEL, Zelis N, Stassen P, Van Twist DJL, De Leeuw PW, Van Kuijk S, et al. Validating the RISE up score for predicting prognosis in patients with COVID-19 in the emergency department: A retrospective study. *BMJ Open* 2021;11:45141. <https://doi.org/10.1136/bmjopen-2020-045141>.
- [6] Gupta RK, Harrison EM, Ho A, Docherty AB, Knight SR, van Smeden M, et al. Development and validation of the ISARIC 4C Deterioration model for adults hospitalised with COVID-19: a prospective cohort study. *Lancet Respir Med* 2021;9:349–59. [https://doi.org/10.1016/S2213-2600\(20\)30559-2](https://doi.org/10.1016/S2213-2600(20)30559-2).
- [7] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020;369:26. <https://doi.org/10.1136/bmj.m1328>.
- [8] On NHC& SA of TCM. Diagnosis and Treatment Protocol for Novel Coronavirus Pneumoniae. (trial version 7). 2020.
- [9] Team RC. R: A language and environment for statistical computing. 2017.
- [10] Karagiannidis C, Mostert C, Hentschker C, Voshaar T, Malzahn J, Schillinger G, et al. Case characteristics, resource use, and outcomes of 10 021 patients with COVID-19 admitted to 920 German hospitals: an observational study. *Lancet Respir Med* 2020;8:853–62. [https://doi.org/10.1016/S2213-2600\(20\)30316-7](https://doi.org/10.1016/S2213-2600(20)30316-7).

- [11] Zaim S, Chong JH, Sankaranarayanan V, Harky A. COVID-19 and Multiorgan Response. *Curr Probl Cardiol* 2020;45. <https://doi.org/10.1016/j.cpcardiol.2020.100618>.
- [12] Lazzaroni MG, Piantoni S, Masneri S, Garrafa E, Martini G, Tincani A, et al. Coagulation dysfunction in COVID-19: The interplay between inflammation, viral infection and the coagulation system. *Blood Rev* 2021;46. <https://doi.org/10.1016/j.blre.2020.100745>.
- [13] Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, et al. Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in Wuhan, China. *Clin Infect Dis* 2020;71:762–8. <https://doi.org/10.1093/cid/ciaa248>.
- [14] Li H, Liu L, Zhang D, Xu J, Dai H, Tang N, et al. SARS-CoV-2 and viral sepsis: observations and hypotheses. *Lancet* 2020;395:1517–20. [https://doi.org/10.1016/S0140-6736\(20\)30920-X](https://doi.org/10.1016/S0140-6736(20)30920-X).
- [15] Rubsamen R, Burkholz S, Massey C, Brasel T, Hodge T, Wang L, et al. Anti-IL-6 Versus Anti-IL-6R Blocking Antibodies to Treat Acute Ebola Infection in BALB/c Mice: Potential Implications for Treating Cytokine Release Syndrome. *Front Pharmacol* 2020;11. <https://doi.org/10.3389/fphar.2020.574703>.
- [16] Frank K, Paust S. Dynamic Natural Killer Cell and T Cell Responses to Influenza Infection. *Front Cell Infect Microbiol* 2020;10:425. <https://doi.org/10.3389/fcimb.2020.00425>.
- [17] Tang Y, Liu J, Zhang D, Xu Z, Ji J, Wen C. Cytokine Storm in COVID-19: The Current Evidence and Treatment Strategies. *Front Immunol* 2020;11:1708. <https://doi.org/10.3389/fimmu.2020.01708>.
- [18] Du RH, Liang LR, Yang CQ, Wang W, Cao TZ, Li M, et al. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: A prospective cohort study. *Eur Respir J* 2020;55. <https://doi.org/10.1183/13993003.00524-2020>.
- [19] Zhu Z, Cai T, Fan L, Lou K, Hua X, Huang Z, et al. Clinical value of immune-inflammatory parameters to assess the severity of coronavirus disease 2019. *Int J Infect Dis* 2020;95:332–9. <https://doi.org/10.1016/j.ijid.2020.04.041>.
- [20] Mare TA, Treacher DF, Shankar-Hari M, Beale R, Lewis SM, Chambers DJ, et al. The diagnostic and prognostic significance of monitoring blood levels of immature neutrophils in patients with systemic inflammation. *Crit Care* 2015;19. <https://doi.org/10.1186/s13054-015-0778-z>.
- [21] Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, et al. Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell* 2020;182:1419-1440.e23. <https://doi.org/10.1016/j.cell.2020.08.001>.
- [22] Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, et al. Suppressible myeloid cells are a hallmark of severe COVID-19. *MedRxiv* 2020;19:2020.06.03.20119818. <https://doi.org/10.1101/2020.06.03.20119818>.

