Development and internal validation of a predictive model for COVID-19 mass screening based on symptoms and a simple olfactory test

Youcef Azeli (youcefazeli@gmail.com)
Sant Joan University Hospital and Sistema d'Emergencies Mèdiques de Catalunya
https://orcid.org/0000-0003-3558-124X

Alberto Fernández
Rovira i Virgili University https://orcid.org/0000-0002-1241-1646

Federico Capriles
Servei d'Urgències. Hospital Universitari Sant Joan

Wojciech Rojewski
Sant Joan University Hospital

Vanesa Lopez-Madrid
Sant Joan University Hospital

David Lissner
CUAP Reus. Gerència Territorial Camp de Tarragona, Institut Català de la Salut

Rosa Maria Serrano
Sant Joan University Hospital

Cristina Rey-Reñones
Unitat de Suport a la Recerca Tarragona-Reus, Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol),

Marta Civit
CUAP Reus, Institut Català de la Salut

Josefina Casellas
Sant Joan University Hospital

Abdel Ouahabi
Sant Joan University Hospital

Maria Foglia-Fernández
ORL Service Sant Joan University Hospital

Salvador Sarrá
Sant Joan University Hospital

Eduard Llobet
MINOS-EMaS, Universitat Rovira i Virgili https://orcid.org/0000-0001-6164-4342

Article

Keywords: testing, mass screening, COVID-19, COVIDGel Test

Posted Date: June 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-622297/v1

Abstract

The early detection of symptoms and rapid testing are the basis of an efficient screening strategy to control COVID-19 transmission. Most COVID-19 patients show olfactory dysfunction and in many cases this is the first symptom. This study aims to develop a machine learning COVID-19 predictive tool based on symptoms and a simple olfactory test, which consists of identifying the smell of an aromatized hydroalcoholic gel (CovidGel Test). A multi-centre population-based prospective study was carried out in the city of Reus (Catalonia, Spain). A total of 519 patients were included, 386 (74.4%) had at least one symptom and 133 (25.6%) were asymptomatic. A classification tree model including sex, age, relevant symptoms and the CovidGel Test results obtained a sensitivity of 0.97 (95% CI 0.91–0.99), a specificity of 0.39 (95% CI 0.34–0.44) and an AUC of 0.87 (95% CI 0.83–0.92). This shows that the CovidGel Test is a promising mass screening tool for predicting COVID-19.

Introduction

Since the first cases of severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) were diagnosed in December 2019 in the Chinese city of Wuhan, the coronavirus disease 2019 (COVID-19) has spread rapidly (1). The strategies applied in the vast majority of countries to control the virus have been ineffective. On 18 May, 2021, 162 million cases were diagnosed and 3.3 million deaths occurred worldwide (2). The first results concerning the safety and effectiveness of different types of vaccines have raised optimism in the scientific community due to the possibility of controlling COVID-19 (3). However, key elements that may affect the effectiveness of vaccines, such as the time needed for their global introduction, the duration of vaccine immunity, and the consequences of the SARS-COV-2 variants, especially in regions with fewer heath resources, are still unknown (4) (5). Recently the emergence of a new variant in India has brought the country's healthcare system to the edge of catastrophe (6).

The early detection of symptoms suggestive of infection, rapid and efficient testing, contact tracing and isolation are the basis of an effective screening strategy to control transmission of COVID-19 and decrease the disease burden on healthcare systems. The Achilles' heel in the fight against this disease is the large number of patients who are asymptomatic or have only a few symptoms that are difficult to differentiate from a common cold, but who are nonetheless able to transmit the disease (7)(8). It was estimated that 51.9% of SARS-COV-2 infected cases were asymptomatic or had only 1 or 2 symptoms suggestive of COVID-19 (9). The reference diagnostic tool for COVID-19 is reverse transcriptase polymerase chain reaction (RT-PCR). Its accessibility may be limited for low-resource healthcare systems and its cost and time requirements preclude its use as a mass triage tool. Recently a screening tool based on a machine learning model including clinical features and symptoms has been constructed to prioritize testing for COVID-19 (10). It was found that a predictive model for COVID-19 that included the combination of symptoms and wearable sensor data performed better than a model based on symptoms alone (11).

Olfactory dysfunction (OD) in COVID-19 patients has recently been described as one of the most prevalent symptoms (12) of the disease and could be used as a means of screening to help identify people who should self-isolate (13). A symptom predictive model for Covid-19 based on a smartphone app including age, sex, loss of smell and taste, persistent cough, severe fatigue and skipped meals obtained a sensitivity of 65% (14). At the time of diagnosis, a recent prospective study found that 31% of patients affected by COVID-19 presented OD (15). Between 11.8% and 23% of cases presented OD before any other symptoms (12) (16). A recent study showed that a simplified clinical test based on the University of Pennsylvania Smell Identification Test was able to detect unperceived OD in COVID-19 patients (15).

Hydroalcoholic gels are widely distributed as they are one of the main strategies for decreasing virus transmission (17). Fragrance essential oils such as lavender, eucalyptus and lemon make them more pleasant and can enhance their anti-viral effect (18)(19). These features make an aromatized hydroalcoholic gel a good candidate for being used as part of a simple, fast and cost-effective large-scale olfactory screening test.

The aim of this study was to develop and validate, using cross-validation techniques, a machine learning predictive model for COVID-19 mass screening using symptoms and a simple olfactory test based on an aromatized hydroalcoholic gel, which could be especially useful when testing resources are limited.

