



Article

The Effect of Naturally Acquired Immunity on Mortality Predictors: A Focus on Individuals with New Coronavirus

Mónica Queipo ^{1,2}, Jorge Mateo ^{3,4,*}, Ana María Torres ^{3,4} and Julia Barbado ^{1,2,5,*}

¹ Autoimmunity and Inflammation Research Group, Río Hortega University Hospital, 47012 Valladolid, Spain; mqueipo@saludcastillayleon.es

² Cooperative Research Network Focused on Health Results—Advanced Therapies (RICORS TERAV), 28220 Madrid, Spain

³ Medical Analysis Expert Group, Institute of Technology, University of Castilla-La Mancha, 13001 Cuenca, Spain

⁴ Medical Analysis Expert Group, Instituto de Investigación Sanitaria de Castilla-La Mancha (IDISCAM), 45071 Toledo, Spain

⁵ Internal Medicine, Río Hortega University Hospital, 47012 Valladolid, Spain

* Correspondence: jorge.mateo@uclm.es (J.M.); jbarbadoa@saludcastillayleon.es (J.B.)

Abstract: Background/Objectives: The spread of the COVID-19 pandemic has spurred the development of advanced healthcare tools to effectively manage patient outcomes. This study aims to identify key predictors of mortality in hospitalized patients with some level of natural immunity, but not yet vaccinated, using machine learning techniques. **Methods:** A total of 363 patients with COVID-19 admitted to Río Hortega University Hospital in Spain between the second and fourth waves of the pandemic were included in this study. Key characteristics related to both the patient's previous status and hospital stay were screened using the Random Forest (RF) machine learning technique. **Results:** Of the 19 variables identified as having the greatest influence on predicting mortality, the most powerful ones could be identified at the time of hospital admission. These included the assessment of severity in community-acquired pneumonia (CURB-65) scale, age, the Glasgow Coma Scale (GCS), and comorbidities, as well as laboratory results. Some variables associated with hospitalization and intensive care unit (ICU) admission (acute renal failure, shock, PRONO sessions and the Acute Physiology and Chronic Health Evaluation [APACHE-II] scale) showed a certain degree of significance. The Random Forest (RF) method showed high accuracy, with a precision of >95%. **Conclusions:** This study shows that natural immunity generates significant changes in the evolution of the disease. As has been shown, machine learning models are an effective tool to improve personalized patient care in different periods.

Keywords: immunity; COVID-19; predictors; mortality; machine learning; Random Forest



Academic Editors: Monika Gudowska-Sawczuk and Monika Zajkowska

Received: 18 February 2025

Revised: 19 March 2025

Accepted: 24 March 2025

Published: 27 March 2025

Citation: Queipo, M.; Mateo, J.; Torres, A.M.; Barbado, J. The Effect of Naturally Acquired Immunity on Mortality Predictors: A Focus on Individuals with New Coronavirus. *Biomedicines* **2025**, *13*, 803. <https://doi.org/10.3390/biomedicines13040803>

Copyright: © 2025 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

When a sufficient number of individuals are immunized against a pathogen, the probability of transmission between the infected and susceptible population decreases due to the interruption of the chain of transmission [1]. In the case of a contagious, direct contact-transferable pathogen that induces long-term immunity, this is maximized if the population has a random pattern of interaction [2]. Diseases such as rubella, measles or pertussis are under control, and others such as smallpox are even eradicated today thanks to herd immunity, either by natural means or through vaccination strategies [3].

Natural immunity, which develops after infection with COVID-19, and vaccine-induced immunity, achieved through controlled exposure to viral antigens, are key mechanisms for protection against the virus. While natural immunity depends on factors like the initial viral load and may be less predictable, vaccines offer more uniform protection and mitigate the risks associated with severe disease [4]. Moreover, recent studies suggest that the combination of both types of immunity can provide more robust and long-lasting protection [5].

This concept gained significant visibility during the coronavirus disease-2019 (COVID-19) pandemic, as the need to achieve herd immunity became a critical objective to minimize the clinical and societal impact of the health emergency. Early estimates indicated that 70% of the population would need immunization to achieve this goal [6]. Data from populations with full vaccination schedules suggest that this threshold was reached between the third and fourth waves of the pandemic (June–September 2021) in regions, such as Europe, Oceania, the Middle East, and North America, while Asia and Africa reached these rates later, between late 2021 and mid-2022 [7].

Well before vaccination campaigns began, however, a substantial percentage of the population had already acquired natural immunity. Studying the characteristics of individuals infected during this period—before vaccination but after the first wave—is of profound clinical interest. Understanding their natural immune responses, the variability in disease severity, treatment protocols, and virus epidemiology provides invaluable insights for managing future outbreaks or similar health crises.

Severe cases of COVID-19 are characterized by an inflammatory response and a cytokine storm that affects all cells of the immune system [8], especially lymphocytes. This has been shown to cause a dysregulation that can lead to an uncontrolled immune response, lung tissue damage or even multi-organ failure [9]. The initial symptoms of the disease, as reported at the onset of the pandemic, included fever, dyspnea, pneumonia and a dry cough [10–12]. Nevertheless, differences in clinical presentation across pandemic waves have been observed, such as reduced intensive care unit (ICU) admission rates [13] or decreased mortality [14,15]. It is evident that this progression can be attributed, firstly, to a heightened comprehension of the disease, facilitated by scientific research and the implementation of social policies. However, it is also associated with the modification of immunological characteristics at both the individual and population levels, owing to the progressive development of immunity, as a growing number of individuals recover from the infection and achieve immunological protection.

Comparing COVID-19 with earlier respiratory infections caused by coronaviruses, such as severe acute respiratory syndrome coronavirus (SARS-CoV) (2002–2003) and Middle Eastern respiratory syndrome coronavirus (MERS-CoV) (2012), offers additional perspective. Both SARS-CoV and MERS-CoV share similarities with SARS-CoV-2 in transmission routes [16] and clinical features, though with distinct differences in transmission rates and global impacts [17]. Lessons learned from studies on these viruses provide a framework for analyzing COVID-19 data and underscore the utility of advanced analytical techniques such as machine learning (ML).

In the context of data analysis from a limited sample of patients during a pandemic, advanced data analysis techniques are increasingly being used. Among artificial intelligence (AI) tools, machine learning (ML) has seen a surge in utilization in studies dealing with voluminous datasets. This branch of AI and computational analysis entails the employment of data and algorithms that emulate human information processing capabilities while enhancing efficiency [18]. Its efficacy has been demonstrated in several areas of knowledge, including biology [19] or medicine [20–22]. Furthermore, these machine learning tools have been extensively utilized in research related to the novel coronavirus, SARS-CoV-2,

particularly in the study of mortality [23,24] and/or severity [25,26], the study of populations with specific characteristics (smokers [27], cancer patients [28,29], etc.) and biomarker analysis [30,31].

A notable benefit of leveraging ML models over conventional statistical tools is their capacity to generate precise predictions while exhibiting high levels of scalability and adaptability. This capability enables the identification of patterns within voluminous datasets. This is particularly relevant in the context of a public health problem that was not only unknown in its early stages but has shown a remarkable evolution over different time periods. Consequently, the objective of this study is to identify key predictors of mortality risk in a cohort of patients hospitalized for COVID-19 during a period when vaccines were not yet available but immunity obtained by natural means was present.

For this purpose, the specific ML model to be validated is the Random Forest (RF). This method might be utilized systematically as a risk evaluation tool in any population, enabling the derivation of conclusions in a relatively brief time frame with data from a limited number of patients (150–300). This approach would enable the development of customized protocols for diverse populations, such as health centers or cities, with the objective of optimizing resource utilization and ensuring a highly personalized healthcare experience.

2. Materials and Methods

2.1. Data Source and Description

The clinical data used in this study were taken from the electronic medical record system of the Río Hortega University Hospital in Valladolid (Spain). Data from 363 patients hospitalized with polymerase chain reaction (PCR)-confirmed COVID-19 infection between the beginning of the second wave (2 May 2020) and the end of the fourth wave (10 June 2021) were obtained from this platform. The information used in this study corresponds to the patients' hospital stay from the time of admission to the emergency department until discharge from the hospital. Each patient was given an anonymous code to protect their privacy, and all patients gave informed consent. This study was conducted according to the principles of Helsinki and was approved by the Ethics Committee of the University Hospital of Río Hortega.

Data were collected, retrospectively reviewed, and manually entered into a pre-designed database. These data included demographics, comorbidities, chronic treatments, date of admission, date of discharge or death and cause, symptoms on admission, date of COVID-19 diagnosis and virus variant, if available, chest X-ray on admission, community acquired pneumonia severity scales (CURB-65), sequential organ failure assessment (SOFA), acute physiology and chronic health disease classification system (APACHE-II), Glasgow Coma Scale (GCS), vital signs, laboratory data, need for intensive care unit (ICU) admission and date (if applicable), need and type of ventilatory support, co-infections, nosocomial infections (specimen typing), complications during hospitalization, previous COVID-19 episodes, COVID-19 specific treatment, and number and type of COVID-19 vaccines received.

