



# Developing a Training Web Application for Improving the COVID-19 Diagnostic Accuracy on Chest X-ray

P. Menéndez Fernández-Miranda<sup>1,6</sup> · P. Sanz Bellón<sup>1,6</sup> · A. Pérez del Barrio<sup>1,6</sup> · L. Lloret Iglesias<sup>2</sup> · P. Solís García<sup>3</sup> · F. Aguilar-Gómez<sup>2</sup> · D. Rodríguez González<sup>2</sup> · J. A. Vega<sup>4,5</sup>

Received: 19 May 2020 / Revised: 6 November 2020 / Accepted: 11 January 2021 / Published online: 8 March 2021  
© Society for Imaging Informatics in Medicine 2021

## Abstract

In December 2019, a new coronavirus known as 2019-nCoV emerged in Wuhan, China. The virus has spread globally and the infection was declared pandemic in March 2020. Although most cases of coronavirus disease 2019 (COVID-19) are mild, some of them rapidly develop acute respiratory distress syndrome. In the clinical management, chest X-rays (CXR) are essential, but the evaluation of COVID-19 CXR could be a challenge. In this context, we developed *COVID-19 TRAINING*, a free Web application for training on the evaluation of COVID-19 CXR. The application included 196 CXR belonging to three categories: *non-pathological*, *pathological compatible with COVID-19*, and *pathological non-compatible with COVID-19*. On the training screen, images were shown to the users and they chose a diagnosis among those three possibilities. At any time, users could finish the training session and be evaluated through the estimation of their diagnostic accuracy values: sensitivity, specificity, predictive values, and global accuracy. Images were hand-labeled by four thoracic radiologists. Average values for sensitivity, specificity, and global accuracy were .72, .64, and .68. Users who achieved better sensitivity registered less specificity ( $p < .0001$ ) and those with higher specificity decreased their sensitivity ( $p < .0001$ ). Users who sent more answers achieved better accuracy ( $p = .0002$ ). The application *COVID-19 TRAINING* provides a revolutionary tool to learn the necessary skills to evaluate COVID-19 on CXR. Diagnosis training applications could provide a new original manner of evaluation for medical professionals based on their diagnostic accuracy values, and an efficient method to collect valuable data for research purposes.

**Keywords** Chest X-ray · COVID-19 · Diagnostic accuracy values · Medical application · Medical education · Training on diagnosis

## Background

In December 2019, a new coronavirus named 2019-nCoV, also known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was isolated in the airway epithelial cells of a cluster of patients with pneumonia of unknown cause in Wuhan, China [1]. Since then, the infection caused by 2019-nCoV has rapidly spread globally [2] affecting more than 4 million people in about 215 countries with more than 280.000 reported deaths to date [3]. On the 11th of March 2020, the World Health Organization (WHO) declared the novel coronavirus outbreak a global pandemic [4].

Most patients with coronavirus disease 2019 (COVID-19) are mild cases, whose symptoms are usually self-limiting and recover within 2 weeks [5]. However, others progress rapidly and develop acute respiratory distress syndrome (ARDS) and septic shock, eventually resulting in multiple

✉ P. Menéndez Fernández-Miranda  
pablomenendezfernandezmiranda@gmail.com

✉ J. A. Vega  
javega@uniovi.es

<sup>1</sup> Departamento de Radiología, Hospital Universitario “Marqués de Valdecilla”, Santander, Spain

<sup>2</sup> Grupo de Computación Avanzada y e-Ciencia, Instituto de Física de Cantabria, (IFCA), Consejo Superior de Investigaciones Científicas (CSIC), Santander, Spain

<sup>3</sup> Wroclaw, Poland

<sup>4</sup> Departamento de Morfología y Biología Celular, Universidad de Oviedo, Oviedo, Spain

<sup>5</sup> Facultad de Ciencias de La Salud, Universidad Autónoma de Chile, Santiago, Chile

