Factors Associated With Death in Confirmed Cases of COVID-19 in the State of Rio De Janeiro

Marcella Cini Oliveira (✉ cini.marcella@gmail.com)

Faculty of Medicine, Federal University of Rio de Janeiro

Tatiana de Araujo Eleuterio

Nursing School, State University of Rio de Janeiro

Allan Bruno de Andrade Corrêa

Physics Institute, Federal University of Rio de Janeiro

Lucas Dalsenter Romano da Silva

Department of Preventive Medicine, Institute of Collective Health Studies, Federal University of Rio de Janeiro

Renata Coelho Rodrigues

Faculty of Medicine, Federal University of Rio de Janeiro

Bruna Andrade de Oliveira

Faculty of Medicine, Federal University of Rio de Janeiro

Marlos Melo Martins

Department of Child Neurology, Martagão Gesteira Institute of Childcare and Pediatrics, Federal University of Rio de Janeiro

Carlos Eduardo Raymundo

Institute of Collective Health Studies, Federal University of Rio de Janeiro

Roberto de Andrade Medronho

Faculty of Medicine, Federal University of Rio de Janeiro

Research Article

Keywords: COVID-19, SARS-CoV-2, XGBoost, Machine Learning, Pandemic, Coronavirus infection, Coronavirus death

Posted Date: April 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-435549/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Infectious Diseases on July 16th, 2021. See the published version at <https://doi.org/10.1186/s12879-021-06384-1>.

Factors associated with death in confirmed cases of COVID-19 in the state of Rio de Janeiro.

Full name, department, institution, city, and country of the authors

***Marcella Cini Oliveira, M.C.O.** (<https://orcid.org/0000-0001-5305-6823>),
Faculty of Medicine, Federal University of Rio de Janeiro, Rio de Janeiro,
Brazil.

Tatiana de Araujo Eleuterio, T.A.E. (<https://orcid.org/0000-0002-8043-2350>),
Nursing School, State University of Rio de Janeiro, Rio de Janeiro, Brazil.

Allan Bruno de Andrade Corrêa, A.B.A.C. Physics Institute, Federal
University of Rio de Janeiro, Rio de Janeiro, Brazil.

Lucas Dalsenter Romano da Silva, L.D.R.S. (<https://orcid.org/0000-0001-6078-4247>), Department of Preventive Medicine, Institute of Collective Health
Studies, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

Renata Coelho Rodrigues, R.C.R. (<https://orcid.org/0000-0001-5014-4991>),
Faculty of Medicine, Federal University of Rio de Janeiro, Rio de Janeiro,
Brazil.

Bruna Andrade de Oliveira, B.A.O. (<https://orcid.org/0000-0003-0640-288X>)
Faculty of Medicine, Federal University of Rio de Janeiro, Rio de Janeiro,
Brazil.

Marlos Melo Martins, M.M.M. (<https://orcid.org/0000-0002-5511-4764>),
Department of Child Neurology, Martagão Gesteira Institute of Childcare and
Pediatrics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

Carlos Eduardo Raymundo, C.E.R. (<http://orcid.org/0000-0002-4150-4403>),
Institute of Collective Health Studies, Federal University of Rio de Janeiro, Rio
de Janeiro, Brazil.

Roberto de Andrade Medronho, R.A.M. (<http://orcid.org/0000-0003-4073-3930>), Institute of Collective Health Studies, Faculty of Medicine, Federal
University of Rio de Janeiro, Rio de Janeiro, Brazil.

*** Corresponding author**

E-mail of the corresponding author: cini.marcella@gmail.com

Abstract

Background: We analyzed the factors associated with death in confirmed cases of COVID-19 in the state of Rio de Janeiro. This cross-sectional study evaluated the association between individual demographic, clinical, and epidemiological variables and the outcome (death) using data from the Unified Health System information systems.

Methods: We used the extreme boosting gradient (XGBoost) model to analyze the data, which uses decision trees weighted by the estimation difficulty. To evaluate the relevance of each independent variable, we used the SHapley Additive exPlanations (SHAP) metric. From the probabilities generated by the XGBoost model, we transformed the data to the logarithm of odds to estimate the odds ratio for each independent variable.

Results: This study showed that older individuals of black race/skin color with heart disease or diabetes who had dyspnea or fever were more likely to die.

Conclusions: The early identification of patients who may progress to a more severe form of the disease can help improve the clinical management of patients with COVID-19 and is thus essential to reduce the lethality of the disease.

Keywords: COVID-19, SARS-CoV-2, XGBoost, Machine Learning, Pandemic, Coronavirus infection, Coronavirus death

Background

In December 2019, cases of pneumonia of unknown cause occurred in Wuhan, Hubei Province, China. On January 7, 2020, a new beta-coronavirus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified (1-3). Like severe acute respiratory syndrome coronavirus and Middle East respiratory syndrome coronavirus, SARS-CoV-2 causes a lower respiratory infection called coronavirus disease 2019 (COVID-19). The World Health Organization (WHO) declared the disease a public health emergency of international concern on January 30, 2020. On February 3, 2020, Brazil announced the disease as a public health emergency of national concern (4), and on February 25, 2020, the country confirmed its first case (5).

On March 2, 2021, the world had 114,140,104 confirmed cases and 2,535,520 deaths, and Brazil had 10,551,259 confirmed cases and 254,942 deaths (6). COVID-19 can occur asymptotically, as influenza-like illness, or as more severe forms, which characterize severe acute respiratory syndrome (SARS). The most common clinical manifestations include fever, dry cough, diarrhea, vomiting, and muscle pain or fatigue (7,8). The average incubation period is 5.2 days (9). Transmission can occur by droplets and aerosols or by contact with contaminated objects (10).

Its mortality rate is higher in individuals over 80 years of age and in people with comorbidities, so these constitute the risk group for severe forms of the disease (11). The gold standard for laboratory confirmation of SARS-CoV-2 infection is reverse transcription-polymerase chain reaction (RT-PCR). The

performance of mass RT-PCR testing, contact tracing, and isolation of positive individuals is essential to control the spread of the virus (12-16).

This study aims to analyze the factors associated with death in confirmed cases of COVID-19 in the state of Rio de Janeiro and considers the importance of defining these factors, the objective of several studies in the literature, for better case management (17-22).

Methods

Design

The design was a cross-sectional study evaluating the association between individual demographic, clinical, and epidemiological predictor variables and the outcome (death) among confirmed cases of COVID-19 in the state of Rio de Janeiro using data gathered from the Unified Health System (SUS) information systems.

Data collection

We gathered the confirmed influenza-like illness cases caused by COVID-19 from the e-SUS NOTIFICA system of the Ministry of Health. We obtained the confirmed SARS cases by COVID-19 from the Flu Epidemiological Surveillance Information System (*SIVEP-Gripe*) of the Ministry of Health. We linked these two databases. The consistency analysis removed any duplicates from the resulting database. Next, we linked this database with the Mortality Information System database of the Ministry of

Health. The study covered the period from March 5 (the first confirmed case in the state of Rio de Janeiro) to September 15, 2020. During this period, the state of Rio de Janeiro confirmed 243,509 cases of COVID-19, of which 178,231 (73.2%) were of influenza-like illness (ILI) and 65,278 (26.8%) were of SARS. All cases included in this study met the ILI and SARS criteria defined by the Ministry of Health (23).

The individual predictor variables selected covered demographic, clinical, and epidemiological characteristics (Table 1).

