

An online platform for COVID-19 diagnostic screening using a machine learning algorithm

Erito Marques de Souza Filho^{1*} , Rodrigo de Souza Tavares¹ , Bruno José Dembogurski¹ ,
Alice Helena Nora Pacheco Gagliano² , Luiz Carlos de Oliveira Pacheco² ,
Luiz Gabriel de Resende Nora Pacheco² , Filipe Braida do Carmo¹ , Leandro Guimarães Marques Alvim¹ ,
Alexandra Monteiro³ 

SUMMARY

OBJECTIVE: COVID-19 has brought emerging public health emergency and new challenges. It configures a complex panorama that has been requiring a set of coordinated actions and has innovation as one of its pillars. In particular, the use of digital tools plays an important role. In this context, this study presents a screening algorithm that uses a machine learning model to assess the probability of a diagnosis of COVID-19 based on clinical data.

METHODS: This algorithm was made available for free on an online platform. The project was developed in three phases. First, a machine learning risk model was developed. Second, a system was developed that would allow the user to enter patient data. Finally, this platform was used in teleconsultations carried out during the pandemic period.

RESULTS: The number of accesses during the period was 4,722. A total of 126 assistances were carried out from March 23, 2020, to June 16, 2020, and 107 satisfaction survey returns were received. The response rate to the questionnaires was 84.92%, and the ratings obtained regarding the satisfaction level were higher than 4.8 (on a 0–5 scale). The Net Promoter Score was 94.4.

CONCLUSION: To the best of our knowledge, this is the first online application of its kind that presents a probabilistic assessment of COVID-19 using machine learning models exclusively based on the symptoms and clinical characteristics of users. The level of satisfaction was high. The integration of machine learning tools in telemedicine practice has great potential.

KEYWORDS: COVID-19. Machine learning. Diagnosis. Telemedicine.

INTRODUCTION

The emerging public health emergency and the new challenges imposed by COVID-19 have configured a complex panorama¹ that has been requiring a set of coordinated actions and has innovation as one of its pillars. In particular, the use of digital tools plays an important role. The use of telemedicine was encouraged as an alternative form of care². On the contrary, artificial intelligence (AI) tools can be of fundamental value. This encompasses a complex framework of sophisticated mathematical-computational models that allow the construction of algorithms to be used to emulate the realization of various human tasks, such as pattern recognition, problem-solving, and language processing³. There is an AI subarea called machine learning (ML), which offers the possibility of learning from experience gained with large databases collected and properly processed³. Several applications using ML have been proposed

with very promising results^{4,5}. In this context, in this study, we developed an application software that uses an ML risk model to estimate the probability of a diagnosis of COVID-19 based on clinical data. This application was made available for free on an online platform.

Application software

In this section, a summary of the system—registered with the Brazilian Institute of Intellectual Property—is presented. It consists of four steps: (1) symptoms, in which a user inputs the patient's clinical data regarding symptoms; (2) radiography, an optional step, such that the user submits a frontal radiograph of the patient's chest; (3) general information, such that this information is related to the patient's demographic; and (4) an estimate, which presents a probability of a patient having COVID-19.

¹Universidade Federal Rural do Rio de Janeiro – Nova Iguaçu (RJ), Brazil.

²Serviços de Exames Ambulatoriais do Coração – Niterói (RJ), Brazil.

³Universidade do Estado do Rio de Janeiro – Rio de Janeiro (RJ), Brazil.

*Corresponding author: mederitomarques@gmail.com

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.

Received on October 24, 2022. Accepted on January 20, 2023.

Before entering the phases described previously, a screen with general information about the system is first presented. Upon entering the platform link (Figure 1), the user accesses a home screen, which has an option to click to find out more information about the system (“Learn More” button), which brings information about what the project is, its motivations, and specifics about the programming part or start entering your clinical information (“Start” button).

Symptoms

By clicking on “Start,” the patient has access to a screen, in which he must select the symptoms he has had in the past 15 days.

Radiography

At this stage, the user sends a photo of the patient’s radiograph. This should be a frontal chest X-ray. The image can be taken from a smartphone or in its original format. Some procedures are informed so that more than one image is not placed in the photo and that it is taken close and vertically.

General information

On the subsequent screen, users must inform about gender, age (if over 60 years), and the presence of comorbidities and must also accept the terms and conditions and the privacy policy. If you click on “Previous,” the user is taken to the screen before the one he is browsing.