Results

Characteristics of the study population

During the study period 3788 patients underwent RT-PCR to diagnose COVID-19 at one of the study health centres. The inclusion of cases and RT-PCRs performed per week at the centres while participating in the study can be consulted in Fig. 1 of the supplementary material. A total of 626 patients were initially included in the study protocol. Of these, 107 patients were excluded because of incomplete data or exclusion criteria as shown in Fig. 1. The final analysis of the study included 519 patients, out of whom 341 patients (65.7%) were from primary care and 179 (34.3%) were from the hospital Emergency Department. According to the criteria for carrying out a RT-PCR test, 386 (74.4%) had at least one symptom suggestive of COVID-19, 118 (22.7%) were asymptomatic and were close contacts of a COVID-19 case, and 15 (2.9%) were asymptomatic and were tested for unknown reasons. A positive RT-PCR was found in 117 patients (22.5%) and a negative RT-PCR was found in 402 patients (77.5%).

The mean (SD) age of the study population was 42.3 (16.3) years, the age range was between 18 and 98 years and 48% were male. Table 1 shows the background and clinical characteristics of the study population. A greater percentage of males among the COVID-19 positive cases was found compared to the COVID-19 negative patients (58.1% vs. 45%) with an absolute difference of 13.09% (95% CI 2.92 to 23.27, P = 0.02).

Regarding medical background, a higher percentage of active smokers was found among COVID-19 negative patients (9.6% vs. 18%) with an absolute difference of 8.41% (95% CI -14.98 to -1.83; P = 0.04), and a higher percentage of patients treated with chronic corticotherapy was found among COVID-19 positive patients (9.6% vs. 1.3%) with an absolute difference of 8.3% (95% CI 2.81 to 13.79; P < 0.01).

The majority of the patients presented a mild disease severity, adding up to a total of 334 cases (64.5%). A higher percentage of moderate severity cases was found among COVID-19 positive patients (18.8 % vs. 2%). No patients died during hospital admission. The most frequent diagnosis in the total study population was upper respiratory tract infection (12.7%). Pneumonia was diagnosed in 18% of the COVID-19 positive patients and in 1.7% of the COVID-19 negative patients with an absolute difference of 16.2% (95% CI 9.1 to 23.2).

COVID-19 symptoms

The mean (SD) number of days of the symptom evolution was 5.8 (5.6) for the COVID-19 positive patients and 5.1 (12.1) for the COVID-19 negative patients with an absolute difference of 0.75 (95% CI -1.35 to 2.84; P = 0.48). The symptoms most strongly associated with COVID-19 were OD and GD (OR 5.78; 95% CI 2.76–12.12, P < 0.01 and OR 5.78; 95% CI 3.03–11.04, P < 0.01, respectively) followed by fever (OR 3.03; 95% CI 1.98–4.65, P < 0.01). Fever, dry cough, asthenia, myalgia, headache, diarrhoea, OD, and GD were the eight symptoms associated with COVID-19. Table 2 shows the reported symptoms and olfactory test results in the population study.

Olfactory test results and diagnostic values of symptoms and olfactory test

In the total population study, the *CovidGel test 1* was positive in 267 patients (51.4%) and negative in 252 patients (48.6%). Among patients with a positive olfactory test result 112 cases (41.9%) identified the gel smell as alcohol, 57 cases (21.3%) as cologne, 27 cases (10.1%) as aromatic herbs, 10 cases (3.7%) as non-citrus fruits (3.7%), 6 cases (2.2%) as alcoholic beverages (2.2%) and 22 cases (8.2%) as other responses. In 25 cases (9.4%) participants reported that they “didn’t smell anything at all” and in 8 cases a “don’t know” response (3%) was reported. Among patients with a negative olfactory test result, 207 cases (82.1%) identified the gel smell as lemon, 26 cases (10.3%) as citrus, 13 cases (5.1%) as orange, 2 cases (0.8%) as tangerine, 2 cases (0.8%) as citronella, and 2 cases (0.8%) as lime.

Among the 117 patients diagnosed with COVID-19, the CovidGel Test 1 was positive in 74 (63.2%) and among the 402 non-COVID-19 patients, the CovidGel Test 1 was positive in 193 patients (48%). A positive CovidGel test 1 was associated with COVID-19 (OR: 1.86; 95% CI 1.22–2.85, P < 0.01). A positive CovidGel Test 1 and 2 had a lower association with COVID-19 (OR: 1.78; 95% CI 1.17–2.69, P < 0.01). The response “do not smell anything at all” was strongly associated with COVID-19 (OR: 4.06; 95% CI 1.8–9.17). Table 3 shows the olfactory test results in relation to symptoms. Among the 13 asymptomatic COVID-19 positive patients, 10 (76.9%) had a positive CovidGel Test result. CovidGel Test 1 positive results or positive results for both the CovidGel Test 1 and 2 in asymptomatic patients were associated with COVID-19 (OR: 3.94; 95% CI 1.03–15.03 and OR: 6.19; 95% CI 1.62–23.73 respectively).

Table 4 shows the diagnostic values of the relevant symptoms, the combination of symptoms and olfactory test for predicting COVID-19.

Results of the machine learning predictive model

Table 5 shows the results of the different classification trees constructed with machine learning according to the variables introduced in the model. By only introducing the relevant symptoms into the model, the sensitivity was 0.86 (95 CI 0.79–0.92), the specificity was 0.37 (95% CI 0.33–0.42) and the AUC was 0.86 (0.81–0.9) for the total population study, and 0.97 (95%CI 0.92–0.99), 0.11 (95% CI 0.07–0.15) and 0.89 (95% CI 0.81–0.9) respectively for the symptomatic population. The sensitivity and specificity obtained was 0.94 (95%CI 0.88–0.98) and 0.32 (95% CI 0.28–0.37) when the olfactory test was introduced into the model for the total study population. The constructed sensitive classification tree only took into account the result of the CovidGel Test 1 and ignored the result of the CovidGel Test 2. Considering other clinical variables, the model also included sex and age, reaching a sensitivity of 0.97 (0.91–0.99), a specificity of 0.39 (0.34–0.44) and an AUC of 0.87 (95% CI 0.83–0.92) for the total study population, and 0.98 (95%CI 0.93-1), 0.31 (95% CI 0.26–0.37) and 0.89 (95% CI 0.84–0.93) for the symptomatic population, respectively.

