



Prediction of SARS-CoV-2 infection with a Symptoms-Based model to aid public health decision making in Latin America and other low and middle income settings

Andrea Ramírez Varela^{a,*}, Sergio Moreno López^a, Sandra Contreras-Arrieta^a, Guillermo Tamayo-Cabeza^a, Silvia Restrepo-Restrepo^a, Ignacio Sarmiento-Barbieri^a, Yuldor Caballero-Díaz^a, Luis Jorge Hernandez-Florez^a, John Mario González^a, Leonardo Salas-Zapata^b, Rachid Laajaj^a, Giancarlo Buitrago-Gutierrez^c, Fernando de la Hoz-Restrepo^d, Martha Vives Florez^a, Elkin Osorio^b, Diana Sofía Ríos-Oliveros^b, Eduardo Behrentz^a

^a Universidad de los Andes, Bogotá, Colombia

^b Secretaría Distrital de Salud de Bogotá, Bogotá, Colombia

^c Instituto de Investigaciones Clínicas, Universidad Nacional de Colombia, Bogotá, Colombia

^d Universidad Nacional de Colombia, Bogotá, Colombia

ARTICLE INFO

Keywords:

SARS-CoV-2

COVID-19

Logistic model

Machine learning

Symptoms

Anosmia

ABSTRACT

Symptoms-based models for predicting SARS-CoV-2 infection may improve clinical decision-making and be an alternative to resource allocation in under-resourced settings. In this study we aimed to test a model based on symptoms to predict a positive test result for SARS-CoV-2 infection during the COVID-19 pandemic using logistic regression and a machine-learning approach, in Bogotá, Colombia. Participants from the CoVIDA project were included. A logistic regression using the model was chosen based on biological plausibility and the Akaike Information criterion. Also, we performed an analysis using machine learning with random forest, support vector machine, and extreme gradient boosting. The study included 58,577 participants with a positivity rate of 5.7%. The logistic regression showed that anosmia ($aOR = 7.76$, 95% CI [6.19, 9.73]), fever ($aOR = 4.29$, 95% CI [3.07, 6.02]), headache ($aOR = 3.29$, 95% CI [1.78, 6.07]), dry cough ($aOR = 2.96$, 95% CI [2.44, 3.58]), and fatigue ($aOR = 1.93$, 95% CI [1.57, 2.93]) were independently associated with SARS-CoV-2 infection. Our final model had an area under the curve of 0.73. The symptoms-based model correctly identified over 85% of participants. This model can be used to prioritize resource allocation related to COVID-19 diagnosis, to decide on early isolation, and contact-tracing strategies in individuals with a high probability of infection before receiving a confirmatory test result. This strategy has public health and clinical decision-making significance in low- and middle-income settings like Latin America.

1. Introduction

The SARS-CoV-2 pandemic has been one of the most significant public health challenges in modern history. By April 2022, more than 500 million cases have been reported, with over 6 million deaths globally (Johns Hopkins University, 2022). The COVID-19 pandemic has posed a significant challenge in terms of the readiness of healthcare systems to mobilize resources for infection diagnosis, contact tracing,

and supporting adequate infrastructure for treatment. Particularly for diagnosis, since the pandemic's start, low- and middle-income countries reported inequities and delays in access to testing, time-to-test consultation, and test results turnaround (Lau et al., 2020). Evidence shows that in high-income countries, test turnaround times can be <3 days, (McGarry et al., 2021; National Health Services, 2021) in contrast to middle-low- and low-income countries, where test turnaround times can be between 8 and 10 days, depending on socioeconomic status (Laajaj

* Corresponding author at: School of Medicine, Universidad de los Andes, Cra 7 #116-05, 110111 Bogotá, Colombia.

E-mail address: an-rami2@uniandes.edu.co (A. Ramírez Varela).

<https://doi.org/10.1016/j.pmedr.2022.101798>

Received 22 December 2021; Received in revised form 16 March 2022; Accepted 17 April 2022

Available online 20 April 2022

2211-3355/© 2022 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

et al., 2021). Even more, a modeling study suggests that delays of over 3 days from symptom onset to test result turnaround significantly decrease the effectiveness of nonpharmacological strategies such as contact tracing (Kretzschmar et al., 2020).

Given these limitations in testing, other methods for rapid diagnostics have been used. Tools such as imaging have been conducted as an alternative to clinical testing for COVID-19 diagnosis and disease severity. These analyses have included both chest X-ray (CXR) and computed tomography (CT) images combining various methods, including machine learning (ML) neural networks, among others (Chen et al., 2021; Li et al., 2020a; Singh et al., 2021; Wang et al., 2022). However, most interventions related to diagnostic imaging require some level of face-to-face medical care, implying an additional challenge to healthcare systems in low-to-middle income countries with accessibility issues such as Colombia and most of Latin America.

Therefore, clinical guidelines have provided common symptoms that are suggestive of the infection. However, given the low predictive capacity of these symptoms if presented alone, it may lead to isolating people with a very low probability of infection, or no isolation at all. Due to this, recent literature suggests to use symptoms clusters to improve detection of COVID-19. This strategy is particularly helpful for clinical decision-making and testing in primary healthcare settings, with scarce diagnostic resources, in particular during pandemic peaks with high community transmission when mass unlimited testing is not possible (Li et al., 2020b; Long et al., 2020; Mercer and Salit, 2021). Also, in the case of telemedicine, this can amplify healthcare access without greater infrastructure and overcome difficulties in patient care (Hincapié et al., 2020; Monkowski et al., 2019).

Some studies have approached COVID-19 diagnosis using symptoms-based models. To our knowledge, two studies have been conducted in Latin America regarding SARS-CoV-2 infection prediction. The first study used a database from a symptoms-tracking app in Brazil with serological testing for SARS-CoV-2. Symptoms such as loss of smell/anosmia and shortness of breath were independently associated with SARS-CoV-2 infection with a negative predictive value (NPV) of 93% (Dantas et al., 2021). The second study was also conducted in Brazil, assessing national seroprevalence and its association with reported symptoms such as anosmia, fever, and body ache through a conditional inference tree analysis. In this study, those participants that did not report these symptoms had a positivity rate of 0.8%, compared to 18.3% of those with loss of smell and fever (Menezes et al., 2021).

In settings with low testing capacity, such as Colombia and Latin America, as shown before, where there is low adherence to non-pharmacological interventions such as isolation because of socioeconomic reasons, a model for predicting SARS-CoV-2 infection may be an alternative to resource allocation, including diagnostic testing capacities and prioritization in healthcare and contact tracing. The CoVIDA project was an intensified epidemiological surveillance study for SARS-CoV-2 performed in Bogotá, the most populated city in Colombia (with 7 million inhabitants). Over 55,000 RT-PCR tests were performed to increase the city's testing capacity during the first two pandemic waves. RT-PCR tests conducted by the CoVIDA project allowed the local health authorities to identify positive cases in individuals with mild symptoms that may have been missed by traditional epidemiological surveillance that focuses mainly on people with moderate to severe COVID-19 (Varela et al., 2021). Therefore, in this study, we aimed to test a prediction model based on symptoms to predict a positive RT-PCR test result for SARS-CoV-2 infection among high mobility working adult participants of the CoVIDA project during the 1st year of the COVID-19 pandemic in Bogotá, Colombia.