The laboratory tests whose results were used in this study were performed at the same hospital center using the following instruments: DXH900 Beckman Coulter Diagnosis (Brea, CA, USA) for whole blood samples, AU5820 Beckman Coulter Diagnosis for serum biochemistry samples and Gem5000 Werfen (Barcelona, Spain) for blood gasometry analysis. The data were entered into the aforementioned hospital electronic data storage system prior to their use in this study. The laboratory parameters considered relevant to this study were the following: leukocytes, neutrophils, lymphocytes, monocytes, eosinophils, basophils, erythrocytes, hemoglobin, hematocrit, mean corpuscular volume (M.C.V.), platelets, D-dimer, prothrombin activity (PT), international normalized

ratio (I.N.R.), activated partial thromboplastin time (aPTT), aPTT ratio, derived fibrinogen, sodium, potassium, chloride, glucose, urea, creatinine, estimated glomerular filtration rate (CKD-EPI 2009), alanine aminotransferase (ALT/GPT), aspartate aminotransferase (AST/GOT), gamma glutamyl transferase (GGT), total bilirubin, alkaline phosphatase, lactate dehydrogenase (LDH), phosphate, C-reactive protein, procalcitonin, albumin, pH, $p\text{CO}_2$, $p\text{O}_2$, HCO_3 , FIO_2 , $p\text{O}_2/\text{FIO}_2$, O_2 gradient Aa and lactate.

These variables were selected after a thorough literature review to identify the most critical factors during the pandemic. The focus was on those that could significantly aid in evaluating disease severity, immune response, and associated risk factors. Additionally, the chosen parameters were those that could be quickly and routinely measured both at the time of hospital admission and throughout the patient's stay in the hospital.

2.2. Machine Learning Methods

In this study, the Random Forest algorithm was developed, an ensemble method that employs the bagging aggregation approach to construct multiple decision trees independently, thereby reducing variance and enhancing the robustness of the model. Random Forest is based on bootstrap sampling of the training data to build a set of independent trees, where each tree is trained on a random sample of the original dataset. Furthermore, at each tree node, a random subset of features is selected instead of considering all variables, introducing an additional source of randomness and minimizing correlation between the trees, thereby increasing the accuracy of the ensemble [32,33].

During the training process, each tree makes decisions independently, and the results are combined using a majority voting scheme for classification. Feature importance is calculated by analyzing the decrease in accuracy or Gini index when the values of a feature are randomly permuted, thereby identifying the most relevant variables for the model. Finally, the model's performance was evaluated using specific metrics such as accuracy, sensitivity, specificity, and the area under the receiver operating characteristic (ROC) curve (AUC), enabling validation of its predictive capability and generalization to unseen data.

In the proposed Random Forest algorithm, a set of decision trees $\{T_1, T_2, \dots, T_m\}$ is constructed using a bagging aggregation approach. To build each tree T_i , the following steps were performed:

Bootstrap sampling: Given a training dataset with n observations and p features, a subset of data D_i is generated by selecting n random samples with replacement from the original dataset. This sampling technique allows certain data points to appear multiple times in D_i , while others may not appear at all.

Random feature selection: At each node of every tree, instead of evaluating all p features, a random subset of k features is selected, where $k = \sqrt{p}$. This reduces the correlation between individual trees, thereby increasing the model's generalization ability.

Splitting criterion at each node: Each node is split according to an impurity reduction criterion, which can be either entropy or the Gini index for classification. In our study, we use the Gini index. The Gini impurity G of a node with a proportion p_k of elements belonging to class k is defined as:

$$G = 1 - \sum_{k=1}^K p_k^2$$

Combination of trees: Once the trees are trained, the prediction of the Random Forest is obtained through aggregation. For a set of trees $\{T_1, T_2, \dots, T_m\}$, the final prediction \hat{y} is calculated using majority voting:

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \dots, T_m(x)\}$$

Feature importance: The importance of each feature was measured by evaluating the change in the splitting criterion when the feature was randomly permuted in the dataset. For importance based on the Gini index, if permuting a specific feature increased the impurity of the tree nodes, that feature was considered important.

Model optimization and tuning: In order to improve the predictive performance of the Random Forest model and avoid overfitting, hyperparameter tuning was performed using a Bayesian optimization strategy. This approach was preferred over grid search due to its efficiency in navigating large parameter spaces with fewer evaluations. The optimization was carried out with a fivefold cross-validation strategy applied to the training data. The hyperparameters included the number of trees in the forest ($n_estimators$), the maximum depth of the trees (max_depth), the minimum number of samples required to split an internal node ($min_samples_split$), and the number of features considered at each split ($max_features$). The optimization objective was to maximize the average cross-validated AUC. This procedure allowed us to determine an optimal configuration that balances model complexity and generalization capacity, minimizing the risk of overfitting.

Model evaluation: The model's performance was evaluated using metrics, such as accuracy, sensitivity, specificity, and the AUC.

In this study, the proposed method underwent extensive evaluation, comparing it with various machine learning techniques for classifying COVID-19 patients. The algorithms included in the comparative analysis were Gaussian Naive Bayes (GNB) [34,35], k-Nearest Neighbors (KNNs) [36,37], Bayesian Linear Discriminant Analysis (BLDA) [38,39], Support Vector Machines (SVM) [40], decision trees (DT) [41,42], and the novel system developed in this study, Random Forest (RF) [32,43,44]. The implementation and evaluation of the models were conducted using MATLAB's Statistics and Machine Learning Toolbox (version 2024a).

To mitigate overfitting, a fivefold cross-validation strategy was applied. The data were split into two subsets, assigning 70% for training and 30% for testing, ensuring the independence of patient groups in each set. Figure 1 schematically illustrates the study workflow, which begins with patient selection and database creation, followed by the training and validation phases of the machine learning models.

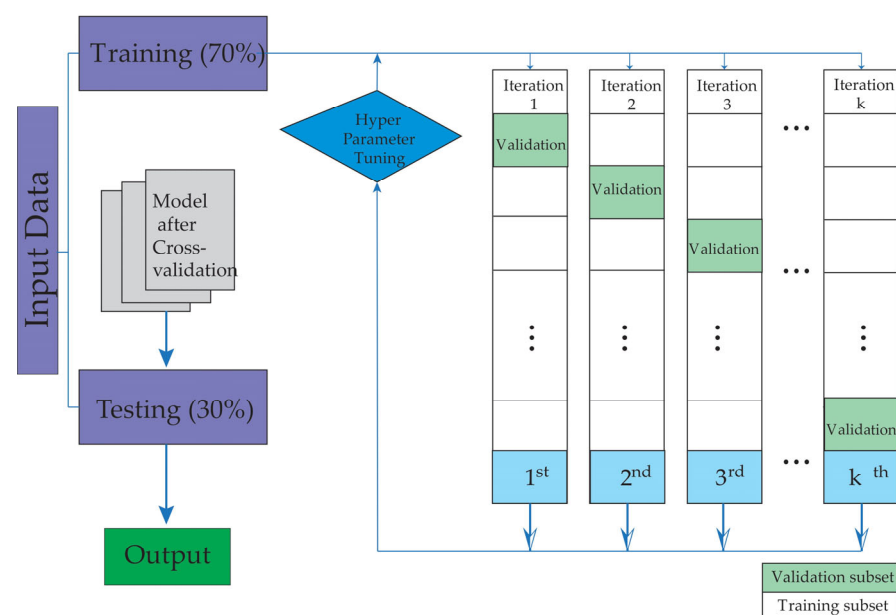


Figure 1. The figure illustrates the training and validation stages used in this study.

3. Results

A total cohort of 363 patients was studied, corresponding to the second (135 patients), third (115 patients) and fourth (113 patients) waves of the pandemic, when mass vaccination campaigns had not yet begun. The study identified 40 deceased patients, resulting in an overall mortality rate of 11%.

The medical treatments and protocols used during the second to fourth waves of COVID-19 were stable during this period and followed hospital, local health and ministry of health protocols, based on WHO indications.

The information obtained is related to the patient's previous condition, the time of admission, and also the evolution of the patient during the hospital stay. More detailed data on demographics, comorbidities, chronic treatments, symptoms on admission, ICU stay, need for ventilatory support or complications during hospital stay, among others, are shown in Tables 1–4.

Table 1. Parameters related to patient history and demographics.