Table 1. Individual variables selected

Demographic	Age
	Sex
	Race/skin color, self-reported
	Municipality of residence
Clinical	Fever
	Cough
	Sneezing
	Runny nose
	Sore throat
	Dyspnea
	Loss of smell
	Loss of taste
	Headache
	Muscle pain
	Diarrhea
Epidemiological	Respiratory disease
	Kidney disease
	Diabetes
	Immunodeficiency/immunosuppression
	Heart disease

Data analysis

We used Microsoft Excel 365 software to handle the database, while we performed the statistical analyses with the software R x64 4.0.0 through RStudio Desktop and Python 3.8.5.

In the descriptive analysis of the data, we categorized the quantitative predictor variable age according to age groups (in years): 0 to 9, 10 to 19, 20 to 29, 30 to 39, 40 to 49, 50 to 59, 60 to 69, 70 to 79, and 80 years or older. For the modeling, we did not categorize patient age.

The bivariate descriptive analysis used bar graphs, pie charts, or box plots for the numerical predictor variable (age) and contingency tables for categorical predictor variables. We used Pearson's chi-squared test to analyze categorical predictor variables and the Wilcoxon test for the numerical predictor variables, both with a level of significance of 5%.

We checked the database for missing data. We kept variables considered epidemiologically relevant but with a high percentage of missing data in the analysis, creating a category named "missing" for each predictor variable. They were "kidney diseases", "immunodeficiency/immunosuppression", "diabetes", "heart disease", "dyspnea", "cough", "fever", "headache", "sneezing", "odynophagia", "muscle pain", "diarrhea", "loss of smell", "loss of taste", and "runny nose".

Due to a large number of categories of the predictor variable "municipality of residence" (91 municipalities), we grouped the residences into regions defined by the state government of Rio de Janeiro (24). The exception was the city of Rio de Janeiro, which has more than one-third of the state's population and more than half of the deaths, so we analyzed it as an individual category, called "Metropolitan Region I - capital".

Then, through the evaluation with several machine learning algorithms, we applied an Extreme Boosting Gradient (XGBoost), as it showed the best performance for the dataset. To meet the assumption of balanced data for a good

performance of machine learning models, we balanced the predicted variable (outcome) using the synthetic minority oversampling technique (SMOTE) (25) and edited nearest neighbor (ENN) (26) techniques. The XGBoost model uses decision trees weighted by the estimation difficulty (27). Compared to machine learning models in general, the advantages of this model are the better generalization of the results, the possibility of visualizing all decision trees, and the reduction of the bias and variance of a single tree.

To evaluate the relevance of each predictor variable, we used the SHapley Additive exPlanations (SHAP) metric (28), which can take negative or positive values, and the closer it is to 0, the less the predictor variable influences the predicted variable. The results show the mean of the modules of the importance of each predictor variable.

Also, we used sensitivity, specificity, accuracy, positive predictive value, negative predictive value, Youden's index, receiver operating characteristic (ROC) curves, and Matthews' correlation coefficient (MCC) to evaluate the goodness of fit (29).

From the probabilities generated by the XGBoost model, we transformed the data to the logarithm of odds, where odds is the ratio between the probability of death and non-death attributed to the categories of each predictor variable for each individual. We used the logarithm of odds to estimate the odds ratio, according to the formula below:

$$Odds\ ratio_i(OR_i) = \frac{e^{mean(\log(odds_{int}))}}{e^{mean(\log(odds_{ref}))}}$$

Where:

int = Category of interest of the predictor variable

ref = Category of reference of the predictor variable

For the numerical variable age, we used the categories of interest described earlier in this section to obtain the summarized results.

We calculated the confidence interval according to Wayne W. LaMorte (30) and estimated the total number of individuals by category and outcome using the following formula:

$$95\% \text{ confidence interval} = e^{(\log(OR_i) \pm [1.96 \times SE(\log(OR_i))])}$$

Where:

$$SE(\log(OR_i)) = \sqrt{\frac{1}{p_{int} \times n_{int}} + \frac{1}{(1 - p_{int}) \times n_{int}} + \frac{1}{p_{ref} \times n_{ref}} + \frac{1}{(1 - p_{ref}) \times n_{ref}}}$$

p = estimated probability of death for the category

n = number of individuals in the category

Results

Between March 10 and September 15, 2020, the state of Rio de Janeiro reported 243,509 confirmed cases of COVID-19, of which 37.3% were residents of the municipality of Rio de Janeiro. Of the total number of cases, 178,231 (73.2%) were classified as ILI, and 65,278 (26.8%) were classified as SARS.

Table 2 shows the incidence, mortality, and lethality of confirmed cases of COVID-19 by age group. The disease incidence was higher in the age group 80 years and older, followed by 40 to 49 years and 30 to 39 years. Mortality and lethality increased with increasing age, except in children younger than ten years.

195 **Table 2. Distribution of incidence, mortality, and lethality by age group**

Age group	Deaths	Cases	Population ^a	Incidence ^b	Mortality ^b	Lethality
0 to 9 years	60	5,105	2,224,713	229.47	2.70	1.18
10 to 19 years	34	7,478	2,193,282	340.95	1.55	0.45
20 to 29 years	210	30,426	2,638,726	1,153.06	7.96	0.69
30 to 39 years	603	53,654	2,692,631	1,992.62	22.39	1.12
40 to 49 years	1,400	51,913	2,476,855	2,095.92	56.52	2.70
50 to 59 years	2,526	40,729	2,146,886	1,897.12	117.66	6.20
60 to 69 years	4,154	27,034	1,645,437	1,642.97	252.46	15.37
70 to 79 years	4,425	15,273	888,202	1,719.54	498.20	28.97
80 years and older	4,622	10,812	459,457	2,353.21	1,005.97	42.75

196 Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September
197 2020

198 ^a Population estimated by the IBGE for 2020

199 ^b per 100,000 inhabitants
200

201 Table 3 shows the predictor variables' distribution as a function of
202 disease severity (ILI and SARS). There was a higher frequency of cases among
203 the age group of 40-49 years (21.3%), among females (52.4%), and among
204 brown individuals (32.9%), followed by white individuals (32.6%). The most
205 frequent signs and symptoms were cough, fever, dyspnea, and odynophagia. The
206 most prevalent comorbidities were cardiovascular disease and diabetes mellitus.
207 Regarding disease evolution, the lethality was 7.4%.

208 ILI was more frequent among younger individuals, the most common
209 age group being 30-39 years (24.4%); the median age at ILI diagnosis was 41
210 years, and it was more frequent among females (52.6%) and brown individuals
211 (34.1%), followed by white individuals (33.1%). The most frequent signs and
212 symptoms were cough, fever, and odynophagia. The frequency of comorbidities
213 was 5.8% for cardiovascular disease and 3.0% for diabetes mellitus. Only 1.2%
214 of ILI cases progressed to death.

215 SARS was more frequent in individuals older than 60 years (38.3%); the
216 median age was 53. Women had the highest frequency (52.0%), but concerning

217 race/color, white individuals had the highest frequency (31.4%), followed by
 218 brown individuals (29.7%). Dyspnea (80.2%), cough (65.7%), and fever
 219 (60.3%) were more frequent than in cases of ILI. The rate of associated
 220 comorbidities was also higher than that in ILI, with 23.8% cardiovascular
 221 disease and 15.6% diabetes mellitus. The lethality observed in cases of SARS
 222 was 24.5%. All predictor variables were significantly different between ILI and
 223 SARS cases.