2. Materials and methods

2.1. Study design and population

This study used the data from the CoVIDA project, a large sentinel

epidemiological surveillance study conducted in Bogotá, Colombia, from April 2020 to March 2021. This project was created to detect transmission patterns among high mobility asymptomatic or mild symptomatic populations due to their occupation (healthcare workers, public transportation workers, employees in public markets, grocery stores, food delivery, construction, cleaning and other home services, education, informal workers, police, military forces, firefighters, and other essential services). Further description of the CoVIDA project sampling and testing allocation methods can be found elsewhere (Varela et al., 2021). After participants accepted the informed consent via a telephone call, they completed a questionnaire about sociodemographic characteristics, including sex, age, socioeconomic strata, contact with a confirmed case of COVID-19, comorbidities, and self-reported symptoms related to SARS-CoV-2 infection (sore throat, dry cough, fatigue, anosmia, diarrhea, fever, dyspnea, confusion, headache, myalgias, dysgeusia, chills, vomiting/nausea, rhinorrhea) (Instituto Nacional de Salud, 2020; Ministerio de Salud y Protección Social de Colombia, 2020). In the case of a positive test, the CoVIDA contact center informed the participants of the results and provided with recommendations. Positive test results were provided to the Colombian health authorities according to national guidelines.

This study was approved by the ethics committee of Universidad de los Andes (Act No. 1278 of 2020). Informed consent was obtained via telephone call, in order to comply with physical distancing.

2.2. Specimen collection and testing

The CoVIDA testing centers performed the RT-PCR test for SARS-CoV-2 infection following two testing models: (a) home visit, for mild-symptomatic participants, and (b) for asymptomatic patients, drive/walk-through testing. Participants underwent laboratory testing for SARS-CoV-2 using RT-PCR tests with nasopharyngeal swab sampling. The samples were processed by the GenCore Sequencing Center of the Universidad de los Andes following the international Berlin protocol and using the U-TOP™ COVID-19 detection kit for one-step real-time RT-PCR (Corman et al., 2020).

2.3. Statistical analysis

We reported descriptive results of the complete database using absolute and relative frequencies and central tendency measurements according to the distribution of the continuous variables. We compared the distribution of categorical sociodemographic characteristics and symptoms by the test result (positive or negative) using the Chi-square test or the Fisher exact test (in variables with reporting frequencies <5) for bivariate analysis. For continuous variables, we used the U-Mann Whitney test. A *p* value of 0.05 or lower was considered statistically significant.

2.4. Variable selection for prediction models

The selection of variables for our prediction model used in the logistic regression is presented in Fig. 1A. As we included many symptoms and biological features in the questionnaire, we aimed to choose the best model with the combination of variables that offered the best diagnostic accuracy using variables selection. First, the choice of variables was based on biological plausibility regarding symptoms used to predict SARS-CoV-2 infection according to the literature. Second, we included all possible interactions between variables (Sauerbrei and Royston, 2011) using the fractional multinomial polynomials method for the complete database. Variables selected for the final models were based on the leaps-and-bounds algorithm which allowed us to reach the best combination of variables as predictors for the model using the best Akaike information criterion to reduce false-positive rates (Lindsey and Sheather, 2015). We calculated the logistic regression's odds ratios (ORs) and 95% confidence intervals (95% CI).

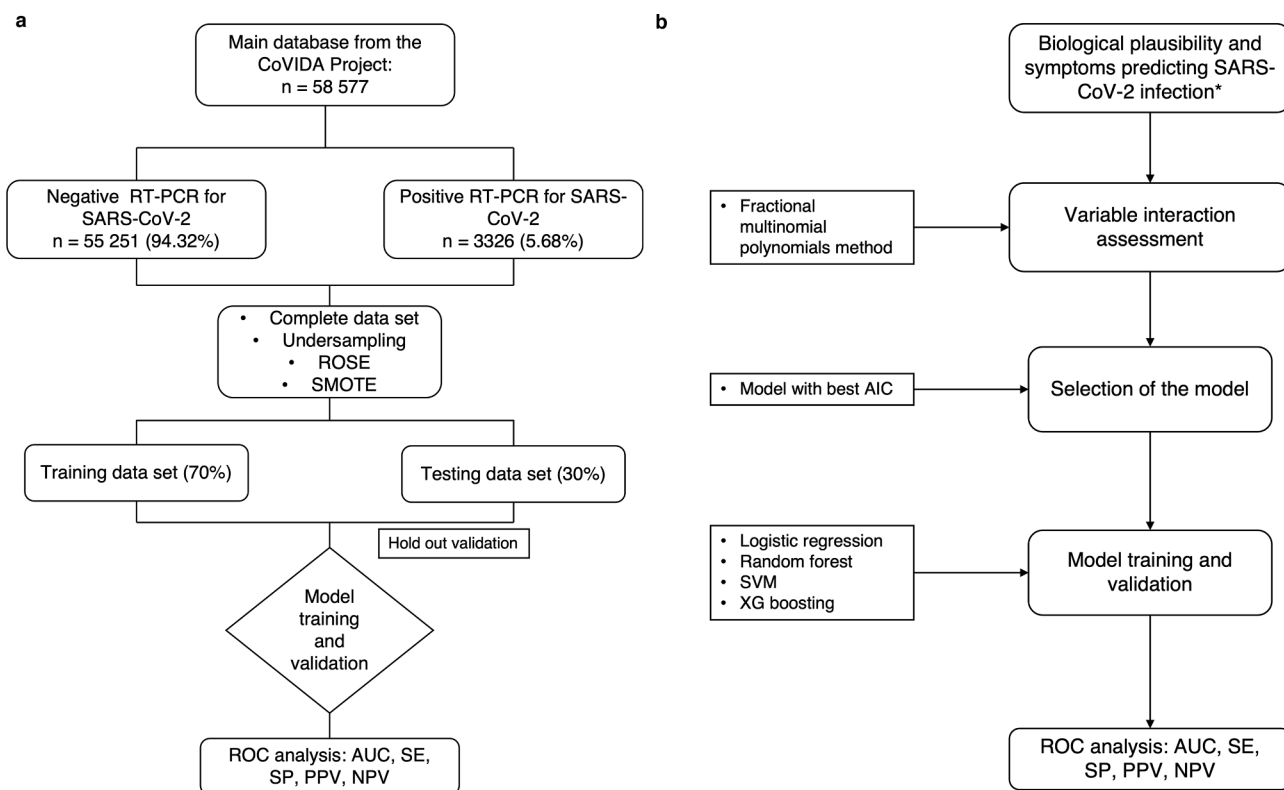


Fig. 1. Modeling framework for the analysis of symptoms association and SARS-CoV-2 prediction. a) Variable's selection for the final logistic regression model; b) Sampling procedure and hold out validation for logistic regression model and the ML approach.