By clicking on “Analyze,” the system outputs the probability that the patient in those conditions has the disease. Figure 2

shows an example of a user aged 60 years, without comorbidities who presented symptoms of arthralgia, hyposmia, cough, fever, and chills. The tool indicated an 80% probability that a user with these clinical characteristics had the disease.

METHODS

The project was developed in three phases. First, a risk model was developed. Second, a system was developed which would allow the user to enter patient data. Finally, this platform was used in teleconsultations carried out during the pandemic period.

Phase 1: machine learning risk predictor

The ML risk predictor works from the patient’s symptom characteristics and, optionally, with imaging characteristics of their lung radiograph. The risk model constructed is based on a Bayesian model. In the case of radiography data entry, ML is used to predict pulmonary anomalies. The information used for building a risk predictor was obtained from scientific articles, such as the selection of the most common symptoms and the presence of other important clinical characteristics. Regarding radiography, a convolutional neural network is trained for learning anomaly lung features from images of a public radiography database⁶. Submitting the chest X-ray, however, is optional and was not considered in phase 3. The evaluation of the model’s performance was performed using a set of traditional metrics that include sensitivity, specificity, accuracy, and the area under the receiver operating characteristic curve (AUC).

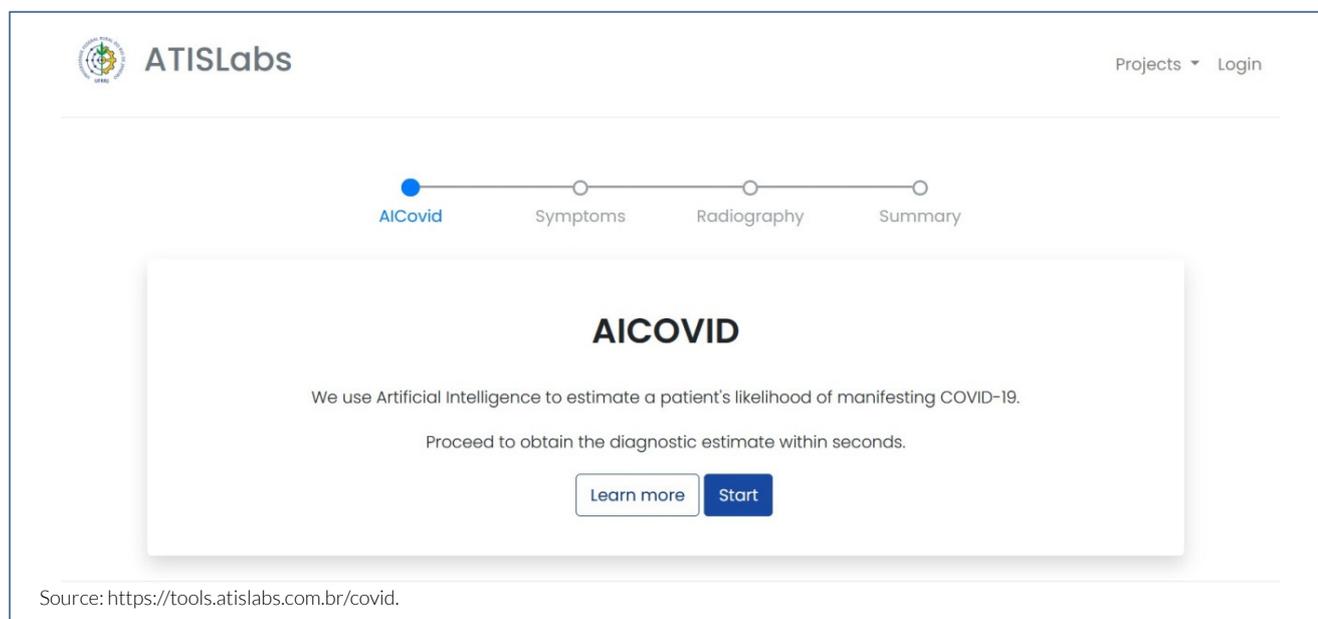


Figure 1. Machine learning tool presentation screen.

Phase 2: online platform development

In this phase, we develop an online platform freely available at <https://tools.atislabs.com.br/covid>. This platform can be consulted in both Portuguese and English. It allows the patient to report the symptoms presented in the last 15 days as well as their clinical data. It returns the probability of having the diagnosis of COVID-19. Probability value less than 50% was considered low probability, between 50 and 69% medium probability, and 70% and above high probability. The platform does not solicit or store user registration information, nor does it collect information from the devices or devices used to access the website and the application. However, the user is required to agree to its terms of use and privacy, which include the purpose of the tool, its limitations, and compliance with the General Data Protection Law. The authors declare no conflicts of interest. The study was conducted following the Declaration of Helsinki and followed ethical standards.