The specific classification tree built took into account the relevant symptoms and age and obtained a sensitivity of 0.29 (0.21–0.38), a specificity of 0.95 (95% CI 0.92–0.97) and an AUC of 0.85 (0.8–0.89). The resulting sensitive and specific ROC curve is shown in Fig. 2 of the supplementary material.

Discussion

The combination of symptoms and a simple olfactory test based on identifying the smell of a hydroalcoholic gel made it possible to develop a predictive model with high sensitivity, which has important clinical implications.

A predictive model based on symptoms reported on a smartphone-based app obtained a sensitivity of 0.65 (95% CI 0.62–0.67) and an AUC of 0.76 (95% CI 0.74–0.78) to predict COVID-19 (14). Another predictive model using machine learning based on symptoms, gender, age and close contacts obtained a sensitivity between 0.85 and 0.87 depending on the possible working points and an AUC of 0.86 (95% CI 0.85–0.87) (10). The different results of our model, depending on the variables included, show similar or even higher diagnostic values with respect to those models proposed as population screening. The model presented has the advantage that it includes asymptomatic patients and does not include close contacts in its variables as this could be difficult to determine in a situation of community transmission. To our knowledge, this is the first model that includes an olfactory test built using a prospective population-based study.

The high sensitivity in our model was obtained thanks to the low false negative rate of the olfactory test among asymptomatic COVID-19 positive patients.

The CovidGel Test obtained a sensitivity almost twice as high as a more complex olfactory test for predicting COVID-19 based on identifying the smell of three scented paper strips and a 4-item scale intensity rate (15). In addition, the simplicity of the CovidGel test means it can be implemented as a self-test, making it a more suitable population screening olfactory test than any reported so far. The wide distribution of the CovidGel due to its low cost also contributes to improving the disease situational awareness of the population.

Our work has some limitations. A high percentage of patients identified the smell of the CovidGel as alcohol. The alcoholic matrix of the gel could hinder olfactory recognition, explaining the low specificity found. Moreover, the patient's capacity to identify smell may decrease in an uncomfortable situation. Thus, an emergency department scenario during a pandemic and the logical concerns about having COVID-19 may have had an influence on the olfactory test results. In order to improve the specificity of the CovidGel test our research group could initiate a study to enhance the performance of different scents in the alcohol matrix. The data collection of this study was robust, but in some cases the second CovidGel test was not performed, mainly due to the overload of nursing work. To minimize the missing data it was decided to exclude these patients from the final analysis after verifying that there was no impact on the results.

It is important to highlight that in our study no side effects related to the inhalation of the hydroalcoholic gel were reported. One study described that repeated exposure to a hydroalcoholic gel by inhalation does not increase blood ethanol levels (20). The side effects described in the literature are related to the occurrence of dermatitis or are due to the ingestion of the gel (21) (22).

There are diverse possible applications of our predictive model. In a situation of community transmission, the early detection of symptoms suggestive of infection and the study of close contacts are the basis of case detection and would reduce the burden of COVID-19 on the healthcare system. Mass testing strategies using diagnostic tests such as antigen tests or PCR are used in situations where the incidence of the disease is high; however, these are difficult to organize and have a high cost and therefore cannot be maintained over long time periods. Our predictive model could be useful to quickly rule out non-infected patients and for selecting the population that could benefit from a more expensive diagnostic test such as antigen testing or PCR. It could also be especially useful for controlling transmission in those regions where testing resources are limited due to scarce economic resources or logistical difficulties.

This predictive model will be the basis of an online questionnaire that can be used as a mass screening, which has been patented (EP 21 382 524.3) and is available upon request. This questionnaire could include a question about the possibility of being a close contact of a COVID-19 case in order to adapt to local health policies. The CovidGel Test can be implemented in a small format using sachets for individual use or in a dispenser, so it could be used as a screening test prior to granting access to a public building or workplace. The introduction of different scents depending on the local culture could be useful for facilitating the reutilization of the predictive tool. The effectiveness of its implementation in different

settings should be tested by performing external validations; therefore, the collaboration of the scientific community is encouraged.

Conclusion

A machine learning predictive model for COVID-19 using symptoms and a simple olfactory test based on an aromatized hydroalcoholic gel showed high sensitivity for diagnosing COVID-19. The capacity of this predictive model to detect infected SARS-COV-2 patients among asymptomatic patients makes it a promising tool for the fight against COVID-19. This predictive model could be especially useful for mass screening when testing resources are limited.

Methods

Study design and setting

This is a population-based prospective cohort study conducted following the transparent TRIPOD Statement for transparent reporting of multivariable prediction models (23). The study was carried out in the Emergency Department of Sant Joan University Hospital of Reus, which is the reference hospital of the region, and in all five primary care centres of the public health network distributed in five basic health areas of the city of Reus.

The municipality of Reus is located in the Mediterranean area, has a surface area of 52.82 km^2 and at the beginning of 2020 it had a population of 106,168 inhabitants and a density of 2010.0 inhabitants/km^2 (24). This study was approved by the Ethics Committee of the Pere i Virgili Health Research Institute (Ref: 120/2020) and the IDIAP Jordi Gol Clinical Research Ethics Committee (Codi: 20/114-PCV). This study did not receive funding. All study participants were required to sign an informed consent form.