Given the low prevalence of the event (an overall positivity rate of 5.7% observed in the CoVIDA project), (Varela et al., 2021) resampling and oversampling procedures were performed to approach to the data imbalance. We performed an independent sensitivity analysis using the following datasets: (a) the complete database; (b) an undersampling dataset resulting from a random resampling using a 4:1 ratio of participants with negative and positive RT-PCR results, respectively; (c) an oversampling dataset obtained by the ROSE method; and d) an oversampling dataset obtained by SMOTE. We performed a holdout validation of the model obtained by randomly splitting all datasets into training and testing datasets in a 70:30 ratio (see Fig. 1B). Analyses were performed in all datasets. Complete definition of sample balancing techniques can be found in [Supplementary Note 1](#).

2.5. Machine learning analysis

Sensitivity analysis to assess the diagnostic performance of the prediction model was conducted using multiple ML methods: random forest (RF), support vector machine (SVM), and Extreme gradient boosting (XG boost). An RF model was performed using a classification approach, using GINI reduction based on Breiman's random forest algorithm. The SVM method followed a based Kernel linear approach. The XG boosting model was trained in the training set with the hyperparameter settings described in [Supplementary Note 2](#). Each ML approach included variables to the prediction model in order of importance. We assessed graphically the importance of each variable included in ML models.

We assessed the performance of the obtained models with all datasets using the following parameters: area under the curve (AUC), sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), and total accuracy (participants correctly classified). We conducted analyses using Stata 16.0 and R version 4.0.5.

3. Results

Data used in this study comes from the CoVIDA project, an intensified epidemiological surveillance study that performed 58,577 RT-PCR tests for SARS-CoV-2 from April 2020 to March 2021. A positivity rate (response variable) of 5.7% was observed in the CoVIDA sample (Varela et al., 2021).

3.1. Demographic characteristics among participants

[Table 1](#) shows sociodemographic characteristics and symptoms related to SARS-CoV-2 infection. The median age of the sample was 36 years (IQR = 28–48). Participants with a positive test result showed a median age of 35 years (IQR = 27–47), while those with a negative test result showed a median of 36 years (IQR = 28–49). Positive RT-PCR for SARS-CoV-2 was more frequent in 30- to 59-year-olds, with 56.5%. Regarding the sex of participants, 50.8% were female. The 18.4% of the participants reported contact with a COVID-19 confirmed case. The cumulative positivity rate for those with contact with a confirmed case was 7.4%, compared to the 5.3% of participants who did not report contact. The most frequent comorbidities were arterial hypertension and smoking, with 7.4% and 5.1%, respectively (see [Table 1](#)).

Asymptomatic participants accounted for 87.5% of the sample. Of these asymptomatic people, 3.6% had a positive RT-PCR result. Among the symptomatic participants, 20.2% had a positive test result. The most frequently reported symptoms among the entire cohort were sore throat (7.2%), dry cough (5.8%), fatigue (4.6%), diarrhea (3.2%), and anosmia (3.1%). In participants with a positive test result, the most frequent symptoms reported were dysgeusia (81.3%), anosmia (41.6%), fever (37.9%), dyspnea (27.8%), dry cough (26.2%), and fatigue (26.1%; see [Table 1](#)).

Table 1

Sociodemographic characteristics and symptoms related to SARS-CoV-2 infection (N = 58,556).

Variable	Total N = 58,556	Positive n = 3,325	Negative n = 55,231	p-value
Age (years) median (IQR)	36 (28–48)	35 (27–47)	36 (28–49)	< -001*
Age (years) (n, %)				-001*
<18	235 (0.4)	30 (0.90)	205 (0.4)	
18–29	17,718 (30.3)	1,096 (32.9)	16,622 (30.1)	
30–59	35,119 (59.9)	1,880 (56.5)	33,239 (60.2)	
>60	5,484 (9.4)	319 (9.6)	5,165 (9.4)	
Sex (n, %)				-183
Female	29,736 (50.8)	1,652 (5.6)	28,084 (94.4)	
Male	28,794 (49.2)	1,673 (5.8)	27,121 (94.2)	
Contact with a COVID-19 confirmed case (n, %)				< -001*
Yes	10,758 (18.4)	796 (7.4)	9,962 (92.6)	
No	47,798 (81.6)	2,529 (5.3)	45,269 (94.7)	
Comorbidities (n, %)				
Arterial hypertension	4,352 (7.4)	228 (5.2)	4,124 (94.8)	-193
Smoking	2,969 (5.1)	133 (4.5)	2,836 (95.5)	-004*
Obesity	2,551 (4.4)	139 (5.5)	2,412 (94.6)	-609
Asthma	2,311 (3.9)	104 (4.5)	2,207 (95.5)	-007*
Diabetes mellitus	1,379 (2.4)	71 (5.2)	1,308 (94.9)	-478
Chronic obstructive lung disease	295 (0.5)	19 (6.4)	276 (93.6)	-652
Symptoms related to SARS-CoV-2 (n, %)				
Asymptomatic	51,254 (87.5)	1,852 (3.6)	49,402 (96.4)	< -001*
Symptomatic	7,302 (12.5)	1,473 (20.2)	5,829 (79.8)	< -001*
Sore throat	4,230 (7.2)	790 (18.7)	3,440 (81.3)	< -001*
Dry cough	3,407 (5.8)	893 (26.2)	2,514 (73.8)	< -001*
Fatigue	2,713 (4.6)	707 (26.1)	2,006 (73.9)	< -001*
Anosmia	1,802 (3.1)	749 (41.6)	1,053 (58.4)	< -001*
Diarrhea	1,871 (3.2)	413 (22.1)	1,458 (77.9)	< -001*
Fever	1,169 (2.0)	444 (37.9)	725 (62.0)	< -001*
Dyspnea	1,376 (2.6)	383 (27.8)	993 (72.2)	< -001*
Confusion	413 (0.7)	119 (28.8)	294 (71.2)	< -001*
Headache	254 (0.4)	50 (19.7)	204 (80.3)	< -001*
Myalgias	68 (0.1)	15 (22.1)	53 (77.9)	< -001*
Dysgeusia	16 (0.03)	13 (81.3)	3 (18.8)	< -001*†
Chills	25 (0.04)	2 (8.0)	23 (92.0)	-616†
Vomiting/nausea	25 (0.04)	3 (12.0)	22 (88.0)	-172†
Rhinorrhea	13 (0.02)	0 (0)	13 (100.0)	-376†

*p value < .05.

†Fisher exact test.