	Deceased (40 Patients, 11%)		Alive (323 Patients, 89%)		Global (363 Patients, 100%)	
	(n)	(%)	(n)	(%)	(n)	(%)
Sex						
Men	33	83	198	61	231	64
Women	7	18	125	39	132	36
Comorbidities						
Hypertension	24	60	144	45	168	46
Diabetes mellitus	10	25	68	21	78	21
Ischemic cardiopathy	7	18	17	5	24	7
Cardiac insufficiency	1	3	6	2	7	2
COPD (Chronic emphysema/bronchitis)	1	3	17	5	18	5
Asthma	1	3	17	5	18	5
Dyslipidemia	16	40	100	31	116	32
Coagulopathies	7	18	34	11	41	11
Renal insufficiency	5	13	19	6	24	7
Active tumors	1	3	16	5	17	5
Immune-mediated diseases	2	5	38	12	40	11
Overweight	1	3	4	1	5	1
Obesity	3	8	30	9	33	9
Chronic treatments	32	80	183	57	215	59
Antihypertensives	21	53	115	36	136	37
Beta-blockers	10	25	38	12	48	13
Diuretics	8	20	37	11	45	12
Antidiabetics	8	20	58	18	66	18
Anti-aggregants	11	28	33	10	44	12
Anticoagulants	6	15	34	11	40	11
Lipid lowering agents	15	38	90	28	105	29
Chemotherapy	1	3	2	1	3	1
Immunosuppressive chronic treatment	3	8	21	7	24	7
Antiretrovirals	0	0	1	0	1	0
Antivirals	0	0	1	0	1	0
Vaccination status						
1 dose	2	5	15	5	17	5
2 doses	0	0	6	2	6	2
3 doses	0	0	0	0	0	0
4 doses	0	0	0	0	0	0

Table 2. Clinical characteristics of patients at the time of hospital admission.

	Deceased (40 Patients, 11%)		Alive (323 Patients, 89%)		Global (363 Patients)	
	(n)	(%)	(n)	(%)	(n)	(%)
Symptoms on admission						
Dyspnea	19	48	172	53	191	53
Chest discomfort	1	3	44	14	45	12
Cough	21	53	202	63	223	61
Rhinorrhea	0	0	10	3	10	3
Loss of smell (anosmia)	0	0	49	15	49	13
Loss of taste (ageusia)	2	5	48	15	50	14
Odynophagia	0	0	12	4	12	3
Myalgia	4	10	57	18	61	17
Fever	26	65	225	70	251	69
Dysthermia	3	8	27	8	30	8
Headache	2	5	33	10	35	10
Nausea/Vomiting	5	13	47	15	52	14
Diarrhea	3	8	65	20	68	19
Asthenia	7	18	101	31	108	30
Confusion	9	23	8	2	17	5
Dizziness	2	5	19	6	21	6
Sputum	6	15	45	14	51	14
Diagnosis on admission						
Respiratory distress	6	15	21	7	27	7
Acute respiratory insufficiency	20	50	127	39	147	40
Multiorgan failure	2	5	0	0	2	1

Table 3. Patient characteristics associated with hospitalization.

	Deceased (40 Patients, 11%)		Alive (323 Patients, 89%)		Global (363 Patients)	
	(n)	(%)	(n)	(%)	(n)	(%)
ICU admission	21	53	61	19	82	23
Invasive mechanical ventilation (IMV)	19	48	56	17	75	21
High Flow Nasal Cannula (HFNC)	4	10	29	9	33	9
Non-invasive mechanical ventilation (NIV)	12	30	30	9	42	12
Prone sessions	12	30	28	9	40	11
Extracorporeal membrane oxygenation (ECMO)	1	3	4	1	5	1
At least one previous episode of COVID	1	3	2	1	3	1
Nosocomial infections						
Viral	0	0	2	1	2	1
Bacterial	11	28	30	9	41	11
Fungal	3	8	3	1	6	2
Complications during hospital stay						
Acute renal failure	13	33	16	5	29	8
Cardiac	9	23	15	5	24	7
Arrhythmias	9	23	15	5	24	7
Gastrointestinal	5	13	22	7	27	7
Increased transaminases	5	13	10	3	15	4
Ileus	0	0	9	3	9	2
Mesenteric ischemia	0	0	1	0	1	0
Subocclusion	0	0	2	1	2	1
Neurological	3	8	19	6	22	6
Delirium	1	3	15	5	16	4
Encephalopathy	1	3	0	0	1	0
Peripheral neuropathy	1	3	4	1	5	1
Coagulopathies	9	23	14	4	23	6
Deep vein thrombosis	0	0	2	1	2	1

Table 3. Cont.

	Deceased (40 Patients, 11%)		Alive (323 Patients, 89%)		Global (363 Patients)	
	(n)	(%)	(n)	(%)	(n)	(%)
Pulmonary thromboembolism	3	8	7	2	10	3
Stroke	1	3	2	1	3	1
Bleeding	5	13	3	1	8	2
Respiratory distress (ARDS)	15	38	36	11	51	14
Shock	5	13	1	0	6	2
Treatment for COVID-19 during hospital stay						
Oxygen	32	80	173	54	205	56
Corticosteroids	26	66	142	44	168	46
Rendesivir	2	5	26	8	28	8
Ceftriaxone	21	53	142	44	163	45
Azithromycin	15	38	111	34	126	35
Heparin	8	20	112	35	120	33
Cause of death (if applicable)						
Multiorgan failure	17	43	0	0	17	5
Respiratory distress	1	3	0	0	1	0
Respiratory failure	11	28	0	0	11	3

Table 4. Numerical baseline clinical characteristics of patients. Mean and standard deviation (SD) were calculated.

	Deceased (40 Patients, 11%)		Alive (323 Patients, 89%)		Global (363 Patients)	
	Mean	SD	Mean	SD	Mean	SD
Age	79.53	10.92	64.40	16.18	66.08	16.37
Leucocytes ($\times 10^3 \mu\text{L}$)	7.99	5.12	6.96	4.86	7.07	4.90
Lymphocytes ($\times 10^3 \mu\text{L}$)	0.71	0.36	1.53	5.60	1.43	5.28
Neutrophils ($\times 10^3 \mu\text{L}$)	6.42	5.23	5.05	3.80	5.20	4.00
Monocytes ($\times 10^3 \mu\text{L}$)	0.50	0.29	0.61	0.81	0.60	0.77
Eosinophils ($\times 10^3 \mu\text{L}$)	0.03	0.09	0.02	0.07	0.02	0.07
Basophils ($\times 10^3 \mu\text{L}$)	0.14	0.74	0.02	0.04	0.03	0.25
Erythrocytes ($\times 10^6 \mu\text{L}$)	4.40	0.68	4.72	0.70	4.69	0.70
Hemoglobin (g/dL)	13.41	1.87	13.79	1.83	13.75	1.83
Hematocrit (%)	39.62	7.00	41.52	5.36	41.30	5.59
M.C.V. (fL)	90.78	9.99	87.81	7.25	88.15	7.64
Platelets ($\times 10^3 \mu\text{L}$)	163.52	80.76	186.62	79.00	184.02	79.42
Glucose (mg/dL)	144.83	62.46	138.84	62.98	139.50	62.86
Urea (mg/dL)	71.72	40.20	43.57	26.37	46.77	29.60
Creatinine (mg/dL)	1.37	0.76	1.37	5.33	1.37	5.02
Estimated glomerular filtrate (CKD-EPI 2009) (mL/min/1.73 m ²)	56.88	23.86	73.79	20.22	71.84	21.33
Sodium (mmol/L)	135.85	4.04	134.91	3.39	135.02	3.48
Potassium (mmol/L)	4.14	0.58	4.20	2.82	4.19	2.66
Chloride (mmol/L)	101.85	3.77	101.13	5.70	101.21	5.52
Total bilirubin (mg/dL)	0.76	0.42	1.64	9.60	1.54	9.05
Aspartate aminotransferase (AST/GOT) (U/L)	72.77	100.68	48.48	36.15	51.20	48.22

Table 4. Cont.