Phase 3: use of the platform in teleconsultation service

The ML tool was tested under the “Seacor Digital” project from March 23, 2020 to June 16, 2020. The project’s idea was to help patients receive medical care from a distance in order to avoid unnecessary crowds, which was mainly due to the fear of many patients when seeking in-person care. Teleconsultations were carried out on a platform developed for this purpose and lasted 30 min. If the physician felt the need, the patient was referred to the emergency room for immediate care or was instructed to

carry out a face-to-face consultation. In addition, consultation with specialists was carried out, if necessary. If COVID-19 was suspected, the ML tool was used to estimate the probability of a diagnosis of the disease. Patients with other clinical suspicions received general guidance on COVID-19 and were instructed to use the ML tool in case of symptoms. Patients were invited to answer an anonymous survey to assess their satisfaction with the project. They were asked to evaluate five aspects of care: ease of scheduling; the professional’s interest in their health status; punctuality of the health professional; resolution of the main complaint by the professional; and access to the virtual room (on a scale of 0–5). In this survey, it was also asked how much (on a scale of 0–10) the patient would recommend the project to a friend or relative. This allowed the calculation of the Net Promoter Score (NPS). It was published in 2003 by Frederick Reichheld⁷. The idea is simple and consists of asking the customer what grade he recommends the service of a particular company, on a scale of 0–10. A customer who gives a grade of 9 or 10 is called a promoter. Grade 7 or 8 was classified as neutral, and a score of 6 or below represents a detractor. The NPS is the result of the percentage difference between promoters and detractors.

RESULTS

The tool was advertised on the main television channels in Brazil. The number of accesses in the period was 4,722 (16,400 pageviews). In the “Seacor Digital” project, 126 assistances were

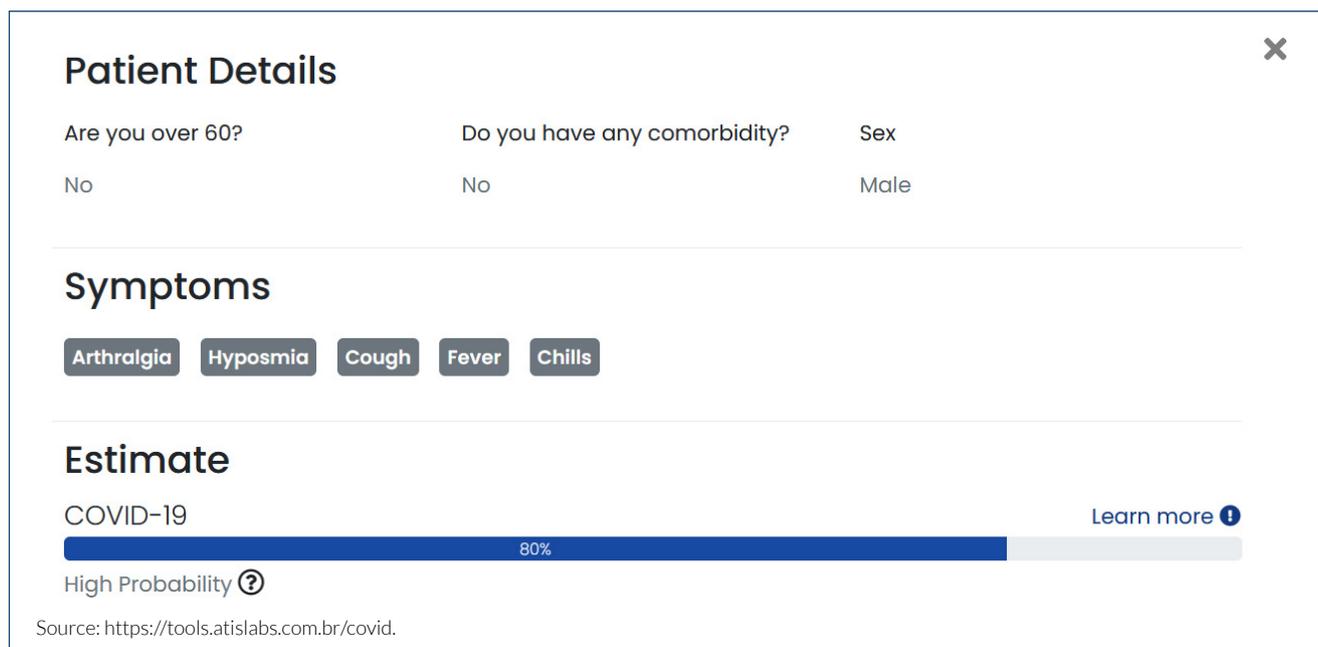


Figure 2. Output screen with probability estimation.

carried out and 107 satisfaction survey returns were received. Professional's interest in their health status and resolution of the main complaint received the maximum grade, while all other items received a score of 4.83 (ease of scheduling), 4.85 (access), and 4.92 (punctuality of the health professional). There were 101 promoters and 6 neutral. The NPS was 94.4 (Figure 3).

