



Article

# Combining Deep Phenotyping of Serum Proteomics and Clinical Data via Machine Learning for COVID-19 Biomarker Discovery

Antonio Paolo Beltrami <sup>1,2</sup>, Maria De Martino <sup>1</sup>, Emiliano Dalla <sup>1</sup>, Matilde Clarissa Malfatti <sup>1</sup>,  
Federica Caponnetto <sup>1</sup>, Marta Codrich <sup>1,2</sup>, Daniele Stefanizzi <sup>1</sup>, Martina Fabris <sup>2</sup>, Emanuela Sozio <sup>2</sup>,  
Federica D'Aurizio <sup>2</sup>, Carlo E. M. Pucillo <sup>1</sup>, Leonardo A. Sechi <sup>1</sup>, Carlo Tascini <sup>1,2,\*</sup>, Francesco Curcio <sup>1,2</sup>,  
Gian Luca Foresti <sup>3</sup>, Claudio Piciarelli <sup>3</sup>, Axel De Nardin <sup>3</sup>, Gianluca Tell <sup>1,\*</sup> and Miriam Isola <sup>1</sup>

<sup>1</sup> Department of Medicine (DAME), University of Udine, 33100 Udine, Italy

<sup>2</sup> Academic Hospital of Udine (ASUFC), 33100 Udine, Italy

<sup>3</sup> Department of Mathematics, Informatics and Physics (DMIF), University of Udine, 33100 Udine, Italy

\* Correspondence: carlo.tascini@uniud.it (C.T.); gianluca.tell@uniud.it (G.T.);

Tel.: +39-0432-559355 (C.T.); +39-0432-494311 (G.T.)

**Abstract:** The persistence of long-term coronavirus-induced disease 2019 (COVID-19) sequelae demands better insights into its natural history. Therefore, it is crucial to discover the biomarkers of disease outcome to improve clinical practice. In this study, 160 COVID-19 patients were enrolled, of whom 80 had a “non-severe” and 80 had a “severe” outcome. Sera were analyzed by proximity extension assay (PEA) to assess 274 unique proteins associated with inflammation, cardiometabolic, and neurologic diseases. The main clinical and hematochemical data associated with disease outcome were grouped with serological data to form a dataset for the supervised machine learning techniques. We identified nine proteins (i.e., CD200R1, MCP1, MCP3, IL6, LTBP2, MATN3, TRANCE,  $\alpha$ 2-MRAP, and KIT) that contributed to the correct classification of COVID-19 disease severity when combined with relative neutrophil and lymphocyte counts. By analyzing PEA, clinical and hematochemical data with statistical methods that were able to handle many variables in the presence of a relatively small sample size, we identified nine potential serum biomarkers of a “severe” outcome. Most of these were confirmed by literature data. Importantly, we found three biomarkers associated with central nervous system pathologies and protective factors, which were downregulated in the most severe cases.

**Keywords:** proximity extension assay; inflammation; cardiometabolic; neurologic disease; COVID-19



**Citation:** Beltrami, A.P.; De Martino, M.; Dalla, E.; Malfatti, M.C.; Caponnetto, F.; Codrich, M.; Stefanizzi, D.; Fabris, M.; Sozio, E.; D'Aurizio, F.; et al. Combining Deep Phenotyping of Serum Proteomics and Clinical Data via Machine Learning for COVID-19 Biomarker Discovery. *Int. J. Mol. Sci.* **2022**, *23*, 9161. <https://doi.org/10.3390/ijms23169161>

Academic Editors: Jacek Z. Kubiak and Malgorzata Kloc

Received: 15 July 2022

Accepted: 11 August 2022

Published: 15 August 2022

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



**Copyright:** © 2022 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

Since its first outbreak in Wuhan in 2019, the global pandemic of COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has affected more than 300 million people and caused more than 5.5 million deaths (<https://covid19.who.int>, accessed on 15 January 2022) [1,2]. From a clinical standpoint, COVID-19 is a challenge due to the heterogeneity of both individual susceptibilities to SARS-CoV-2 infection and disease severity, which ranges from an asymptomatic form to a critical illness with high rates of lethal complications [3–6].

Part of this heterogeneity could be attributed to the high mutation rates of SARS-CoV-2, a single-stranded enveloped RNA virus. In analogy with other coronaviruses (CoV), with which it shares 79% of sequence identity, the Spike (S) protein of SARS-CoV-2 is required for cell infection via the Angiotensin Converting Enzyme (ACE)-2 receptor [7]. Crucial mutational events, involving the coding sequence of the S protein, enable bat coronavirus to infect human cells and possibly account for the zoonotic transfer [2,8]. Recent studies have demonstrated the high complexity of the SARS-CoV-2 transcriptome owing to numerous

recombination events, including the presence of different ORFs bearing fusions, deletions, and/or frameshift mutations [9]. When these events also involve the coding region of the Spike protein, novel, more infectious variants emerge that could escape the immune system, as in the case of Omicron (B.1.1.529) [10].

Another component responsible for the heterogeneity of the clinical presentation could be ascribed to host features. Aside from ethnicity and social determinants, age, sex, smoking habits, and comorbidities, such as diabetes, obesity, cardiovascular and cerebrovascular disease, chronic kidney disease (CKD), chronic obstructive pulmonary disease (COPD), and chronic liver disease and malignancy have been associated with COVID-19 poor outcome [5,11]. Additionally, the genetic background of the individual has been associated both with susceptibility and disease severity. Among the most important elements, we should acknowledge genomic variants affecting the expression of factors critically involved in: viral entrance (e.g., *ACE2*, *Transmembrane Serine Protease 2-TMPRSS-*, *Cathepsin B* and *L*, the blood group antigens A and O), innate antiviral response (e.g., *Oligoadenylate Synthetase-OAS-* and *Interferon (IFN) alpha/beta Receptor 2-IFNAR2-*), inflammation (e.g., *Dipeptidyl Dipeptidase 9-DPP9-*, *Tyrosine Kinase 2-TYK2-*, *Toll-like Receptor 7-TLR7-*, and *C-C Chemokine Receptor Type 2-CCR2-*), and immunity (e.g., *Human Leukocyte Antigen E-HLA-E-*, *Killer Cell Lectin-like Receptor C2-KLRC2-*) [11–13]. Finally, the persistence of memory T cells cross-reactive with other coronaviruses has been evoked as a potentially protective factor against SARS-CoV-2 infection [6].

Critically ill patients are characterized by a hyperinflammatory state associated with endothelitis, hypercoagulability, organ damage, and death [4]. Indeed, a dysregulated immune response is associated with patient outcomes and, in extreme cases, can also trigger multi inflammatory syndrome in children (MISC) [14]. Indeed, literature data show that SARS-CoV-2 infection causes an increased secretion of Interleukin (IL)-1 $\beta$ , Interferon (IFN)- $\gamma$ , Interferon Gamma-Induced Protein 10 (IP-10), Monocyte Chemoattractant Protein (MCP)-1, IL-4, and IL-10 [15], while disease severity has been associated with an elevation of IL-2, IL-7, Granulocyte Colony-Stimulating Factor (GCSF), Macrophage Inflammatory Protein-1 Alpha (MIP-1A), Tumor Necrosis Factor alpha (TNF- $\alpha$ ), IL-10, IP-10, and MCP-1 [15–17]. The cardiovascular system is also crucial in dictating the outcome of COVID-19 patients, and cardiac damage markers, such as troponins and brain-type natriuretic peptide (BNP), have been associated with increased mortality [7]. Last, several neurological (e.g., anosmia and ageusia) and psychiatric symptoms have been described in COVID-19 patients, most of which are independent from a direct affection of the central nervous system (CNS) by the virus but are considered to be secondary to the immune reaction [18,19]. CNS involvement is also relevant for the post-acute sequelae of SARS-CoV-2 infection (PASC), a spectrum of symptoms experienced by COVID-19 patients that persist after the resolution of the infection, which is characterized by the high prevalence of neurologic symptoms [19,20]. However, novel studies investigating the evidence of CNS involvement in COVID-19 patients are needed [21].

Identifying new biomarkers associated with disease severity is both crucial to help understand the natural history of COVID-19 and to prospectively identify patients at high risk of developing a critical illness in order to target the allocation of resources, the escalation of care, and inclusion in experimental clinical trials to those who are in most need.

Analytically, there is the need to increase the number of biomolecules assessed per volume of sample to maximize the information obtained and spare biological material [22]. Proximity extension assay (PEA) is a high-throughput technology developed for the high-multiplex analysis of proteins, allowing a successful detection and quantification of several biomarkers [23] by incubating a low amount of the biological sample. To date, some comprehensive studies identifying sera biomarkers correlated with COVID-19 disease severity have been published [24–29], but only few of them [25] have employed machine learning techniques to reach a holistic view that summarizes an extensive clinical characterization (including demographic, comorbidity, clinical, and hematochemical data) of the patients

with a large number of potential biomarkers in order to identify those that more effectively classify patients with adverse prognosis.

Given these premises, we decided to compare the serum levels of 274 proteins associated with inflammation, cardiometabolic, and neurologic diseases employing PEA analysis in a case series of 160 COVID-19 patients, dichotomized based on COVID-19 severity. Age, comorbidities, and vaccination status significantly differed between the two groups of patients. Analyses were carried out employing machine learning techniques and nine proteins were identified as being able to classify, when combined with relative lymphocyte and neutrophil counts, disease severity. Of note, neurology-associated protein biomarkers were among the strongest predictors of COVID-19 severity.

## 2. Results

### 2.1. Patients' Characteristics

Between February and September 2021, we enrolled 160 subjects infected by SARS-CoV-2, including 80 patients who were either paucisymptomatic or affected by a mild to moderate form of the disease (here referred to as “non-severe”) and 80 patients who were either affected by a severe to critical COVID-19 at disease onset or who had moderate disease at the onset that worsened and required admission to the ICU (here referred to as “severe”). Demographic, risk factor, and comorbidity data of the enrolled population are summarized in Table 1. Enrolled patients were mainly males, older than 65 years, on average overweight (median BMI 28.2), and had a median Charlson comorbidity index (CCI) of 3. Importantly, having been conducting the enrollment at the beginning of the vaccine campaign, 85% of patients had still not been vaccinated. Although the genotypization of the SARS-CoV-2 strain was conducted only in a minority of patients, available data indicate that they were affected mostly by either the alpha (B.1.1.7) or delta (B.1.167.2) variants of concern (VOC), in consistence with the survey conducted by the Istituto Superiore di Sanità (ISS, Italy) and the data of strain prevalence obtained at our institution during the same timeframe [30,31]. The “severe” patients were characterized by significantly older age, higher CCI, and lower vaccination rate in comparison to the “non-severe” ones.

**Table 1.** Summary of baseline characteristics of enrolled patients. Baseline demographic, clinical features (comorbidities), and vaccination status of the enrolled patients ( $n = 160$ ). Patients were stratified according to the severity of the disease into a “non-severe” group and a “severe” one. Data are presented as either percentage or median and interquartile range (IQR). Results of the comparison between “non-severe” vs. “severe” patients are shown in the right column ( $p$ -value). Significant results are shown in bold.

	Total ( $N = 160$ )	Non-Severe ( $N = 80$ )	Severe ( $N = 80$ )	$p$ -Value
<b>Gender, <math>n</math> (%)</b>				
Female	59 (36.9)	33 (41.2)	26 (32.5)	0.251
Male	101 (63.1)	47 (58.8)	54 (67.5)	
<b>Age, years, median (IQR)</b>	67 (56–76)	61 (50–74)	70 (63–77)	<b>0.001</b>
<b>BMI, median (IQR)</b>	28.2 (24.9–31.2)	27.8 (24.9–31.1)	28.5 (25.5–31.5)	0.541
<b>Smoking habit, <math>n/N</math> (%)</b>				
Non-smoker	79/107 (73.8)	41/53 (77.4)	38/54 (70.4)	0.411
Smoker	28/107 (26.2)	12/53 (22.6)	16/54 (29.6)	
<b>Charlson comorbidity index, median (IQR)</b>	3 (1–5)	2 (1–5)	4 (2–5.5)	<b>0.007</b>
<b>Comorbidities, <math>n/N</math> (%)</b>				
Hypertension	79/155 (51.0)	39/76 (51.3)	40/79 (50.6)	0.932
Obesity	64/134 (47.8)	30/66 (45.4)	34/68 (50.0)	0.598
Diabetes	31/155 (20.0)	13/76 (17.1)	18/79 (22.8)	0.377
COPD	16/156 (10.3)	5/77 (6.5)	11/79 (13.9)	0.126
Cardiovascular disease	57/156 (36.5)	27/77 (35.1)	30/79 (38.0)	0.706
Liver disease	5/155 (3.2)	1/76 (1.3)	4/79 (5.1)	0.187
Renal impairment	13/156 (8.3)	6/77 (7.8)	7/79 (8.9)	0.809
Immunocompromised	14/147 (9.5)	6/67 (9.0)	8/80 (10.0)	0.830
<b>COVID-19 vaccination, <math>n</math> (%)</b>	24 (15.0)	20 (25.0)	4 (5.0)	<b>&lt;0.001</b>

The clinical outcome of the enrolled patients is summarized in Table 2. As expected, the clinical presentation of the whole case study was heterogeneous, including paucisymptomatic and critical patients, requiring invasive mechanical ventilation (IMV) in the Intensive Care Unit (ICU). The median duration of in-hospital stay was 9 days, and the mean rate of in-hospital death was 56%. The duration of in-hospital stay and the necessity for Continuous Positive Airway Pressure (CPAP) and IMV were significantly higher in the “severe” group of patients. Lastly, while no patient in the “non-severe” group was either admitted to the ICU or died, 51% and 56% of those in the “severe” group met these unfavorable outcomes.

**Table 2.** Summary of clinical characteristics of enrolled patients. Stratification of the enrolled patients according to the disease severity, necessity for invasive and non-invasive ventilation, admission to the intensive care unit, the duration of hospitalization, death rate, and the duration of infection. Data are presented as either percentage or median and interquartile range. Results of the comparison between non-severe vs. severe patients are shown in the right column (*p*-value). Significant results are shown in bold.