Participants

The study included consecutive patients undergoing RT-PCR for the first time to rule out COVID-19 infection who consulted the hospital emergency department or their primary care centre between 15 June and 11 September, 2020. Patients were tested for presenting symptoms suggestive of COVID-19 or for being close contacts of a confirmed COVID-19 case. Close contacts were considered those persons who had shared an area with a positive case at a distance of less than 2 meters, for more than 15 minutes, without protection and from 48 hours prior to the onset of symptoms.

The study did not include patients under 18 years of age, patients who did not sign the informed consent form, and patients with pathologies or conditions that may interfere with the olfactory function, such as any degree of cognitive impairment, Parkinson's disease, chronic rhinosinusopathy, head trauma, nasal obstruction, treatment with high concentrations of oxygen, acute respiratory failure, patients with an altered state of consciousness, or who use inhaled corticosteroids.

CovidGel test development

A multidisciplinary cooperation was established for creating a hydro-alcoholic hand sanitizing gel that meets current requirements in terms of its composition (25).

Based on the literature and habits of our Mediterranean study population, it was determined that the most suitable odoriferous substance was lemon (26). Tests were carried out with different concentrations of lemon essential oil and lemon fragrances of chemical origin. The composition of the gel was adapted to attenuate the smell of alcohol. A study was carried out to determine the most effective composition with and without thickener. Gas chromatography and mass spectrometry were used to obtain semi-quantitative results. A headspace sampling technique was used to establish the effectiveness of the volatile odoriferous substance that evaporated from the hydroalcoholic gel at 37 ºC. Finally, two hydroalcoholic gels with increasing concentrations of lemon essential oil were created as an olfactory test.

Description of the olfactory test

The olfactory test was performed by appropriately trained primary care and emergency nurses before the sample for SARS-COV-2 RT-PCR was collected. Therefore, both the patient and the healthcare personnel did not know the patient's infection status. Firstly, the test consisted of applying 1 ml of 0.3% CovidGel using a dispenser onto the patient's palm. Then the patient rubbed the gel into their hands and waited for 3 seconds. The patient was then asked to smell their hands and to "please, identify the smell of this gel". The answer was recorded on the basic data collection sheet regardless of the result. If the answer was not lemon or if it was inconclusive, the same test was repeated after 30 seconds with the 0.5% CovidGel. The olfactory test was considered negative if the patient recognized a citrus fruit, and the olfactory test was considered positive if the patient could not smell the gel or did not recognize a citrus fruit.

Data collection

A data collection sheet was completed by the attending nurse before taking the sample for the RT-PCR Test. It included the results of the two olfactory tests when both were performed. It also included age, gender, duration of symptoms (in days), and a yes/no questionnaire to check for symptoms such as fever, dry cough, dyspnoea, anorexia, myalgia, headache, diarrhoea, asthenia, productive cough, sore throat, OD or gustatory dysfunction (GD), others or no symptoms. The RT-PCR test for detecting SARS-COV-2 was considered the gold standard for diagnosis. During our study, the RT-PCR was performed by trained personnel according to the technical considerations of the manufacturer using a double sampling of the pharynx and the nose. The conservation of the sample and the transfer to the laboratory followed the channels of the usual clinical practice of the centre. RT-PCR tests were carried out with the VIASURE SARS-COV-2 Real Time PCR Detection Kit (CerTest Biotec, Zaragoza, Spain), or with the Procleix1 method in a Panther automated extractor and amplifier (Grifols Laboratories, Barcelona, Spain). Once all the data collection sheets were completed, the medical digital records were consulted and the RT-PCR test results were recorded, as well as the patient's background, evolution and discharge diagnosis. Regarding the severity of the disease, the patients attended and discharged immediately were considered as mild, those admitted to the hospital as moderate and those requiring ICU during hospitalization as severe.

This study was conducted at the beginning of the second wave of COVID-19 in our region (27). The 14-day cumulative incidence of COVID-19 cases in the city of Reus increased gradually from 0.9 cases/100,000 inhabitants on 15 June to 376.09 cases/100,000 inhabitants on 24 August (28).

Statistical analysis

The quantitative variables used in this study were described using the mean, the standard deviation, the median and the first and third quartiles. The differences between means and their corresponding 95% confidence interval (CI) were also used to compare groups of patients. Categorical variables were described using the number of cases, percentages and 95% CI. Comparisons between groups of patients were performed using Student's T test for quantitative variables, while the chi-squared test was used for categorical variables. Groups of patients were also compared in terms of the risk difference and odds ratio of the binary variables, and their corresponding 95% CI. All tests were two-tailed and p-values lower than 0.05 were considered statistically significant.

Diagnostic values in terms of sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio and negative likelihood ratio, as well as their corresponding 95% CI, were calculated for the binary variables and smell tests.

The symptoms that proved to be statistically significant in a logistic regression predictive model, were fever, dry cough, myalgia, headache, diarrhoea, asthenia, altered sense of smell, and altered sense of taste. These symptoms were defined as relevant. The productive cough variable was also included as a relevant symptom. The number of relevant symptoms was counted for each patient and this new variable was used to develop the model based on classification trees using a recursive partitioning algorithm (29). The internal model validation was carried out using cross validation techniques in machine learning, avoiding overfitting over-adjustment during the construction of the classification trees. These classification trees were built using the following parameters: the splitting index was the Gini coefficient; the minimum number of patients in any node of a tree for a split to be attempted was set at 30; the minimum number of patients in any terminal node of a tree was set at 10; node splits were only attempted if they improved the fit by a factor of 0.01; and the number of cross-validations to be run was set at 10.