<sup>6</sup> Departamento Morfología y Biología Celular, Universidad de Oviedo, Oviedo, Spain

organ failure [6]. At the time of this writing, fever and cough were the main clinical manifestations, followed by dyspnea, myalgia or weakness, and chest tightness [7]. Furthermore, a variable percentage of patients report decreased smell function or even anosmia and dysgeusia [8]. Currently, COVID-19 is being diagnosed using molecular detection methods such as the reverse-transcriptase polymerase chain reaction (RT-PCR) test, regarded as the standard of reference [9–11]. Nevertheless, despite RT-PCR COVID-19 testing has a specificity of 100% [11], it shows a potentially high false negative rate deteriorating the sensitivity, and then it is not a definitive diagnostic method [10, 12]. Other techniques for the detection of COVID-19 are being used to increase the efficiency in the diagnosis, especially medical imaging modalities such as chest X-rays (CXR) and computed tomography (CT) scans [9, 10, 13].

In patients with a high level of clinical suspicion of COVID-19 and negative RT-PCR, CXR can be key to identify false negatives to RT-PCR COVID-19 testing as CXR abnormalities may appear before eventually testing positive on RT-PCR [13]. Moreover, in populations around the world with limited access to reliable real-time molecular diagnostic methods, the utilization of CXR for early disease detection also plays a capital role [14]. Additionally, imaging is also critical in assessing the severity and disease progression in a COVID-19 infection [15]. Thus, physicians evaluating COVID-19 on medical images should be aware of the imaging manifestations and radiological features that have been well-described [16].

However, the interpretation of chest radiographs is a challenging task, requiring experience and expertise [17–19]. The American College of Radiology (ACR) recommends that qualified radiologists be available to interpret all radiographs obtained in the Emergency Departments (ED) [20], and previous studies have reported suboptimal performance in the interpretation of CXR by ED physicians compared with expert radiologists [21–23]. Importantly, the number of chest radiographs per ED visit have increased during the last decades [24] which is a practical limitation with regard to the full-time availability of expert radiologists [25]. This situation has deteriorated

since the pandemic outbreak appeared in the worldwide clinical scenario.

In the context of COVID-19 pandemic and the high demand of CXR reporting, many CXR are being interpreted by nonexpert physicians who have been forced to acquire the competence of detecting and evaluating the radiological features of COVID-19. Being aware of it, we have developed a free Web application to ease the learning process of interpreting COVID-19 CXR for physicians, residents, students, and anybody else interested in acquiring this competence. As far as we know, this application called *COVID-19 TRAINING* (<https://xray.covid.ifca.es/en>) is the first available tool that allows a single user to calculate their diagnostic accuracy values: sensitivity, specificity, positive predictive value, negative predictive value, and global accuracy. The *COVID-19 TRAINING* application can set a precedent for further applications because this is the first time that a physician is considered as a single diagnostic tool by themselves. Usually, diagnostic accuracy values are calculated for diagnostic techniques or collectives of physicians but not for a single professional. In our opinion, providing this kind of tool to physicians can help them to evaluate performance by setting specific metrics, and to follow up their progression in their diagnostic efficiency.

## Methods

### Patients Recruitment

We collected 196 CXR in our Institution belonging to 33 females (35.48%) and 60 males (64.51%). The average age of these patients was 61.43 years with a standard deviation (SD) of 17.20 years and a range from 19 to 88 (Table 1).