3.2. Model training

Given the low prevalence of the event (an overall positivity rate of 5.7% observed in the CoVIDA project), (Varela et al., 2021) resampling and oversampling were performed to address the data imbalance. The multivariable fractional polynomial algorithm with the complete database suggested an interaction between dry cough and anosmia. However, this interaction variable was not included in the final model by the leaps and bounds algorithm. The variables that were recommended by the algorithm for the logistic regression model were age, socioeconomic strata, contact with confirmed COVID-19 case, arterial hypertension, and symptoms related to SARS-CoV-2 (fever, anosmia, dry cough, fatigue, and headache), as shown in the equation below:

$$\text{Probability of a positive RT PCR test} = \frac{e^{\text{prediction}}}{1 + e^{\text{prediction}}}, \text{ where}$$

$$\begin{aligned} \text{prediction} = & -2.83 + (0.004 \times \text{age}) + (1.15 \\ & \times \text{socioeconomic strata 1}) + (0.99 \\ & \times \text{socioeconomic strata 2}) + (0.82 \\ & \times \text{socioeconomic strata 3}) + (0.20 \\ & \times \text{socioeconomic strata 4}) + (0.24 \\ & \times \text{socioeconomic strata 5}) + (0.25 \\ & \times \text{Contact with a confirmed case}) - (0.23 \\ & \times \text{arterial hypertension}) + (1.45 \times \text{fever}) + (2.05 \\ & \times \text{anosmia}) + (1.08 \times \text{dry cough}) + (0.66 \times \text{fatigue}) + (1.19 \\ & \times \text{headache}) \end{aligned}$$

Table 2 presents the unadjusted and adjusted odds ratios obtained in the logistic regression. The contact with a confirmed COVID-19 case was associated with higher odds of SARS-CoV-2 infection. Regarding symptoms, anosmia showed the highest odds (aOR = 7.76, 95% CI [6.19, 9.73]) of having a positive test compared to a participant who did not report it. Fever had an aOR of 4.29 (CI 95% 3.07–6.02), headache an aOR of 3.29 (CI 95% 1.78–6.07), dry cough an aOR of 2.96 (CI 95% 2.44–3.58), and fatigue an aOR of 1.93 (CI 95% 1.57–2.93). Fig. 2A presents the adjusted ORs obtained in the logistic regression model in a forest plot. A diagnostic performance assessment of the model was conducted in the testing dataset of the undersampling database. AUC was 0.73, with an SE of 26%, SP of 98%, PPV of 73%, and NPV of 86%. With the logistic regression model, 85% of participants were correctly classified. Fig. 2B shows the logistic regression model's AUC. The variables included in each model and the logistic regression estimates for the complete database, ROSE dataset, and SMOTE dataset are presented in Supplementary Tables 1–3.

Comparison of the diagnostic performance of logistic regression and ML methods is shown in Table 3. Similar AUCs were obtained using the complete database, the SMOTE dataset and the undersampling dataset. The ROSE dataset had higher AUC, SE, and PPV but lower SP and proportion of correctly classified individuals. Similar SP and PPV values were obtained using all datasets. Variable importance for each dataset using the ML approach is presented in Supplementary Figs. 1 through 4. Anosmia was considered the primary classification variable when using the complete and undersampling datasets.

4. Discussion

The main finding in this study is the high prediction capacity of this symptoms-based model. Our model correctly classified 8 out of 10 participants with SARS-CoV-2 with an AUC of 0.73. We found that an individual with anosmia alone, has over seven times the risk of having a

Table 2

Logistic regression with undersampling dataset (n = 14,475).

Variable	Unadjusted OR	95% CI	p-value	Adjusted OR	95% CI	p-value
Age	0.99	[0.99, 1.00]	.108	1.00	[0.99, 1.01]	.052
Socioeconomic strata						
Low-low	3.71	[2.42, 5.70]	< .001*	3.16	[1.98, 5.02]	< .001
Low	3.13	[2.16, 4.52]	< .001*	2.70	[1.82, 5.02]	< .001
Middle-low	2.74	[1.91, 3.95]	< .001*	2.27	[1.81, 4.01]	< .001
Middle	1.37	[0.93, 2.02]	.103	1.22	[0.81, 1.84]	.337
Middle-high	1.30	[0.84, 2.02]	.231	1.27	[0.79, 2.01]	.332
Contact with confirmed COVID-19	1.62	[1.45, 1.82]	< .001*	1.27	[1.12, 1.46]	< .001*
Arterial hypertension	0.78	[0.64, 0.96]	.020*	0.79	[1.12, 1.46]	.058
Symptoms related to SARS-CoV-2						
Fever	15.75	[11.81, 21.01]	< .001*	4.29	[3.07, 6.02]	< .001*
Anosmia	17.19	[13.97, 20.91]	< .001*	7.76	[6.19, 9.73]	< .001*
Dry cough	8.02	[6.90, 9.33]	< .001*	2.96	[2.44, 3.58]	< .001*
Fatigue	7.09	[6.03, 8.33]	< .001*	1.93	[1.57, 2.93]	< .001*
Headache	4.85	[2.80, 8.37]	< .001*	3.29	[1.78, 6.07]	< .001*

*p value < 0.05.

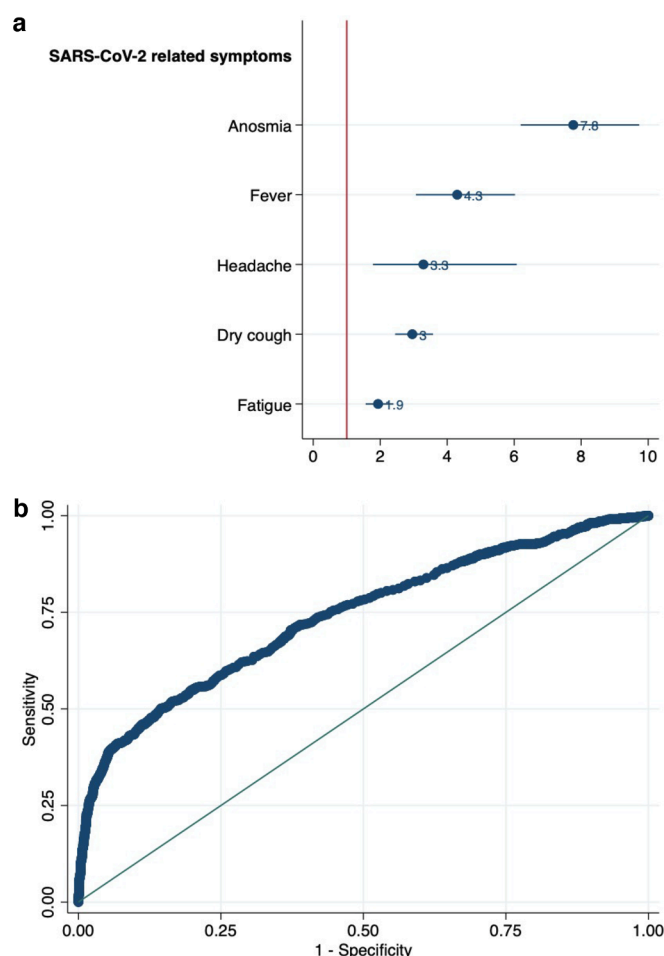


Fig. 2. Logistic regression model obtained under BIC criterion. a) Forest plot for association between sociodemographic characteristics, COVID-19 related symptoms and SARS-CoV-2 positive RT-PCR test; b) ROC curve for prediction of SARS-CoV-2 positive RT-PCR test result.

positive RT-PCR test for SARS-CoV-2 relative to those without this symptom. Fever, headache, fatigue, and dry cough were also important symptoms in predicting SARS-CoV-2 infection. Combining classical statistical methods such as logistic regression and more modern ones such as machine learning with RF, SVM, and XG boosting, aided both in robustness and interpretation for the results. A model that uses symptoms to predict SARS-CoV-2 infection may aid clinical decision-making in robustness and interpretation for the results. A model that uses symptoms to predict SARS-CoV-2 infection may aid clinical decision-making in robustness and interpretation for the results. A model that uses symptoms to predict SARS-CoV-2 infection may aid clinical decision-making in robustness and interpretation for the results.