	Total ( <i>N</i> = 160)	Non-Severe ( <i>N</i> = 80)	Severe ( <i>N</i> = 80)	<i>p</i> -Value
<b>Acute COVID-19 severity, <i>n</i> (%)</b>				
<b>Paucisymptomatic</b>	16 (10.0)	16 (20.0)	0 (0.0)	
Mild	55 (34.4)	55 (68.7)	0 (0.0)	<b>&lt;0.001</b>
Moderate	14 (8.7)	9 (11.3)	5 (6.3)	
Severe	71 (44.4)	0 (0.0)	71 (88.7)	
Critical	4 (2.5)	0 (0.0)	4 (5.0)	
<b>CPAP, <i>n</i> (%)</b>	69 (43.1)	1 (1.2)	68 (85.0)	<b>&lt;0.001</b>
<b>Invasive mechanical ventilation (IMV), <i>n</i> (%)</b>	45 (28.1)	0 (0.0)	45 (56.2)	<b>&lt;0.001</b>
<b>Length of in-hospital stay, days, median (IQR)</b>	9 (5–15.5)	5.5 (4–9)	14 (9–19)	<b>&lt;0.001</b>
<b>ICU, <i>n</i> (%)</b>	51 (31.9)	0 (0.0)	51 (63.7)	<b>&lt;0.001</b>
<b>In-hospital death, <i>n</i> (%)</b>	56 (35.0)	0 (0.0)	56 (70.0)	<b>&lt;0.001</b>
<b>Duration of nasal swab positivity days, median (IQR)</b>	19 (14–24)	17 (12–20)	21.5 (17.5–29.5)	<b>0.008</b>

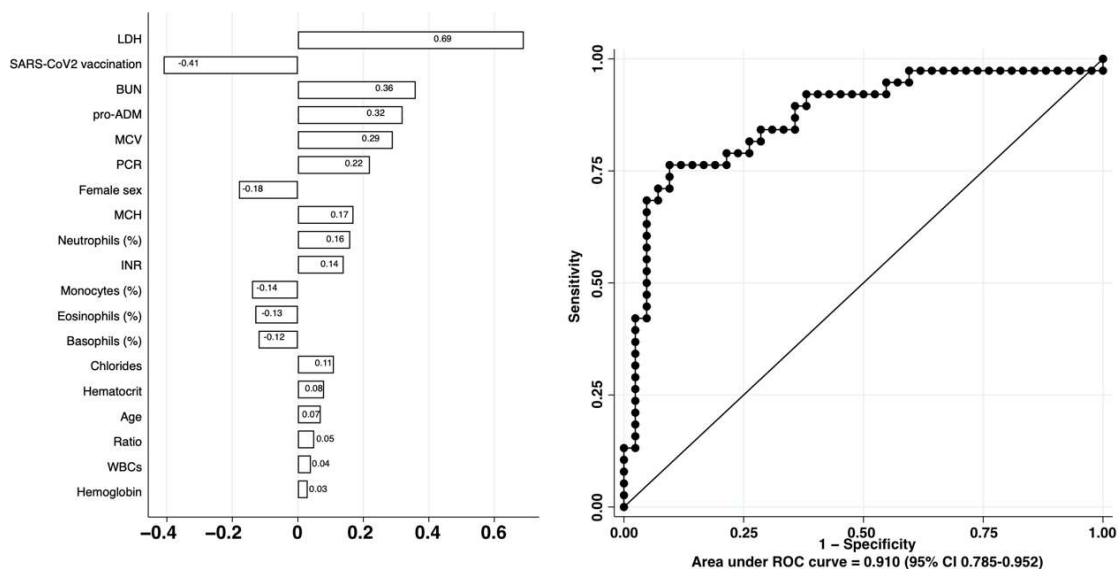
These data indicate that the two subsets of “severe” and “non-severe” patients have distinct, non-overlapping demographic and clinical baseline features with opposing clinical outcomes.

## 2.2. Clinical and Hematochemical Variables Associated with Patient Outcome

Next, we analyzed 32 different hematochemical variables collected at the time of enrollment (Supplementary Table S2) describing the hematologic, coagulation cascade, liver, kidney, and cardiac compensation of the enrolled patients together with 26 additional demographic and anamnestic variables (Supplementary Table S2). The identification, among the 58 variables, of the strongest classifiers of COVID-19 severity was assessed by employing an elastic net logistic regression approach with cross-validation. As shown in Figure 1, vaccination against SARS-CoV-2 and female sex had a protective effect, while patient age was weakly associated with a worse outcome.

Hematological abnormalities associated with unfavorable outcomes mostly involved the white blood cell (WBC) lineage. Specifically, while the WBC count was weakly associated with a worse outcome, relative neutrophilia, relative monocytopenia, and a relative reduction of basophils and eosinophils were more strongly associated with a worse clinical outcome. The defects also involved the red blood cell lineage, where increased mean corpuscular volume and mean cell hemoglobin were associated with disease severity. These data are consistent with what has already been reported in the literature [32,33]. Concerning the coagulation cascade, we observed that prothrombin time (PT) and activated partial thromboplastin time (aPTT) increases were associated with disease severity. Furthermore, kidney dysfunction (increased blood urea nitrogen—BUN) and electrolyte imbalance were

both associated with adverse outcomes in our cohort. Last, we found, in line with literature data, that severe patients are characterized by increased levels of the markers of cell damage (e.g., lactate dehydrogenase—LDH) [34], vascular permeability, stabilization of microcirculation (e.g., mid regional pro-adrenomedullin—MR-ProADM) [35,36], and inflammatory markers (e.g., C reactive protein—CRP) [37]. The model built by employing the above-described parameters showed an excellent ability to classify the patients of the independent dataset, with an area under the curve (AUC) of 0.868 (95% CI 0.785–0.952).



**Figure 1.** Elastic net logistic regression with hematochemical variables. Receiver Operating Characteristic (ROC), Area Under the Curve (AUC) with its 95% confidence interval (CI) are 0.868 (95% CI 0.785–0.952).

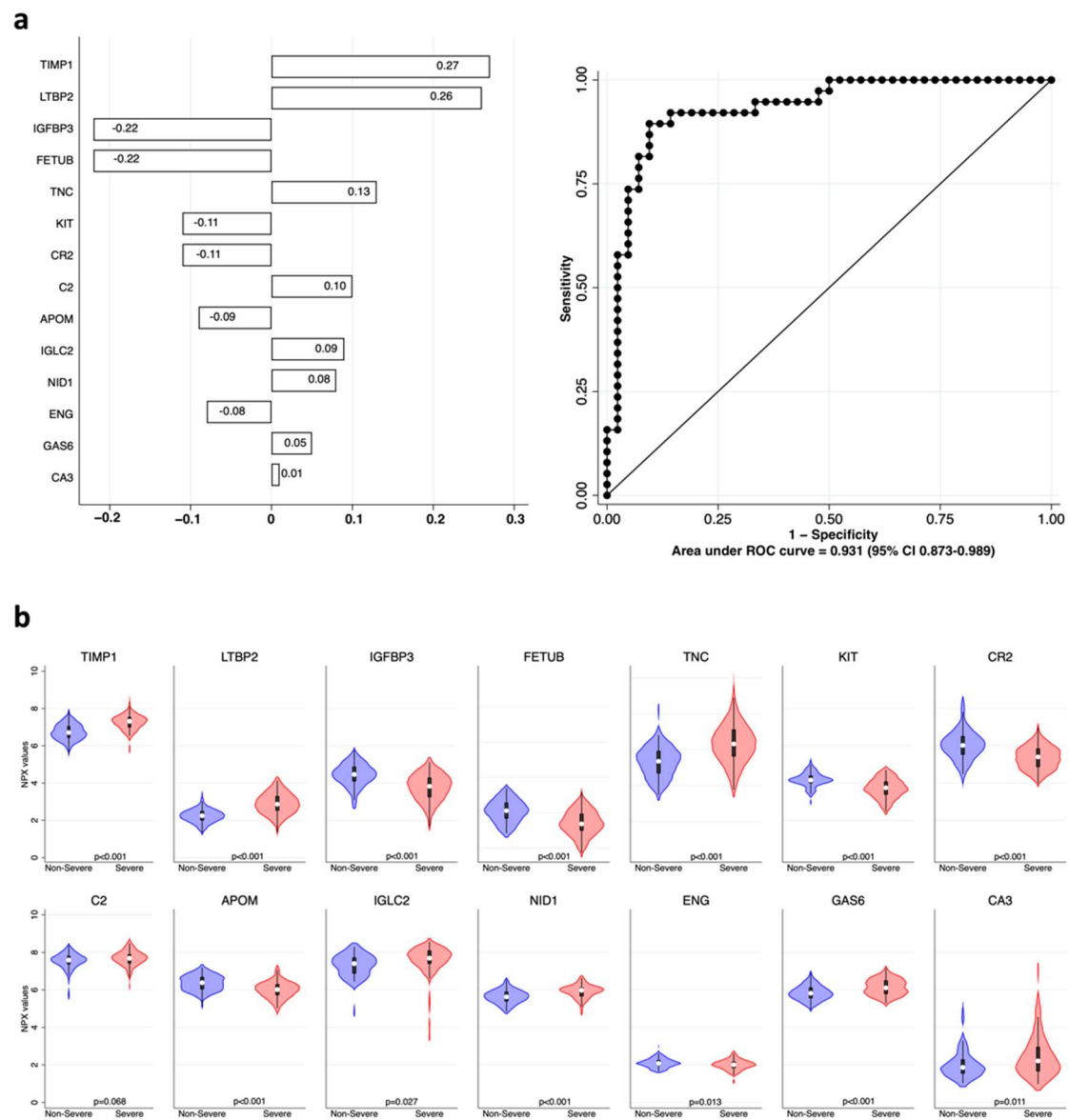
### 2.3. Identification of Serum Proteins Associated with Clinical Outcome

To add new pieces of information concerning the role of the immune, cardiovascular, and nervous systems in COVID-19 pathophysiology, we analyzed patient sera for the expression of 274 different proteins.

First, cardiometabolic-related proteins were analyzed to identify classifiers of disease severity using elastic net logistic regression with cross-validation. Results are shown in Figure 2 and literature data confirming our findings are quoted in Table 3. The complete model containing the 13 markers showed, by ROC analysis, an excellent ability to correctly classify the remaining 80 patients of the dataset, with an AUC of 0.931 (95% CI 0.873–0.989).

Next, we analyzed inflammatory biomarkers, employing the same approach. Results are shown in Figure 3 and literature data confirming our findings are quoted in Table 3. The model containing the eight markers was able to correctly classify the remaining 80 patients with an AUC of 0.906 (95% CI 0.832–0.980).

The last analyzed panel included neurology-related protein biomarkers. Results are shown in Figure 4 and literature data confirming our findings are quoted in Table 3. Again, the complete model had an excellent ability to classify severe cases with an AUC of 0.918 (95% CI 0.853–0.948).

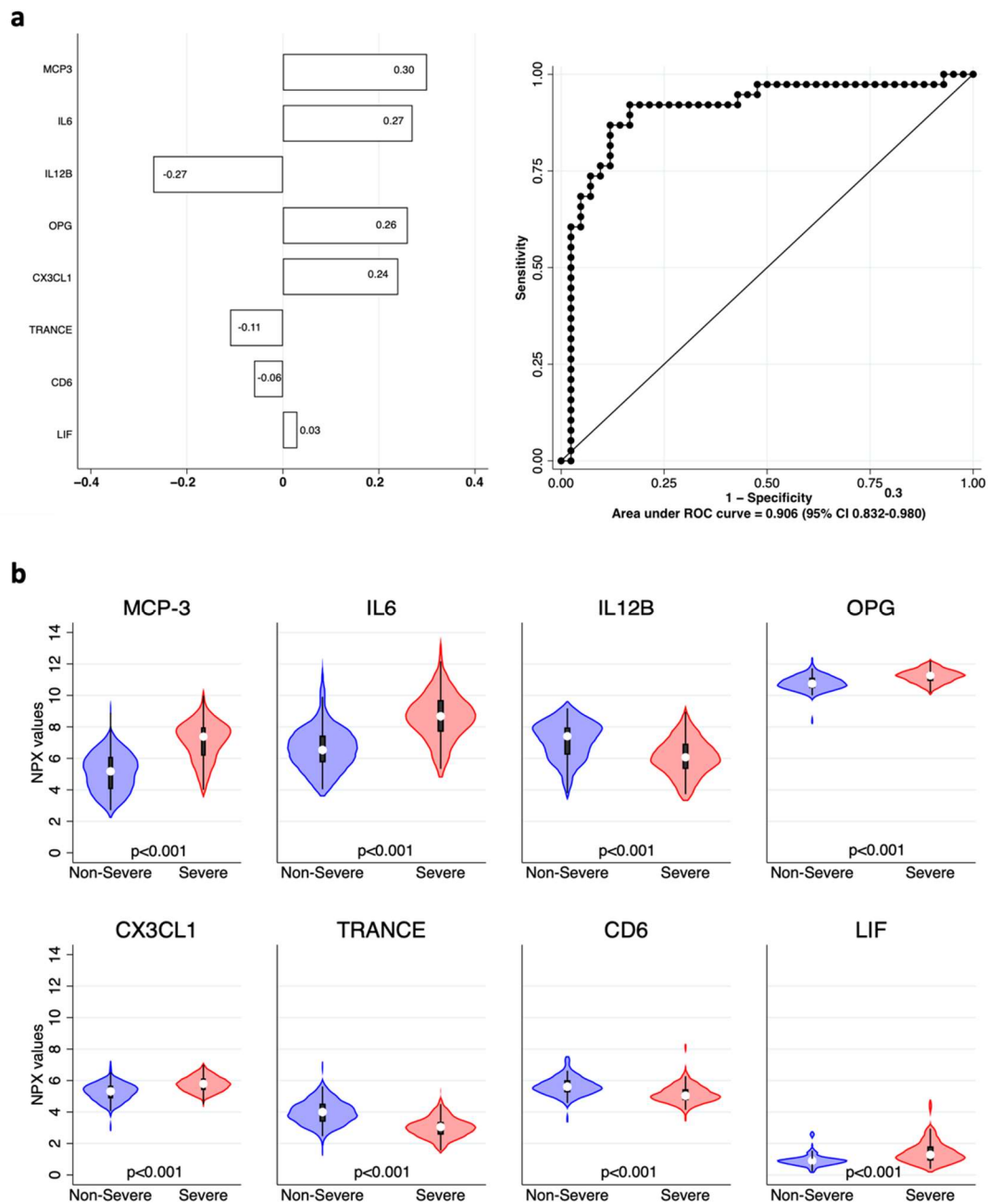


**Figure 2.** Cardiometabolic biomarkers associated with COVID-19 severity. **(a)** Elastic net logistic regression with cardiometabolic variables. Coefficients of each variable are shown in the left panel as horizontal bars. The right panel shows that the receiver operating characteristic (ROC) and area under the curve (AUC), with its 95% confidence interval (CI) of the model, are 0.931 (95% CI 0.873–0.989). **(b)** Violin plots showing the distribution of the normalized protein expression (NPX) values of each variable included in the model in “non-severe” (blue) vs. “severe” (red) patients; *p* values for each comparison are shown at the bottom of each plot.