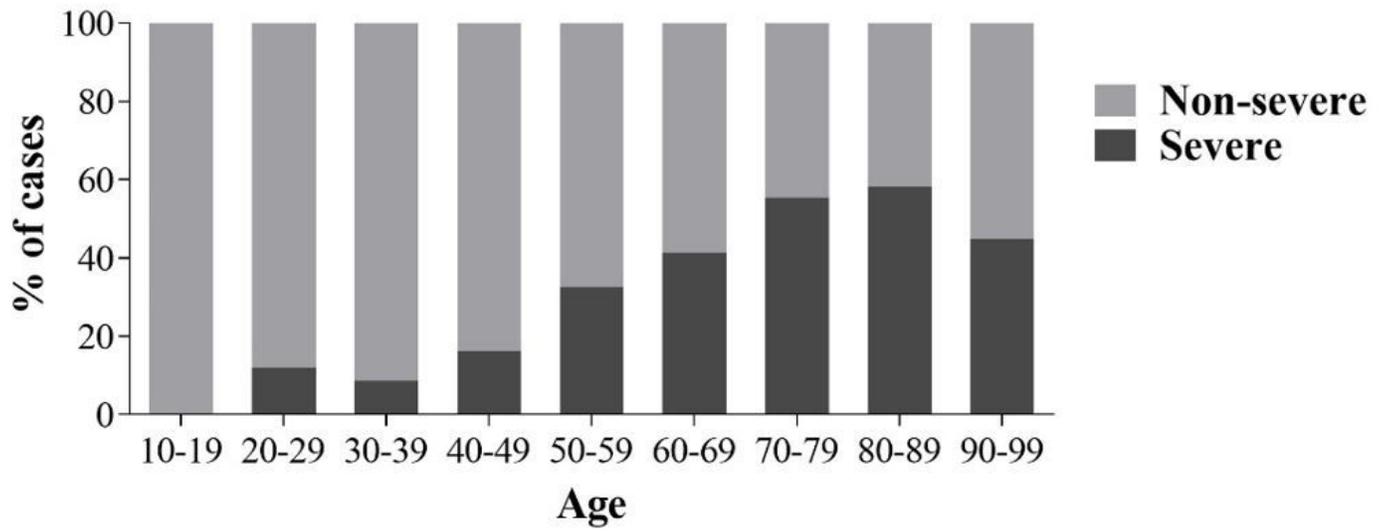
- [23] Carissimo G, Xu W, Kwok I, Abdad MY, Chan YH, Fong SW, et al. Whole blood immunophenotyping uncovers immature neutrophil-to-VD2 T-cell ratio as an early marker for severe COVID-19. *Nat Commun* 2020;11:1–12. <https://doi.org/10.1038/s41467-020-19080-6>.
- [24] Zini G, Bellesi S, Ramundo F, d’Onofrio G. Morphological anomalies of circulating blood cells in COVID-19. *Am J Hematol* 2020;95:870–2. <https://doi.org/10.1002/ajh.25824>.
- [25] Bronte V, Zanovello P. Regulation of immune responses by L-arginine metabolism. *Nat Rev Immunol* 2005;5:641–54. <https://doi.org/10.1038/nri1668>.
- [26] Linssen J, Ermens A, Berrevoets M, Seghezzi M, Previtali G, Brugge S van der S van der, et al. A novel haemocytometric covid-19 prognostic score developed and validated in an observational multicentre european hospital-based study. *Elife* 2020;9:1–37. <https://doi.org/10.7554/eLife.63195>.
- [27] Babapoor-Farrokhran S, Gill D, Walker J, Rasekhi RT, Bozorgnia B, Amanullah A. Myocardial injury and COVID-19: Possible mechanisms. *Life Sci* 2020;253. <https://doi.org/10.1016/j.lfs.2020.117723>.
- [28] Shi W, Lv J, Lin L. Coagulopathy in COVID-19: Focus on vascular thrombotic events. *J Mol Cell Cardiol* 2020;146:32–40. <https://doi.org/10.1016/j.yjmcc.2020.07.003>.
- [29] Opoka-Winiarska V, Grywalska E, Roliński J. Could hemophagocytic lymphohistiocytosis be the core issue of severe COVID-19 cases? *BMC Med* 2020;18. <https://doi.org/10.1186/s12916-020-01682-y>.