224

225 **Table 3. Distribution of confirmed cases of ILI and SARS, according to the**
 226 **individual variables selected.**

Variables	Overall		ILI		SARS		p-value
	n	%	n	%	n	%	
Age group							0.000
0-9 years	5,105	2.1%	4,369	2.5%	736	1.1%	-
10-19 years	7,478	3.1%	6,442	3.6%	1,036	1.6%	-
20-29 years	30,426	12.5%	25,014	14.0%	5,412	8.3%	-
30-39 years	53,654	22.0%	43,466	24.4%	10,188	15.6%	-
40-49 years	51,913	21.3%	40,163	22.5%	11,750	18.0%	-
50-59 years	40,729	16.7%	29,536	16.6%	11,193	17.1%	-
60-69 years	27,034	11.1%	16,751	9.4%	10,283	15.8%	-
70-79 years	15,273	6.3%	7,356	4.1%	7,917	12.1%	-
80 years or older	10,812	4.5%	4,055	2.3%	6,757	10.4%	-
No Information	1,085	40.0%	1,079	60.0%	6	0.0%	-
Sex							0.011
Female	127,677	52.4%	93,740	52.6%	33,937	52.0%	-
Male	115,832	47.6%	84,491	47.4%	31,341	48.0%	-
Race/skin color							0.000
Brown	80,162	32.9%	60,806	34.1%	19,356	29.7%	-
Asian	11,875	4.9%	9,808	5.5%	2,067	3.2%	-
White	79,495	32.6%	58,986	33.1%	20,509	31.4%	-
Indigenous	491	0.2%	441	0.2%	50	0.1%	-
Black	15,802	6.5%	11,007	6.2%	4,795	7.3%	-
Not reported	55,684	22.9%	37,183	20.9%	18,501	28.3%	-
Presence of signs and symptoms							-
Fever	93,102	38.3%	53,721	30.1%	39,381	60.3%	0.000
Cough	108,434	44.5%	65,565	36.8%	42,869	65.7%	0.000
Headache	33,020	13.6%	24,667	13.8%	8,353	12.8%	0.000

Sneezing	1,636	0.7%	1,249	0.7%	387	0.6%	0.005
Odynophagia	42,797	17.6%	28,276	15.9%	14,521	22.2%	0.000
Dyspnea	52,376	21.5%	0	0.0%	52,376	80.2%	0.000
Runny nose	18,139	7.4%	13,931	7.8%	4,208	6.4%	0.000
Diarrhea	12,801	5.3%	9,059	5.1%	3,742	5.7%	0.000
Muscle pain	23,015	9.5%	16,163	9.1%	6,852	10.5%	0.004
Loss of smell	21,764	8.9%	16,272	9.1%	5,492	8.4%	0.000
Loss of taste	15,478	6.4%	11,826	6.6%	3,652	5.6%	0.000
Presence of comorbidities							-
Cardiovascular disease	25,845	10.6%	10,296	5.8%	15,549	23.8%	0.000
Diabetes mellitus	15,500	6.4%	5,333	3.0%	10,167	15.6%	0.000
Chronic kidney disease	2,203	0.9%	530	0.3%	1,673	2.6%	0.000
Chronic respiratory disease	6,318	2.6%	3,036	1.7%	3,282	5.0%	0.000
Immunodeficiency/immunosuppression	2,527	1.0%	1,226	0.7%	1,301	2.0%	0.000
Evolution							0.000
Death	18,076	0.074	2,097	0.012	15,979	0.245	-
Non-death	225,433	0.926	176,134	0.988	49,299	0.755	-
Total	243,509	100.0%	178,231	73.19%	65,278	26.81%	-

227 Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September

228 2020

229 Table 4 shows the relationship between the predictor variables and the
230 outcome (death). There was a higher proportion of elderly age (over 60 years
231 old), male sex, and black race/skin color among deaths than among non-deaths.
232 The signs and symptoms that stood out were dyspnea, fever, and cough.
233 Dyspnea was present in 56.5% of cases of death and in 18.7% of non-deaths.
234 Fever was present in 48.3% of the cases of death and in 37.4% of the cases of
235 non-death. Cough was present in 49.3% of cases of death and 44.1% of cases of
236 non-death. The most common comorbidities in both groups were cardiovascular
237 disease, at 34.1% in cases of death and 8.7% in cases of non-death, and diabetes
238 mellitus, at 24.0% in cases of death and 5.0% in cases of non-death.

239 The number of missing data of predictor variables were as follows: 1,085
240 (0.4%) for age, 3,506 (1.4%) for signs and symptoms, 396 (0.2%) for chronic
241 respiratory disease, 55,684 (22.9%) for skin color/race, 19,006 (7.8%) for
242 cardiovascular disease, 21,701 (8.9%) for diabetes mellitus, 25,870 (10.6%) for

243 chronic kidney disease, 26,181 (10.8%) for
 244 immunodeficiency/immunosuppression, and 273 (0.1%) for municipality of
 245 residence.

246 There was a statistically significant association between the outcome and
 247 the predictor variables listed in Table 4.

248

249 **Table 4. Distribution of COVID-19 outcomes according to the individual**
 250 **variables selected.**

Variables	Death		Non-death		p-value
	n	%	n	%	
Age group					0.000
0-9 years	60	0,3%	5,045	2.2%	-
10-19 years	34	0,2%	7,444	3.3%	-
20-29 years	210	1,2%	30,216	13.4%	-
30-39 years	603	3,3%	53,051	23.5%	-
40-49 years	1,400	7,7%	50,513	22.4%	-
50-59 years	2,526	14,0%	38,203	16.9%	-
60-69 years	4,154	23,0%	22,880	10.1%	-
70-79 years	4,425	24,5%	10,848	4.8%	-
80 years or older	4,622	25,6%	6,190	2.7%	-
Sex					0.000
Female	8,003	44,3%	119,674	53.1%	-
Male	10,073	55,7%	105,759	46.9%	-
Race/color					0.000
Brown	4,480	24,8%	75,682	33.6%	-
Asian	182	1,0%	11,693	5.2%	-
White	5,041	27,9%	74,454	33.0%	-
Indigenous	11	0,1%	480	0.2%	-
Black	1,479	8,2%	14,323	6.4%	-
Missing	6,883	38,08%	48,801	21.6%	-
Presence of signs and symptoms					-
Fever	8,723	48,3%	84,379	37.4%	0.000
Cough	8,908	49,3%	99,526	44.1%	0.000
Headache	503	2,8%	32,517	14.4%	0.000
Sneezing	10	0,1%	1,626	0.7%	0.000
Odynophagia	1,560	8,6%	41,237	18.3%	0.000
Dyspnea	10,204	56,5%	42,172	18.7%	0.000
Runny nose	233	1,3%	17,906	7.9%	0.000

Diarrhea	143	0,8%	12,658	5.6%	0.000
Muscle pain	803	4,4%	22,212	9.9%	0.000
Loss of smell	194	1,1%	21,570	9.6%	0.000
Loss of taste	99	0,5%	15,379	6.8%	0.000
Presence of comorbidities					-
Cardiovascular disease	6,163	34,1%	19,682	8.7%	0.000
Diabetes mellitus	4,341	24,0%	11,159	5.0%	0.000
Chronic kidney disease	923	5,1%	1,280	0.6%	0.000
Chronic respiratory disease	845	4,7%	5,473	2.4%	0.000
Immunodeficiency/immunosuppression	464	2,6%	2,063	0.9%	0.000
Total	18,076	7,4%	225,433	92.6%	-

Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September

2020

The historical case series per epidemiological week showed a proportional increase in cases of ILI relative to those of SARS starting from epidemiological week 19 (Figure 1a). Figure 1b shows a proportional increase in deaths compared to non-deaths between the 13th and 20th epidemiological weeks, with a maximum of 2,465 deaths (14.7%) in week 18. After that, the death percentage dropped.