	Deceased (40 Patients, 11%)		Alive (323 Patients, 89%)		Global (363 Patients)	
	Mean	SD	Mean	SD	Mean	SD
Alanine aminotransferase (ALT/GPT) (U/L)	51.83	75.99	42.17	40.45	43.21	45.59
Lactate dehydrogenase (LDH) (U/L)	411.46	201.74	325.42	128.12	335.40	140.97
Albumin (g/dL)	3.53	0.40	3.82	0.38	3.79	0.39
C-reactive protein (mg/dL)	115.47	70.71	78.23	69.85	82.22	70.79
Procalcitonin (ng/mL)	0.81	2.93	0.25	0.92	0.31	1.30
D-dimer (ng/mL)	1500.51	2568.30	1155.00	2318.82	1192.10	2345.08
Fibrinogen (Derived) (mg/dL)	688.53	171.63	667.34	161.70	669.71	162.71
Ratio (TP)	2.03	4.62	1.90	7.06	1.91	6.83
Ratio (TTPA)	29.83	3.87	30.78	5.24	30.67	5.11
pH	7.41	0.07	7.43	0.07	7.42	0.07
pCO ₂ (mmHg)	36.91	9.16	34.75	7.05	35.05	7.39
pO ₂ (mmHg)	64.20	24.83	60.80	23.28	61.27	23.48
Bicarbonate (CO ₃ H [−]) (mmol/L)	23.01	4.61	23.07	3.10	23.06	3.34
FIO ₂ (%)	26.86	10.08	24.18	11.32	24.57	11.16
pO ₂ /FIO ₂	256.10	86.47	279.83	109.36	276.21	106.32
Arterial/alveolar O ₂ gradient (mmHg)	72.19	55.01	65.13	76.85	66.11	74.11
Lactate (mmol/L)	1.89	1.03	1.72	1.16	1.74	1.15
Days in hospital	20.59	17.44	14.92	18.34	15.54	18.30
Days elapsed between PCR and hospital admission	3.13	5.80	4.00	4.75	3.91	4.87
Number of consolidations	2.57	1.27	2.47	1.55	2.48	1.51
Number of opacities	3.41	1.76	3.18	1.69	3.20	1.69
Curb 65 Scale value	1.63	0.81	0.74	0.81	0.84	0.86
Temperature (°C)	36.87	1.08	36.70	0.94	36.72	0.96
Systolic blood pressure (SBP)	135.41	23.07	131.94	22.62	132.32	22.66
Diastolic blood pressure (DBP)	63.31	16.10	71.92	16.58	70.98	16.72
Heart rate	87.56	19.81	90.90	17.51	90.54	17.78
Respiratory rate	25.32	7.67	23.01	6.95	23.34	7.07
Glasgow Coma Scale value	14.43	1.32	14.86	0.78	14.81	0.87
SOFA scale Value	4.45	1.23	4.07	1.63	4.17	1.54
APACHE-II scale value	12.80	4.69	8.31	3.23	9.53	4.16
Number of COVID-19 vaccine doses	1.00	0.00	1.29	0.46	1.26	0.45

Several ML methods were used to identify risk patterns in a population with confirmed COVID-19 infection. The goal was to determine which algorithm provided the best predictive results. This study presents performance metrics for these ML methods, including balanced accuracy, recall, specificity, precision, Mathew's correlation coefficient (MCC), F1 score, kappa, AUC, and degenerated Youden's index (DYI), as shown in Tables 5 and 6.

The data clearly indicate that the proposed method, RF, is the one with the highest acquisition and recall value. The RF algorithm consistently provides a positive predictive value of greater than 95%, demonstrating consistent performance.

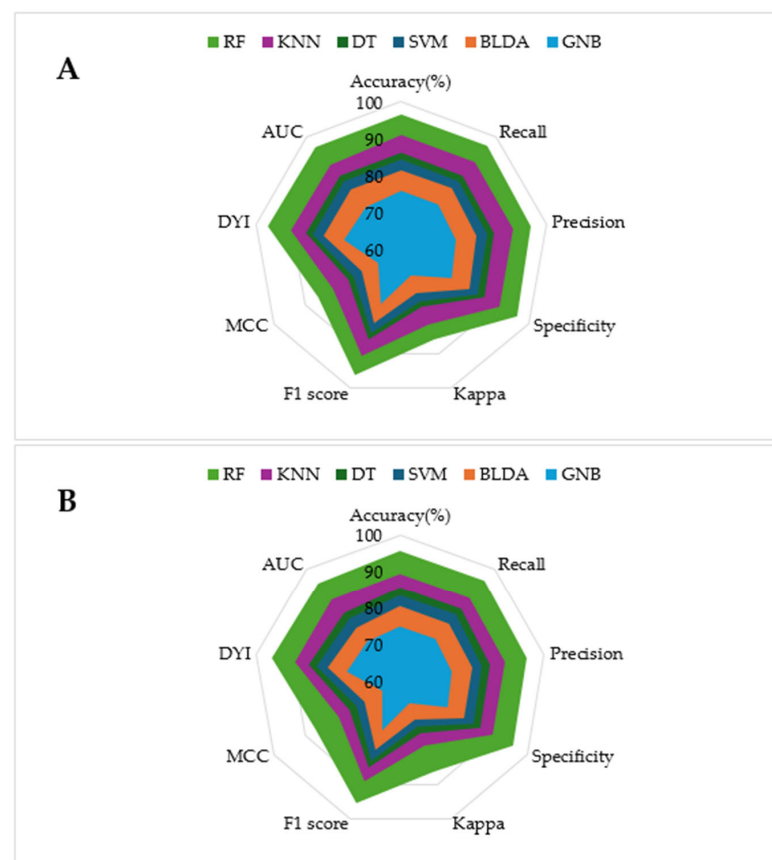
Table 5. The final results of accuracy, recall, precision, specificity and kappa are shown in the table.

	Accuracy (%)	Recall	Precision	Specificity	Kappa
SVM	83.32	83.42	82.73	83.22	73.39
BLDA	80.37	80.47	79.78	80.27	71.41
DT	85.52	85.62	84.98	85.42	75.46
GNB	74.82	74.91	74.35	74.73	66.67
KNN	89.24	89.31	89.14	89.18	79.01
RF	95.83	95.92	95.15	95.73	86.32

Table 6. The final results of F1 score, MCC, DYI, AUC and AUC (%) are shown in the table.

	F1 Score	MCC	DYI	AUC	AUC (%)
SVM	83.07	73.93	83.32	82	0.82
BLDA	80.12	71.31	80.37	79	0.79
DT	85.29	75.94	85.52	84	0.84
GNB	74.63	65.83	74.82	74	0.74
KNN	89.23	79.48	89.24	89	0.89
RF	95.53	86.83	95.83	95	0.95

The model training subset and the test subset both present high scores for all the metrics, with slightly lower scores for the test subset. This consistency is attributed to the algorithm reaching an optimal level of training without overfitting or underfitting. As shown in the radar plots in Figure 2, the RF model covers a larger area compared to the other methods tested, which is an example of a well-balanced model with high generalization capability, ensuring accurate outputs for new inputs.

**Figure 2.** The figure shows the radar plots comparing multiple performance metrics for specific models. Each axis represents a key performance metric. The radial axis ranges from 0 (minimum relevance) to 100% (maximum relevance). (A) Training phase. (B) Test phase.

Furthermore, the ROC curve was generated by plotting the sensitivity and specificity measures for each threshold to evaluate the classification capabilities of the different ML algorithms. The results are displayed in Figure 3. Again, the proposed RF-based system covers a larger area, indicating superior predictive accuracy.

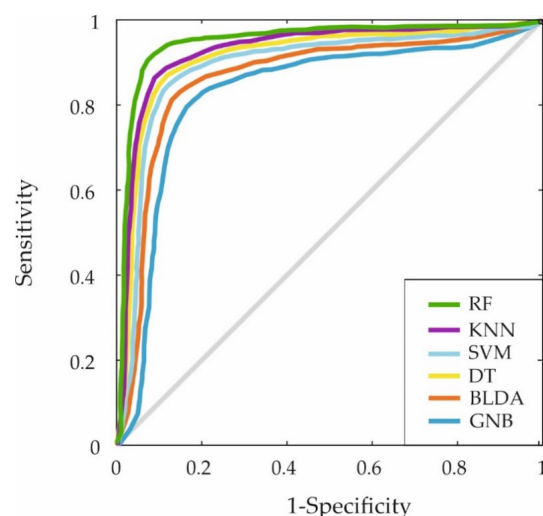


Figure 3. The figure shows the ROC curves for the six machine learning predictors. This curve shows the trade-off between sensitivity (true positive rate) and 1-specificity (false positive rate) for the ML models mentioned: RF, KNN, SVM, DT, BLDA and GNB.

Figure 4 shows the most clinically relevant parameters contributing to mortality in patients hospitalized for COVID-19 after the first wave and before mass vaccination, according to the proposed RF model. Listed in descending order of relevance, the aforementioned parameters are age, CURB-65 scale on admission, urea, procalcitonin, estimated glomerular filtrate, lymphocytes, C-reactive protein, acute renal failure, albumin, PRONO sessions, oxygen requirements, Glasgow Coma Scale, ischemic heart disease, invasive mechanical ventilation (IMV), creatinine, shock, noninvasive mechanical ventilation (NIMV), APACHE-II scale and chronic renal failure.