**Table 3.** Summary of elastic net logistic regression analyses of biomarkers associated with disease severity.

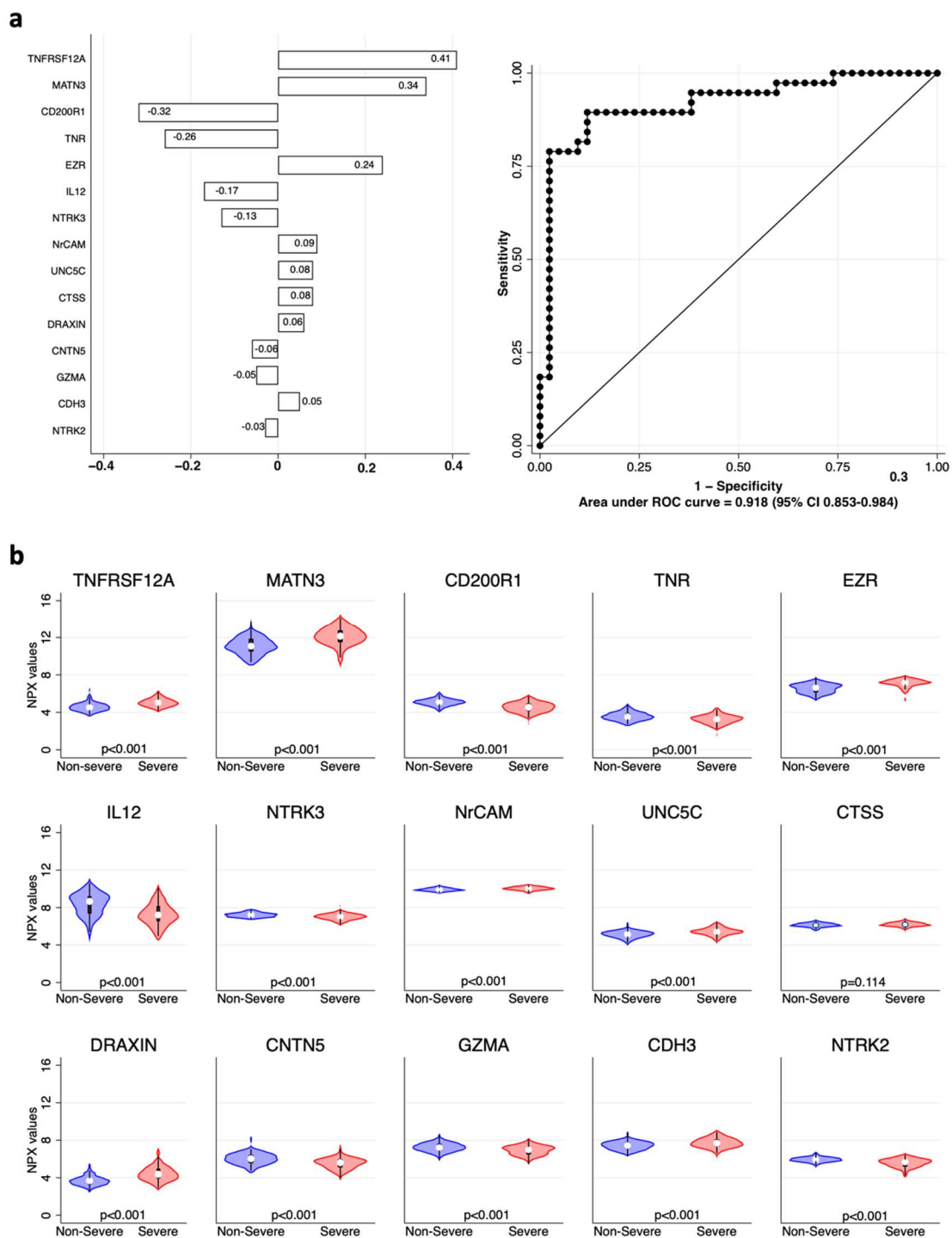
	Short Name	Biomarker Category	Association with Outcome	Literature Data
Tissue inhibitor of metalloprotease 1	TIMP-1	Cardio	Direct	[38]
Latent-Transforming Growth Factor Beta-Binding Protein 2	LTBP2	Cardio	Direct	[39]
Insulin-like Growth Factor Binding Protein 3	IGFBP3	Cardio	Inverse	[40]
Fetuin B	FETUB	Cardio	Inverse	[41]
Tenascin-C	TNC	Cardio	Direct	[42]
KIT proto-oncogene receptor tyrosine kinase	KIT	Cardio	Inverse	
Complement C3d Receptor 2	CR2	Cardio	Inverse	
Complement component 2	C2	Cardio	Direct	[43]
Apolipoprotein M	APOM	Cardio	Inverse	[44]
Immunoglobulin Lambda Constant 2	IGLC2	Cardio	Direct	[45]
Nidogen-1	NID1	Cardio	Direct	[46]
Endoglin	ENG	Cardio	Inverse	[47]
Growth Arrest Specific 6	GAS6	Cardio	Direct	[48]
Carbonic Anhydrase	CA3	Cardio	Direct	[49]
Monocyte Chemoattractant Protein 3	MCP3	Inflammatory	Direct	[50]
Interleukin 6	IL6	Inflammatory	Direct	[16,51]
Interleukin 12B	IL12B	Inflammatory	Inverse	[52]
Osteoprotegerin	OPG	Inflammatory	Direct	[53,54]
C-X3-C Motif Chemokine Ligand 1	CX3CL1	Inflammatory	Direct	[54]
Tumor necrosis factor ligand superfamily member 11	TNFSF11 or TRANCE, RANKL, OPG	Inflammatory	Inverse	[54]
T-cell differentiation antigen CD6	CD6	Inflammatory	Inverse	[54]
Leukemia Inhibitory Factor	LIF	Inflammatory	Direct	[55]
Tumor Necrosis Factor Receptor Superfamily member 12A	TNFRSF12A	Neurology	Direct	[56]
Matrilin 3	MATN3	Neurology	Direct	[54]
CD200 Receptor 1	CD200R1	Neurology	Inverse	[54,57]
Tenascin R	TNR	Neurology	Inverse	
Ezrin	EZR	Neurology	Direct	[54]
Interleukin 12	IL12	Neurology	Inverse	[52]
Neurotrophic Receptor Tyrosine Kinase 3	NTRK3	Neurology	Inverse	[54]
Neuronal Cell Adhesion Molecule	NrCAM	Neurology	Direct	[58]
Netrin receptor	UNC5C	Neurology	Direct	[54] †
Cathepsin S	CTSS	Neurology	Direct	[54]
Dorsal Inhibitory Axon Guidance Protein	DRAXIN	Neurology	Direct	[54]
Contactin 5	CNTN5	Neurology	Inverse	[54]
Granzyme A	GZMA	Neurology	Inverse *	[54]
Cadherin 3	CDH3	Neurology	Direct	[54] †
Neurotrophic Receptor Tyrosine Kinase 2	NTRK2	Neurology	Inverse	[54]

Cardiometabolic, inflammatory, and neurology disease biomarkers associated with disease severity, according to the elastic net logistic regression analyses. Columns indicate, for each biomarker, its full name, its short name, whether a direct or inverse relationship between the biomarker level and disease severity was identified, and literature data supporting our findings. Biomarkers highlighted in grey have not yet been associated with severe COVID-19 by other authors. \* an opposite, yet significant association was described in the literature. † literature results do not reach significance.



**Figure 3.** Inflammatory biomarkers associated with COVID-19 severity. (a) Elastic net logistic regression with inflammatory variables. Coefficients of each variable are shown in the left panel as horizontal bars. The right panel shows the ROC and AUC of 0.906 (95% CI 0.832–0.980). (b) Violin plots showing the distribution of the normalized protein expression (NPX) values of each variable included in the model in “non-severe” (blue) vs. “severe” (red) patients;  $p$  values for each comparison are shown at the bottom of each plot.



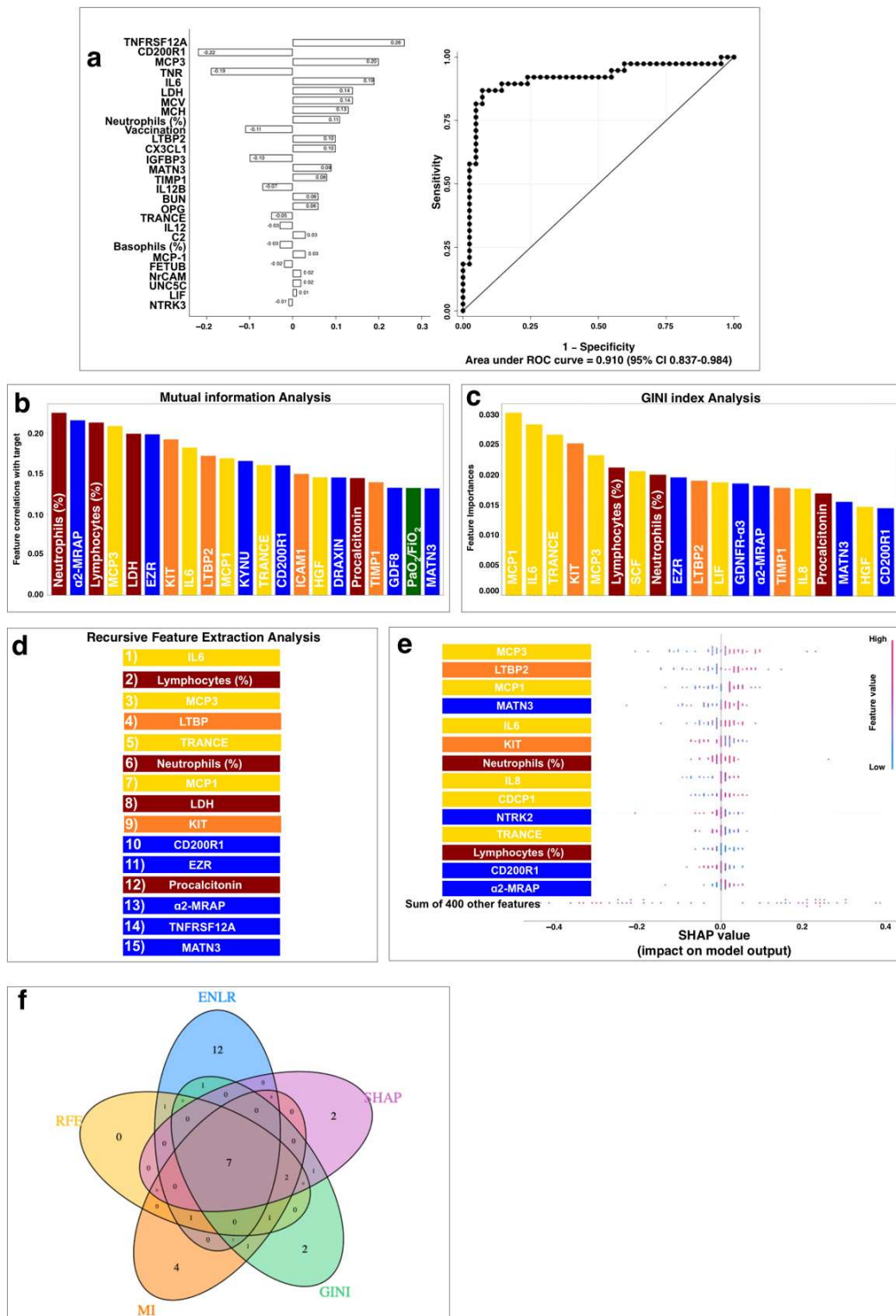


**Figure 4.** Neurology biomarkers associated with COVID-19 severity. (a) Elastic net logistic regression with neurology-related variables. Coefficients of each variable are shown in the left panel as horizontal bars. The right panel shows the ROC and AUC of 0.918 (95% CI 0.853–0.984). (b) Violin plots showing the distribution of the normalized protein expression (NPX) values of each variable included in the model in “non-severe” (blue) vs. “severe” (red) patients;  $p$  values for each comparison are shown at the bottom of each plot.

#### 2.4. Correlation Discovery Analysis

Last, we aimed to identify the most relevant parameters, able to stratify COVID-19 severity by combining the clinical, demographic, anamnestic, and hematochemical data of each patient with the biomarkers of the three different panels.

For this purpose, we first employed the elastic net logistic regression approach with the cross-validation described previously. Results are shown in Figure 5a and Table 4. The complete model was characterized by an 89.2% sensitivity (95% CI 74.6–97), an 88.4% specificity (95% CI 74.9–96.1), an 86.8% positive predictive value (95% CI 71.9–95.6), a 90.5% negative predictive value (95% CI 77.4–97.3), and 0.776 Youden’s J statistic.



**Figure 5.** Correlation discovery analysis. (a) Elastic net logistic regression with hematochemical, cardiometabolic, inflammatory, and neurology-related variables. ROC and AUC with its 95% confidence

interval (CI) are 0.910 (95% CI 0.837–0.984). (b) Results of the mutual information analysis. On the *x*-axis, the dataset parameters are provided ranked based on their MI value related to the target variable. (c) Ranking of the dataset parameters based on the GINI index values obtained by the adopted random forest mode. (d) The ranking of the features obtained through the RFE algorithm. (e) The graph was obtained through the SHAP analysis approach. On each row of the graph, the value of the corresponding feature for each instance of the dataset is represented by a dot. The color of the dot indicates how large the value of the feature is in that instance (blue dot: small value, red dot: large value). Furthermore, the position of the dot with respect to the central vertical line indicates whether that feature led the model to classify the patient as a severe case (dot on the right side of the line) or not (dot on the left side). As a clarifying example, we can see that for the MCP3 parameter higher values are correlated with a severe condition for the patient, as the red dots are mostly on the right side of the graph, while the instances in which the MCP3 value is low are often correlated with a non-severe condition for the patient. In (b–e), the parameters are color-coded to represent their category (red: hemochemical; yellow: inflammatory biomarkers; orange: cardiometabolic biomarkers; blue: neurological biomarkers; green: clinical data). (f) Venn diagram of the variables shared between the five models. The eight variables shared by all the models are Neutrophil count, MCP3, IL6, TRANCE, MCP1, CD200R1, MATN3, and LTBP2; KIT and a-MRAP are shared by four models.