Figure 1. The proportion of cases per epidemiological week according to clinical evolution (ILI/SARS) (a) and outcome (b).

Concomitantly, there was a progressive increase in cases in younger age groups throughout the study period (Figure 2).

Figure 2. The proportion of cases per epidemiological week according to age group.

Figure 3 illustrates the comparison of the age distribution according to clinical evolution - ILI vs. SARS (a) and death vs. non-death (b). ILI was more frequent in younger individuals and SARS in older individuals. The comparison between the age groups of individuals according to the outcome showed a higher concentration of elderly individuals in the group that progressed to death. In both cases, there was a statistically significant difference.

Figure 3. Age distribution according to clinical evolution (ILI/SARS) (a) and outcome (b).

The geographical distribution of confirmed cases and deaths per municipality of residence is shown in Table 5. The incidence for the entire state was equal to 1,522.9 cases per 100,000 inhabitants, the mortality rate was 112.9 deaths per 100,000 inhabitants, and the lethality rate was 7.4%. We observed the highest incidences in Varre-Sai, Macaé, and Natividade, with rates of 6,321.9, 3,799.7, and 3,626.8 cases per 100,000 inhabitants, respectively. The municipalities with the highest lethality rates were Nilópolis, São João do Meriti, and Rio de Janeiro, with 14.3%, 13.3%, and 11.9%, respectively. Rio de Janeiro, Iguaba Grande, and Tanguá recorded the highest mortality rates, with 170.7, 148.8, and 136.7 deaths per 100,000 inhabitants, respectively.

Regarding the regions of the state of Rio de Janeiro, Metropolitan Region I had the highest lethality (10.5%) and mortality (139.0 deaths per 100,000 inhabitants), while the highest incidence was in the Costa Verde Region (2,840.2 cases per 100,000 inhabitants).

Table 5. Geographic distribution of cases by regions and municipalities of the state of Rio de Janeiro.

(table at the end of the manuscript)

Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020.

The values in Figure 4 illustrate the mean values of the SHAP metric of the most relevant predictor variables in defining the odds of death of the XGBoost model. Immunodeficiency/immunosuppression, age, region of residence, and presence of dyspnea were the characteristics with the most significant impact on the outcome.

Figure 4. Means of the SHAP metric of the predictor variables of the XGBoost model.

Table 6 shows the evaluation measures of the XGBoost model. Its high sensitivity (89%) and specificity (89%) indicate the model's ability to identify cases of death and non-death, respectively. The accuracy revealed 89% of hits, while the negative predictive value showed that 99% of the cases predicted as non-deaths were correct. The positive predictive value of 30% may reflect the data imbalance since more than 90% of the cases had a non-death outcome.

Table 6. Evaluation measures of the XGBoost model.

Sensitivity	89%
Specificity	89%

Accuracy	89%
Youden's index	0.78
The area under the ROC curve	94%
Positive predictive value	30%
Negative predictive value	99%
MCC	48%

Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

Figure 5 shows the ROC curve, which appears very close to the upper left corner, with an area under the curve of 94%, revealing that the model performed excellently in predicting the outcome.

Figure 5. Receiver operating characteristic (ROC) curve of the XGBoost model.

Table 7 shows the odds ratio of death with the respective confidence intervals of the predictor variables. Note that the variables with the highest odds of death were metropolitan region I - capital, metropolitan region I - without capital, and metropolitan region II (compared to the northwest fluminense region); the presence of dyspnea, presence of fever, presence of diabetes, presence of heart disease (compared to absence); black race/color (compared to white color/race); male sex (compared to female); and age over 30 years (compared to age range 0 to 9 years). The variables associated with lower odds of death were coastal lowlands region (compared to the northwest fluminense region); the presence of headache, presence of odynophagia, presence of muscle pain, presence of diarrhea, presence of loss of smell, and presence of runny nose (compared to absence); and Asian race/color (compared to white color/race).

Table 7. Odds ratios and respective 95% confidence intervals were obtained from the results of the *XGBoost* model.

(table at the end of the manuscript)

Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

Discussion

In the present study, the mortality and lethality of COVID-19 increased with age, showing that the disease tends to be more severe in the older population. The Chinese Center for Disease Control and Prevention (31) analyzed 44,672 confirmed cases of the disease and found similar findings. However, the Chinese study found much lower lethality in all age groups than in this study. In the age group of 70 to 79 years, the lethality of the present study was 3.6 times higher than the lethality in the Chinese study (28.97% *versus* 8.0%), while in the age group of 80 years and over, the lethality was 2.9 times higher (42.75% *versus* 14.8%).

Older people tend to have lower immunity and, therefore, greater susceptibility to severe forms of infectious diseases and a higher frequency of comorbidities such as cardiovascular diseases, diabetes, and chronic lung diseases. These factors may explain the higher frequency of severe form (SARS) cases in addition to higher lethality among people over 60 years. Advanced age as a risk factor for death has been described in Noor et al (19) and Albitar et al (20).

The higher incidence of COVID-19 in the age group between 30 and 49 years was probably attributable to their greater exposure to SARS-CoV-2 in public transport or at work since the need to generate income prevented many people from staying on lockdown. Conversely, the evolution of the frequency of cases of SARS and ILI over time showed a proportional increase in ILI relative to SARS, starting from epidemiological week 19. Over the same time, we observed an increase in the frequency of cases in younger age groups, which may be related to lower adherence to social distancing in this population group. There was a higher frequency of female patients in this study. Williamson et al. (17) Cobre et al. (18) found similar results. We show the association of black race/skin color with higher odds of death than white race/skin color. Williamson et al. (17) found a higher risk of death in black individuals than in white individuals, even after adjusting for other factors (hazard ratio 1.48 (95% CI 1.29-1.69)).

Dyspnea, cough, and fever were the most commonly found symptoms, being more frequent in SARS cases and cases of death. Fever is a clinical sign that is part of the set of organic inflammatory responses to the presence of an infectious agent and is mediated mainly by cytokines. These inflammatory mediators can cause tissue damage and organ dysfunction by stimulating free radicals and other toxic components, as cited in the study by Chang et al. (32). The latter retrospectively followed 211 patients and investigated risk factors for the progression of COVID-19. Therefore, fever may be a clinical sign associated with greater disease severity. Dyspnea, in turn, is the main symptom of SARS, a severe form of the disease where more deaths are expected. Fever and dyspnea were associated with a higher chance of death. At the same time, other reported

symptoms, such as headache, odynophagia, muscle pain, diarrhea, loss of smell, and runny nose, had an association with a lower chance of death. Maciel et al. reported dyspnea as a factor associated with COVID-19 death, while cough, fever, and other symptoms were protective factors (33).