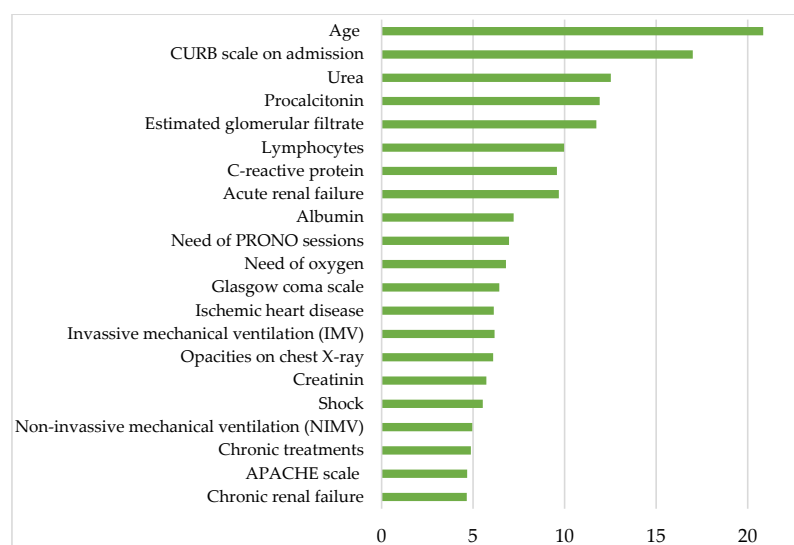


Figure 4. The histogram shows the 21 most relevant parameters contributing to the mortality in COVID-19 hospitalized patients on the y-axis. The x-axis represents the statistical weight or importance assigned of each parameter in predicting mortality. More details about the units in which the parameters are presented in Tables 1–4.

4. Discussion

Despite significant progress in the fight against COVID-19 and the time we have been coexisting with the virus after the global emergency, the pandemic remains a global concern. The emergence of new variants and the increase in cases during certain periods underscore the need to continue to research and improve our response strategies. In addition, the pandemic has highlighted weaknesses in health systems and the importance of researching more effective interventions and protocols to provide higher-quality care.

In this study, the analysis of the most relevant parameters for predicting the risk of mortality in hospitalized COVID-19 patients is organized into different categories for clarity and understanding. These categories include demographic factors, clinical features observed on admission, laboratory parameters, and variables monitored during hospitalization. This categorization highlights the complementary roles these parameters play in predicting mortality and enhances the study's contribution to clinical decision making.

The most relevant parameters for predicting the risk of mortality in this group of hospitalized patients with COVID-19 during the period from the second to the fourth wave of the pandemic were identified. At that time, there was some immunity following infections in previous waves, and the population had not yet been vaccinated with more than one dose or had not yet been vaccinated at all. Among the key variables identified, the most powerful parameters were those that could be obtained at the time of hospital admission, either previous comorbidities or the clinical features assessed in the emergency department.

Thus, the strongest predictor among demographic factors is age, a conclusion common to many studies [45,46]. It should be emphasized that, although the majority of the patients in this cohort were older adults, this does not diminish the importance of age as a predictive factor. Rather than merely reflecting the higher mortality commonly associated with older populations, age itself plays a crucial role in the progression and outcome of COVID-19. This is supported by extensive evidence in the medical literature highlighting the physiological and immunological changes associated with aging that increase susceptibility to severe disease. To ensure the robustness of the analysis, statistical techniques were employed, including multivariate models, to isolate the independent contribution of age and confirm its predictive power while accounting for the influence of other clinical and pathologic parameters.

Similar trends were observed in SARS-CoV and MERS, where older age was also a significant risk factor for mortality, although the overall case fatality rates differed markedly, SARS-CoV at ~10% and MERS-COV at ~35%, compared to COVID-19's lower fatality rate but higher transmissibility [47]. However, Philipps and Carver [48] concluded that it is not so much the age itself but what comes with advanced age that is a risk factor for mortality, namely comorbidities, chronic treatments and social situation. The results of this study support their theory, in that some of the factors that emerge as strong predictors of mortality are related to chronic diseases, such as ischemic heart disease and chronic renal failure.

The CURB-65 scale, a hallmark in the results, includes age as a variable. This scale has been evaluated in other studies as an adequate predictor of mortality in patients prior to hospitalization with COVID-19 pneumonia [49–51], and it is also commonly used to determine the severity in patients with any type of pneumonia. In addition to demographic factors, clinical features such as those included in the CURB-65 scale, such as urea levels or low level of consciousness, have considerable predictive power for mortality. In this study, the CURB-65 scale is a good parameter, and since it can be determined at the time of admission, this information can be used to perform rapid and efficient triage. Furthermore, a comparison with other viruses reveals that this scale, which focuses on clinical severity, has also been used to effectively manage pneumonia-like symptoms [52,53].

Laboratory parameters obtained at the time of hospitalization, such as urea levels, C-reactive protein, procalcitonin, albumin and/or lymphocyte count, are also good predictors of mortality according to the model used. Mohammadi et al. studied more than 1000 patients and concluded that elevated creatinine and urea levels were associated with poor prognosis in patients with COVID-19 [54]. Other studies, such as those of Al-Shajlawi et al. [55] or Singh and Singh [56], also include elevated C-reactive protein and procalcitonin levels, low albumin levels and low lymphocyte count as good predictors of mortality. Elevated urea levels, together with low lymphocyte levels, may be related to a severe systemic inflammatory response and consequent inflammation. C-reactive protein and procalcitonin are known markers of inflammation and sepsis, and their elevation reflects the severity of the inflammatory response. On the other hand, albumin, a protein produced by the liver, is normally decreased in states of chronic inflammation and physiological stress due to its role in the acute-phase response. A low lymphocyte count, or lymphopenia, may indicate immune dysfunction, which is common in severe viral infections such as COVID-19. All of these data would be consistent with a state of advanced inflammation and organ damage, explaining the association with an increased risk of mortality.

Moreover, this study shows that certain in-hospital monitoring variables, including acute renal failure and shock, have predicting significance. These belong to the category of monitoring parameters, which track the disease progression during hospitalization and provide critical insights into mortality risk. The study by Yüksel et al. [57] associates the aforementioned imbalances in laboratory parameters with the development of acute renal failure. At the same time, studies such as that by Qureshi et al. [58] have shown that non-surviving patients have a higher incidence of septic shock, among other pathologies. This underscores the importance of continuous monitoring and early intervention in COVID-19 patients with these alterations.

Additional monitored variables, such as the need for oxygen therapy, prone positioning sessions (PRONO), or mechanical ventilation, have also shown predictive significance for mortality risk as the disease progresses. Studies show they hold predictive significance for mortality risk as the disease progresses during ICU stay [59]. In turn, the APACHE-II scale has been shown to be a good predictor of mortality in some studies, although there are contradictions in this regard [59–62]. These include physiologic and biochemical parameters assessed on admission to the ICU that reflect the systemic response to severe physiologic stress common in severe cases of COVID-19. The higher relevance of APACHE-II in our study may be related to its ability to better capture multiorgan dysfunction in the early stages, providing a more robust prediction of mortality.

The data management approach used in this study involved an RF machine learning model, which was based on either bootstrap aggregation or bagging. This model has exhibited several advantageous characteristics, including high generalizability, stability and interpretability, along with a low risk of overfitting. Other studies have used the same method to analyze diseases, such as cardiovascular disease [63], interstitial lung disease [64], breast [65] or pancreatic cancer [66], autoimmune diseases [63] or neurological diseases [67] or neurodegenerative diseases [68]. The model can, therefore, be used as a strategic tool to quickly address clinical questions with relatively little data.

The application of the aforementioned model in the analysis of the variables with the most significant influence on the prediction of mortality in the cohort of hospitalized patients with COVID-19 during the second and third waves reveals a discrepancy in the results compared to those of analogous studies conducted in the first wave [11–13,69–71]. In these studies, patients had not yet developed any form of immunity. In contrast, the present study demonstrates that once a population achieves a certain level of immunity, additional predictive variables can be obtained at the time of admission. Furthermore,

during the patient's hospital stay, variables related to the stay itself were identified, and these variables can be used in decision making.

5. Conclusions

After the first wave of the recent COVID-19 pandemic, the size of the population that has overcome the infection and developed natural immunity is increasing. Therefore, it is crucial to study how natural immunity affects the evolution of the disease in order to improve general knowledge about it and improve preparedness for potential new health challenges.

The present study demonstrates that predictive variables for mortality risk, and, therefore, severity, can be identified in a given population at the time of hospital admission using an RF machine learning model. In this cohort, the CURB-65 scale and age were the strongest predictors of mortality, along with other laboratory tests. These results underscore the need to emphasize not only the early patient assessment but also the personalized analysis of the population and situation, given their capacity for change.

The present study shows that the increase in the number of immunized patients, even in the absence of herd immunity, generates a significant modification in the evolution of the disease, as well as the mortality risk parameters. This underscores the necessity for a more customized approach to patient care, employing distinct protocols tailored to specific groups and temporal periods. In this regard, the employment of machine learning tools emerges as a highly valuable and efficient resource.