**Table 4.** Summary of biomarkers emerging from the correlation discovery analyses.

	Short Name	Model	Literature Data
C-X3-C Motif Chemokine Ligand 1	CX3CL1	ENLR	[54]
CD200 Receptor 1	CD200R1	ENLR, MI, GINI, SHAP, RFE	[54,57]
Complement component 2	C2	ENLR	[43]
CUB domain-containing protein 1	CDCP1	SHAP	[54]
Dorsal Inhibitory Axon Guidance Protein	DRAXIN	MI	[54]
Ezrin	EZR	MI, GINI, RFE	[54]
Fetuin B	FETUB	ENLR	[41]
Glial Cell Derived Neurotrophic Factor Receptor $\alpha$ -3	GDNFR- $\alpha$ 3	GINI	[54]
Growth/Differentiation factor 8	GDF8	MI	[54]
Hepatocyte Growth Factor	HGF	MI, GINI	[26,55]
Insulin-like Growth Factor Binding Protein 3	IGFBP3	ENLR	[40]
Intercellular Adhesion Molecule 1	ICAM1	MI	[52] †
Interleukin 12	IL12	ENLR	[52]
Interleukin 12B	IL12B	ENLR	[52]
Interleukin 6	IL6	ENLR, MI, GINI, SHAP, RFE	[16,51]
Interleukin 8	IL8	GINI, SHAP	[17,39,55]
KIT proto-oncogene receptor tyrosine kinase	KIT	MI, GINI, SHAP, RFE	
Kynureninase	KYNU	MI	[54]
Latent-Transforming Growth Factor Beta-Binding Protein 2	LTBP2	ENLR, MI, GINI, SHAP, RFE	[39,58]
Leukemia Inhibitory Factor	LIF	ENLR, GINI	[55]
Matrilin 3	MATN3	ENLR, MI, GINI, SHAP, RFE	[54]
Monocyte Chemoattractant Protein 1	MCP1	ENLR, MI, GINI, SHAP, RFE	[17,24]
Monocyte Chemoattractant Protein 3	MCP3	ENLR, MI, GINI, SHAP, RFE	[50]
Netrin receptor	UNC5C	ENLR	[54] †
Neuronal Cell Adhesion Molecule	NrCAM	ENLR	[58]
Neurotrophic Receptor Tyrosine Kinase 2	NTRK2	SHAP	[54]
Neurotrophic Receptor Tyrosine Kinase 3	NTRK3	ENLR	[54]
Osteoprotegerin	OPG	ENLR	[54]

Table 4. Cont.

	Short Name	Model	Literature Data
Stem Cell Factor	SCF	GINI	[17]
Tenascin R	TNR	ENLR	
Tissue inhibitor of metalloprotease 1	TIMP-1	ENLR, MI, GINI	[38]
Tumor necrosis factor ligand superfamily member 11	TNFSF11 or TRANCE, RANKL, OPGL	ENLR, MI, GINI, SHAP, RFE	[54]
Tumor Necrosis Factor Receptor Superfamily member 12A	TNFRSF12A	ENLR, RFE	[56]
$\alpha$ 2-microglobulin receptor-associated protein	$\alpha$ 2-MRAP	MI, GINI, SHAP, RFE	[54]

Correlation discovery analyses results. Table summarizing biomarkers associated with disease severity in the analyses conducted on the complete dataset, including demographic, clinical, anamnestic, hematochemical, and immunometric parameters. The columns show in which of the five models the biomarker was associated with the severity of the disease and which literature data confirm our results. Biomarkers highlighted in grey have not yet been associated with severe COVID-19 by other authors. ENLR: elastic net logistic regression; MI: mutual information analysis; GINI: GINI index analysis; SHAP: Shapley additive explanations analysis; RFE: recursive feature extraction analysis. † the literature results do not reach significance.

As machine and deep learning techniques are becoming more and more prominent in the medical field [59] with notable examples, especially in automated diagnosis [60] and biomedical images anomaly detection [61], we provided a complementary approach, which relies on four different feature correlation algorithms aiming to highlight the correlation between each of the parameters present in the complete dataset, including the 26 demographic, clinical, and anamnestic parameters; the 32 hematochemical and immunometric parameters together with the plasma proteomics data of 274 different proteins; and the target variable indicating the severity of the each patient's conditions. Both model-independent and model-dependent approaches were employed.

Results of the model-independent mutual information (MI) analysis [62] are reported in Figure 5b and Table 4. Specifically, MI identified four hematochemical parameters (i.e., relative neutrophil and lymphocyte counts, LDH, and procalcitonin), the clinical parameter PaO<sub>2</sub>/FiO<sub>2</sub>, seven neurology associated markers, five inflammatory markers, and four cardiometabolic factors associated with outcome.

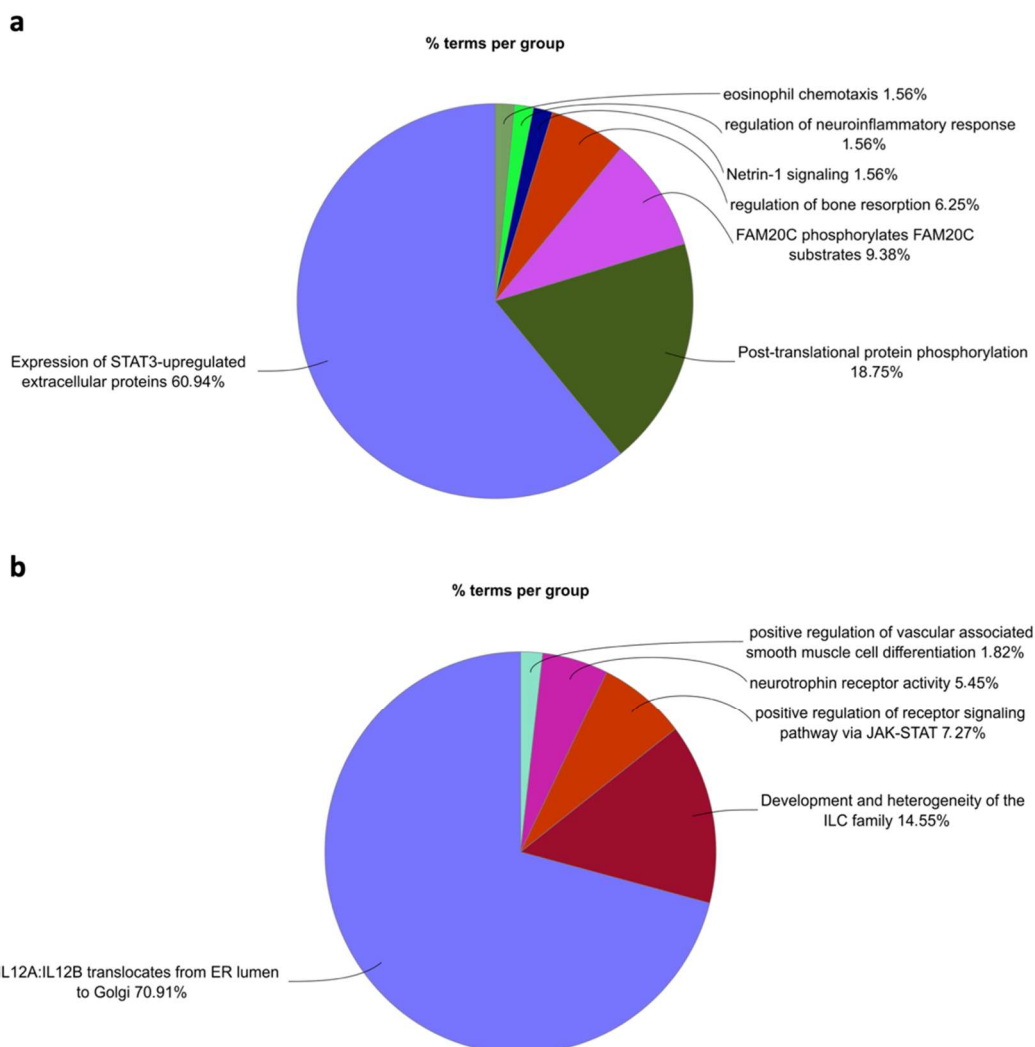
The results obtained by three model-dependent approaches are reported in Figure 5c–e and Table 4. Specifically, GINI index analysis (Figure 5c) identified two hematochemical parameters (i.e., relative neutrophil and lymphocyte counts) together with eight inflammatory markers, five neurology associated markers, and three cardiometabolic factors associated with outcome. Recursive feature extraction analysis (RFE, Figure 5d) identified four hematochemical parameters (i.e., relative neutrophil and lymphocyte counts, LDH, and procalcitonin), combined with five neurology-associated markers, and two cardiometabolic factors associated with outcome. Last, according to the Shapley additive explanations (SHAP, Figure 5e) analysis, two hematochemical parameters (i.e., relative neutrophil and lymphocyte counts), combined with six inflammatory markers, four neurology-associated markers, and two cardiometabolic factors were correlated with the outcome. Intriguingly, no clinical parameter emerged from model-dependent approaches.

Finally, we assessed which variables were associated with disease severity that were shared in all the above-mentioned approaches (Figure 5f and Table 4). The relative Neutrophil count, together with the inflammatory markers MCP3 [50], IL6 [16,51], TRANCE [63], and MCP1 [24]; the neurology-associated markers CD200R1 [54] and MATN3 [54]; and the cardiometabolic marker LTBP2 [39] emerged in every analysis conducted. Lymphocyte count, KIT, and  $\alpha$ 2-MRAP [54] were shared by the four different feature correlation algorithms.

### 2.5. Functional Enrichment Analysis of Proteins Associated with the Disease Outcome

To associate a functional role to the proteins discriminating the “severe” patients from the “non-severe” ones, potential biomarkers emerging from the cardiometabolic, inflammation, and neurology panels were subjected to the ClueGO functional enrichment analysis.

The elastic net logistic regression model built by employing the proteins profiled with the three panels together with clinical, demographic, and hematochemical data (Figure 5a) included, except for MCP1, all the predictors emerging from the analyses conducted by separately employing every single panel. Therefore, we decided to group the proteins emerging from the first three models (Figures 2–4) and to conduct the functional analysis, distinguishing those positively associated with disease severity from those negatively associated with a “severe” COVID-19. As shown in Figure 6a, most of the proteins positively associated with disease severity pertained to the functional terms “expression of STAT-3 up-regulated extracellular proteins”, followed by “Post-translational protein phosphorylation”, “FAM20C phosphorylates FAM20C substrates”, “Regulation of bone resorptions”, “Netrin-1 signaling”, “Regulation of neuroinflammatory response”, and “Eosinophil chemotaxis”. Employing the same dataset to query the Molecular Signature Database (MSigDB) [64], we observed that the upregulated genes were associated with the hallmarks: “inflammatory response”, “allograft rejection”, and “epithelial to mesenchymal transition”. Consistently, several mesenchymal cell types emerged.



**Figure 6.** Functional enrichment analysis of proteins associated with a “severe” outcome. Pie charts representing the most enriched functional terms ( $p_{adj} \leq 0.05$ ) related to proteins whose coefficients were positively (a) and negatively (b) associated with the “severe” class. The Cytoscape [65] app ClueGO [66] was used to query the following functional databases: GO (BiologicalProcess and ImmuneSystemProcess), KEGG, REACTOME, CLINVAR, WikiPathways, and CORUM-3.0.

Conversely, proteins negatively associated with disease severity (Figure 6b) are enriched in the functional terms: “IL12A-IL12B translocate from ER lumen to Golgi”, “Development and heterogeneity of the ILC family”, “Positive regulation of receptor signaling via JAK-STAT”, “Neurotrophin receptor activity”, and “Positive regulation of vascular associated smooth muscle cell differentiation”. The MSigDB analysis of the terms inversely associated with severity identified several lymphocyte-related terms.

### 3. Discussion

Here we present the results of an extensive analysis of the serum proteome of 160 adult patients affected by the SARS-CoV-2 infection during a phase of the COVID-19 pandemic (i.e., February–September 2021) dominated by the VOC Alpha and Delta [30,31]. In this study we extensively used PEA, an emerging technology that can simultaneously measure, with high sensitivity and specificity, 92 proteins across 96 samples, using only 1  $\mu$ L of serum [67]. Given the involvement of both inflammation [3] and cardiac injury (triggered both by direct [68] and indirect mechanisms [69]) in the pathophysiology of COVID-19, two commercial panels containing 92 “inflammation” and “cardiometabolic” biomarkers were employed. Since a large number of neurological manifestations (e.g., headache, confusion, neuroinflammatory, and cerebrovascular disease) have been described to occur in the acute and chronic phase of COVID-19 [20,70], we additionally employed a commercial panel containing 92 “neurology” biomarkers.

Furthermore, we employed machine learning (ML) techniques to identify the strongest predictors of disease severity, by aggregating 26 clinical, demographic, and anamnestic parameters (including the major risk factors for severe COVID-19 and the vaccination status of the enrolled patients) and 32 hematochemical and immunometric parameters (comprising markers of inflammation—CRP [71], myocardial injury—cardiac Troponin [72], severe bacterial infection—procalcitonin [72], and emerging markers of sepsis—MR-proADM [72], which have been associated with COVID-19 outcome) together with the plasma proteomics data of 274 different proteins. Although several works aiming at identifying the biomarkers of severe COVID-19 have been published [24–27,29,45], some of which have also taken advantage of PEA technology, we believe that our work has two major strengths: (1) this is one of the largest cohorts of adult patients profiled so far, after [25,27,29], and (2) ML approaches were employed to explore which of the clinical, hematochemical, and biomarker parameters characterizing patients at admission are able to stratify outcomes.