The city of Rio de Janeiro had the highest mortality within the state. This may have been due to the better laboratory diagnostic capacity for COVID-19, which would lead to more accurate classification of the underlying cause of death. Also, there is better access of the population to health services in the city, facilitating the confirmation of the diagnosis on time. The difficulty of laboratory diagnosis of COVID-19 and access to health services in most cities of the state may have contributed to inaccuracy in the diagnosis of COVID-19 (34). A less robust health system, especially when associated with a lower socioeconomic level, seems to be a risk factor for higher lethality due to COVID-19 (21).

In this study, people with heart disease had a higher chance of death. Zheng et al. (34) described the association of some comorbidities with higher lethality from COVID-19 in a meta-analysis. Cardiac tissue has a higher number of angiotensin-converting enzyme receptors (ACE2), which are involved in the endocytosis of SARS-CoV-2, than other tissues, which could allow the virus to damage cardiac tissue directly. Mikami et al. analyzed 6,493 hospitalized patients with a confirmed diagnosis of COVID-19 and found a higher risk of death in patients with elevated serum troponin levels (35).

This study found a higher chance of death in people with diabetes. Hyperglycemia secondary to diabetes mellitus can lead to immune dysfunction

through the impairment of humoral and cellular functions and the antioxidant system. Besides, diabetic patients are more vulnerable to nosocomial infections (36). These factors may be associated with a higher chance of death in diabetic patients with COVID-19.

Some limitations of our study should be noted. The *RedCap* platform has received input of ILI notifications until March 27, 2020, and the e-SUS Notifica has received input of ILI notifications after that. The notifications of SARS are made by sending the scanned investigation form, which feeds the Influenza Epidemiological Surveillance Information System (SIVEP-Gripe). In this context, the reported cases refer to those who sought care in health units in the state of Rio de Janeiro and entered into these information systems. The instructions given to the population were to seek medical care in case of signs of severity, which might have generated a selection bias. Another factor that should be taken into account is the missing data for some variables, which may have generated information bias.

Conclusions

COVID-19 is a disease that can evolve into severe forms leading to death, especially in certain population groups. This study showed that older individuals of black race/skin color with heart disease or diabetes who had dyspnea or fever were more likely to die. The present study aimed to contribute to the early identification of COVID-19 patients who may progress to a more severe form of the disease, improve the clinical management of patients with COVID-19, and reduce the disease's lethality.

438

439 **List of abbreviations**

440 ACE2 – Angiotensin-converting enzyme receptors

441 COVID-19 – Coronavirus disease 2019

442 ENN – Edited nearest neighbor

443 XGBoost – Extreme gradient boosting

444 ILI - Influenza-like illness

445 MCC – Matthews' correlation coefficient

446 ROC – Receiver operating characteristic

447 RT-PCR – Reverse transcription polymerase chain reaction

448 SARS – Severe acute respiratory syndrome

449 SARS-CoV-2 - severe acute respiratory syndrome coronavirus 2

450 SHAP – SHapley Additive exPlanations

451 SIVEP-Gripe – Flu Epidemiological Surveillance Information System

452 SMOTE – Synthetic minority oversampling technique

453 SUS – Unified Health System

454 WHO – World Health Organization

455

456 **Declarations**

457 **Ethics approval and consent to participate/ publish**

458 This study is a cross-sectional study from a database that contains information

459 that does not allow for individual identification. Therefore, an ethics committee

460 did not need approval as cited in CONEP/Brazil, Brazilian committee
461 responsible for ethical evaluation in clinical research. The attached document in
462 “Related Files” supports this statement.

464 **Availability of data and materials**

465 The databases and materials used in this manuscript may be made available
466 upon request from interested researchers. To that end, please contact the
467 corresponding author, Marcella Cini Oliveira, through the e-mail
468 cini.marcella@gmail.com.

470 **Competing Interests**

471 The authors declare that they have no competing interests.

473 **Funding**

474 The National Council for Scientific and Technological Development (CNPq)
475 and the Federal University of Rio de Janeiro (UFRJ) are responsible for granting
476 the Scientific Initiation scholarship to co-author Marcella Cini Oliveira, through
477 the Institutional Program for Scientific Initiation Scholarships (PIBIC).

478 The Institute of Epidemiology and Collective Health of the Federal University
479 of Rio de Janeiro (IESC-UFRJ) and the DS / CAPES - 2019 Program are
480 responsible for granting the doctoral scholarship to co-author Lucas Dalsenter
481 Romano da Silva.

482 Research Support Foundation of the State of Rio de Janeiro (FAPERJ) is
483 responsible for acquiring the funds for translation and publication of the
484 manuscript.

486 **Author Contributions**

487 M.C.O. contributed to conceptualization, data curation, formal analysis,
488 investigation, methodology, project administration, software, supervision,
489 visualization, and writing – original draft, writing – review & editing, and was a
490 major contributor in writing the manuscript.

491 T.A.E. contributed to conceptualization, data curation, formal analysis,
492 investigation, methodology, software, supervision, visualization, and writing –
493 original draft, writing – review & editing.

494 A.B.A.C. contributed to conceptualization, data curation, formal analysis,
495 investigation, methodology, software, visualization, and writing – original draft,
496 writing – review & editing.

497 L.D.R.S. contributed to conceptualization, data curation, formal analysis,
498 investigation, visualization, and writing – original draft, writing – review &
499 editing.

500 R.C.R. contributed conceptualization, formal analysis, investigation,
501 visualization, and writing – original draft, writing – review & editing.

502 B.A.O. contributed to conceptualization, formal analysis, investigation,
503 visualization, and writing – original draft, writing – review & editing.

504 M.M.M. contributed to conceptualization, formal analysis, investigation, writing
505 – review & editing.

506 C.E.R. contributed to data curation, software.

507 R.A.M. contributed to conceptualization, formal analysis, investigation, project
508 administration, supervision, visualization, and writing – original draft, writing –
509 review & editing.

510 All authors read and approved the final manuscript.

511 **Acknowledgements**

512 Not Applicable.

514 **References**

- 515 1. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and
 516 outcomes of critically ill patients with SARS-CoV-2 pneumonia in
 517 Wuhan, China: a single-centered, retrospective, observational study.
 518 *Lancet Respir Med* [Internet]. 2020;8(5):475–81. Available from:
 519 [http://dx.doi.org/10.1016/S2213-2600\(20\)30079-5](http://dx.doi.org/10.1016/S2213-2600(20)30079-5)

- 520 2. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological
 521 and clinical characteristics of 99 cases of 2019 novel coronavirus
 522 pneumonia in Wuhan, China: a descriptive study. *Lancet* [Internet].
 523 2020;395(10223):507–13. Available from:
 524 [http://dx.doi.org/10.1016/S0140-6736\(20\)30211-7](http://dx.doi.org/10.1016/S0140-6736(20)30211-7)

- 525 3. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical
 526 Characteristics of 138 Hospitalized Patients with 2019 Novel
 527 Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA - J Am Med*
 528 *Assoc.* 2020;323(11):1061–9.

- 529 4. Croda J, Oliveira WK de, Frutuoso RL, Mandetta LH, Baia-Da-silva DC,
 530 Brito-Sousa JD, et al. Covid-19 in Brazil: Advantages of a socialized
 531 unified health system and preparation to contain cases. *Rev Soc Bras Med*
 532 *Trop.* 2020;53(April):2–7.

- 533 5. de Jesus JG, Sacchi C, Candido D da S, Claro IM, Sales FCS, Manuli ER,
 534 et al. Importation and early local transmission of covid-19 in Brazil, 2020.
 535 *Rev Inst Med Trop Sao Paulo.* 2020;62(April):1–5.