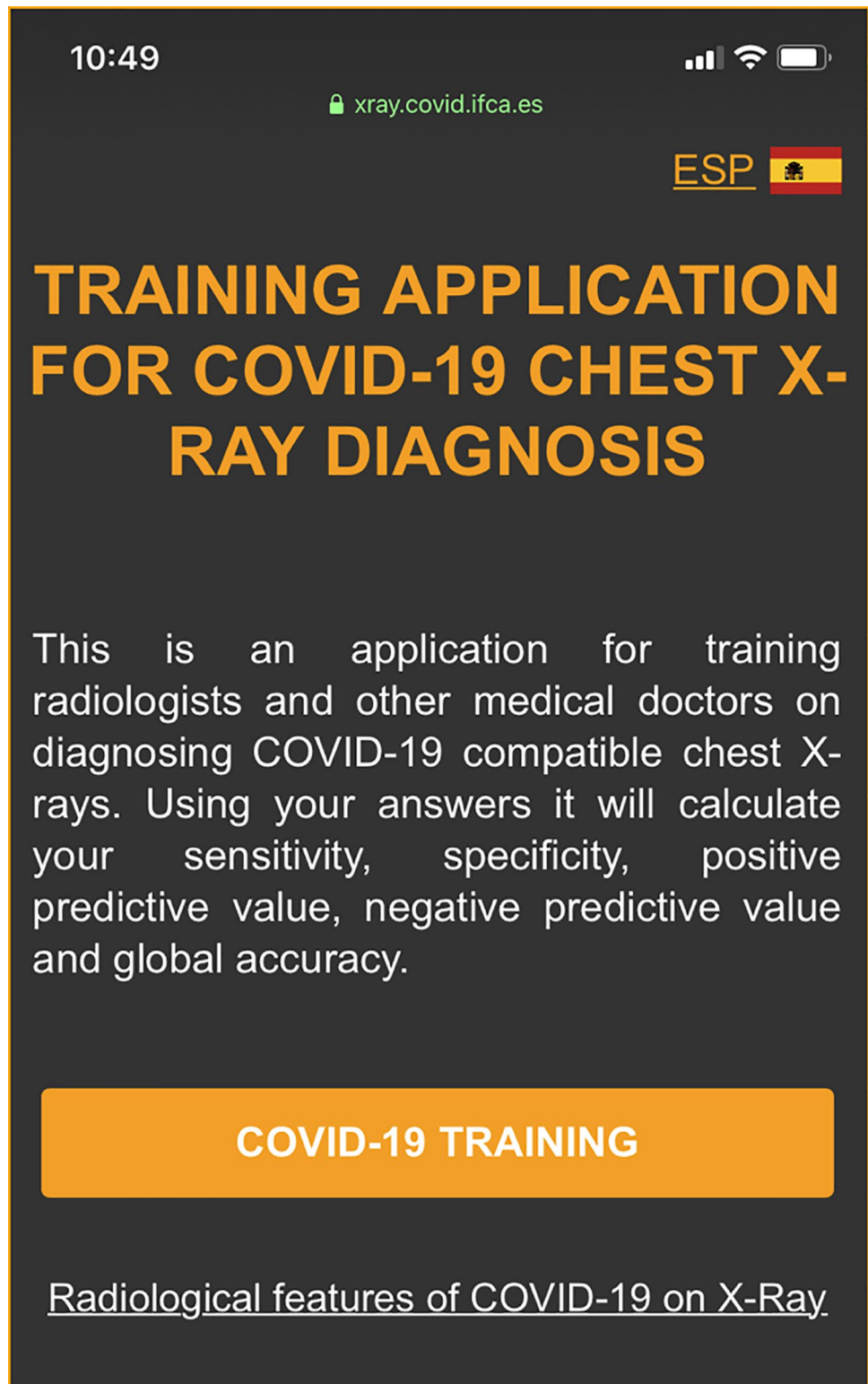
The chest radiographs were all hand-labeled and classified by four expert thoracic radiologists into 3 categories: *non-pathological*, *pathological compatible with COVID-19*, and *pathological non-compatible with COVID-19*. Classification and inclusion criteria for each category were as follows: (1) *non-pathological*: (a) the CXR was obtained before the appearance of SARS-CoV-2 virus; and (b) four reports made