To analyze our database, we first employed an elastic net logistic approach with cross-validation, a method that performs well even in the case of highly correlated predictors and in the presence of many variables over a smaller sample size. We additionally employed four different ML approaches, both model-independent and model-dependent. The main feature of model agnostic approaches is that they do not rely on a ML model to identify the correlation between the dataset parameters, but they work only on the data itself. The only model-independent approach we selected was the one based on the correlation identification through mutual information (MI) analysis [62]. Concerning model-dependent approaches, they rely on a ML model to infer the correlation between each parameter and the target variable. Here, we specifically selected RFE analysis, feature selection through GINI Index, and SHAP analysis. Each of these provides a higher degree of descriptiveness, which allowed us to gain a deeper understanding of the highlighted correlations. Since these approaches rely on a ML model, the quality of the parameter correlation predictions highly depends on the performance of the latter for the task at hand, in this case, the classification of patients based on the severity of their conditions. The model we selected for our analysis is a random forest [73], consisting of 200 decision trees, while the metrics we adopted to assess its performance during the classification task are, as for the previously described statistical approaches, the ROC–AUC calculated by employing a 10-fold-cross-validation process on the dataset. The obtained result is a ROC–AUC of 0.942 (95% CI 0.935–0.950), which allowed us to proceed with further correlation analysis with confidence.

To identify the variables most strongly associated with COVID-19 severity, we finally compared the results of the five different approaches and observed that the relative neutrophil and lymphocyte count, together with the inflammatory markers MCP3 [50], IL6 [16], TRANCE [63], and MCP1 [24]; the neurology-associated markers CD200R1 [54],  $\alpha$ 2-MRAP [54], and MATN3 [54]; and the cardiometabolic markers KIT and LTBP2 [39] emerged in at least four of the analyses conducted.

From a pathophysiological standpoint, the dysfunction of both the innate and adaptive immune responses characterize patients with severe COVID-19 [74,75]. Although increased neutrophils and neutrophil-to-lymphocyte ratio have been described in cases of severe COVID-19 [38,76], the active role of neutrophils in COVID-19 pathophysiology has been debated [77]. However, some pieces of evidence support their importance; SARS-CoV-2 can directly induce the release of neutrophil extracellular traps (NETs) from donor-derived neutrophils [78]. These web-like chromatin structures have a role in viral clearance, but excessive NET levels exacerbate inflammation in acute respiratory distress syndrome and have been associated with COVID-19 complications, ranging from thrombosis to CNS involvement [77]. Granulocytes and monocytes activated by a continuous and weak signal (e.g., cytokines) have an immature phenotype and produce immunosuppressive and anti-inflammatory factors [79]. Indeed, the presence of immature neutrophils with reduced functional properties is associated with COVID-19 severity [38,76]. Consistently, most of the patients with severe outcome had microbial superinfections [76], a finding consistent with the recurrence of procalcitonin variable in most of the ML models in our dataset. The profound alterations of the hematopoietic stem cell maturation are observed in COVID-19 patients (i.e., the expansion of the “granulocyte-monocyte progenitor” and the “erythroid progenitor” cell pools, at the expense of the lymphoid cell pool) [80]. In line with this, the emergence of monocytes expressing low levels of HLA-DR and anti-inflammatory molecules, together with neutrophils expressing the immune checkpoint protein PD-L1 [81] coupled with the exhaustion and anergy of T cells, are all hallmarks of severe COVID-19 [82]. Consistently, the relative Lymphocyte count was identified to be associated with disease severity by all the feature correlation algorithms we employed. Moreover, elastic net logistic regression analysis showed an inverse relationship between CR2 plasma levels and disease severity. CR2 (also known as CD21) is present on B lymphocytes, where it is required for their activation. CD21 shedding follows B lymphocyte activation [83], thus a soluble form of CD21 can be found in plasma, reflecting the activity of B-cells [84]. Our results therefore suggest that COVID-19 severity is higher in those patients that show a lower B-cell response. To the best of our knowledge, this association has not been described in the literature yet.

Concerning inflammatory biomarkers associated with disease severity, a hyperinflammatory response with high levels of interferons, cytokines, and chemokines characterizes patients with bad prognosis [15,85]. In line with this, in our cohort, IL6, MCP1, and MCP3 emerged among the strongest independent predictors of disease severity. More complex is the involvement of TRANCE (alias RANKL) in modulating COVID-19 severity. Indeed, this cytokine and its soluble decoy receptor OPG may play a relevant role in the context of acquired immunity [86]. Although these are the main regulators of bone metabolism, TRANCE also enhances the immune response by promoting dendritic cell viability and function [86]. Indeed, the stimulation of dendritic cells with TRANCE triggers an anti-viral CD8<sup>+</sup> T cell cytotoxic response [87]. Therefore, it is not surprising that we and other authors found TRANCE levels to be inversely associated with disease severity [56].

One of the most innovative findings of our work is that three biomarkers that are also altered in neurological diseases (CD200R1,  $\alpha$ 2-MRAP, and MATN3) are among the strongest predictors of a negative outcome in the complete model. This finding is supported by the identification, through the functional enrichment analysis of proteins associated with the disease outcome and with “Regulation of neuroinflammatory response” and “neurotrophin receptor activity” terms. Intriguingly, several neurological and psychiatric symptoms have been described in COVID-19 patients, most of which are independent of a direct affection

of the central nervous system (CNS) by the virus but are considered to be secondary to the immune reaction [18]. Consistently, we observed an inverse relationship between CD200R1 serum levels and COVID-19 severity. CD200R1 is the receptor for CD200 and has been involved in the inhibition of neuroinflammatory pathways in aging, stroke, and multiple sclerosis [88]. CD200R is also expressed by cells of innate immunity (e.g., monocytes and neutrophils) and regulates their function and maturation [89]. Importantly, the expression of CD200R1 on infiltrating lymphocytes dictates the survival of mice exposed to brain injury by regulating the post-stroke immune suppression and vulnerability of animals to superinfections [90]. Interestingly, the direct injection of the S1 subunit of the SARS-CoV-2 Spike protein in the brain of mice induces neuroinflammation and, after an initial increase, reduces the levels of *Cd200r1* under the level of control animals [19]. Concerning the other neurological markers positively associated with a negative outcome,  $\alpha$ 2-MRAP (also known as LDL receptor-related protein associated protein -1 or LRPAP1) is involved in the trafficking of LDL receptor family members. Specifically, it regulates the amount of LRP that is expressed in the liver and the brain. LRP binds ApoE and  $\alpha$ 2-macroglobulin, a protein involved in the clearance of  $\beta$ -amyloid (A $\beta$ ). Consistently, LRP-1 is involved in the export of A $\beta$  from the brain [91], while LRPAP1 gene polymorphisms are associated with late-onset Alzheimer's disease [92]. Importantly, viral infections can drive A $\beta$  formation or deposition, providing a possible link between COVID-19, neuroinflammation, and neurodegeneration [93]. Although recent literature data confirms the association of MATN3 with COVID-19 severity [54], its pathophysiological role is not clear. An additional neurology associated marker that emerged only from the elastic net logistic regression analysis and has not been reported in the literature yet is Tenascin R (TNR). This is a member of the tenascin family of extracellular proteins that is specifically expressed in the CNS. Although another proinflammatory member of this family (i.e., Tenascin C) has been identified in the exosomes of COVID-19 patients [42], no data on Tenascin R could be found in COVID-19-related literature. TNR is part of the highly specialized perineuronal networks of the extracellular matrix, which repel adjacent dendrites and axons to maintain the established synaptic networks [94]. Reduced TNR levels have been described in the cerebro spinal fluid of patients affected by amyotrophic lateral sclerosis (ALS), reflecting the loss of immunoreactivity in the spinal cord of ALS patients. This finding suggests that the control of TNR levels is important to prevent CNS abnormalities [95]. These data are consistent with ours, since patients affected by severe COVID-19 have significantly reduced TNR levels.

Concerning the cardiometabolic biomarker panel, while LTBP2 was higher in patients with a "severe" outcome, KIT showed the opposite behavior. LTBP2 expression increases in response to myocardial stressors (e.g., pressure overload [96] or isoproterenol [97]) and has been associated with both cardiac [96] and pulmonary fibrosis [39]. Consistently, in patients with COVID-19, LTBP2 levels are related to pulmonary fibrosis [39] and are associated with disease severity [98]. In adulthood, the KIT receptor and its ligand SCF are mainly expressed by stem cells and mastocytes and regulate their function [99]. Intriguingly, while hematopoietic stem cells express low levels of ACE2, this receptor, which is involved in SARS-CoV-2 virus entry, is upregulated during erythroid differentiation and parallels KIT expression. Consistently, erythroid progenitors are susceptible to coronavirus infection and possibly account for the increased circulation of nucleated red blood cells following SARS-CoV-2 infection [33]. Regarding mastocytes, lung biopsies have shown an abundant infiltrate of KIT<sup>+</sup> cells in COVID-19 patients and SARS-CoV-2-infected African green monkeys, suggesting the early recruitment of mastocyte progenitors to the alveolar septa. This morphologic finding was also coupled with interstitial and alveolar edema, suggesting a relevant pathophysiological role for this cell type [100,101]. How these alterations relate to our findings could not be predicted at this time. However, to the best of our knowledge, the serum levels of KIT have not yet been associated with COVID-19 patient outcomes.

By combining a novel high-throughput proteomics approach with ML techniques in a patient population deeply characterized from the clinical standpoint, we identified



potential serum biomarkers of a “severe” outcome of COVID-19-affected patients. Our study was limited by its retrospective nature and the limited number of enrolled patients. Although we did not have a validation cohort and we could not find any open access datasets containing the same clinical and hematochemical information we employed, we must underline that most of the biomarkers that emerged from our analysis are confirmed by many, unrelated authors in the literature, further supporting the solidity of our approach. However, to our knowledge, this is the one of the largest studies that used extensive PEA analysis applied to such a large cohort ( $n = 160$ ) of COVID-19-affected patients. One of the major novelties of this work is the demonstration that the markers of neurological disease are among the strongest predictors of COVID-19 severity, when combined with relevant laboratory data, such as lymphocyte and neutrophil counts. Our findings are hypothesis-generating and stimulate studies aimed at understanding the pathophysiological role of neuroinflammation in the longer-term outcome of COVID-19 patients.

## 4. Materials and Methods

### 4.1. Examined Cohort

#### 4.1.1. Patient Enrollment and Ethics

The study, authorized by the Regional Ethic Committee (2020-Os-033; Em. Sost. N. 1 versione 1 data 16 August 2021), was conducted according to the declaration of Helsinki and signed informed consent was collected from enrolled patients. Inclusion criteria were age > 18 years and nasopharyngeal swab positivity for the SARS-CoV-2 genome.

#### 4.1.2. Study Design

This work aims at identifying serum proteins discriminating “non-severe” from “severe” COVID-19 patients, in analogy with recently published works [27]. We classified as “severe”, on top of severe and critical patients according to the WHO classification [102], also those patients presenting a moderate disease at onset that worsened over time, requiring admission to the intensive care unit at a later stage.

#### 4.1.3. Clinical Data

Relevant patient data were collected and extracted by a team of physicians (E.S., C.T.) from the hospital electronic health records (INSIEL, Trieste, Italy), pseudonymized, and recorded on a cloud-based clinical data management platform (Castor, The Netherlands and the USA). Immunocompromised patients were defined as patients with permanent dysfunction of the immune response resulting from immunosuppressive medication or comorbidities, such as AIDS, or malignancies that cause neutropenia (neutrophil counts  $\leq 500$  cells/mm<sup>3</sup>). Patients with “renal impairment” were defined as having kidney damage or glomerular filtration rate (GFR) < 60 mL/min/1.73 m<sup>2</sup> for 3 months or more, irrespective of the cause.

### 4.2. Samples

#### 4.2.1. Blood Sampling and Serum Storage

Blood samples, drawn within 1 day (median value) of hospital admission, were collected into a 5 mL serum tube with a clot activator and gel separator (Vacuette, Greiner Bio-One, Nürtingen, Germany), immediately sent to the core laboratory to be processed, and kept at  $-80$  °C until used.

#### 4.2.2. Proximity Extension Assay (PEA)

Sera were analyzed for 92 inflammation-related protein biomarkers (Olink Target 96 Inflammation), 92 cardiometabolic-related (Olink Target 96 Cardiometabolic), and 92 neurology-related protein biomarkers (Olink Target 96 Neurology) employing the proximity extension assay (PEA) technique of Olink<sup>®</sup> Proteomics (Uppsala, Sweden) [23]. The quality control of the analyzed samples, data normalization, and the quantification of protein levels was performed with the Olink Analyze R package [103], using NPX data

files generated from Olink NPX Manager as input. Levels of proteins were described as normalized protein expression (NPX), consisting of data normalization to the extension control (known standard), log<sub>2</sub>-transformation, and level adjustment using the plate control (plasma sample). We performed exploratory analysis using the ‘olink\_qc\_plot’ and ‘olink\_pca\_plot’ functions, generating QC and PCA plots that allowed us to easily identify missing and/or problematic samples.

#### 4.2.3. Statistical Analyses

Descriptive statistics for categorical variables are presented as number (percent) and for continuous variables as mean  $\pm$  standard deviation (SD) or median (interquartile range (IQR)). Normality was assessed using Shapiro–Wilk test. Comparisons between categorical variables were performed using the Chi-square or Fisher’s exact test, as appropriate. Comparisons between continuous variables were performed using the *t*-test or Mann–Whitney U-test, as appropriate.

An elastic net logistic regression algorithm was used to build different models to predict a “non-severe” or “severe” outcome [104,105]. The elastic net logistic regression is a regularization model that combines, through a linear combination of LASSO and Ridge methods, both L<sub>1</sub> and L<sub>2</sub> penalties. This model performs a variable selection, forming a subset of predictors, each one matched with a regression coefficient. Variable importance is ordered using the absolute value of the regression coefficient, a higher value showing a bigger contribution to the model. The dataset was split into testing and training set with a 1:1 proportion. A 10-fold cross-validation was applied to the training set to tune the hyper-parameters  $\lambda$ , determining the amount of shrinkage and  $\alpha$ , and explaining the presence of L<sub>1</sub> and L<sub>2</sub> penalties. The model was trained for the clinical and hematochemical data, inflammatory, cardiometabolic, and neurological PEA panels separately and jointly. The performance of these model was evaluated on the testing set using the receiver operating curve (ROC) and the area under the curve (AUC), with its 95% confidence interval, compared using Long’s test. Analyses were performed using STATA 17. Table S3 provides hyperparameter and model validation information.