- 536 6. WHO. WHO Coronavirus Disease (COVID-19) Dashboard [Internet].
 537 2021 [cited 2021 Mar 2]. Available from:
 538 [https://covid19.who.int/?gclid=EAIaIQobChMI54-](https://covid19.who.int/?gclid=EAIaIQobChMI54-s182f6wIVkYSRCh23mgziEAAYASAAEgK5U_D_BwE)
 539 [s182f6wIVkYSRCh23mgziEAAYASAAEgK5U_D_BwE](https://covid19.who.int/?gclid=EAIaIQobChMI54-s182f6wIVkYSRCh23mgziEAAYASAAEgK5U_D_BwE)

- 540 7. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of
 541 patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*
 542 (London, England). 2020;395(10223):497–506.

- 543 8. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al.
 544 World Health Organization declares global emergency: A review of the
 545 2019 novel coronavirus (COVID-19). *Int J Surg* [Internet].
 546 2020;76(February):71–6. Available from:
 547 <https://doi.org/10.1016/j.ijssu.2020.02.034>

- 548 9. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission
 549 Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia.
 550 *N Engl J Med.* 2020;382(13):1199–207.

- 551 10. Adhikari SP, Meng S, Wu Y, Mao Y, Ye R, Wang Q, et al. A scoping

552 review of 2019 Novel Coronavirus during the early outbreak period:
553 Epidemiology, causes, clinical manifestation and diagnosis, prevention
554 and control. 2020;1–12.

555 11. Varghese GM, John R, Manesh A, Karthik R, Abraham OC. Clinical
556 management of COVID-19. J Dent Educ [Internet]. 2012;76(11):401–10.
557 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23144490>

558 12. ECDC. An overview of the rapid test situation for COVID-19 diagnosis
559 in the EU / EEA. Eur Cent Dis Prev Control. 2020;(April):1–3.

560 13. Tang Y, Schmitz JE, Persing DH, Stratton CW. Laboratory Diagnosis of
561 COVID-19: Current Issues and Challenges. J Clin Microbiol.
562 2020;58(6):1–9.

563 14. Vashist SK. In vitro diagnostic assays for COVID-19: Recent advances
564 and emerging trends. Diagnostics. 2020;10(4).

565 15. Bruno R, Mondelli M, Brunetti E, Matteo A Di, Seminari E, Maiocchi L,
566 et al. Performance of VivaDiag COVID-19 IgM/IgG Rapid Test is
567 inadequate for diagnosis of COVID-19 in acute patients referring to
568 emergency room department. J Med Virol. 2020;92(10):1724–7.

569 16. Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody
570 responses to SARS-CoV-2 in patients with COVID-19. Nat Med.
571 2020;26(6):845–8.

572 17. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE,
573 et al. Factors associated with COVID-19-related death using
574 OpenSAFELY. Nature [Internet]. 2020;584(7821):430–6. Available
575 from: <http://dx.doi.org/10.1038/s41586-020-2521-4>

576 18. Cobre A de F, Böger B, Fachi MM, Vilhena R de O, Domingos EL,
577 Tonin FS, et al. Risk factors associated with delay in diagnosis and
578 mortality in patients with covid-19 in the city of Rio de Janeiro, Brazil.
579 Ciênc. Saúde Coletiva. 2020;25:4131–40.

580 19. Noor FM, Islam MM. Prevalence and Associated Risk Factors of
581 Mortality Among COVID-19 Patients: A Meta-Analysis. J Community
582 Health [Internet]. 2020;45(6):1270–82. Available from:
583 <https://doi.org/10.1007/s10900-020-00920-x>

584 20. Albitar O, Ballouze R, Ooi JP, Ghadzi SMS. Risk factors for mortality of
585 COVID-19 patients. medRxiv. 2020;(January).

586 21. Mohammed M, Muhammad S, Mohammed FZ, Mustapha S, Sha'aban A,
587 Sani NY, et al. Risk Factors Associated with Mortality Among Patients
588 with Novel Coronavirus Disease (COVID-19) in Africa. J Racial Ethn
589 Heal Disparities. 2020;6–11.

590 22. Escosteguy CC, Eleuterio T de A, Pereira AGL, Marques MRVE,
591 Brandão AD, Batista JPM. COVID-19: a cross-sectional study of

592 suspected cases admitted to a federal hospital in Rio de Janeiro, Brazil,
593 and factors associated with hospital death. *Epidemiol. Serv. Saúde*.
594 2021;30(1):e2020750. Available from: [http://dx.doi.org/10.1590/s1679-](http://dx.doi.org/10.1590/s1679-49742021000100023)
595 49742021000100023

596 23. Rio de Janeiro. State Health Department. Coronavirus disease (COVID-
597 19), updated information from the technical note - SVS/SES-RJ number
598 2020/01 (fourth update). Technical Note SVS/SES-RJ [Internet]. 2020.
599 Available from: [http://patydoalferes.rj.gov.br/wp-](http://patydoalferes.rj.gov.br/wp-content/uploads/2020/08/9-ATUALIZACAO.pdf)
600 content/uploads/2020/08/9-ATUALIZACAO.pdf

601 24. CEPERJ. Regions of the state of Rio de Janeiro [Internet]. [cited 2021
602 Mar 11]. Available from:
603 <http://www.ceperj.rj.gov.br/Conteudo.asp?ident=81>

604 25. Mansourifar H, Shi W. Deep synthetic minority over-sampling technique.
605 arXiv. 2020;16:321–57.

606 26. Wilson DL. Asymptotic Properties of Nearest Neighbor Rules Using
607 Edited Data. *IEEE Trans Syst Man Cybern*. 1972;2(3):408–21.

608 27. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *J Assoc*
609 *Physicians India*. 1994;42(8):665.

610 28. Štrumbelj E, Kononenko I. Explaining prediction models and individual
611 predictions with feature contributions. *Knowl Inf Syst*. 2014;41(3):647–
612 65.

613 29. Matthews BW. Comparison of the predicted and observed secondary
614 structure of T4 phage lysozyme. *BBA - Protein Struct*. 1975;405(2):442–
615 51.

616 30. Wayne W. LaMorte. Confidence Interval for an Odds Ratio [Internet].
617 Comparing Frequencies. 2018 [cited 2021 Mar 11]. p. 8. Available from:
618 [https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-](https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717_ComparingFrequencies/PH717_ComparingFrequencies8.html#headingtaglink_1)
619 QuantCore/PH717_ComparingFrequencies/PH717_ComparingFrequenci
620 es8.html#headingtaglink_1

621 31. Novel Coronavirus Pneumonia Emergency Response Epidemiology
622 Team. The epidemiological characteristics of an outbreak of 2019 novel
623 coronavirus diseases (COVID-19) in China. *Chinese Cent Dis Control*
624 *Prev Wkly* [Internet]. 2020;41(2):145–51. Available from:
625 <https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.003>

626 32. Chang MC, Park YK, Kim BO, Park D. Risk factors for disease
627 progression in COVID-19 patients. *BMC Infect Dis*. 2020;20(1):4–9.

628 33. Leonor Maciel E, Jabor P, Etereldes Goncalves J, Tristão-Sá R, Rita de
629 Cássia Duarte L, Reis-Santos B, et al. Factors associated with COVID-19
630 hospital deaths in Espírito Santo, Brazil, 2020. *Epidemiol. Serv. Saúde*.
631 2020;29(4):1–11.

34. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. 2020;(January).
35. Mikami T, Miyashita H, Yamada T, Harrington M, Steinberg D, Dunn A, et al. Risk Factors for Mortality in Patients with COVID-19 in New York City. *J Gen Intern Med*. 2021;36(1):17–26.
36. Klonoff DC, Umpierrez GE. Letter to the Editor: COVID-19 in patients with diabetes: Risk factors that increase morbidity. *Metabolism* [Internet]. 2020;108:154224. Available from: <https://doi.org/10.1016/j.metabol.2020.154224>

Figure 1. The proportion of cases per epidemiological week according to clinical evolution (ILI/SARS) (a) and outcome (b).

Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

Figure 2. The proportion of cases per epidemiological week according to age group.

Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

Figure 3. Age distribution according to clinical evolution (ILI/SARS) (a) and outcome (b).

Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

Figure 4. Means of the SHAP metric of the predictor variables of the XGBoost model.

Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

662 **Figure 5. Receiver operating characteristic (ROC) curve of the XGBoost**
663 **model.**

664 Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September
665 2020

666
667
668
669
670

671 **Table 5. Geographic distribution of cases by regions and municipalities of**
672 **the state of Rio de Janeiro.**

Region/Municipality	Population	Death	Non-death	Cases	Lethality	Mortality (100 thousand inhabitants)	Incidence (100 thousand inhabitants)
Central-South Fluminense Region	272,407	157	5,620	5,777	2.7%	57.6	2,120.7
Areal	11,423	6	180	186	3.2%	52.5	1,628.3
Comendador Levy Gasparian	8,180	4	149	153	2.6%	48.9	1,870.4
Engenheiro Paulo de Frontin	13,237	4	135	139	2.9%	30.2	1,050.1
Mendes	17,935	3	173	176	1.7%	16.7	981.3
Miguel Pereira	24,642	12	241	253	4.7%	48.7	1,026.7
Paraíba do Sul	41,084	21	869	890	2.4%	51.1	2,166.3
Paty do Alferes	26,539	7	167	174	4.0%	26.4	655.6
Sapucaia	17,525	22	373	395	5.6%	125.5	2,253.9
Três Rios	77,432	59	2,624	2,683	2.2%	76.2	3,465.0
Vassouras	34,410	19	709	728	2.6%	55.2	2,115.7
Costa Verde Region	243,500	222	6,694	6,916	3.2%	91.2	2,840.2
Angra dos Reis	169,511	150	5,170	5,320	2.8%	88.5	3,138.4
Mangaratiba	36,456	48	699	747	6.4%	131.7	2,049.0
Paraty	37,533	24	825	849	2.8%	63.9	2,262.0
Coastal Lowlands Region	810,666	577	12,126	12,703	4.5%	71.2	1,567.0
Araruama	112,008	63	1,224	1,287	4.9%	56.2	1,149.0
Armação dos Búzios	27,560	15	423	438	3.4%	54.4	1,589.3
Arraial do Cabo	27,715	6	184	190	3.2%	21.6	685.5
Cabo Frio	186,227	139	2,222	2,361	5.9%	74.6	1,267.8
Cachoeiras de Macacu	54,273	34	558	592	5.7%	62.6	1,090.8
Casimiro de Abreu	35,347	19	898	917	2.1%	53.8	2,594.3
Iguaba Grande	22,851	34	576	610	5.6%	148.8	2,669.5
Rio Bonito	55,551	55	1,805	1,860	3.0%	99.0	3,348.3
Rio das Ostras	105,676	101	1,875	1,976	5.1%	95.6	1,869.9

São Pedro da Aldeia	87,875	42	1,201	1,243	3.4%	47.8	1,414.5
Saquarema	74,234	56	1,010	1,066	5.3%	75.4	1,436.0
Silva Jardim	21,349	13	150	163	8.0%	60.9	763.5
Middle Paraíba Region	855,193	480	13,998	14,478	3.3%	56.1	1,693.0
Barra do Pirai	94,778	51	1,023	1,074	4.7%	53.8	1,133.2
Barra Mansa	177,813	117	2,210	2,327	5.0%	65.8	1,308.7
Itatiaia	28,783	6	351	357	1.7%	20.8	1,240.3
Pinheiral	22,719	10	462	472	2.1%	44.0	2,077.6
Pirai	26,314	16	648	664	2.4%	60.8	2,523.4
Porto Real	16,592	11	355	366	3.0%	66.3	2,205.9
Quatis	12,793	1	144	145	0.7%	7.8	1,133.4
Resende	119,769	56	1,932	1,988	2.8%	46.8	1,659.9
Rio Claro	17,425	5	280	285	1.8%	28.7	1,635.6
Rio das Flores	8,561	2	21	23	8.7%	23.4	268.7
Valença	71,843	17	566	583	2.9%	23.7	811.5
Volta Redonda	257,803	188	6,006	6,194	3.0%	72.9	2,402.6
Metropolitan Region I	9,920,734	13,793	117,483	131,276	10.5%	139.0	1,323.2
Belford Roxo	469,332	292	8,546	8,838	3.3%	62.2	1,883.1
Duque de Caxias	855,048	749	8,320	9,069	8.3%	87.6	1,060.6
Itaguaí	109,091	116	2,164	2,280	5.1%	106.3	2,090.0
Japeri	95,492	37	397	434	8.5%	38.7	454.5
Magé	227,322	236	3,474	3,710	6.4%	103.8	1,632.0
Mesquita	168,376	168	1,344	1,512	11.1%	99.8	898.0
Nilópolis	157,425	188	1,124	1,312	14.3%	119.4	833.4
Nova Iguaçu	796,257	603	5,564	6,167	9.8%	75.7	774.5
Paracambi	47,124	39	677	716	5.4%	82.8	1,519.4
Queimados	137,962	83	2,341	2,424	3.4%	60.2	1,757.0
Rio de Janeiro	6,320,446	10,787	79,981	90,768	11.9%	170.7	1,436.1
São João de Meriti	458,673	442	2,876	3,318	13.3%	96.4	723.4
Seropédica	78,186	53	675	728	7.3%	67.8	931.1
Metropolitan Region II	1,914,974	1,668	33,696	35,364	4.7%	87.1	1,846.7
Guapimirim	51,483	50	1,608	1,658	3.0%	97.1	3,220.5
Itaboraí	218,008	225	4,313	4,538	5.0%	103.2	2,081.6
Maricá	127,461	134	3,391	3,525	3.8%	105.1	2,765.6
Niterói	487,562	486	12,354	12,840	3.8%	99.7	2,633.5
São Gonçalo	999,728	731	11,248	11,979	6.1%	73.1	1,198.2
Tanguá	30,732	42	782	824	5.1%	136.7	2,681.2
Northeast Fluminense Region	317,493	160	8,070	8,230	1.9%	50.4	2,592.2
Aperibé	10,213	5	188	193	2.6%	49.0	1,889.7
Bom Jesus do Itabapoana	35,411	10	736	746	1.3%	28.2	2,106.7
Cambuci	14,827	2	181	183	1.1%	13.5	1,234.2
Italva	14,063	7	348	355	2.0%	49.8	2,524.4
Itacara	22,899	17	333	350	4.9%	74.2	1,528.5
Itaperuna	95,841	77	2,671	2,748	2.8%	80.3	2,867.2
Laje do Muriaé	7,487	3	242	245	1.2%	40.1	3,272.3