Anosmia has been reported as a common symptom in COVID-19 patients (Lechien et al., 2020; Saniasiaya et al., 2021). This symptom has been the main feature addressed by several prediction models (Callejon-Leblic et al., 2021; Lechien et al., 2020; Saniasiaya et al., 2021). While most studies have used a binary categorization for anosmia, some studies used a visual analog scale (VAS) assessment of symptoms resulting in better diagnostic performance (Callejon-Leblic et al., 2021; Gerkin et al., 2021). Anosmia has also been used to predict local healthcare stress and to develop public health policies to prevent strain in healthcare systems (Pierron et al., 2020).

Despite the potential of anosmia alone in predicting a positive SARS-CoV-2 test result, we found that a combination of symptoms (anosmia, fever, headache, dry cough, and fatigue) and clinical features provided the best accuracy. The study conducted by Menezes et al. used conditional inference tree analyses to identify which combinations of symptoms were most likely to predict positive test results. Changes in smell or taste, fever, and body aches had the best diagnostic performance (Menezes et al., 2021). Other symptoms such as cough, fever, fatigue, malaise or body aches, sore throat, and headache have also been included in various models. In a systematic review evaluated the diagnostic accuracy of over 80 signs and symptoms related to SARS-CoV-2 infection, anosmia alone, ageusia alone, and anosmia and ageusia combined had the best results, including sensitivities below 50% but specificities over 90%. Various combinations of symptoms assessed in this review, primarily including fever and cough with other symptoms, had an SP above 80% but at the cost of very low SE (<30%) (Struyf et al., 2020).

While our model had an AUC of 0.73, with an SP of 98% and PPV of 73%, the diagnostic performance of the models reported in the literature

Table 3
Diagnostic performance of prediction models.

Dataset	Variable	AUC	SE	SP	PPV	NPV	Correctly classified
Complete data set	Logistic regression	.71	.10	.99	.57	.95	.95
	RF	.66	.04	.99	.54	.94	.95
	SVM	.59	.02	.99	.48	.95	.94
	XG boosting	.73	.07	.99	.55	.95	.95
Undersampling	Logistic regression	.73	.26	.98	.73	.86	.85
	RF	.81	.28	.98	.72	.86	.85
	SVM	.73	.29	.98	.74	.86	.85
	XG boosting	.77	.09	.99	.88	.83	.84
SMOTE	Logistic regression	.72	.22	.94	.71	.66	.67
	RF	.87	.69	.96	.91	.83	.86
	SVM	.66	.24	.94	.71	.66	.67
	XG boosting	.90	.98	.99	.97	.81	.86
ROSE	Logistic regression	.74	.47	.88	.79	.62	.67
	RF	.81	.65	.84	.80	.71	.75
	SVM	.73	.47	.88	.79	.62	.67
	XG boosting	.77	.57	.81	.76	.65	.69

has varied highly. Other studies reported highly accurate models, including only symptoms-related variables with the high AUC (Callejon-Leblic et al., 2021; Roland et al., 2020). Other models that have included symptoms such as fatigue, cough, fever, and respiratory and gastrointestinal symptoms have had slightly lower accuracy (La Torre et al., 2020; Lan et al., 2020; Menni et al., 2020; Tudrej et al., 2020).

Diagnostic imaging has also been used for COVID-19 prediction diagnosis. Deep learning models using X-rays (XR) and computed tomography (CT) radiomics have shown accuracy with AUC ranging from 0.87 (Wang et al., 2021) to 0.99 (Chen et al., 2021; Li et al., 2020a; Singh et al., 2022; 2021). Other models, using radiomics and clinical information to predict severity of the disease, had higher AUC (0.897 compared to 0.847 of radiomics and 0.767 clinical variables alone) (Purkayastha et al., 2021). However, accessibility to XR or CT is problematic in low resourced settings. Countries like Colombia, have highly dispersed rural areas where primary care does not include clinical or imaging testing for SARS-CoV-2. In highly populated cities such as Bogotá, healthcare infrastructure remains insufficient, especially for management of mild cases that do not require fixed medical attention but may overwhelm healthcare facilities for more severe cases.

Classical epidemiological methods may be limited for assessing prediction models for larger populations, especially when the low prevalence of the outcome causes an imbalance in the response variable. ML methods have been applied in population-based systems to complement classical statistical methods given their capabilities of processing large datasets, detecting patterns, and analyzing trends to provide important information to public health (Zeng et al., 2021). Recent literature has demonstrated the use of ML in research during the COVID-19 pandemic; specifically, on its advantages in diagnostics in middle-low- and low-income countries (Li et al., 2020c; Naseem et al., 2020). In fact, ML methods applied in this paper are highly used in tabular data, including XG boosting with promising results in diagnostic medicine and clinical data on COVID-19 (Li et al., 2020c). We also included two traditional ML methods such as random forest and SVM that have previously shown appropriate diagnostic performance when assessing severity in the COVID-19 pandemic (Alotaibi et al., 2021; Kumar et al., 2021) and in other medical contexts such as genomics (Ogut et al., 2011). Nevertheless, methods such as logistic regression provide interpretation input to assess risk in a more intuitive way for clinicians and decision makers than ML methods. Hence, both methods can complement prediction in public health-related events such as COVID-19.

Other studies have approached COVID-19 with similar ML methods, including resampling strategies to address the imbalance in the response variables such as was our case. In the specific case of Dantas et al., a large sample and low outcome prevalence (positivity rate of 11.8%) led them to evaluate several methods to get to a better prediction model. In

their case, they used the up-sampling balancing strategy (Dantas et al., 2021). Other studies have used ML methods for prediction diagnosis. The model proposed by Zoabi et al. used a large dataset provided by the Israeli Ministry of Health and included symptoms such as cough, fever, headache, sore throat, and shortness of breath, as well as variables such as age, sex, and history of contact with a confirmed case of COVID-19 with an AUC of 0.86 (Zoabi et al., 2021). Nonetheless, this model did not include anosmia (unlike our study), which has proved to be a relevant symptom in other prediction models.