Author Contributions: Conceptualization, M.Q. and J.B.; methodology, M.Q., J.B., A.M.T. and J.M.; software, A.M.T. and J.M.; validation, M.Q., J.B. and J.M.; formal analysis, M.Q., J.B., A.M.T. and J.M.; investigation, M.Q., J.B., A.M.T. and J.M.; writing—original draft, M.Q., J.B., A.M.T. and J.M.; writing—review and editing, J.B., A.M.T. and J.M.; visualization, M.Q., A.M.T. and J.M.; supervision, J.M.; project administration, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: Research was funded by the Institute of Technology (University of Castilla-La Mancha, Spain), the Río Hortega University Hospital (Valladolid, Spain), and the University of Valladolid [60-195041] (Spain).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Río Hortega University Hospital (Ref.: 23-PI186).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets employed and analyzed in the current study are accessible upon reasonable request from the corresponding author. We do not have the patients' permission to publish the data collected in this study in open access.

Acknowledgments: Chair of Artificial Intelligence, sponsored by Bayer.

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

Abbreviations

The following abbreviations are used in this manuscript:

COVID-19	Coronavirus Disease 19
ICU	Intensive Care Unit
PCR	Polymerase Chain Reaction
AI	Artificial Intelligence
ML	Machine learning
RF	Random Forest

CURB	Confusion, Urea, Respiratory rate, Blood pressure
SOFA	Sequential Organ Failure Assessment
APACHE-II	Acute Physiology and Chronic Health Evaluation
MCV	Mean Corpuscular Volume
PT	Prothrombin time
INR	International normalized Ratio
aPTT	Activated Partial Thromboplastin Time
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
ALT/GPT	Alanine Aminotransferase/Glutamate Pyruvate Transaminase
AST/GOT	Aspartate Aminotransferase/Glutamate Oxaloacetate Transaminase
GGT	Gamma-Glutamyl Transferase
LDH	Lactate Dehydrogenase
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
GNB	Gaussian Naïve Bayes
KNN	K-Nearest Neighbors
BLDA	Bayesian Linear Discriminant Analysis
SMV	Support Vector Machine
DT	Decision Tree
COPD	Chronic Obstructive Pulmonary Disease
IMV	Invasive Mechanical Ventilation
HFNC	High-Flow Nasal Cannula
NIV	Non-Invasive Mechanical Ventilation
ECMO	Extracorporeal Membrane Oxygenation
ARDS	Acute Respiratory Distress Syndrome
SD	Standard Deviation
CRP	C-Reactive Protein
SBP	Systolic Blood Pressure
DBP	Diastolic Blood Pressure
MCC	Mathew's Correlation Coefficient
DYI	Degenerated Youden's Index

References

1. Suryawanshi, Y.N.; Biswas, D.A. Herd Immunity to Fight Against COVID-19: A Narrative Review. *Cureus* **2023**, *15*. [\[CrossRef\]](#)
2. Clemente-Suárez, V.J.; Hormeño-Holgado, A.; Jiménez, M.; Benitez-Agudelo, J.C.; Navarro-Jiménez, E.; Perez-Palencia, N.; Maestre-Serrano, R.; Laborde-Cárdenas, C.C.; Tornero-Aguilera, J.F. Dynamics of Population Immunity Due to the Herd Effect in the COVID-19 Pandemic. *Vaccines* **2020**, *8*, 236. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Fine, P.; Eames, K.; Heymann, D.L. "Herd Immunity": A Rough Guide. *Clin. Infect. Dis.* **2011**, *52*, 911–916. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Hussain, A.; Yang, H.; Zhang, M.; Liu, Q.; Alotaibi, G.; Irfan, M.; He, H.; Chang, J.; Liang, X.-J.; Weng, Y.; et al. mRNA Vaccines for COVID-19 and Diverse Diseases. *J. Control. Release* **2022**, *345*, 314–333. [\[CrossRef\]](#)
5. Zhang, M.; Hussain, A.; Yang, H.; Zhang, J.; Liang, X.-J.; Huang, Y. mRNA-Based Modalities for Infectious Disease Management. *Nano Res.* **2023**, *16*, 672–691. [\[CrossRef\]](#)
6. Omer, S.B.; Yildirim, I.; Forman, H.P. Herd Immunity and Implications for SARS-CoV-2 Control. *JAMA* **2020**, *324*, 2095. [\[CrossRef\]](#)
7. Mathieu, E.; Ritchie, H.; Rodés-Guirao, L.; Appel, C.; Gavrilov, D.; Giattino, C.; Hasell, J.; Macdonald, B.; Dattani, S.; Beltekian, D.; et al. Total COVID-19 Vaccine Doses Administered. In *COVID-19 Pandemic*; Data Adapted from Official Data Collated by Our World in Data; World Health Organization: Geneva, Switzerland, 2020.
8. Ruan, Q.; Yang, K.; Wang, W.; Jiang, L.; Song, J. Clinical Predictors of Mortality Due to COVID-19 Based on an Analysis of Data of 150 Patients from Wuhan, China. *Intensive Care Med.* **2020**, *46*, 846–848. [\[CrossRef\]](#)
9. Fajgenbaum, D.C.; June, C.H. Cytokine Storm. *N. Engl. J. Med.* **2020**, *383*, 2255–2273. [\[CrossRef\]](#)
10. Iftimie, S.; López-Azcona, A.F.; Vallverdú, I.; Hernández-Flix, S.; De Febrer, G.; Parra, S.; Hernández-Aguilera, A.; Riu, F.; Joven, J.; Andreychuk, N.; et al. First and Second Waves of Coronavirus Disease-19: A Comparative Study in Hospitalized Patients in Reus, Spain. *PLoS ONE* **2021**, *16*, e0248029. [\[CrossRef\]](#)