**Table 1** Characteristics of the patients included in the application and number of X-rays for each category

Attribute	Non-pathological	Pathological compatible with COVID-19	Pathological non compatible with COVID-19	Total
Gender, <i>n</i> (%)				
Female	9 (42.86)	10 (23.81)	14 (46.67)	33 (35.48)
Male	12 (57.14)	32 (76.19)	16 (53.33)	60 (64.51)
Age, mean $\pm$ SD	61.04 $\pm$ 16.74	64.64 $\pm$ 15.16	54.07 $\pm$ 17.69	61.43 $\pm$ 17.20
Range	31–84	30–88	19–82	19–88
CXR, <i>n</i> (%)	45 (22.96)	121 (61.73)	30 (15.31)	196 (100)

CXR chest X-ray

**Fig. 1** Welcome page screen, as viewed from a mobile device



by the four expert thoracic radiologists confirm the absence of any pathological findings on the CXR; (2) *pathological compatible with COVID-19*: (a) the CXR has four reports made by the four expert thoracic radiologists indicating

pathological findings compatible with COVID-19; (b) the patients have a *subsequent confirmation* of the disease by a positive RT-PCR COVID-19 testing; and (c) progression of the radiological findings reported previously were observed

on further CXR; (3) *pathological non-compatible with COVID-19*: (a) the CXR was obtained before the appearance of SARS-CoV-2 virus; and (b) the CXR has four reports made by the four expert thoracic radiologists confirming pathological findings.

The exclusion criteria for the three categories were as follows: (a) low-quality image, (b) improper alignment of the X-ray tube to the film, and (c) did not meet any of the other inclusion criteria.

The chest radiographs were obtained with different equipment (portable and conventional machines), in different projections (posterior-anterior and anterior–posterior) and with different patient positions (standing and supine decubitus), in order to reproduce the most realistic clinical scenario. Data about patient distribution for each category is shown in Table 1. This distribution tries to guarantee that the application shows the user a proportional number of images from each category with a slight predominance of the category *pathological compatible with COVID-19*, which are the most interesting images for the users. In addition, the application also resembles the real clinical context where females are less frequently infected than males [26].

## Data Access and Anonymization

This work was approved by the Ethical Committee of our Institution, and all the data collected to develop the application were fully de-identified before transfer to the development team.

## Application Programming and Launching

Subject information and links to the corresponding images are stored in a SQLite database. The application has been implemented using Python 3 [27] for the backend and

Flask, a Python framework, for generating the client side or frontend. Some external libraries were used in order to expand the possibilities of the frontend, achieving multilanguage (currently supporting Spanish and English) or more structured styles hierarchy. Some other good practices based on the Clean Code principles are being applied. All the code is publicly available in an open access repository [28]. User management is supported by an Authentication and Authorization Infrastructure based on Open ID Connect standard.

A public beta version was published on the 11th of April 2020, and the 24th of April we launched the final version.