The main strengths of our study were the large sample size and the comparison of several statistical approaches to develop a symptoms-based prediction model. The inclusion of several symptoms in the questionnaire applied to the participants allowed assessing relevant clinical characteristics. Our model's high SP and NPV could be advantageous in redirecting the limited testing resources to those patients with a higher probability of having COVID-19. Such a strategy provides elements for clinical decision making, especially in countries such as Colombia that require diagnostic test results for case definition. In the case of saturation of the public health systems, a triage testing strategy could also be helpful in assessing which individuals have the highest probability of being infected.

Also, some limitations of our study should be considered. The first was that symptoms were self-reported via telephone survey. In this sense, the participant could have reported apparent COVID-19-related symptoms that were unspecific and biased symptoms by the information spread by media and other sources. Also, the test result and self-reported symptoms could not have reflected the specific moment of the infection and affected the model's accuracy.

5. Conclusions

Our study produced a symptoms-based model with multiple statistical approaches that correctly identified over 85% of participants. This model can be used to strengthen epidemiological surveillance protocols, to prioritize resource allocation related to COVID-19 diagnosis, to decide on early isolation, and for contact tracing strategies in individuals with a high probability of infection and before receiving a confirmatory test result. This strategy has public health and clinical decision-making significance in low- and middle-income settings like Latin America.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The CoVIDA study was funded through donors managed by the philanthropy department of Universidad de los Andes. We thank the CoVIDA logistic group, allies, and participants. Also, to Dr. German Otálora, head of Medical Department at Universidad de los Andes. We thank members of the CoVIDA working group, Paola Rueda, Daniela Rodríguez, Yessica Daniela Campaz Landazabal, Marylin Hidalgo, Juliana Quintero, Paola Betancourt, Pablo Rodríguez, Andrés Felipe Patiño, Jose David Pinzon Ortiz, Maribel Rincón, Alejandro Segura, Rodrigo Rodríguez and Leonardo Salas Zapata.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pmedr.2022.101798>.