In order to obtain different values of sensitivity and specificity in the resulting classification trees, distinct costs of false positives and false negatives were used in the loss matrix parameter of the classification tree algorithm. In particular, for the sensitive classification tree the cost for false negatives was set at 1:8, while the cost for false positives was kept constant at 1.

Several predictive models were analysed to handle missing data in the study protocol. A data-complete analysis was adopted over other strategies due to the low relevance of the missing data in the final results of the predictive machine learning model. All statistical analyses were performed using R software version 4.0.

Declarations

Acknowledgements

We would like to thank the company LUCTA for selflessly contributing to the creation and production of the CovidGel. Special thanks to Javier Marín Hernández, the Fragrance director of LUCTA. We greatly appreciate the support of the Ernesto Ventós Foundation. Thanks to the entire team of doctors, nurses and clinical assistants of the Emergency Department of the University Hospital Sant Joan de Reus who were able to carry out their work and also help in this research project at such a difficult time. Thanks to Mauricio Blotta, Leonor Arjona, Maria del Carmen Lareu, Luis Erwin Rodríguez, Sofia Ascanio, Laia Boters, Esther Duarte, Alba Diez, Ona Benach, Nuria Garcia, Andrea Simón, María Casanovas, Aurora del Pleguezuelos, María Francisca Mestre, Cristina Mañero, Anna María García, María Jesús Sevillano , María Teresa Xicola, Anna Baúles, Felicitas Garcia, Esther Lidia Marti, Tania Fabregat, Ana Dueñas, Mar Angels, Alba Rofes, Cristina Rodríguez, María Zaragoza, Carlos Olguin , Ana Català, Andrea Martos , Marta Pont, Camila Alejandra Hidalgo, Lourdes Fernández, Victoria Hidalgo, Marta Taberner, Giuliana Correa, Ana Vicente, Ignacio Escrigas, Jose Francisco López and Laia Mir . Thanks to all the collaborators from the Institut Català de la Salut: Jonathan Caselles Martinez, Judith Alonso Sansó, Raquel Navas Nuñez, Laia Sabaté Miró, Lucia Virginia Ndjoli Malonga, Laura Aubí Masip, Eva Torne Rivas, Maria Mercè Vizcarro López, Maria del Carme Barceló Prats, Maria Isabel Tomsen Fajardo, María Cinta Prats Bartra, Irene Pijuán Robledo, Maria Jesús Sánchez Lozano and Blanca Canela Llauradó. E.L. is a recipient of the 2018 ICREA Academia Award.

Funding

No funding was received for this work.

Author contributions

Y.A and F.C conceived the idea. Y.A and E.L. designed the study and the first CovidGel formula,

W.R., V. LM., D. L., RM. S., C. RR, M. C., J. C., A. O., M. FF., collected the study data. M. FF., C.RR and S.S. helped with technical support. Y.A, E.L. and A.F. analysed and interpreted the data results. A.F built the machine learning model. Y.A. wrote the manuscript. All authors revised the manuscript.

Competing interests

None to declare

References

1. Zhu, N., et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 382, 727-733 (2020).

2. World Health Organization. COVID-19 Weekly Epidemiological Update https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—18-may-2021(2021).

3. Rubin, E.J., Baden, L.R. & Morrissey, S. Audio Interview: Covid-19 Vaccine Fundamentals. N Engl J Medicine. 383, e146 (2020).

4. Speiser, D.E., & Bachmann, M.F. COVID-19: Mechanisms of Vaccination and Immunity. Vaccines. 8,404 (2020).

5. Lauring, A.S. & Hodcroft, E.B. Genetic Variants of SARS-CoV-2—What Do They Mean?, JAMA. 325,529-531(2021).

6. The Lancet. India's COVID-19 emergency. The Lancet. 397, 1683 (2021).

7. Rapid and frequent testing. Nat Biomed Eng. 4, 1121-1122 (2020).

8. Gandhi, M., Yokoe, D.S. & Havlir, D.V. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. N Engl J Med. 382, 2158-2160 (2020).

9. Pollán, M., et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. The Lancet. 396, 535-544 (2020).

10. Zoabi, Y., Deri-Rozov, S., & Shomron, N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. npj Digit Med. 4, 3 (2021).

11. Quer, G,, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med. 27, 73-77 (2021).

12. Lechien, J.R., et al. 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 of Otorhino-Laryngology. 277, 2251-2261(2020)

13. Spinato, G., et al. Alterations in Smell or Taste in Mildly Symptomatic Outpatients With SARS-CoV-2 Infection. JAMA. 323, 2089 (2020).

14. Menni, C., et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nature Medicine. 26, 1037-1040 (2020).

15. Villerabel, C., et al. Diagnostic Value of Patient-Reported and Clinically Tested Olfactory Dysfunction in a Population Screened for COVID-19. JAMA Otolaryngol Head Neck Surg. 147:271-279 (2021).

16. Kaye, R., et al. COVID-19 Anosmia Reporting Tool: Initial Findings. Otolaryngology–Head and Neck Surg. 163, 132-134 (2020).

17. Golin, A.P., Choi, D. & Ghahary, A. Hand sanitizers: A review of ingredients, mechanisms of action, modes of delivery, and efficacy against coronaviruses. American Journal of Infection Control. 48, 1062-1067 (2020).

18. Astani, A,, Reichling, J. & Schnitzler. P. Comparative study on the antiviral activity of selected monoterpenes derived from essential oils: Antiviral activity of monoterpenes derived from essential oils. Phytotherapy Research. 24, 673-679 (2010).

19. da Silva JKR, Figueiredo PLB, Byler KG, & Setzer WN. Essential Oils as Antiviral Agents, Potential of Essential Oils to Treat SARS-CoV-2 Infection: An In-Silico Investigation. International Journal of Molecular Sciences. 21, 3426 (2020).