## Figures



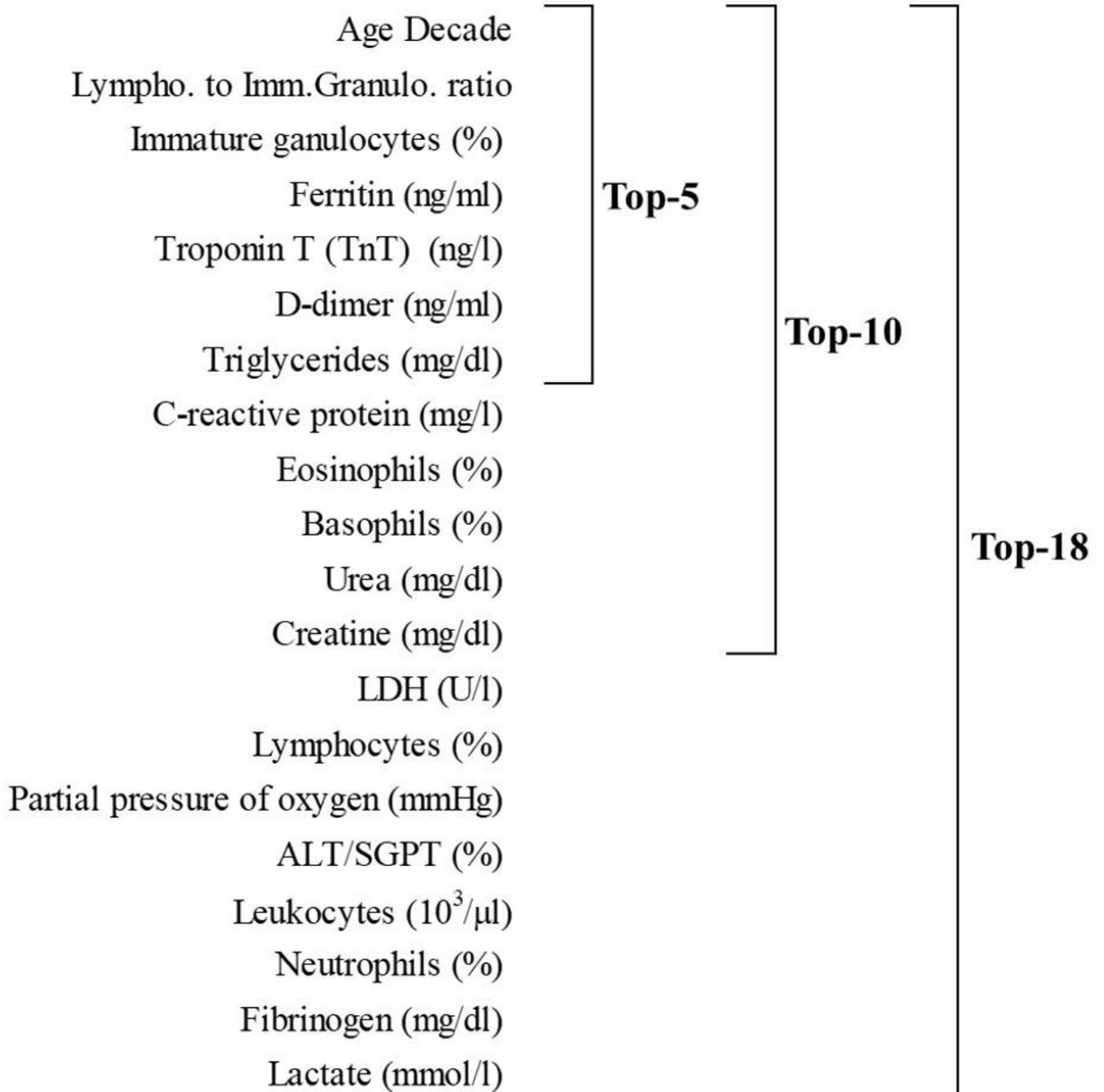
**Figure 1**

Schematic representation of the classifier analysis.