#### 4.2.4. Correlation Discovery Analysis

All the algorithms employed for the correlation discovery analysis have been implemented in Python 3.7.9 by relying on two main tools, the scikit-learn library [106] and the SHAP package [107]. The former was adopted for the development of the random forest model as well as for the recurrent feature extraction, mutual information, and GINI Index approach, while the latter was employed to implement the SHAP component of the analysis.

The only model-independent approach is MI [26]. MI between two random variables is a non-negative value, which measures the dependency between the variables. It is equal to zero if and only if two random variables are independent, and higher values mean higher dependency. For this reason, a higher MI value indicates a higher correlation between the selected parameter and the target variable. Concerning model-dependent approaches:

1. *Recursive feature extraction (RFE) analysis* [108]: an algorithm that at each step removes the features that have the lowest correlation with the target variable for the considered task and then retrains the model on the remaining parameters. The features can then be ranked based on the removal order, where the last feature to be removed is placed in 1st position regarding importance;
2. *Feature selection through GINI Index* [73]: this approach is very similar to the already presented MI one. The main difference is represented by the metric adopted, which in this case is the GINI Index, a measure that represents the amount of information with which each feature contributes to determining the final output of the model. Compared to the RFE algorithm it not only provides a ranking of the features but also a measure of correlation for each;

3. *SHAP (Shapley additive explanations) analysis [107]*: it is the most descriptive of the proposed approaches, as it not only allows the measure of how much a certain feature contributes to the outcome to be obtained but also to which of the outcomes the feature leads the model towards.

The data used for the analysis were kept consistent with the ones described in the previous sections to provide a meaningful comparison between the different approaches. For the data preprocessing, we followed a three-step pipeline. The first step consisted of the removal of those features which, instead of being a predictor of the patient outcome were depending on it (e.g., patient death, hospitalization of the patient in the Intensive Care Unit). Following this first step, we relied on a data imputing procedure to introduce a placeholder where the value for the corresponding feature was missing. Finally, we standardized the data to avoid the results of the analysis being biased towards the features characterized by larger-scale values.

#### 4.2.5. Functional Enrichment Analysis

To characterize the functional role of the most interesting, clinically-relevant identified biomarkers, the Cytoscape app ClueGO was used to query the following functional databases: GO (BiologicalProcess and ImmuneSystemProcess), KEGG, REACTOME, CLINVAR, WikiPathways, and CORUM-3.0. A two-sided hypergeometric test (corrected using the Benjamini–Hochberg method to control the false discovery rate, adjusted  $p \leq 0.05$ ) was used to determine the probability that each functional term was assigned to the gene sets due to chance alone.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23169161/s1>.

**Author Contributions:** Conception and design of the work, A.D.N., A.P.B., C.E.M.P., C.P., C.T., E.D., F.C. (Francesco Curcio), G.L.F., G.T., L.A.S., M.C., M.C.M., M.D.M. and M.I.; patient enrollment, C.T. and E.S.; acquisition of samples and biobanking, A.P.B., D.S., F.C. (Federica Caponnetto), F.D., M.F. and M.C.M.; clinical data collection, A.P.B., C.T., E.S., F.C. (Federica Caponnetto), F.D. and M.F.; data analysis, A.D.N., C.P., E.D., G.L.F., M.D.M. and M.I.; interpretation of data, A.P.B., C.E.M.P., C.T., E.D., E.S., F.C. (Federica Caponnetto), F.C. (Francesco Curcio), G.T., L.A.S., M.C. and M.C.M.; drafted the work, A.D.N., A.P.B., C.P., F.C. (Federica Caponnetto), G.L.F., G.T., M.D.M. and M.I.; substantively revised it, A.D.N., C.E.M.P., C.P., C.T., E.S., F.C. (Francesco Curcio), G.L.F., G.T., L.A.S., M.C., M.C.M. and M.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** G.T. was funded to carry out this work by the Italian Ministry of the Research and University (FISR2020IP\_01563) and by C.I.B. (Bando Progetti Competitivi MIUR D.M.n.442 del 10 August 2020 Art.7.1 All.3).

**Institutional Review Board Statement:** A biobank for the collection of sera of patients referring to the Academic Hospital of Udine (ASUFC) was authorized by the Regional Ethics Committee in 2020 (2020-Os-033) and amended in 2021 to continue enrollment (Em. Sost. N. 1 versione 1 data 16 August 2021). Access to patient sample and relevant clinical data for the present study was authorized by the Regional Ethics Committee (Identificazione di biomarcatori circolanti per la stratificazione prognostica dei pazienti COVID 19; ID 361 and Emendamento n°1).

**Informed Consent Statement:** All patients signed an informed consent form at enrollment.

**Data Availability Statement:** Raw data are provided in Supplementary Table S1.

**Acknowledgments:** We would like to thank Catia Mio for sharing SARS-CoV-2 variant screening data.

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

## References

1. Rodriguez-Morales, A.J.; Cardona-Ospina, J.A.; Gutierrez-Ocampo, E.; Villamizar-Pena, R.; Holguin-Rivera, Y.; Escalera-Antezana, J.P.; Alvarado-Arnez, L.E.; Bonilla-Aldana, D.K.; Franco-Paredes, C.; Henao-Martinez, A.F.; et al. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Med. Infect. Dis.* **2020**, *34*, 101623. [CrossRef] [PubMed]
2. Lake, M.A. What we know so far: COVID-19 current clinical knowledge and research. *Clin. Med.* **2020**, *20*, 124–127. [CrossRef] [PubMed]
3. Brodin, P. Immune determinants of COVID-19 disease presentation and severity. *Nat. Med.* **2021**, *27*, 28–33. [CrossRef] [PubMed]
4. Karaderi, T.; Bareke, H.; Kunter, I.; Seytanoglu, A.; Cagnan, I.; Balci, D.; Barin, B.; Hocaoglu, M.B.; Rahmioglu, N.; Asilmaz, E.; et al. Host Genetics at the Intersection of Autoimmunity and COVID-19: A Potential Key for Heterogeneous COVID-19 Severity. *Front. Immunol.* **2020**, *11*, 586111. [CrossRef]
5. Li, X.; Zhong, X.; Wang, Y.; Zeng, X.; Luo, T.; Liu, Q. Clinical determinants of the severity of COVID-19: A systematic review and meta-analysis. *PLoS ONE* **2021**, *16*, e0250602. [CrossRef]
6. Kundu, R.; Narean, J.S.; Wang, L.; Fenn, J.; Pillay, T.; Fernandez, N.D.; Conibear, E.; Koycheva, A.; Davies, M.; Tolosa-Wright, M.; et al. Cross-reactive memory T cells associate with protection against SARS-CoV-2 infection in COVID-19 contacts. *Nat. Commun.* **2022**, *13*, 80. [CrossRef]
7. Aleksova, A.; Gagno, G.; Sinagra, G.; Beltrami, A.P.; Janjusevic, M.; Ippolito, G.; Zumla, A.; Fluca, A.L.; Ferro, F. Effects of SARS-CoV-2 on Cardiovascular System: The Dual Role of Angiotensin-Converting Enzyme 2 (ACE2) as the Virus Receptor and Homeostasis Regulator-Review. *Int. J. Mol. Sci.* **2021**, *22*, 4526. [CrossRef]
8. Andersen, K.G.; Rambaut, A.; Lipkin, W.I.; Holmes, E.C.; Garry, R.F. The proximal origin of SARS-CoV-2. *Nat. Med.* **2020**, *26*, 450–452. [CrossRef]
9. Kim, D.; Lee, J.Y.; Yang, J.S.; Kim, J.W.; Kim, V.N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914–921.e10. [CrossRef]
10. Araf, Y.; Akter, F.; Tang, Y.D.; Fatemi, R.; Parvez, M.S.A.; Zheng, C.; Hossain, M.G. Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. *J. Med. Virol.* **2022**, *94*, 1825–1832. [CrossRef]
11. Pereira, N.L.; Ahmad, F.; Byku, M.; Cummins, N.W.; Morris, A.A.; Owens, A.; Tuteja, S.; Cresci, S. COVID-19: Understanding Inter-Individual Variability and Implications for Precision Medicine. *Mayo Clin. Proc.* **2021**, *96*, 446–463. [CrossRef] [PubMed]
12. Fallerini, C.; Daga, S.; Mantovani, S.; Benetti, E.; Picchiotti, N.; Francisci, D.; Paciosi, F.; Schiaroli, E.; Baldassarri, M.; Fava, F.; et al. Association of Toll-like receptor 7 variants with life-threatening COVID-19 disease in males: Findings from a nested case-control study. *eLife* **2021**, *10*, e67569. [CrossRef] [PubMed]
13. Velavan, T.P.; Pallerla, S.R.; Ruter, J.; Augustin, Y.; Kremsner, P.G.; Krishna, S.; Meyer, C.G. Host genetic factors determining COVID-19 susceptibility and severity. *EBioMedicine* **2021**, *72*, 103629. [CrossRef] [PubMed]
14. Ganguly, M.; Nandi, A.; Banerjee, P.; Gupta, P.; Sarkar, S.D.; Basu, S.; Pal, P. A comparative study of IL-6, CRP and NT-proBNP levels in post-COVID multisystem inflammatory syndrome in children (MISC) and Kawasaki disease patients. *Int. J. Rheum. Dis.* **2022**, *25*, 27–31. [CrossRef]
15. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [CrossRef]
16. Fabris, M.; Del Ben, F.; Sozio, E.; Beltrami, A.P.; Cifu, A.; Bertolino, G.; Caponnetto, F.; Cotrufo, M.; Tascini, C.; Curcio, F. Cytokines from Bench to Bedside: A Retrospective Study Identifies a Definite Panel of Biomarkers to Early Assess the Risk of Negative Outcome in COVID-19 Patients. *Int. J. Mol. Sci.* **2022**, *23*, 4830. [CrossRef]
17. Herr, C.; Mang, S.; Mozafari, B.; Guenther, K.; Speer, T.; Seibert, M.; Srikakulam, S.K.; Beisswenger, C.; Ritzmann, F.; Keller, A.; et al. Distinct Patterns of Blood Cytokines Beyond a Cytokine Storm Predict Mortality in COVID-19. *J. Inflamm. Res.* **2021**, *14*, 4651–4667. [CrossRef]
18. Finsterer, J.; Scorza, F.A. Clinical and Pathophysiologic Spectrum of Neuro-COVID. *Mol. Neurobiol.* **2021**, *58*, 3787–3791. [CrossRef]
19. Frank, M.G.; Nguyen, K.H.; Ball, J.B.; Hopkins, S.; Kelley, T.; Baratta, M.V.; Fleshner, M.; Maier, S.F. SARS-CoV-2 spike S1 subunit induces neuroinflammatory, microglial and behavioral sickness responses: Evidence of PAMP-like properties. *Brain Behav. Immun.* **2022**, *100*, 267–277. [CrossRef]
20. Peluso, M.J.; Sans, H.M.; Forman, C.A.; Nylander, A.N.; Ho, H.E.; Lu, S.; Goldberg, S.A.; Hoh, R.; Tai, V.; Munter, S.E.; et al. Plasma Markers of Neurologic Injury and Inflammation in People with Self-Reported Neurologic Postacute Sequelae of SARS-CoV-2 Infection. *Neurol. Neuroimmunol. Neuroinflamm.* **2022**, *9*. [CrossRef]
21. Spudich, S.; Nath, A. Nervous system consequences of COVID-19. *Science* **2022**, *375*, 267–269. [CrossRef] [PubMed]
22. F.D.A. Coronavirus (COVID-19) Update: Serological Tests. Available online: <https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-serological-tests> (accessed on 7 May 2020).
23. Olink. Proximity Extension Assay (PEA) technology. Available online: <https://www.olink.com/our-platform/our-pea-technology/> (accessed on 4 March 2022).
24. Bauer, W.; Weber, M.; Diehl-Wiesenecker, E.; Galtung, N.; Prpic, M.; Somasundaram, R.; Tauber, R.; Schwenk, J.M.; Micke, P.; Kappert, K. Plasma Proteome Fingerprints Reveal Distinctiveness and Clinical Outcome of SARS-CoV-2 Infection. *Viruses* **2021**, *13*, 2456. [CrossRef] [PubMed]