11. Chen, N.; Zhou, M.; Dong, X.; Qu, J.; Gong, F.; Han, Y.; Qiu, Y.; Wang, J.; Liu, Y.; Wei, Y.; et al. Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study. *Lancet* **2020**, *395*, 507–513. [\[CrossRef\]](#)
12. Soriano, V.; Ganado-Pinilla, P.; Sanchez-Santos, M.; Gómez-Gallego, F.; Barreiro, P.; De Mendoza, C.; Corral, O. Main Differences between the First and Second Waves of COVID-19 in Madrid, Spain. *Int. J. Infect. Dis.* **2021**, *105*, 374–376. [\[CrossRef\]](#)
13. Buttenschøn, H.N.; Lynggaard, V.; Sandbøl, S.G.; Glassou, E.N.; Haagerup, A. Comparison of the Clinical Presentation across Two Waves of COVID-19: A Retrospective Cohort Study. *BMC Infect. Dis.* **2022**, *22*, 423. [\[CrossRef\]](#) [\[PubMed\]](#)
14. The MO-COVID19 Working Group; Coloretti, I.; Farinelli, C.; Biagioni, E.; Gatto, I.; Munari, E.; Dall'Ara, L.; Busani, S.; Meschiari, M.; Tonelli, R.; et al. Critical COVID-19 Patients through First, Second, and Third Wave: Retrospective Observational Study Comparing Outcomes in Intensive Care Unit. *J. Thorac. Dis.* **2023**, *15*, 3218–3227. [\[CrossRef\]](#) [\[PubMed\]](#)
15. San Martín-López, J.V.; Mesa, N.; Bernal-Bello, D.; Morales-Ortega, A.; Rivilla, M.; Guerrero, M.; Calderón, R.; Farfán, A.I.; Rivas, L.; Soria, G.; et al. Seven Epidemic Waves of COVID-19 in a Hospital in Madrid: Analysis of Severity and Associated Factors. *Viruses* **2023**, *15*, 1839. [\[CrossRef\]](#)
16. Wu, Z.; Harrich, D.; Li, Z.; Hu, D.; Li, D. The Unique Features of SARS-CoV-2 Transmission: Comparison with SARS-CoV, MERS-CoV and 2009 H1N1 Pandemic Influenza Virus. *Rev. Med. Virol.* **2021**, *31*, e2171. [\[CrossRef\]](#)
17. Zhou, H.; Yang, J.; Zhou, C.; Chen, B.; Fang, H.; Chen, S.; Zhang, X.; Wang, L.; Zhang, L. A Review of SARS-CoV2: Compared with SARS-CoV and MERS-CoV. *Front. Med.* **2021**, *8*, 628370. [\[CrossRef\]](#)
18. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*; Morgan Kaufmann Publishers: Burlington, VT, USA, 2022.
19. Hesami, M.; Alizadeh, M.; Jones, A.M.P.; Torkamaneh, D. Machine Learning: Its Challenges and Opportunities in Plant System Biology. *Appl. Microbiol. Biotechnol.* **2022**, *106*, 3507–3530. [\[CrossRef\]](#)
20. Haug, C.J.; Drazen, J.M. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *N. Engl. J. Med.* **2023**, *388*, 1201–1208. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Wu, H.-T.; Liao, C.-C.; Peng, C.-F.; Lee, T.-Y.; Liao, P.-H. Exploring the Application of Machine Learning to Identify the Correlations between Phthalate Esters and Disease: Enhancing Nursing Assessments. *Health Inf. Sci. Syst.* **2024**, *13*, 10. [\[CrossRef\]](#)
22. Rueda, R.; Fabello, E.; Silva, T.; Genzor, S.; Mizera, J.; Stanke, L. Machine Learning Approach to Flare-up Detection and Clustering in Chronic Obstructive Pulmonary Disease (COPD) Patients. *Health Inf. Sci. Syst.* **2024**, *12*, 50. [\[CrossRef\]](#)
23. Barough, S.S.; Safavi-Naini, S.A.A.; Siavoshi, F.; Tamimi, A.; Ilkhani, S.; Akbari, S.; Ezzati, S.; Hatamabadi, H.; Pourhoseingholi, M.A. Generalizable Machine Learning Approach for COVID-19 Mortality Risk Prediction Using on-Admission Clinical and Laboratory Features. *Sci. Rep.* **2023**, *13*, 2399. [\[CrossRef\]](#)
24. Kourmpanis, N.; Liaskos, J.; Zoulias, E.; Mantas, J. Predicting Mortality in COVID-19 Patients Using 6 Machine Learning Algorithms. In *Studies in Health Technology and Informatics*; Mantas, J., Gallos, P., Zoulias, E., Hasman, A., Househ, M.S., Charalampidou, M., Magdalinou, A., Eds.; IOS Press: Nieuwe Hemweg, Amsterdam, The Netherlands, 2023; ISBN 978-1-64368-400-0.
25. Alballa, N.; Al-Turaiqi, I. Machine Learning Approaches in COVID-19 Diagnosis, Mortality, and Severity Risk Prediction: A Review. *Inform. Med. Unlocked* **2021**, *24*, 100564. [\[CrossRef\]](#)
26. Liu, L.; Song, W.; Patil, N.; Sainlaire, M.; Jasuja, R.; Dykes, P.C. Predicting COVID-19 Severity: Challenges in Reproducibility and Deployment of Machine Learning Methods. *Int. J. Med. Inform.* **2023**, *179*, 105210. [\[CrossRef\]](#)
27. Sharifi-Kia, A.; Nahvijou, A.; Sheikhtaheri, A. Machine Learning-Based Mortality Prediction Models for Smoker COVID-19 Patients. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 129. [\[CrossRef\]](#)
28. Navlakha, S.; Morjaria, S.; Perez-Johnston, R.; Zhang, A.; Taur, Y. Projecting COVID-19 Disease Severity in Cancer Patients Using Purposefully-Designed Machine Learning. *BMC Infect. Dis.* **2021**, *21*, 391. [\[CrossRef\]](#)
29. Jung, Y.J.; Ahn, J.; Park, S.; Sun, J.-M.; Lee, S.-H.; Ahn, J.S.; Ahn, M.-J.; Cho, S.Y.; Jung, H.A. Machine Learning Prediction of the Case-Fatality of COVID-19 and Risk Factors for Adverse Outcomes in Patients with Non-Small Cell Lung Cancer. *Transl. Cancer Res.* **2024**, *13*, 2587–2595. [\[CrossRef\]](#)
30. Subramanian, D.; Vittala, A.; Chen, X.; Julien, C.; Acosta, S.; Rusin, C.; Allen, C.; Rider, N.; Starosolski, Z.; Annapragada, A.; et al. Stratification of Pediatric COVID-19 Cases Using Inflammatory Biomarker Profiling and Machine Learning. *J. Clin. Med.* **2023**, *12*, 5435. [\[CrossRef\]](#)
31. Cerbulescu, T.; Ignuta, F.; Rayudu, U.S.; Afra, M.; Rosca, O.; Vlad, A.; Loredana, S. Inflammatory Markers and Severity in COVID-19 Patients with Clostridioides Difficile Co-Infection: A Retrospective Analysis Including Subgroups with Diabetes, Cancer, and Elderly. *Biomedicines* **2025**, *13*, 227. [\[CrossRef\]](#)
32. Schonlau, M.; Zou, R.Y. The Random Forest Algorithm for Statistical Learning. *Stata J. Promot. Commun. Stat. Stata* **2020**, *20*, 3–29. [\[CrossRef\]](#)
33. Rhodes, J.S.; Cutler, A.; Moon, K.R. Geometry- and Accuracy-Preserving Random Forest Proximities. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10947–10959. [\[CrossRef\]](#)
34. Jayachitra, S.; Prasanth, A. Multi-Feature Analysis for Automated Brain Stroke Classification Using Weighted Gaussian Naïve Bayes Classifier. *J. Circuit syst. Comp.* **2021**, *30*, 2150178. [\[CrossRef\]](#)

35. Balaji, V.R.; Suganthi, S.T.; Rajadevi, R.; Krishna Kumar, V.; Saravana Balaji, B.; Pandiyan, S. Skin Disease Detection and Segmentation Using Dynamic Graph Cut Algorithm and Classification through Naive Bayes Classifier. *Measurement* **2020**, *163*, 107922. [\[CrossRef\]](#)
36. Uddin, S.; Haque, I.; Lu, H.; Moni, M.A.; Gide, E. Comparative Performance Analysis of K-Nearest Neighbour (KNN) Algorithm and Its Different Variants for Disease Prediction. *Sci. Rep.* **2022**, *12*, 6256. [\[CrossRef\]](#)
37. Zhang, S.; Li, X.; Zong, M.; Zhu, X.; Wang, R. Efficient kNN Classification with Different Numbers of Nearest Neighbors. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 1774–1785. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Zhou, W.; Liu, Y.; Yuan, Q.; Li, X. Epileptic Seizure Detection Using Lacunarity and Bayesian Linear Discriminant Analysis in Intracranial EEG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 3375–3381. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Yuan, S.; Zhou, W.; Li, J.; Wu, Q. Sparse Representation-Based EMD and BLDA for Automatic Seizure Detection. *Med. Biol. Eng. Comput.* **2017**, *55*, 1227–1238. [\[CrossRef\]](#)
40. Pérez, A.; Larrañaga, P.; Inza, I. Supervised Classification with Conditional Gaussian Networks: Increasing the Structure Complexity from Naive Bayes. *Int. J. Approx. Reason.* **2006**, *43*, 1–25. [\[CrossRef\]](#)
41. Patel, H.H.; Prajapati, P. Study and Analysis of Decision Tree Based Classification Algorithms. *Int. J. Comput. Sci. Eng.* **2018**, *6*, 74–78. [\[CrossRef\]](#)
42. Charbuty, B.; Abdulazeez, A. Classification Based on Decision Tree Algorithm for Machine Learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [\[CrossRef\]](#)
43. Jiang, C.; Lv, W.; Li, J. Protein-Protein Interaction Sites Prediction Using Batch Normalization Based CNNs and Oversampling Method Borderline-SMOTE. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2023**, *20*, 2190–2199. [\[CrossRef\]](#)
44. Alshammari, T.S. Applying Machine Learning Algorithms for the Classification of Sleep Disorders. *IEEE Access* **2024**, *12*, 36110–36121. [\[CrossRef\]](#)
45. Andrew, M.K.; Godin, J.; LeBlanc, J.; Boivin, G.; Valiquette, L.; McGeer, A.; McElhaney, J.E.; Hatchette, T.F.; ElSherif, M.; MacKinnon-Cameron, D.; et al. Older Age and Frailty Are Associated with Higher Mortality but Lower ICU Admission with COVID-19. *Can. Geriatr. J.* **2022**, *25*, 183–196. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Conway, B.J.; Kim, J.W.; Brousseau, D.C.; Conroy, M. Heart Disease, Advanced Age, Minority Race, and Hispanic Ethnicity Are Associated with Mortality in COVID-19 Patients. *WMJ Off. Publ. State Med. Soc. Wis.* **2021**, *120*, 152–155.
47. Petrosillo, N.; Viceconte, G.; Ergonul, O.; Ippolito, G.; Petersen, E. COVID-19, SARS and MERS: Are They Closely Related? *Clin. Microbiol. Infect.* **2020**, *26*, 729–734. [\[CrossRef\]](#)
48. Phillips, S.P.; Carver, L.F. Greatest Risk Factor for Death from COVID-19: Older Age, Chronic Disease Burden, or Place of Residence? Descriptive Analysis of Population-Level Canadian Data. *Int. J. Environ. Res. Public Health* **2023**, *20*, 7181. [\[CrossRef\]](#)
49. Satici, C.; Demirkol, M.A.; Sargin Altunok, E.; Gursoy, B.; Alkan, M.; Kamat, S.; Demirok, B.; Surmeli, C.D.; Calik, M.; Cavus, Z.; et al. Performance of Pneumonia Severity Index and CURB-65 in Predicting 30-Day Mortality in Patients with COVID-19. *Int. J. Infect. Dis.* **2020**, *98*, 84–89. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Bradley, J.; Sbaih, N.; Chandler, T.R.; Furmanek, S.; Ramirez, J.A.; Cavallazzi, R. Pneumonia Severity Index and CURB-65 Score Are Good Predictors of Mortality in Hospitalized Patients with SARS-CoV-2 Community-Acquired Pneumonia. *Chest* **2022**, *161*, 927–936. [\[CrossRef\]](#)
51. Alegría-Baños, J.A.; Rosas-Alvarado, M.A.; Jiménez-López, J.C.; Juárez-Muciño, M.; Méndez-Celis, C.A.; Enríquez-De Los Santos, S.T.; Valdez-Vázquez, R.R.; Prada-Ortega, D. Sociodemographic, Clinical and Laboratory Characteristics and Risk Factors for Mortality of Hospitalized COVID-19 Patients at Alternate Care Site: A Latin American Experience. *Ann. Med.* **2023**, *55*, 2224049. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Ilg, A.; Moskowitz, A.; Konanki, V.; Patel, P.V.; Chase, M.; GrosseStreuer, A.V.; Donnino, M.W. Performance of the CURB-65 Score in Predicting Critical Care Interventions in Patients Admitted with Community-Acquired Pneumonia. *Ann. Emerg. Med.* **2019**, *74*, 60–68. [\[CrossRef\]](#)
53. Kaal, A.G.; Op De Hoek, L.; Hochheimer, D.T.; Brouwers, C.; Wiersinga, W.J.; Snijders, D.; Rensing, K.L.; Van Dijk, C.E.; Steyerberg, E.W.; Van Nieuwkoop, C. Outcomes of Community-Acquired Pneumonia Using the Pneumonia Severity Index versus the CURB-65 in Routine Practice of Emergency Departments. *ERJ Open Res.* **2023**, *9*, 00051–02023. [\[CrossRef\]](#)
54. Mohammadi, Z.; Faghih Dinevari, M.; Vahed, N.; Ebrahimi Bakhtavar, H.; Rahmani, F. Clinical and Laboratory Predictors of COVID-19-Related In-Hospital Mortality; a Cross-Sectional Study of 1000 Cases. *Arch. Acad. Emerg. Med.* **2022**, *10*, e49. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Al-Shajlawi, M.; Alsayed, A.R.; Abazid, H.; Awajan, D.; Al-Imam, A.; Basheti, I. Using Laboratory Parameters as Predictors for the Severity and Mortality of COVID-19 in Hospitalized Patients. *Pharm. Pract.* **2022**, *20*, 2721. [\[CrossRef\]](#)
56. Singh, S.; Singh, K. Blood Urea Nitrogen/Albumin Ratio and Mortality Risk in Patients with COVID-19. *Indian J. Crit. Care Med.* **2022**, *26*, 626–631. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Yüksel, E.; Arac, S.; Deniz Altıntaş, D. Predictors of Mortality in COVID-19 Induced Acute Kidney Injury. *Ther. Apher. Dial.* **2022**, *26*, 897–907. [\[CrossRef\]](#) [\[PubMed\]](#)