	Miracema	26,843	2	429	431	0.5%	7.5	1,605.6
	Natividade	15,082	3	544	547	0.5%	19.9	3,626.8
	Porciúncula	17,760	12	593	605	2.0%	67.6	3,406.5
	Santo Antônio de Pádua	40,589	19	1,068	1,087	1.7%	46.8	2,678.1
	São Jose de Ubá	7,003	1	140	141	0.7%	14.3	2,013.4
	Varre-Sai	9,475	2	597	599	0.3%	21.1	6,321.9
North Fluminense Region		849,515	566	15,995	16,561	3.4%	66.6	1,949.5
	Campos dos Goytacazes	463,731	321	4,941	5,262	6.1%	69.2	1,134.7
	Carapebus	13,359	7	278	285	2.5%	52.4	2,133.4
	Cardoso Moreira	12,600	4	348	352	1.1%	31.7	2,793.7
	Conceição de Macabu	21,211	10	593	603	1.7%	47.1	2,842.9
	Macaé	206,728	152	7,703	7,855	1.9%	73.5	3,799.7
	Quissamã	20,242	8	445	453	1.8%	39.5	2,237.9
	São Fidelis	37,543	25	234	259	9.7%	66.6	689.9
	São Francisco de Itabapoana	41,354	21	539	560	3.8%	50.8	1,354.2
	São João da Barra	32,747	18	914	932	1.9%	55.0	2,846.1
Mountain Region		805,627	433	11,498	11,931	3.6%	53.7	1,481.0
	Bom Jardim	25,333	5	156	161	3.1%	19.7	635.5
	Cantagalo	19,830	2	151	153	1.3%	10.1	771.6
	Carmo	17,434	3	162	165	1.8%	17.2	946.4
	Cordeiro	20,430	7	217	224	3.1%	34.3	1,096.4
	Duas Barras	10,930	3	62	65	4.6%	27.4	594.7
	Macuco	5,269	2	65	67	3.0%	38.0	1,271.6
	Nova Friburgo	182,082	86	2,749	2,835	3.0%	47.2	1,557.0
	Petrópolis	295,917	174	2,135	2,309	7.5%	58.8	780.3
	Santa Maria Madalena	10,321	3	103	106	2.8%	29.1	1,027.0
	São Jose do Vale do Rio Preto	20,251	16	439	455	3.5%	79.0	2,246.8
	São Sebastiao do Alto	8,895	1	30	31	3.2%	11.2	348.5
	Sumidouro	14,900	11	139	150	7.3%	73.8	1,006.7
	Teresópolis	163,746	120	5,034	5,154	2.3%	73.3	3,147.6
	Trajano de Moraes	10,289	0	56	56	0.0%	0.0	544.3
	Missing	-	20	253	273	7.3%	-	-
Total		15,990,109	18,076	225,433	243,509	7.4%	112.9	1,522.9

673 Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September

674 2020

675

676

677

678

679

Table 7. Odds ratios and respective 95% confidence intervals were obtained from the results of the *XGBoost* model.

Predictor variable and reference category	Category of interest	Odds ratio (95% CI)
Sex (reference: female)		
	Male	1.029 (1.012; 1.046)
Race/color (reference: white)		
	Asian	0.935 (0.899; 0.973)
	Indigenous	1.029 (0.857; 1.236)
	Brown	1.032 (0.996; 1.069)
	Black	1.033 (1.012; 1.054)
	Missing	1.078 (1.054; 1.102)
Respiratory diseases (reference: absence)		
	Presence	0.982 (0.932; 1.033)
	Missing	0.657 (0.539; 0.800)
Kidney diseases (reference: absence)		
	Presence	1.011 (0.928; 1.102)
	Missing	1.336 (1.300; 1.373)
Diabetes (reference: absence)		
	Presence	1.176 (1.137; 1.218)
	Missing	1.021 (0.992; 1.051)
Heart diseases (reference: absence)		
	Presence	1.073 (1.045; 1.103)
	Missing	1.006 (0.975; 1.037)
Immunodeficiency/Immunosuppression (reference: absence)		
	Presence	1.007 (0.930; 1.091)
	Missing	1.685 (1.639; 1.732)
Age (reference: 0 to 9 years)		
	10-19 years	1.004 (0.934; 1.078)
	20-29 years	1.011 (0.953; 1.073)
	30-39 years	1.061 (1.001; 1.124)
	40-49 years	1.116 (1.053; 1.183)
	50-59 years	1.285 (1.212; 1.363)

60-69 years	1.585 (1.491; 1.684)
70-79 years	1.811 (1.697; 1.933)
80 years or older	1.869 (1.744; 2.002)
Dyspnea (reference: absence)	
Presence	1.309 (1.283; 1.336)
Missing	1.411 (1.313; 1.516)
Cough (reference: absence)	
Presence	1.006 (0.989; 1.022)
Missing	1.411 (1.313; 1.516)
Fever (reference: absence)	
Presence	1.029 (1.011; 1.046)
Missing	1.019 (0.951; 1.092)
Headache (reference: absence)	
Presence	0.909 (0.888; 0.931)
Missing	0.976 (0.911; 1.045)
Sneezing (reference: absence)	
Presence	1.000 (0.905; 1.106)
Missing	1.000 (0.934; 1.071)
Odynophagia (reference: absence)	
Presence	0.933 (0.913; 0.954)
Missing	0.985 (0.919; 1.055)
Muscle pain (reference: absence)	
Presence	0.964 (0.938; 0.992)
Missing	0.985 (0.919; 1.055)
Diarrhea (reference: absence)	
Presence	0.954 (0.920; 0.989)
Missing	0.990 (0.924; 1.060)
Loss of smell (reference: absence)	
Presence	0.913 (0.887; 0.939)
Missing	0.959 (0.896; 1.027)
Loss of taste (reference: absence)	
Presence	0.976 (0.944; 1.010)
Missing	0.989 (0.923; 1.059)
Runny nose (reference: absence)	
Presence	0.955 (0.926; 0.985)
Missing	0.983 (0.918; 1.053)
Region of residence (reference: Northwest Fluminense)	
Central-South Fluminense Region	0.955 (0.893; 1.023)
Costa Verde Region	0.941 (0.883; 1.004)
Coastal Lowlands Region	0.944 (0.893; 0.999)
Middle Paraíba Region	1.000 (0.946; 1.056)
Metropolitan Region I - Capital	1.322 (1.263; 1.384)
Metropolitan Region I - Without Capital	1.269 (1.209; 1.332)
Metropolitan Region II	1.113 (1.060; 1.169)

North Fluminense Region	0.984 (0.933; 1.038)
Mountain Region	0.962 (0.909; 1.018)
Missing	0.942 (0.739; 1.201)

687 Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September
688 2020

689

Figures

Figure 1

The proportion of cases per epidemiological week according to clinical evolution (ILI/SARS) (a) and outcome (b). Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

Figure 2

The proportion of cases per epidemiological week according to age group. Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

Figure 3

Age distribution according to clinical evolution (ILI/SARS) (a) and outcome (b). Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

Figure 4

Means of the SHAP metric of the predictor variables of the XGBoost model. Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020

Figure 5

Receiver operating characteristic (ROC) curve of the XGBoost model. Note: confirmed cases of COVID-19, state of Rio de Janeiro, March–September 2020