20. Huynh-Delerme C, et al. Short Communication: Is Ethanol-Based Hand Sanitizer Involved in Acute Pancreatitis after Excessive Disinfection?—An Evaluation with the Use of PBPK Model. J Toxicol. 2012, 1-7 (2012);2012:1–7.

21. Emadi A & Coberly L. Intoxication of a Hospitalized Patient with an Isopropanol-Based Hand Sanitizer. N Engl J Med. 356, 530-531.

22. Quenan S & Piletta P. Hand dermatitis in healthcare workers: 15-years experience with hand sanitizer solutions. Contact Dermatitis. 84, 339-340 (2020).

23. Collins GS, Reitsma JB, Altman DG, & Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med. 13, 1 (2015).

24. IDESCAT. Statistical Institute of Catalonia. https://www.idescat.cat/emex/?id=431233&lang=en#h1f (2021)

25. European Chemicals Agency. Information on biocides https://echa.europa.eu/es/information-on-chemicals/biocidal-active-substances (2021)

26. Moein ST, et al.. Smell dysfunction: a biomarker for COVID-19. Int Forum of Allergy Rhinology. 10,944-950 (2020).

27. Iftimie S, et al. First and second waves of coronavirus disease-19: A comparative study in hospitalized patients in Reus, Spain. PLOS ONE.16, e0248029 (2021)

28. Gencat. Salut/Dades COVID. https://dadescovid.cat/setmanal?tipus_territori=territori&id_html=up_4_23&tipus=municipi&codi=43123 (2021)

29. Mola F. Classification and Regression Trees Software and New Developments. In: Rizzi A, Vichi M, Bock H-H, editors. Advances in Data Science and Classification [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 1998 [cited 2021 Apr 11]. p. 311–8. (Bock H-H, Gaul W, Opitz O, Schader M, editors.

Studies in Classification, Data Analysis, and Knowledge Organization).
http://link.springer.com/10.1007/978-3-642-72253-0_42

Tables

ble 1. Background and Clinical Characteristics of the Study
pulation

	Total	SARS-CoV2 positive	SARS-CoV2 negative	
riable	**N (%)**	N=117	N=402	**Absolute difference (95% CI). %**
nographic data				
le patients	249 (48)	68 (58.1)	181 (45)	13.09 (2.92 to 23.27)
e. Years	42.3 (16.3)	43.4 (15.95)	41.9 (16.3)	1.51 (-1.82 to 4.83)
kground				
pertension	96 (18.8)	24 (20.9)	72 (18.2)	2.64 (-5.7 to 10.99)
betes	39 (7.6)	9 (7.8)	30 (7.6)	0.23 (-5.33 to 5.79)
lipidaemia	57 (11.2)	16 (13.9)	41 (10.4)	3.53 (-3.47 to 10.54)
oking	82 (16.1)	11 (9.6)	71 (18)	-8.41 (-14.98 to -1.83)
olism	18 (3.5)	8 (7)	10 (2.5)	4.42 (-0.48 to 9.33)
onic bronchopathy	60 (11.8)	13 (11.3)	47 (11.9)	-0.59 (-7.2 to 6.02)
onic heart disease	28 (5.5)	5 (4.3)	23 (5.8)	-1.47 (-5.86 to 2.91)
oplasia	19 (3.7)	4 (3.5)	15 (3.8)	-0.32 (-4.16 to 3.52)
oimmune disease	14 (2.7)	6 (5.2)	8 (2)	3.19 (-1.1 to 7.49)
onic renal failure	4 (0.8)	0 (0)	4 (1)	-1.01 (-2 to -0.03)
onic liver disease	14 (2.7)	5 (4.3)	9 (2.3)	2.07 (-1.94 to 6.08)
othyroidism	26 (5.1)	3 (2.6)	23 (5.8)	-3.21 (-6.93 to 0.5)
esity	43 (8.3)	11 (9.4)	32 (8)	1.44 (-4.47 to 7.35)
onic cortico-therapy	16 (3.1)	11 (9.6)	5 (1.3)	8.3 (2.81 to 13.79)
nunosuppressive rapy	5 (0.98)	2 (1.7)	3 (0.8)	0.96 (-1.55 to 3.48)
ease severity				
d	334 (64.5)	74 (63.2)	260 (64.8)	-1.59 (-11.5 to 8.32)
derate	30 (5.8)	22 (18.8)	8 (2)	16.81 (9.6 to 24.02)
ere	6 (1.2)	2 (1.7)	4 (1)	0.71 (-1.83 to 3.25)
gen therapy	38 (7.34)	21 (17.9)	17 (4.2)	13.71 (6.48 to 20.94)
gnosis				
per respiratory tract ection	66 (12.7)	34 (29.1)	32 (8.0)	21.1 (12.5 to 29.7)
ver respiratory tract ection	22 (4.2)	2 (1.7)	20 (5.0)	-3.3 (-6.4 to -0.10)
eumonia	28 (5.4)	21 (18.0)	7 (1.7)	16.2 (31.5 to 50.5)

ues are median (Standard Deviation) and n (%).