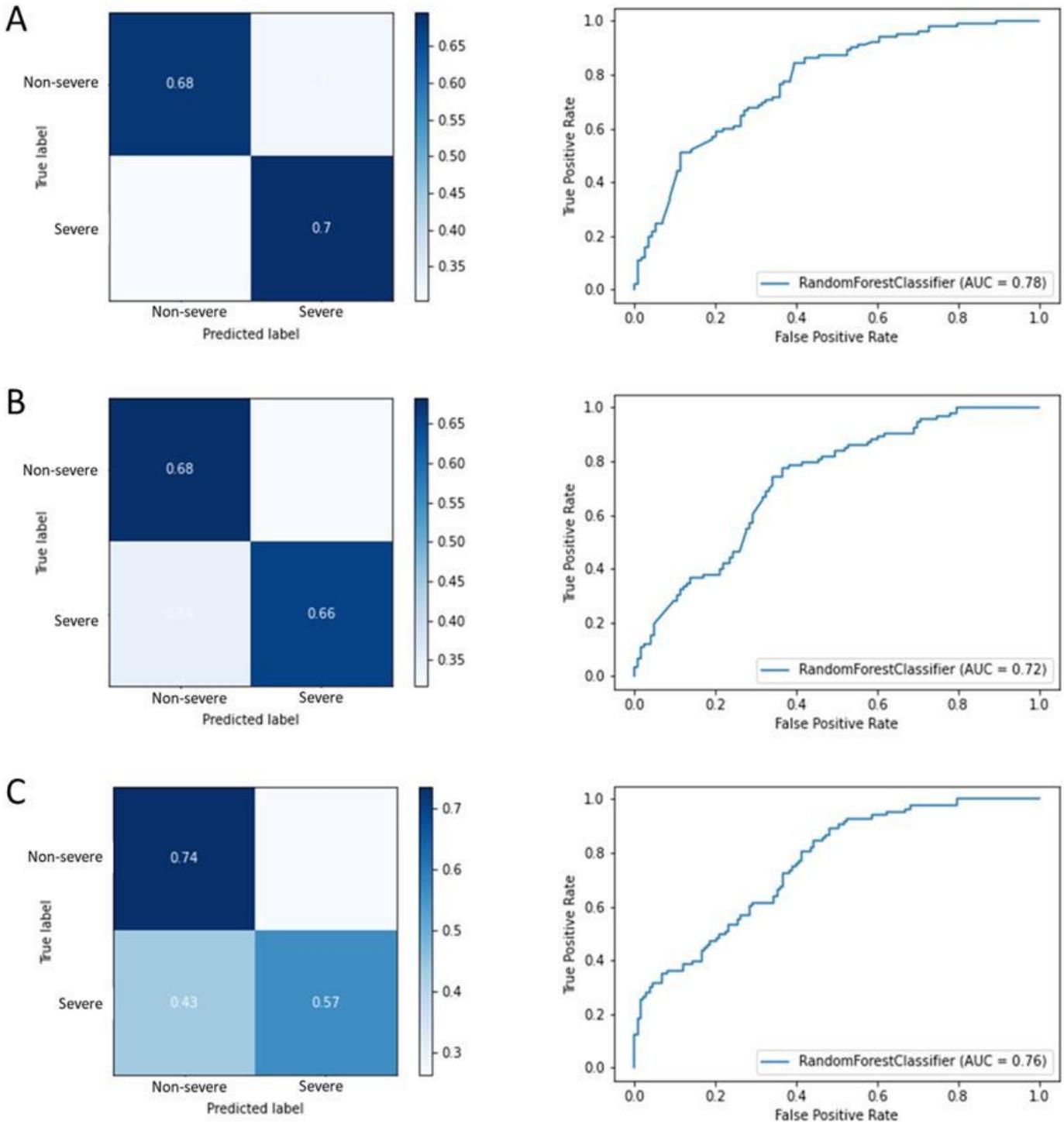
**Figure 2**

Age distribution of COVID-19 severity. For each of the age ranges the percentage of Non-severe and Severe cases are shown. Age is presented in years.



**Figure 3**

List of the variables of the three classifications models.



**Figure 4**

Classification models for the prediction of disease severity. The confusion matrix and ROC curves for the (A) Top-5, (B) Top-10 and (C) Top-18 models are shown.