DISCUSSION

In the present study, a decision-support algorithm capable of evaluating the probability of a diagnosis of COVID-19 in patients with the suspected disease was developed. To the best of our knowledge, this is the first application of its kind that presents a probabilistic assessment of suspected disease using ML models exclusively based on the symptoms and clinical characteristics of users. The application of ML tools to estimate the probability of a disease has been of great value not only within the scope of this study but also in several different clinical contexts. In addition, the tool developed here does not require the use of adequate information for additional exams and, as it is free, allows people to use it free of charge if they have access to the internet. The user's exemption from filling in personal information such as name, telephone, and email, as well as the need to agree to the terms and conditions and the privacy policy add additional security concerning user anonymization and the security of sensitive data, which aligns the needs of transparency and compliance⁸⁻¹⁰.

In this context, the experience of applying the tool in a telemedicine context was very productive. The results obtained

were very promising. The response rate to the questionnaires was 84.92%, and the ratings obtained were higher than 4.8. The worst grades were related to connectivity and access issues. Some important variables in this process may include issues related to infrastructure and digital literacy. Smith and Magnani highlighted that people harmed by the social determinants of health may face difficulties in accessing digital health owing to the lack of means to do so. They also reiterated that in addition to this access barrier, the lack of adequate literacy can contribute to difficulties in understanding the content¹¹. On the contrary, the NPS achieved was high (94.4), which reflects a promoter score. Alismail et al., used NPS in an Outpatient Allergy and Pulmonary Clinic to evaluate its performance. They compared a Tablet-Based Tool (using NPS) and a Traditional Survey Method. The response rate was 37.9% (648 responses) versus 27% (156 responses). They both had similar outcomes regarding patient satisfaction. The NPS was also high and achieved 96%¹².

These experiences show that the use of NPS seems to have potential in a clinical context. Another merit of this study was the integration of ML tools in clinical practice since the tool could be tested in practice. The integration of algorithms into practice brings with it several challenges. In this context, this tool had the advantage of being conceived in such a way that it included an important intersection between the team that would be involved in carrying out the telemedicine care and the development team, which certainly contributed to its improvement. Taha et al., pointed out that telemedicine care can achieve the same status as face-to-face consultations from a quality point of view¹³. Despite this, it is important to emphasize that the study had noteworthy limitations. The first one is that the developed tool was validated in a context with a small number of patients in a retrospective study. The pandemic indeed posed severe difficulties for research, whether from a budgetary point of view or regarding its performance per se; however, it is important to increase the number of patients. Furthermore, validation takes place in a single center, which certainly raises the need for validation in other contexts.

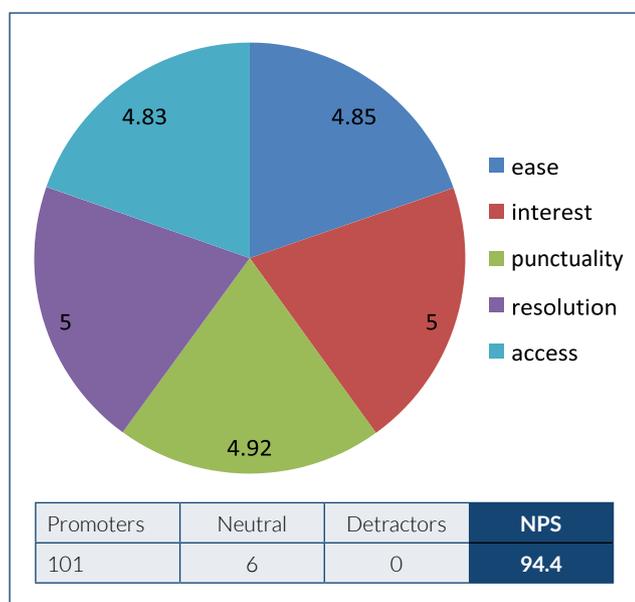


Figure 3. Evaluation of five aspects of care and Net Promoter Score.