ble 2. Patient reported COVID-19 symptoms and Olfactory Test results

	SARS-COV2 positive	SARS-COV2 negative		
	N=117	N=402	Odds Ratio (95% CI)	P-value
mptoms				
ver	59 (50.4)	101 (25.1)	3.03 (1.98-4.65)	0.00
cough	45 (38.5)	73 (18.2)	2.82 (1.8-4.42)	0.00
henia	34 (29.1)	60 (14.9)	2.33 (1.44-3.79)	0.00
algias	30 (25.6)	61 (15.2)	1.93 (1.17-3.17)	0.01
phalea	39 (33.3)	83 (20.6)	1.92 (1.22-3.03)	0.01
rrhoea	35 (29.9)	82 (20.4)	1.67 (1.05-2.65)	0.04
	19 (16.2)	13 (3.2)	5.79 (2.76-12.12)	0.00
	25 (21.4)	18 (4.5)	5.78 (3.03-11.04)	0.00
pnoea	23 (19.7)	54 (13.4)	1.58 (0.92-2.7)	0.13
ductive cough	15 (12.8)	40 (10)	1.33 (0.7-2.5)	0.48
e throat	31 (26.5)	96 (23.9)	1.15 (0.72-1.84)	0.65
norrhoea	6 (5.1)	9 (2.2)	2.36 (0.82-6.77)	0.18
orexia	10 (8.5)	30 (7.5)	1.16 (0.55-2.45)	0.85
mptomatic	13 (11.1)	120 (29.9)	0.29 (0.16-0.54)	0.00
mptoms combination				
and OD	31 (26.5)	24 (6)	5.66 (3.16-10.13)	0.00
ver and dry cough	75 (64.1)	146 (36.3)	3.13 (2.04-4.81)	0.00
ver. dry cough and OD	82 (70.1)	152 (37.9)	3.84 (2.46-5.98)	0.00
actory Test results				
t 1 positive	74 (63.2)	193 (48)	1.86 (1.22-2.85)	0.01
smell at all	13 (11.1)	12 (3)	4.06 (1.8-9.17)	0.00
t 1 and 2 positive	62 (52.9)	156 (38.8)	1.78 (1.17-2.69)	0.01

Olfactory disfunction; GD Gustatory disfunction; Values are n (%)

Table 3. Olfactory test in relation to symptoms

	SARS-COV2 positive	SARS-COV2 negative	Odds ratio (95% CI)
All patients	N=117	N=402	
Test 1 positive	74 (63.2)	193 (48)	1.86 (1.22-2.85)
No smell at all	13 (11.1)	12 (3)	4.06 (1.8-9.17)
Test 1 and 2 positive	62 (52.9)	156 (38.8)	1.78 (1.17-2.69)
Symptomatic patients	N=104	N=282	
Test 1 positive	64 (61.54)	138 (48.94)	1.67 (1.06-2.64)
No smell at all	12 (11.5)	12 (4.3)	2.93 (1.27-6.76)
Test 1 and test 2 positive	50 (50.0)	114 (40.4)	1.47 (0.94-2.32)
Asymptomatic patients	N= 13	N= 120	
Test 1 positive	10 (76.9)	55 (45.8)	3.94 (1.03-15.03)
No smell at all	1 (7.7)	0 (0)	-
Test 1 and 2 positive	10 (76.9)	42 (35)	6.19 (1.62-23.73)

Values are n (%)

e 4 Relevant symptoms and Olfactory Test diagnostic values

toms	Sensitivity (95% CI)	Specificity (95%CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR
r	0.50 (0.41-0.6)	0.75 (0.7-0.79)	0.37 (0.29-0.45)	0.84 (0.8-0.87)	2.01	0.66
ough	0.38 (0.3-0.48)	0.82 (0.78-0.85)	0.38 (0.29-0.48)	0.82 (0.78-0.86)	2.12	0.75
enia	0.29 (0.21-0.38)	0.85 (0.81-0.88)	0.36 (0.27-0.47)	0.8 (0.76-0.84)	1.95	0.83
gias	0.26 (0.18-0.35)	0.85 (0.81-0.88)	0.33 (0.23-0.44)	0.8 (0.76-0.83)	1.69	0.88
alea	0.33 (0.25-0.43)	0.79 (0.75-0.83)	0.32 (0.24-0.41)	0.8 (0.76-0.84)	1.61	0.84
hoea	0.3 (0.22-0.39)	0.8 (0.75-0.83)	0.3 (0.22-0.39)	0.8 (0.75-0.83)	1.47	0.88
	0.16 (0.1-0.24)	0.97 (0.95-0.98)	0.59 (0.41-0.76)	0.8 (0.76-0.83)	5.01	0.87
	0.21 (0.14-0.3)	0.96 (0.93-0.97)	0.58 (0.42-0.73)	0.81 (0.77-0.84)	4.76	0.82
toms ination						
nd GD	0.26 (0.19-0.35)	0.94 (0.91-0.96)	0.56 (0.42-0.7)	0.81 (0.78-0.85)	4.43	0.78
r and dry h	0.64 (0.55-0.73)	0.64 (0.59-0.68)	0.34 (0.28-0.41)	0.86 (0.81-0.9)	1.77	0.56
r. dry cough OD	0.7 (0.61-0.78)	0.62 (0.57-0.67)	0.35 (0.29-0.42)	0.88 (0.83-0.91)	1.85	0.48
tory test ts						
1 positive	0.63 (0.54-0.72)	0.52 (0.47-0.57)	0.28 (0.22-0.33)	0.83 (0.78-0.87)	1.32	0.71
1 and 2 ive	0.53 (0.44-0.62)	0.61 (0.56-0.66)	0.28 (0.23-0.35)	0.82 (0.77-0.86)	1.37	0.77