25. Gutmann, C.; Takov, K.; Burnap, S.A.; Singh, B.; Ali, H.; Theofilatos, K.; Reed, E.; Hasman, M.; Nabeebaccus, A.; Fish, M.; et al. SARS-CoV-2 RNAemia and proteomic trajectories inform prognostication in COVID-19 patients admitted to intensive care. *Nat. Commun.* **2021**, *12*, 3406. [[CrossRef](#)] [[PubMed](#)]
26. Haljasmagi, L.; Salumets, A.; Rumm, A.P.; Jurgenson, M.; Krassohhina, E.; Remm, A.; Sein, H.; Kareinen, L.; Vapalahti, O.; Sironen, T.; et al. Longitudinal proteomic profiling reveals increased early inflammation and sustained apoptosis proteins in severe COVID-19. *Sci. Rep.* **2020**, *10*, 20533. [[CrossRef](#)] [[PubMed](#)]
27. Filbin, M.R.; Mehta, A.; Schneider, A.M.; Kays, K.R.; Guess, J.R.; Gentili, M.; Fenyves, B.G.; Charland, N.C.; Gonye, A.L.K.; Gushterova, I.; et al. Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Rep. Med.* **2021**, *2*, 100287. [[CrossRef](#)]
28. Maucourant, C.; Filipovic, I.; Ponzetta, A.; Aleman, S.; Cornillet, M.; Hertwig, L.; Strunz, B.; Lentini, A.; Reinius, B.; Brownlie, D.; et al. Natural killer cell immunotypes related to COVID-19 disease severity. *Sci. Immunol.* **2020**, *5*, eabd6832. [[CrossRef](#)]
29. Rovas, A.; Buscher, K.; Osiaevi, I.; Drost, C.C.; Sackarnd, J.; Tepaspe, P.R.; Fobker, M.; Kuhn, J.; Braune, S.; Gobel, U.; et al. Microvascular and proteomic signatures overlap in COVID-19 and bacterial sepsis: The MICROCODE study. *Angiogenesis* **2022**. [[CrossRef](#)]
30. Lai, A.; Bergna, A.; Menzo, S.; Zehender, G.; Caucci, S.; Ghisetti, V.; Rizzo, F.; Maggi, F.; Cerutti, F.; Giurato, G.; et al. Circulating SARS-CoV-2 variants in Italy, October 2020-March 2021. *Virol. J.* **2021**, *18*, 168. [[CrossRef](#)]
31. ISS. Press Release N°42/2021 COVID-19, Confirmed Predominance of the DELTA Variant. Available online: [https://www.iss.it/en/cov19-cosa-fa-iss-varianti/-/asset\\_publisher/yJS4xO2fauqM/content/id/5832836?\\_com\\_liferay\\_asset\\_publisher\\_web\\_portlet\\_AssetPublisherPortlet\\_INSTANCE\\_yJS4xO2fauqM\\_redirect=https%3A%2F%2Fwww.iss.it%2Fcov19-cosa-fa-iss-varianti%3Fp\\_p\\_id%3Dcom\\_liferay\\_asset\\_publisher\\_web\\_portlet\\_AssetPublisherPortlet\\_INSTANCE\\_yJS4xO2fauqM%26p\\_p\\_lifecycle%3D0%26p\\_p\\_state%3Dnormal%26p\\_p\\_mode%3Dview%26\\_com\\_liferay\\_asset\\_publisher\\_web\\_portlet\\_AssetPublisherPortlet\\_INSTANCE\\_yJS4xO2fauqM\\_cur%3D0%26\\_r\\_p\\_resetCur%3Dfalse%26\\_com\\_liferay\\_asset\\_publisher\\_web\\_portlet\\_AssetPublisherPortlet\\_INSTANCE\\_yJS4xO2fauqM\\_assetEntryId%3D5832836](https://www.iss.it/en/cov19-cosa-fa-iss-varianti/-/asset_publisher/yJS4xO2fauqM/content/id/5832836?_com_liferay_asset_publisher_web_portlet_AssetPublisherPortlet_INSTANCE_yJS4xO2fauqM_redirect=https%3A%2F%2Fwww.iss.it%2Fcov19-cosa-fa-iss-varianti%3Fp_p_id%3Dcom_liferay_asset_publisher_web_portlet_AssetPublisherPortlet_INSTANCE_yJS4xO2fauqM%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26_com_liferay_asset_publisher_web_portlet_AssetPublisherPortlet_INSTANCE_yJS4xO2fauqM_cur%3D0%26_r_p_resetCur%3Dfalse%26_com_liferay_asset_publisher_web_portlet_AssetPublisherPortlet_INSTANCE_yJS4xO2fauqM_assetEntryId%3D5832836) (accessed on 12 July 2022).
32. Mao, J.; Dai, R.; Du, R.C.; Zhu, Y.; Shui, L.P.; Luo, X.H. Hematologic changes predict clinical outcome in recovered patients with COVID-19. *Ann. Hematol.* **2021**, *100*, 675–689. [[CrossRef](#)]
33. Hueriga Encabo, H.; Grey, W.; Garcia-Albornoz, M.; Wood, H.; Ulferts, R.; Aramburu, I.V.; Kulasekararaj, A.G.; Mufti, G.; Papayannopoulos, V.; Beale, R.; et al. Human Erythroid Progenitors Are Directly Infected by SARS-CoV-2: Implications for Emerging Erythropoiesis in Severe COVID-19 Patients. *Stem Cell Rep.* **2021**, *16*, 428–436. [[CrossRef](#)]
34. Henry, B.M.; Aggarwal, G.; Wong, J.; Benoit, S.; Vikse, J.; Plebani, M.; Lippi, G. Lactate dehydrogenase levels predict coronavirus disease 2019 (COVID-19) severity and mortality: A pooled analysis. *Am. J. Emerg. Med.* **2020**, *38*, 1722–1726. [[CrossRef](#)] [[PubMed](#)]
35. Koyama, T.; Kuriyama, N.; Suzuki, Y.; Saito, S.; Tanaka, R.; Iwao, M.; Tanaka, M.; Maki, T.; Itoh, H.; Ihara, M.; et al. Mid-regional pro-adrenomedullin is a novel biomarker for arterial stiffness as the criterion for vascular failure in a cross-sectional study. *Sci. Rep.* **2021**, *11*, 305. [[CrossRef](#)] [[PubMed](#)]
36. Garcia de Gadiana-Romualdo, L.; Martinez Martinez, M.; Rodriguez Mulero, M.D.; Esteban-Torrella, P.; Hernandez Olivo, M.; Alcaraz Garcia, M.J.; Campos-Rodriguez, V.; Sancho-Rodriguez, N.; Galindo Martinez, M.; Alcaraz, A.; et al. Circulating MR-proADM levels, as an indicator of endothelial dysfunction, for early risk stratification of mid-term mortality in COVID-19 patients. *Int. J. Infect. Dis.* **2021**, *111*, 211–218. [[CrossRef](#)] [[PubMed](#)]
37. Smilowitz, N.R.; Kunichoff, D.; Garshick, M.; Shah, B.; Pillinger, M.; Hochman, J.S.; Berger, J.S. C-reactive protein and clinical outcomes in patients with COVID-19. *Eur. Heart J.* **2021**, *42*, 2270–2279. [[CrossRef](#)] [[PubMed](#)]
38. Metzemaekers, M.; Cambier, S.; Blanter, M.; Vandoooren, J.; de Carvalho, A.C.; Malengier-Devlies, B.; Vanderbeke, L.; Jacobs, C.; Coenen, S.; Martens, E.; et al. Kinetics of peripheral blood neutrophils in severe coronavirus disease 2019. *Clin. Transl. Immunol.* **2021**, *10*, e1271. [[CrossRef](#)]
39. Zou, M.; Zou, J.; Hu, X.; Zheng, W.; Zhang, M.; Cheng, Z. Latent Transforming Growth Factor-beta Binding Protein-2 Regulates Lung Fibroblast-to-Myofibroblast Differentiation in Pulmonary Fibrosis via NF-kappaB Signaling. *Front. Pharmacol.* **2021**, *12*, 788714. [[CrossRef](#)]
40. Kimura, Y.; Nakai, Y.; Shin, J.; Hara, M.; Takeda, Y.; Kubo, S.; Jeremiah, S.S.; Ino, Y.; Akiyama, T.; Moriyama, K.; et al. Identification of serum prognostic biomarkers of severe COVID-19 using a quantitative proteomic approach. *Sci. Rep.* **2021**, *11*, 20638. [[CrossRef](#)]
41. Vollmy, F.; van den Toorn, H.; Zenezini Chiozzi, R.; Zucchetti, O.; Papi, A.; Volta, C.A.; Marracino, L.; Vieceli Dalla Sega, F.; Fortini, F.; Demichev, V.; et al. A serum proteome signature to predict mortality in severe COVID-19 patients. *Life Sci. Alliance* **2021**, *4*, e202101099. [[CrossRef](#)]
42. Sur, S.; Khatun, M.; Steele, R.; Isbell, T.S.; Ray, R.; Ray, R.B. Exosomes from COVID-19 Patients Carry Tenascin-C and Fibrinogen-beta in Triggering Inflammatory Signals in Cells of Distant Organ. *Int. J. Mol. Sci.* **2021**, *22*, 3184. [[CrossRef](#)]
43. Yu, J.; Gerber, G.F.; Chen, H.; Yuan, X.; Chaturvedi, S.; Braunstein, E.M.; Brodsky, R.A. Complement dysregulation is associated with severe COVID-19 illness. *Haematologica* **2021**, *107*, 1095–1105. [[CrossRef](#)]
44. Begue, F.; Tanaka, S.; Mouktadi, Z.; Rondeau, P.; Veeren, B.; Diotel, N.; Tran-Dinh, A.; Robert, T.; Velia, E.; Mavingui, P.; et al. Altered high-density lipoprotein composition and functions during severe COVID-19. *Sci. Rep.* **2021**, *11*, 2291. [[CrossRef](#)] [[PubMed](#)]

45. Lee, J.S.; Han, D.; Kim, S.Y.; Hong, K.H.; Jang, M.J.; Kim, M.J.; Kim, Y.G.; Park, J.H.; Cho, S.I.; Park, W.B.; et al. Longitudinal proteomic profiling provides insights into host response and proteome dynamics in COVID-19 progression. *Proteomics* **2021**, *21*, e2000278. [[CrossRef](#)] [[PubMed](#)]
46. Overmyer, K.A.; Shishkova, E.; Miller, I.J.; Balnis, J.; Bernstein, M.N.; Peters-Clarke, T.M.; Meyer, J.G.; Quan, Q.; Muehlbauer, L.K.; Trujillo, E.A.; et al. Large-Scale Multi-omic Analysis of COVID-19 Severity. *Cell Syst.* **2021**, *12*, 23–40.e7. [[CrossRef](#)]
47. Vieceli Dalla Sega, F.; Fortini, F.; Spadaro, S.; Ronzoni, L.; Zucchetti, O.; Manfrini, M.; Mikus, E.; Fogagnolo, A.; Torsani, F.; Pavasini, R.; et al. Time course of endothelial dysfunction markers and mortality in COVID-19 patients: A pilot study. *Clin. Transl. Med.* **2021**, *11*, e283. [[CrossRef](#)] [[PubMed](#)]
48. Morales, A.; Rojo Rello, S.; Cristobal, H.; Fiz-Lopez, A.; Arribas, E.; Mari, M.; Tutusaus, A.; de la Cal-Sabater, P.; Nicolaes, G.A.F.; Ortiz-Perez, J.T.; et al. Growth Arrest-Specific Factor 6 (GAS6) Is Increased in COVID-19 Patients and Predicts Clinical Outcome. *Biomedicines* **2021**, *9*, 335. [[CrossRef](#)] [[PubMed](#)]
49. Deniz, S.; Uysal, T.K.; Capasso, C.; Supuran, C.T.; Ozensoy Guler, O. Is carbonic anhydrase inhibition useful as a complementary therapy of COVID-19 infection? *J. Enzym. Inhib. Med. Chem.* **2021**, *36*, 1230–1235. [[CrossRef](#)]
50. Yang, Y.; Shen, C.; Li, J.; Yuan, J.; Wei, J.; Huang, F.; Wang, F.; Li, G.; Li, Y.; Xing, L.; et al. Plasma IP-10 and MCP-3 levels are highly associated with disease severity and predict the progression of COVID-19. *J. Allergy Clin. Immunol.* **2020**, *146*, 119–127.e4. [[CrossRef](#)]
51. Sinha, P.; Matthay, M.A.; Calfee, C.S. Is a “Cytokine Storm” Relevant to COVID-19? *JAMA Intern. Med.* **2020**, *180*, 1152–1154. [[CrossRef](#)]
52. Abers, M.S.; Delmonte, O.M.; Ricotta, E.E.; Fintzi, J.; Fink, D.L.; de Jesus, A.A.A.; Zarembek, K.A.; Alehashemi, S.; Oikonomou, V.; Desai, J.V.; et al. An immune-based biomarker signature is associated with mortality in COVID-19 patients. *JCI Insight* **2021**, *6*, e144455. [[CrossRef](#)]
53. Kilic, G.; Bulut, O.; Jaeger, M.; Ter Horst, R.; Koeken, V.; Moorlag, S.; Mourits, V.P.; de Bree, C.; Dominguez-Andres, J.; Joosten, L.A.B.; et al. The Immunological Factors Predisposing to Severe COVID-19 Are Already Present in Healthy Elderly and Men. *Front. Immunol.* **2021**, *12*, 720090. [[CrossRef](#)]
54. Yasar, S.; Colak, C.; Yologlu, S. Artificial Intelligence-Based Prediction of COVID-19 Severity on the Results of Protein Profiling. *Comput. Methods Programs Biomed.* **2021**, *202*, 105996. [[CrossRef](#)] [[PubMed](#)]
55. Guo, J.; Wang, S.; Xia, H.; Shi, D.; Chen, Y.; Zheng, S.; Chen, Y.; Gao, H.; Guo, F.; Ji, Z.; et al. Cytokine Signature Associated with Disease Severity in COVID-19. *Front. Immunol.* **2021**, *12*, 681516. [[CrossRef](#)] [[PubMed](#)]
56. Etter, M.M.; Martins, T.; Kulsvehagen, L.; Poessnecker, E.; Duchemin, W.; Hogan, S.; Diaz, G.S.; Mueller, J.; Chiappini, A.; Rychen, J. Severe Neuro-COVID is associated with peripheral immune signatures, autoimmunity and signs of neurodegeneration: A prospective cross-sectional study. *medRxiv* **2022**. [[CrossRef](#)]
57. Kvedaraite, E.; Hertwig, L.; Sinha, I.; Ponzetta, A.; Hed Myrberg, I.; Lourda, M.; Dzidic, M.; Akber, M.; Klingstrom, J.; Folkesson, E.; et al. Major alterations in the mononuclear phagocyte landscape associated with COVID-19 severity. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2018587118. [[CrossRef](#)] [[PubMed](#)]
58. Filbin, M.R.; Mehta, A.; Schneider, A.M.; Kays, K.R.; Guess, J.R.; Gentili, M.; Fenyves, B.G.; Charland, N.C.; Gonye, A.L.K.; Gushterova, I.; et al. Plasma proteomics reveals tissue-specific cell death and mediators of cell-cell interactions in severe COVID-19 patients. *bioRxiv* **2020**. [[CrossRef](#)]
59. Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. *Brief. Bioinform.* **2018**, *19*, 1236–1246. [[CrossRef](#)]
60. Guncar, G.; Kukar, M.; Notar, M.; Brvar, M.; Cernelc, P.; Notar, M. An application of machine learning to haematological diagnosis. *Sci. Rep.* **2018**, *8*, 411. [[CrossRef](#)]
61. De Nardin, A.; Mishra, P.; Foresti, G.L.; Piciarelli, C. Masked Transformer for image Anomaly Localization. *Int. J. Neural Syst.* **2022**, *32*, 2250030. [[CrossRef](#)]
62. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138. [[CrossRef](#)]
63. Janssen, N.A.F.; Grondman, I.; de Nooijer, A.H.; Boahen, C.K.; Koeken, V.; Matzaraki, V.; Kumar, V.; He, X.; Kox, M.; Koenen, H.; et al. Dysregulated Innate and Adaptive Immune Responses Discriminate Disease Severity in COVID-19. *J. Infect. Dis.* **2021**, *223*, 1322–1333. [[CrossRef](#)]
64. Liberzon, A.; Subramanian, A.; Pinchback, R.; Thorvaldsdottir, H.; Tamayo, P.; Mesirov, J.P. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **2011**, *27*, 1739–1740. [[CrossRef](#)] [[PubMed](#)]
65. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]
66. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W.H.; Pages, F.; Trajanoski, Z.; Galon, J. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25*, 1091–1093. [[CrossRef](#)] [[PubMed](#)]
67. Pai, S.; Ellington, A.D.; Levy, M. Proximity ligation assays with peptide conjugate ‘burrs’ for the sensitive detection of spores. *Nucleic Acids Res.* **2005**, *33*, e162. [[CrossRef](#)]