Reference

- Alotaibi, A., Shiblee, M., Alshahrani, A., 2021. Prediction of severity of COVID-19-infected patients using machine learning techniques. *Computers* 10, 31. <https://doi.org/10.3390/computers10030031>.
- Callejon-Leblic, M.A., Moreno-Luna, R., Del Cuvillo, A., Reyes-Tejero, I.M., Garcia-Villar, M.A., Santos-Peña, M., Maza-Solano, J.M., Martín-Jiménez, D.L., Palacios-García, J.M., Fernández-Velez, C., González-García, J., Sánchez-Calvo, J.M., Solanellas-Soler, J., Sánchez-Gómez, S., 2021. Loss of smell and taste can accurately predict COVID-19 infection: a machine-learning approach. *J. Clin. Med.* 10, 570. <https://doi.org/10.3390/jcm10040570>.
- Chen, H.J., Mao, L., Chen, Y., Yuan, L., Wang, F., Li, X., Cai, Q., Qiu, J., Chen, F., 2021. Machine learning-based CT radiomics model distinguishes COVID-19 from non-COVID-19 pneumonia. *BMC Infect. Dis.* 21, 931. <https://doi.org/10.1186/s12879-021-06614-6>.
- Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., Mulders, D.G., Haagmans, B.L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J.-L., Ellis, J., Zambon, M., Peiris, M., Goossens, H., Reusken, C., Koopmans, M.P., Drosten, C., 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25, 1–13. <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>.
- Dantas, L.F., Peres, I.T., Bastos, L.S.L., Marchesi, J.F., de Souza, G.F.G., Gelli, J.G.M., Baiao, F.A., Maçaira, P., Hamacher, S., Bozza, F.A., Perra, N., 2021. App-based symptom tracking to optimize SARS-CoV-2 testing strategy using machine learning. *PLoS One* 16 (3), e0248920.
- Gerkin, R.C., Ohla, K., Veldhuizen, M.G., Joseph, P.V., Kelly, C.E., Bakke, A.J., Steele, K.E., Farruggia, M.C., Pellegrino, R., Pepino, M.Y., Bouysset, C., Soler, G.M., Pereda-Loth, V., Dibattista, M., Cooper, K.W., Croijmans, I., Di Pizio, A., Ozdener, M.H., Fjaeldstad, A.W., Lin, C., Sandell, M.A., Singh, P.B., Brindha, V.E., Olsson, S.B., Saraiva, L.R., Ahuja, G., Alwashahi, M.K., Bhutani, S., D'Errico, A., Fornazieri, M.A., Golebiowski, J., Dar Hwang, L., Öztürk, L., Roura, E., Spinelli, S., Whitcroft, K.L., Faraji, F., Fischmeister, F.P.S., Heinbockel, T., Hsieh, J.W., Huart, C., Konstantinidis, I., Menini, A., Morini, G., Olofsson, J.K., Philpott, C.M., Pierron, D., Shields, V.D.C., Voznessenskaya, V.V., Albayay, J., Altundag, A., Bensafi, M., Bock, M.A., Calcinoni, O., Fredborg, W., Laudamiel, C., Lim, J., Lundström, J.N., Macchi, A., Meyer, P., Moein, S.T., Santamaría, E., Sengupta, D., Rohlf, Dominguez, P., Yanik, H., Hummel, T., Hayes, J.E., Reed, D.R., Niv, M.Y., Munger, S.D., Parma, V., Boesveldt, S., de Groot, J.H.B., Dinnella, C., Freiherr, J., Laktonova, T., Marino, S., Monteleone, E., Nunez-Parra, A., Abdulrahman, O., Ritchie, M., Thomas-Danguin, T., Walsh-Messinger, J., Al Abri, R., Alizadeh, R., Bignon, E., Cantone, E., Paola Cecchini, M., Chen, J., Dolores Guàrdia, M., Hoover, K. C., Karni, N., Navarro, M., Nolden, A.A., Portillo Mazal, P., Rowan, N.R., Sarabi-Jamab, A., Archer, N.S., Chen, B., Di Valerio, E.A., Feeney, E.L., Frasnelli, J., Hannum, M.E., Hopkins, C., Klein, H., Mignot, C., Mucignat, C., Ning, Y., Ozturk, E. E., Peng, M., Saatci, O., Sell, E.A., Yan, C.H., Alfaro, R., Cecchetto, C., Coureaud, G., Herriman, R.D., Justice, J.M., Kaushik, P.K., Koyama, S., Overdevest, J.B., Pirastu, N., Ramirez, V.A., Roberts, S.C., Smith, B.C., Cao, H., Wang, H., Balungwe Birindwa, P., Baguma, P., 2021. Recent smell loss is the best predictor of COVID-19 among individuals with recent respiratory symptoms. *Chem. Senses* 46, 1–12. <https://doi.org/10.1093/chemse/bjaa081>.
- Hincapié, M.A., Gallego, J.C., Gempeler, A., Piñeros, J.A., Nasner, D., Escobar, M.F., 2020. Implementation and usefulness of telemedicine during the COVID-19 pandemic: a scoping review. *J. Prim. Care Community Health* 11, 215013272098061. <https://doi.org/10.1177/2150132720980612>.
- Instituto Nacional de Salud, 2020. Instructivo para la vigilancia en salud pública intensificada de infección respiratoria aguda asociada al nuevo coronavirus 2019 (COVID-19). Bogotá, Colombia.
- Johns Hopkins University, 2022. COVID-19 Map - Johns Hopkins Coronavirus Resource Center [WWW Document]. accessed 3.11.22. <https://coronavirus.jhu.edu/map.html>.
- Kretzschmar, M.E., Rozhnova, G., Bootsma, M.C.J., van Boven, M., van de Wijgert, J.H.H. M., Bonten, M.J.M., 2020. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Heal.* 5, e452–e459. [https://doi.org/10.1016/S2468-2667\(20\)30157-2](https://doi.org/10.1016/S2468-2667(20)30157-2).
- Kumar, V., Singh, D., Kaur, M., Damaševičius, R., 2021. Overview of current state of research on the application of artificial intelligence techniques for COVID-19. *PeerJ Comput. Sci.* 7, e564. <https://doi.org/10.7717/peerj-cs.564>.
- La Torre, G., Massetti, A.P., Antonelli, G., Fimiani, C., Fantini, M., Marte, M., Faticoni, A., Previte, C.M., Turriziani, O., Pugliese, F., Villari, P., Romano, F., Mastroianni, C.M., 2020. Anosmia and ageusia as predictive signs of COVID-19 in healthcare workers in Italy: a prospective case-control study. *J. Clin. Med.* 9, 2870. <https://doi.org/10.3390/jcm9092870>.
- Laajaj, R., Webb, D., Aristizabal, D., Behrentz, E., Bernal, R., Buitrago, G., Cucunubá, Z., de la Hoz, F., Gaviria, A., Hernández, L.J., De Los Rios, C., Ramírez Varela, A., Restrepo, S., Schady, N., Vives, M., 2021. Understanding how socioeconomic inequalities drive inequalities in SARS-CoV-2 infections [WWW Document]. Univ. los Andes, Repos (accessed 1.21.22).
- Lan, F.Y., Lan, F.Y., Filler, R., Filler, R., Mathew, S., Buley, J., Iliaki, E., Iliaki, E., Bruno-Murtha, L.A., Osgood, R., Osgood, R., Christophi, C.A., Christophi, C.A., Fernandez-Montero, A., Fernandez-Montero, A., Kales, S.N., Kales, S.N., 2020. COVID-19 symptoms predictive of healthcare workers' SARS-CoV-2 PCR results. *PLoS One* 15, 1–12. <https://doi.org/10.1371/journal.pone.0235460>.
- Lau, L.L., Hung, N., Wilson, K., 2020. COVID-19 response strategies: considering inequalities between and within countries. *Int. J. Equity Health* 19, 137. <https://doi.org/10.1186/s12939-020-01254-9>.
- Lechien, J.R., Chiesa-Estomba, C.M., De Siati, D.R., Horoi, M., Le Bon, S.D., Rodríguez, A., Dequanter, D., Blecic, S., El Afia, F., Distinguin, L., Chekkoury-Idrissi, Y., Hans, S., Delgado, L.L., Calvo-Henriquez, C., Lavigne, P., Falanga, C., Barillari, M.R., Cammaroto, G., Khalife, M., Leich, P., Souchay, C., Rossi, C., Journe, F., Hsieh, J., Edjlali, M., Carlier, R., Ris, L., Lovato, A., De Filippis, C., Coppee, F., Fakhry, N., Ayad, T., Saussez, S., 2020. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. *Eur. Arch. Otorhinolaryngol.* 277, 2251–2261. <https://doi.org/10.1007/s00405-020-05965-1>.
- Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., Cao, K., Liu, D., Wang, G., Xu, Q., Fang, X., Zhang, S., Juan, X., Jun, X., 2020a. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology* 296, E65–E71. <https://doi.org/10.1148/radiol.2020200905>.
- Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020b. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 368 (6490), 489–493. <https://doi.org/10.1126/science.abb3221>.
- Li, W.T., Ma, J., Shende, N., Castaneda, G., Chakladar, J., Tsai, J.C., Apostol, L., Honda, C.O., Xu, J., Wong, L.M., Zhang, T., Lee, A., Gnanasekar, A., Honda, T.K., Kuo, S.Z., Yu, M.A., Chang, E.Y., Rajasekaran, M.R., Ongkeko, W.M., 2020c. Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis. *BMC Med. Inform. Decis. Mak.* 20, 247. <https://doi.org/10.1186/s12911-020-01266-z>.
- Lindsey, C., Sheather, S., 2015. Best subsets variable selection in nonnormal regression models. *Stat. J.* 15 (4), 1046–1059.
- Long, Q.-X., Tang, X.-J., Shi, Q.-L., Li, Q., Deng, H.-J., Yuan, J., Hu, J.-L., Xu, W., Zhang, Y., Lv, F.-J., Su, K., Zhang, F., Gong, J., Wu, B., Liu, X.-M., Li, J.-J., Qiu, J.-F., Chen, J., Huang, A.-L., 2020. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* 26, 1200–1204. <https://doi.org/10.1038/s41591-020-0965-6>.
- McGarry, B.E., SteelFisher, G.K., Grabowski, D.C., Barnett, M.L., 2021. COVID-19 test result turnaround time for residents and staff in US nursing homes. *JAMA Intern. Med.* 181, 556. <https://doi.org/10.1001/jamainternmed.2020.7330>.
- Menezes, A.M.B., Victora, C.G., Hartwig, F.P., Silveira, M.F., Horta, B.L., Barros, A.J.D., Mesenburg, M.A., Wehrmeister, F.C., Pellanda, L.C., Dellagostin, O.A., Struchiner, C. J., Burattini, M.N., Barros, F.C., Hallal, P.C., 2021. High prevalence of symptoms among Brazilian subjects with antibodies against SARS-CoV-2. *Sci. Rep.* 11, 1–8. <https://doi.org/10.1038/s41598-021-92775-y>.
- Menni, C., Valdes, A.M., Freidin, M.B., Sudre, C.H., Nguyen, L.H., Drew, D.A., Ganesh, S., Varsavsky, T., Cardoso, M.J., El-Sayed Moustafa, J.S., Visconti, A., Hysi, P., Bowyer, R.C.E., Mangino, M., Falchi, M., Wolf, J., Ourselin, S., Chan, A.T., Steves, C. J., Spector, T.D., 2020. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat. Med.* 26, 1037–1040. <https://doi.org/10.1038/s41591-020-0916-2>.
- Mercer, T.R., Salit, M., 2021. Testing at scale during the COVID-19 pandemic. *Nat. Rev. Genet.* 22, 415–426. <https://doi.org/10.1038/s41576-021-00360-w>.
- Ministerio de Salud y Protección Social de Colombia, 2020. Lineamientos para el uso de pruebas diagnósticas de laboratorio durante la pandemia del SARS-COV-2 (COVID-19) en Colombia.
- Monkowski, D., Rhodes, L.V., Templer, S., Kromer, S., Hartner, J., Pianucci, K., Kincaid, H., 2019. A retrospective cohort study to assess the impact of an inpatient infectious disease telemedicine consultation service on hospital and patient outcomes. *Clin. Infect. Dis.*
- Naseem, M., Akhund, R., Arshad, H., Ibrahim, M.T., 2020. Exploring the potential of artificial intelligence and machine learning to combat COVID-19 and existing opportunities for LMIC: a scoping review. *J. Prim. Care Community Health* 11, 215013272096363. <https://doi.org/10.1177/2150132720963634>.
- National Health Services, 2021. How you get your coronavirus (COVID-19) test result - NHS [WWW Document]. accessed 12.14.21. <https://www.nhs.uk/conditions/coronavirus-covid-19/testing/test-results/how-you-get-your-test-result/>.