Positive predictive value; NPV Negative predictive value; PLR Positive
hood ratio; NLR Negative likelihood ratio

able 5. Results of machine learning classification trees

ensitive tree	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95%CI)	NPV (95%CI)	BA	F1	MCC	AUC (95% CI)
elevant symptoms	0.86 (0.79-0.92)	0.37 (0.33-0.42)	0.29 (0.24-0.34)	0.9 (0.85-0.94)	0.62	0.43	0.21	0.86 (0.81-0.9)
	0.97 (0.92-0.99)	0.11 (0.07-0.15)	0.29 (0.24-0.34)	0.91 (0.76-0.98)	0.54	0.44	0.12	0.89 (0.84-0.93)
elevant symptoms nd olfactory test	0.94 (0.88-0.98)	0.32 (0.28-0.37)	0.29 (0.24-0.34)	0.95 (0.9-0.98)	0.63	0.44	0.25	0.87 (0.83-0.92)
	0.96 (0.9-0.99)	0.23 (0.18-0.28)	0.32 (0.26-0.37)	0.94 (0.86-0.98)	0.60	0.48	0.22	0.89 (0.84-0.93)
elevant symptoms. lfactory test. sex and ge	0.97 (0.91-0.99)	0.39 (0.34-0.44)	0.31 (0.27-0.36)	0.97 (0.94-0.99)	0.68	0.47	0.32	0.87 (0.83-0.92)
	0.98 (0.93-1)	0.31 (0.26-0.37)	0.34 (0.29-0.4)	0.98 (0.92-1)	0.64	0.51	0.30	0.89 (0.84-0.93)
pecific tree								
elevant symptoms nd age	0.29 (0.21-0.38)	0.95 (0.92-0.97)	0.62 (0.48-0.75)	0.82 (0.78-0.85)	0.62	0.40	0.32	0.85 (0.8-0.89)
	0.33 (0.24-0.43)	0.93 (0.89-0.95)	0.62 (0.48-0.75)	0.79 (0.74-0.83)	0.63	0.43	0.32	0.82 (0.77-0.87)

AUC Area under the curve; BA Balanced Accuracy. MCC: Matthews correlation coefficient
The highlighted rows show symptomatic patients and the white rows show the total study population.

Figures

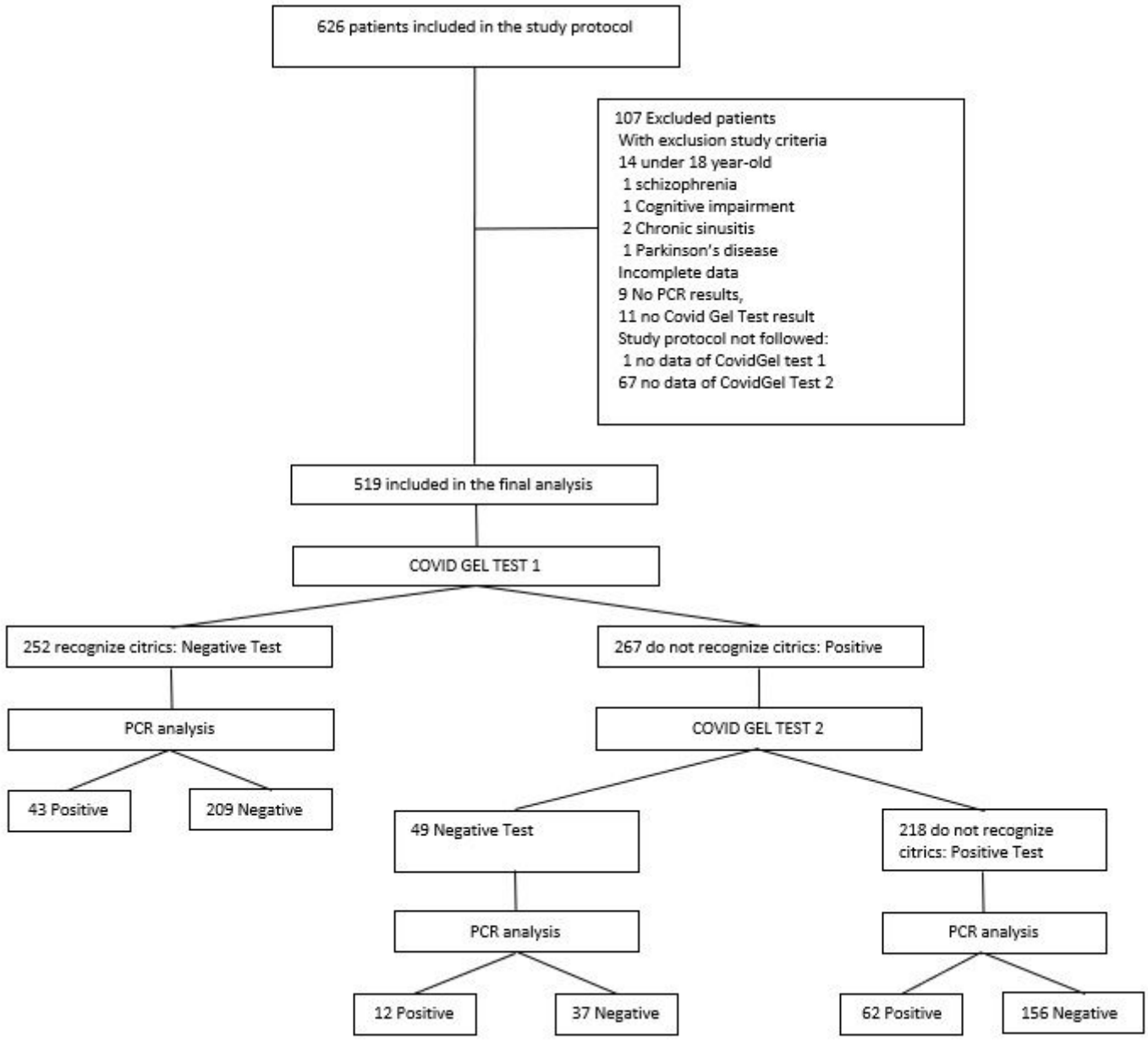

Figure 1

Flow Chart

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.pdf