58. Qureshi, M.A.; Toori, K.U. Raja Mobeen Ahmed Predictors of Mortality in COVID-19 Patients: An Observational Study. *Pak. J. Med. Sci.* **2022**, *39*. [\[CrossRef\]](#)
59. Muñoz Lezcano, S.; Armengol De La Hoz, M.Á.; Corbi, A.; López, F.; García, M.S.; Reiz, A.N.; González, T.F.; Zlatkov, V.Y. Predictors of Mechanical Ventilation and Mortality in Critically Ill Patients with COVID-19 Pneumonia. *Med. Intensiv. (Engl. Ed.)* **2023**, S2173572723001303. [\[CrossRef\]](#)
60. Arunachala, S.; Parthasarathi, A.; Basavaraj, C.K.; Kaleem Ullah, M.; Chandran, S.; Venkataraman, H.; Vishwanath, P.; Ganguly, K.; Upadhyay, S.; Mahesh, P.A. The Validity of the ROX Index and APACHE II in Predicting Early, Late, and Non-Responses to Non-Invasive Ventilation in Patients with COVID-19 in a Low-Resource Setting. *Viruses* **2023**, *15*, 2231. [\[CrossRef\]](#)
61. Vandenbrande, J.; Verbrugge, L.; Bruckers, L.; Geebelen, L.; Geerts, E.; Callebaut, I.; Gruyters, I.; Heremans, L.; Dubois, J.; Stessel, B. Validation of the Acute Physiology and Chronic Health Evaluation (APACHE) II and IV Score in COVID-19 Patients. *Crit. Care Res. Pract.* **2021**, *2021*, 1–9. [\[CrossRef\]](#)
62. Onuk, S.; Sipahioğlu, H.; Karahan, S.; Yeşiltepe, A.; Kuzugüden, S.; Karabulut, A.; Beştepe Dursun, Z.; Akin, A. Cytokine Levels and Severity of Illness Scoring Systems to Predict Mortality in COVID-19 Infection. *Healthcare* **2023**, *11*, 387. [\[CrossRef\]](#)
63. He, J.; Tang, D.; Liu, D.; Hong, X.; Ma, C.; Zheng, F.; Zeng, Z.; Chen, Y.; Du, J.; Kang, L.; et al. Serum Proteome and Metabolome Uncover Novel Biomarkers for the Assessment of Disease Activity and Diagnosing of Systemic Lupus Erythematosus. *Clin. Immunol.* **2023**, *251*, 109330. [\[CrossRef\]](#)
64. Huang, Y.; Ma, S.-F.; Oldham, J.M.; Adegunsoye, A.; Zhu, D.; Murray, S.; Kim, J.S.; Bonham, C.; Strickland, E.; Linderholm, A.L.; et al. Machine Learning of Plasma Proteomics Classifies Diagnosis of Interstitial Lung Disease. *Am. J. Respir. Crit. Care Med.* **2024**, *210*, 444–454. [\[CrossRef\]](#)
65. Safia, S.N. Prediction of Breast Cancer Through Random Forest. *Curr. Med. Imaging* **2023**, *19*, e300922209414. [\[CrossRef\]](#)
66. Sidiropoulos, T.; Dovrolis, N.; Katifelis, H.; Michalopoulos, N.V.; Kokoropoulos, P.; Arkadopoulos, N.; Gazouli, M. Dysbiosis Signature of Fecal Microbiota in Patients with Pancreatic Adenocarcinoma and Pancreatic Intraductal Papillary Mucinous Neoplasms. *Biomedicines* **2024**, *12*, 1040. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Founta, K.; Dafou, D.; Kanata, E.; Sklaviadis, T.; Zanos, T.P.; Gounaris, A.; Xanthopoulos, K. Gene Targeting in Amyotrophic Lateral Sclerosis Using Causality-Based Feature Selection and Machine Learning. *Mol. Med.* **2023**, *29*, 12. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Kelly, J.; Moyeed, R.; Carroll, C.; Luo, S.; Li, X. Blood Biomarker-Based Classification Study for Neurodegenerative Diseases. *Sci. Rep.* **2023**, *13*, 17191. [\[CrossRef\]](#)
69. Khamis, F.; Memish, Z.; Bahrani, M.A.; Dowaiqi, S.A.; Pandak, N.; Bolushi, Z.A.; Salmi, I.A.; Al-Zakwani, I. Prevalence and Predictors of In-Hospital Mortality of Patients Hospitalized with COVID-19 Infection. *J. Infect. Public Health* **2021**, *14*, 759–765. [\[CrossRef\]](#)
70. Aly, M.H.; Rahman, S.S.; Ahmed, W.A.; Alghamedi, M.H.; Al Shehri, A.A.; Alkalkami, A.M.; Hassan, M.H. Indicators of Critical Illness and Predictors of Mortality in COVID-19 Patients. *Infect. Drug Resist.* **2020**, *13*, 1995–2000. [\[CrossRef\]](#)
71. Tian, W.; Jiang, W.; Yao, J.; Nicholson, C.J.; Li, R.H.; Sigurslid, H.H.; Wooster, L.; Rotter, J.I.; Guo, X.; Malhotra, R. Predictors of Mortality in Hospitalized COVID-19 Patients: A Systematic Review and Meta-analysis. *J. Med. Virol.* **2020**, *92*, 1875–1883. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.