- Ogut, J.O., Piepho, H.-P., Schulz-Streeck, T., 2011. A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc.* 5, S11. <https://doi.org/10.1186/1753-6561-5-S3-S11>.
- Pierron, D., Pereda-Loth, V., Mantel, M., Moranges, M., Bignon, E., Alva, O., Kabous, J., Heiske, M., Pacalon, J., David, R., Dinnella, C., Spinelli, S., Monteleone, E., Farruggia, M.C., Cooper, K.W., Sell, E.A., Thomas-Danguin, T., Bakke, A.J., Parma, V., Hayes, J.E., Letellier, T., Ferdenzi, C., Golebiowski, J., Bensafi, M., 2020. Smell and taste changes are early indicators of the COVID-19 pandemic and political decision effectiveness. *Nat. Commun.* 11, 5152. <https://doi.org/10.1038/s41467-020-18963-y>.
- Purkayastha, S., Xiao, Y., Jiao, Z., Thepumnoeysuk, R., Halsey, K., Wu, J., Tran, T.M.L., Hsieh, B., Choi, J.W., Wang, D., Vallières, M., Wang, R., Collins, S., Feng, X., Feldman, M., Zhang, P.J., Atalay, M., Sebro, R., Yang, L., Fan, Y., Liao, W., Bai, H.X., 2021. Machine learning-based prediction of COVID-19 severity and progression to critical illness using CT imaging and clinical data. *Korean J. Radiol.* 22, 1213. <https://doi.org/10.3348/kjr.2020.1104>.
- Roland, L.T., Gurrola, J.G., Loftus, P.A., Cheung, S.W., Chang, J.L., 2020. Smell and taste symptom-based predictive model for COVID-19 diagnosis. *Int. Forum Allergy Rhinol.* 10 (7), 832–838.
- Saniasia, J., Islam, M.A., Abdullah, B., 2021. Prevalence of olfactory dysfunction in Coronavirus Disease 2019 (COVID-19): a meta-analysis of 27,492 patients. *Laryngoscope* 131, 865–878. <https://doi.org/10.1002/lary.29286>.
- Sauerbrei, W., Royston, P., 2011. In: *International Encyclopedia of Statistical Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 899–902.
- Singh, D., Kumar, V., Kaur, M., Jabarulla, M.Y., Lee, H.-N., 2021. Screening of COVID-19 suspected subjects using multi-crossover genetic algorithm based dense convolutional neural network. *IEEE Access* 9, 142566–142580. <https://doi.org/10.1109/ACCESS.2021.3120717>.
- Singh, D., Kumar, V., Kaur, M., Kumari, R., 2022. Early diagnosis of COVID-19 patients using deep learning-based deep forest model. *J. Exp. Theor. Artif. Intell.* 1–11. <https://doi.org/10.1080/0952813X.2021.2021300>.
- Struyf, T., Deeks, J.J., Dinnes, J., Takwoingi, Y., Davenport, C., Leeftang, M.M., Spijker, R., Hooft, L., Emperador, D., Dittich, S., Domen, J., Horn, S.R.A., Van den Bruel, A., 2020. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease. *Cochrane Database Syst. Rev.* 2021. <https://doi.org/10.1002/14651858.CD013665>.
- Tudrej, B., Sebo, P., Lourdaux, J., Cuzin, C., Floquet, M., Haller, D.M., Maisonneuve, H., 2020. Self-reported loss of smell and taste in SARS-CoV-2 patients: primary care data to guide future early detection strategies. *J. Gen. Intern. Med.* 35, 2502–2504. <https://doi.org/10.1007/s11606-020-05933-9>.
- Varela, A.R., Florez, L.J.H., Tamayo-Cabeza, G., Contreras-Arrieta, S., Restrepo, S.R., Laajaj, R., Gutierrez, G.B., Guevara, Y.P.R., Caballero-Díaz, Y., Florez, M.V., Osorio, E., Barbieri, I.S., Sanchez, D.R., Nuñez, L.L., Bernal, R., Oliveros, S.R., Zapata, L.S., Guevara-Suarez, M., Uribe, A.G., Behrentz, E., 2021. Factors associated with SARS-CoV-2 infection in Bogotá, Colombia: Results from a large epidemiological surveillance study. *Lancet Reg. Heal. - Am.* 2, 100048. <https://doi.org/10.1016/j.lana.2021.100048>.
- Wang, G., Liu, X., Shen, J., Wang, C., Li, Z., Ye, L., Wu, X., Chen, T., Wang, K., Zhang, X., Zhou, Z., Yang, J., Sang, Y., Deng, R., Liang, W., Yu, T., Gao, M., Wang, J., Yang, Z., Cai, H., Lu, G., Zhang, L., Yang, L., Xu, W., Wang, W., Olvera, A., Ziyar, I., Zhang, C., Li, O., Liao, W., Liu, J., Chen, W., Chen, W., Shi, J., Zheng, L., Zhang, L., Yan, Z., Zou, X., Lin, G., Cao, G., Lau, L.L., Mo, L., Liang, Y., Roberts, M., Sala, E., Schönlieb, C.-B., Fok, M., Lau, J.-Y.-N., Xu, T., He, J., Zhang, K., Li, W., Lin, T., 2021. A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images. *Nat. Biomed. Eng.* 5, 509–521. <https://doi.org/10.1038/s41551-021-00704-1>.
- Wang, S.-H., Attique Khan, M., Govindaraj, V., Fernandes, L.S., Zhu, Z., Zhang, Y.-D., 2022. Deep rank-based average pooling network for Covid-19 recognition. *Comput. Mater. Contin.* 70, 2797–2813. <https://doi.org/10.32604/cmc.2022.020140>.
- Zeng, D., Cao, Z., Neill, D.B., 2021. Artificial intelligence-enabled public health surveillance—from local detection to global epidemic monitoring and control, in: *Artificial Intelligence in Medicine*. Elsevier, pp. 437–453. <https://doi.org/10.1016/B978-0-12-821259-2.00022-3>.
- Zoabi, Y., Deri-Rozov, S., Shomron, N., 2021. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *NPJ Digit. Med.* 4, 3. <https://doi.org/10.1038/s41746-020-00372-6>.