68. Avolio, E.; Carrabba, M.; Milligan, R.; Kavanagh Williamson, M.; Beltrami, A.P.; Gupta, K.; Elvers, K.T.; Gamez, M.; Foster, R.R.; Gillespie, K.; et al. The SARS-CoV-2 Spike protein disrupts human cardiac pericytes function through CD147 receptor-mediated signalling: A potential non-infective mechanism of COVID-19 microvascular disease. *Clin. Sci.* **2021**, *135*, 2667–2689. [[CrossRef](#)]
69. Tschöpe, C.; Ammirati, E.; Bozkurt, B.; Caforio, A.L.P.; Cooper, L.T.; Felix, S.B.; Hare, J.M.; Heidecker, B.; Heymans, S.; Hubner, N.; et al. Myocarditis and inflammatory cardiomyopathy: Current evidence and future directions. *Nat. Rev. Cardiol.* **2021**, *18*, 169–193. [[CrossRef](#)]
70. Heming, M.; Li, X.; Rauber, S.; Mausberg, A.K.; Borsch, A.L.; Hartlehnert, M.; Singhal, A.; Lu, I.N.; Fleischer, M.; Szepanowski, F.; et al. Neurological Manifestations of COVID-19 Feature T Cell Exhaustion and Dedifferentiated Monocytes in Cerebrospinal Fluid. *Immunity* **2021**, *54*, 164–175.e6. [[CrossRef](#)]
71. Messner, C.B.; Demichev, V.; Wendisch, D.; Michalick, L.; White, M.; Freiwald, A.; Textoris-Taube, K.; Vernardis, S.I.; Egger, A.S.; Kreidl, M.; et al. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst.* **2020**, *11*, 11–24.e4. [[CrossRef](#)]
72. Sozio, E.; Tascini, C.; Fabris, M.; D’Aurizio, F.; De Carlo, C.; Graziano, E.; Bassi, F.; Sbrana, F.; Ripoli, A.; Pagotto, A.; et al. MR-proADM as prognostic factor of outcome in COVID-19 patients. *Sci. Rep.* **2021**, *11*, 5121. [[CrossRef](#)]
73. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2013.
74. Schimke, L.F.; Marques, A.H.C.; Baiocchi, G.C.; de Souza Prado, C.A.; Fonseca, D.L.M.; Freire, P.P.; Rodrigues Placa, D.; Salerno Filgueiras, I.; Coelho Salgado, R.; Jansen-Marques, G.; et al. Severe COVID-19 Shares a Common Neutrophil Activation Signature with Other Acute Inflammatory States. *Cells* **2022**, *11*, 847. [[CrossRef](#)]
75. Guan, W.J.; Ni, Z.Y.; Hu, Y.; Liang, W.H.; Ou, C.Q.; He, J.X.; Liu, L.; Shan, H.; Lei, C.L.; Hui, D.S.C.; et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* **2020**, *382*, 1708–1720. [[CrossRef](#)] [[PubMed](#)]
76. Loyer, C.; Lapostolle, A.; Urbina, T.; Elabbadi, A.; Lavoie, J.P.; Chaigneau, T.; Simoes, C.; Dessajan, J.; Desnos, C.; Morin-Bureau, M.; et al. Impairment of neutrophil functions and homeostasis in COVID-19 patients: Association with disease severity. *Crit. Care* **2022**, *26*, 155. [[CrossRef](#)] [[PubMed](#)]
77. McKenna, E.; Wubben, R.; Isaza-Correa, J.M.; Melo, A.M.; Mhaonaigh, A.U.; Conlon, N.; O’Donnell, J.S.; Ni Cheallaigh, C.; Hurley, T.; Stevenson, N.J.; et al. Neutrophils in COVID-19: Not Innocent Bystanders. *Front. Immunol.* **2022**, *13*, 864387. [[CrossRef](#)] [[PubMed](#)]
78. Veras, F.P.; Pontelli, M.C.; Silva, C.M.; Toller-Kawahisa, J.E.; de Lima, M.; Nascimento, D.C.; Schneider, A.H.; Caetite, D.; Tavares, L.A.; Paiva, I.M.; et al. SARS-CoV-2-triggered neutrophil extracellular traps mediate COVID-19 pathology. *J. Exp. Med.* **2020**, *217*, e20201129. [[CrossRef](#)]
79. Reusch, N.; De Domenico, E.; Bonaguro, L.; Schulte-Schrepping, J.; Bassler, K.; Scholtze, J.L.; Aschenbrenner, A.C. Neutrophils in COVID-19. *Front. Immunol.* **2021**, *12*, 652470. [[CrossRef](#)]
80. Wilk, A.J.; Lee, M.J.; Wei, B.; Parks, B.; Pi, R.; Martinez-Colon, G.J.; Ranganath, T.; Zhao, N.Q.; Taylor, S.; Becker, W.; et al. Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis in COVID-19. *J. Exp. Med.* **2021**, *218*, e20210582. [[CrossRef](#)]
81. Schulte-Schrepping, J.; Reusch, N.; Paclik, D.; Bassler, K.; Schlickeiser, S.; Zhang, B.; Kramer, B.; Krammer, T.; Brumhard, S.; Bonaguro, L.; et al. Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell* **2020**, *182*, 1419–1440.e23. [[CrossRef](#)]
82. Singh, Y.; Trautwein, C.; Fendel, R.; Krickeberg, N.; Berezhnoy, G.; Bissinger, R.; Ossowski, S.; Salker, M.S.; Casadei, N.; Riess, O.; et al. SARS-CoV-2 infection paralyzes cytotoxic and metabolic functions of the immune cells. *Heliyon* **2021**, *7*, e07147. [[CrossRef](#)]
83. Masilamani, M.; Kassahn, D.; Mikkat, S.; Glocker, M.O.; Illges, H. B cell activation leads to shedding of complement receptor type II (CR2/CD21). *Eur. J. Immunol.* **2003**, *33*, 2391–2397. [[CrossRef](#)]
84. Komori, M.; Lin, Y.C.; Cortese, I.; Blake, A.; Ohayon, J.; Cherup, J.; Maric, D.; Kosa, P.; Wu, T.; Bielekova, B. Insufficient disease inhibition by intrathecal rituximab in progressive multiple sclerosis. *Ann. Clin. Transl. Neurol.* **2016**, *3*, 166–179. [[CrossRef](#)]
85. Fajgenbaum, D.C.; June, C.H. Cytokine Storm. *N. Engl. J. Med.* **2020**, *383*, 2255–2273. [[CrossRef](#)] [[PubMed](#)]
86. Walsh, M.C.; Choi, Y. Biology of the RANKL-RANK-OPG System in Immunity, Bone, and Beyond. *Front. Immunol.* **2014**, *5*, 511. [[CrossRef](#)] [[PubMed](#)]
87. Yu, Q.; Gu, J.X.; Kovacs, C.; Freedman, J.; Thomas, E.K.; Ostrowski, M.A. Cooperation of TNF family members CD40 ligand, receptor activator of NF-kappa B ligand, and TNF-alpha in the activation of dendritic cells and the expansion of viral specific CD8+ T cell memory responses in HIV-1-infected and HIV-1-uninfected individuals. *J. Immunol.* **2003**, *170*, 1797–1805. [[CrossRef](#)] [[PubMed](#)]
88. Zhao, S.C.; Heng, X.; Ya-Ping, W.; Di, L.; Wen-Qian, W.; Ling-Song, M.; Chu, Z.H.; Xu, Y. CD200-CD200R1 signaling pathway regulates neuroinflammation after stroke. *Brain Behav.* **2020**, *10*, e01882. [[CrossRef](#)]
89. Wright, G.J.; Cherwinski, H.; Foster-Cuevas, M.; Brooke, G.; Puklavec, M.J.; Bigler, M.; Song, Y.; Jenmalm, M.; Gorman, D.; McClanahan, T.; et al. Characterization of the CD200 receptor family in mice and humans and their interactions with CD200. *J. Immunol.* **2003**, *171*, 3034–3046. [[CrossRef](#)]

90. Ritzel, R.M.; Al Mamun, A.; Crapser, J.; Verma, R.; Patel, A.R.; Knight, B.E.; Harris, N.; Mancini, N.; Roy-O'Reilly, M.; Ganesh, B.P.; et al. CD200-CD200R1 inhibitory signaling prevents spontaneous bacterial infection and promotes resolution of neuroinflammation and recovery after stroke. *J. Neuroinflamm.* **2019**, *16*, 40. [CrossRef]
91. Jaeger, L.B.; Dohgu, S.; Sultana, R.; Lynch, J.L.; Owen, J.B.; Erickson, M.A.; Shah, G.N.; Price, T.O.; Fleegal-Demotta, M.A.; Butterfield, D.A.; et al. Lipopolysaccharide alters the blood-brain barrier transport of amyloid beta protein: A mechanism for inflammation in the progression of Alzheimer's disease. *Brain Behav. Immun.* **2009**, *23*, 507–517. [CrossRef]
92. Sánchez, L.; Alvarez, V.; González, P.; González, I.; Alvarez, R.; Coto, E. Variation in the LRP-associated protein gene (LRPAP1) is associated with late-onset Alzheimer disease. *Am. J. Med. Genet.* **2001**, *105*, 76–78. [CrossRef]
93. Sita, G.; Graziosi, A.; Hrelia, P.; Morroni, F. NLRP3 and Infections: Beta-Amyloid in Inflammasome beyond Neurodegeneration. *Int. J. Mol. Sci.* **2021**, *22*, 6984. [CrossRef]
94. Collins, M.; Bowser, R. Molecular Mechanisms of Amyotrophic Lateral Sclerosis. *Mol. Cell. Ther. Mot. Neuron Dis.* **2017**, 61–99. [CrossRef]
95. Collins, M.A.; An, J.; Hood, B.L.; Conrads, T.P.; Bowser, R.P. Label-Free LC-MS/MS Proteomic Analysis of Cerebrospinal Fluid Identifies Protein/Pathway Alterations and Candidate Biomarkers for Amyotrophic Lateral Sclerosis. *J. Proteome Res.* **2015**, *14*, 4486–4501. [CrossRef] [PubMed]
96. Park, S.; Ranjbarvaziri, S.; Zhao, P.; Ardehali, R. Cardiac Fibrosis Is Associated with Decreased Circulating Levels of Full-Length CILP in Heart Failure. *JACC. Basic Transl. Sci.* **2020**, *5*, 432–443. [CrossRef]
97. Shi, C.; Li, X.; Hong, F.; Wang, X.; Jiang, T.; Sun, B.; Li, S. Latent-transforming growth factor beta-binding protein 2 accelerates cardiac fibroblast apoptosis by regulating the expression and activity of caspase-3. *Exp. Ther. Med.* **2021**, *22*, 1146. [CrossRef]
98. Michels, J.R.; Nazrul, M.S.; Adhikari, S.; Wilkins, D.; Pavel, A.B. Th1, Th2 and Th17 inflammatory pathways synergistically predict cardiometabolic protein expression in serum of COVID-19 patients. *Mol. Omics* **2022**. [CrossRef] [PubMed]
99. Caslin, H.L.; Kiwanuka, K.N.; Haque, T.T.; Taruselli, M.T.; MacKnight, H.P.; Paranjape, A.; Ryan, J.J. Controlling Mast Cell Activation and Homeostasis: Work Influenced by Bill Paul That Continues Today. *Front. Immunol.* **2018**, *9*, 868. [CrossRef] [PubMed]
100. Motta Junior, J.D.S.; Miggiolaro, A.; Nagashima, S.; de Paula, C.B.V.; Baena, C.P.; Scharfstein, J.; de Noronha, L. Mast Cells in Alveolar Septa of COVID-19 Patients: A Pathogenic Pathway That May Link Interstitial Edema to Immunothrombosis. *Front. Immunol.* **2020**, *11*, 574862. [CrossRef]
101. Malone, R.W.; Tisdall, P.; Fremont-Smith, P.; Liu, Y.; Huang, X.P.; White, K.M.; Miorin, L.; Moreno, E.; Alon, A.; Delaforge, E.; et al. COVID-19: Famotidine, Histamine, Mast Cells, and Mechanisms. *Front. Pharmacol.* **2021**, *12*, 633680. [CrossRef]
102. WHO. Living Guidance for Clinical Management of COVID-19. Available online: <https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-2> (accessed on 15 January 2022).
103. Nevola, K.; Sandin, M.; Guess, J.; Forsberg, S.; Cambronero, C.; Pucholt, P. OlinkAnalyze: Facilitate Analysis of Proteomic Data from Olink. Available online: <https://cran.r-project.org/web/packages/OlinkAnalyze/index.html> (accessed on 15 January 2022).
104. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net (vol B 67, pg 301, 2005). *J. R. Stat. Soc. B* **2005**, *67*, 768. [CrossRef]
105. Li, Y.; Schneider, A.M.; Mehta, A.; Sade-Feldman, M.; Kays, K.R.; Gentili, M.; Charland, N.C.; Gonye, A.L.; Gushterova, I.; Khanna, H.K.; et al. SARS-CoV-2 viremia is associated with distinct proteomic pathways and predicts COVID-19 outcomes. *J. Clin. Investig.* **2021**, *131*, e148635. [CrossRef]
106. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
107. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
108. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]