CONCLUSION

Our study developed an online platform for screening patients with suspected signs and symptoms of a disease caused by COVID-19. This tool performs an automated calculation of infection risk probability using an ML algorithm after the user has entered the requested clinical information. It was used successfully in a telemedicine project during the pandemic. The results obtained were very promising (NPS=94.4). Despite the

various challenges present in the digital ecosystem, we believe that the integration of ML tools in telemedicine activities can bring important benefits to patients and contribute to the generation of value in the health chain.

AUTHORS' CONTRIBUTIONS

EMSF: Conceptualization, Data curation, Investigation, Methodology, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. **RST:** Conceptualization, Investigation, Methodology, Software, Writing – review & editing.

BJD: Data curation, Formal Analysis, Investigation, Software. **AHNPG:** Conceptualization, Methodology, Project administration, Writing – original draft. **LCOP:** Conceptualization, Data curation, Project administration. **LGRNP:** Conceptualization, Methodology, Project administration, Writing – original draft. **FBC:** Conceptualization, Data curation, Formal Analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **LGMA:** Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. **AM:** Methodology, Validation, Writing – original draft, Writing – review & editing.

REFERENCES

1. Provenzano BC, Bartholo T, Ribeiro-Alves M, Santos APGD, Mafort TT, Castro MCS, et al. The impact of healthcare-associated infections on COVID-19 mortality: a cohort study from a Brazilian public hospital. *Rev Assoc Med Bras* (1992). 2021;67(7):997-1002. <https://doi.org/10.1590/1806-9282.20210433>
2. Mintz J, Labiste C, DiCaro MV, McElroy E, Alizadeh R, Xu K. Teleophthalmology for age-related macular degeneration during the COVID-19 pandemic and beyond. *J Telemed Telecare*. 2020;28:1357633X20960636. <https://doi.org/10.1177/1357633X20960636>
3. Souza Filho EM, Fernandes FA, Soares CLA, Seixas FL, Santos AASMDD, Gismondi RA, et al. Artificial intelligence in cardiology: concepts, tools and challenges - "the horse is the one who runs, you must be the jockey". *Arq Bras Cardiol*. 2020;114(4):718-25. <https://doi.org/10.36660/abc.20180431>
4. Than MP, Pickering JW, Sandoval Y, Shah ASV, Tsanas A, Apple FS, et al. Machine learning to predict the likelihood of acute myocardial infarction. *Circulation*. 2019;140(11):899-909. <https://doi.org/10.1161/CIRCULATIONAHA.119.041980>
5. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. *J Am Heart Assoc*. 2018;7(13):e008678. <https://doi.org/10.1161/JAHA.118.008678>
6. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Illcus S, Chute C, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *arXiv [Preprint]*. 2019;arXiv:1901.07031. <https://doi.org/10.48550/arXiv.1901.07031>
7. Reichheld FF. The one number you need to grow. *Harv Bus Rev*. 2003;81(12):46-54, 124. PMID: 14712543
8. Hendl T, Chung R, Wild V. Pandemic surveillance and racialized subpopulations: mitigating vulnerabilities in COVID-19 apps. *J Bioeth Inq*. 2020;17(4):829-34. <https://doi.org/10.1007/s11673-020-10034-7>
9. Mbunge E. Integrating emerging technologies into COVID-19 contact tracing: opportunities, challenges and pitfalls. *Diabetes Metab Syndr*. 2020;14(6):1631-6. <https://doi.org/10.1016/j.dsx.2020.08.029>
10. Lucivero F, Hallowell N, Johnson S, Prainsack B, Samuel G, Sharon T. COVID-19 and contact tracing apps: ethical challenges for a social experiment on a global scale. *J Bioeth Inq*. 2020;17(4):835-9. <https://doi.org/10.1007/s11673-020-10016-9>
11. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. *Int J Cardiol*. 2019;292:280-2. <https://doi.org/10.1016/j.ijcard.2019.05.066>
12. Alismail A, Schaeffer B, Oh A, Hamiduzzaman S, Daher N, Song HY, et al. The use of the net promoter score (NPS) in an outpatient allergy and pulmonary clinic: an innovative look into using tablet-based tool vs traditional survey method. *Patient Relat Outcome Meas*. 2019;11:137-42. <https://doi.org/10.2147/PROM.S248431>
13. Taha AR, Shehadeh M, Alshehhi A, Altamimi T, Housser E, Simsekler MCE, et al. The integration of mHealth technologies in telemedicine during the COVID-19 era: a cross-sectional study. *PLoS One*. 2022;17(2):e0264436. <https://doi.org/10.1371/journal.pone.0264436>