## Application Usability and Requirements

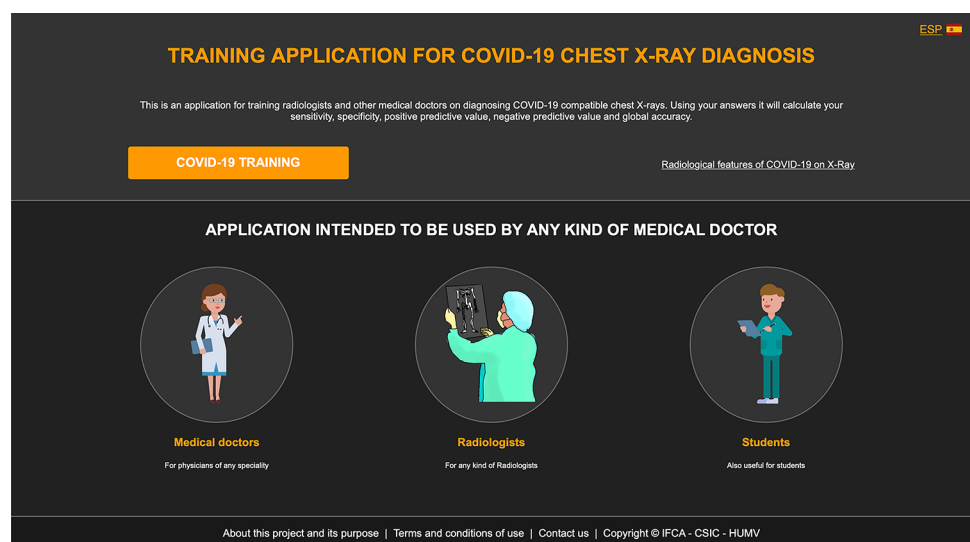
The *COVID-19 TRAINING* application was designed to help all doctors, residents, medical students, and anybody else interested in acquiring the skills to recognize the radiological findings of COVID-19 on CXR. It includes CXR with different levels of difficulty so everyone will be able to join. The application is available for free and adapted both for mobile and tablet (Fig. 1) as well as for computer screens (Fig. 2) in Spanish and English languages.

### Access to the Application

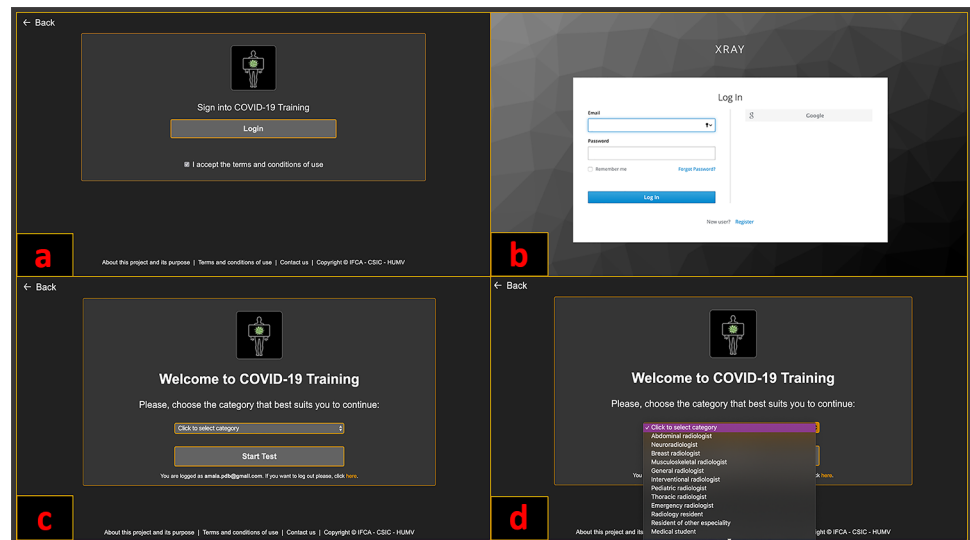
The access and navigation on the application is shown on a video added in the Appendix 1.

*First step:* The first screen that the user finds when opening the application is the welcome page (Figs. 1 and 2). It serves as the starting point of the application and shows a headline and a paragraph describing briefly its functionalities. On this interface two buttons are displayed: “*COVID-19 TRAINING*” that provides easy access to the login screen, and “Radiological features of COVID-19 on

**Fig. 2** Welcome page screen, as viewed from a desktop device



**Fig. 3** **a** Login and **b** authentication screens, and **c**, **d** drop-down menu to select the specialty or professional category of the user, as viewed from a desktop device



X-Ray,” which opens a summary of the radiological COVID-19 findings that the user should be able to identify on CXR before starting the training.

**Second Step:** Login screen requires the user to accept the terms and conditions and to press the “Login” button to access the authentication interface (Fig. 3a). Authentication can be easily done either just entering a Google Gmail or creating a new account by pressing on the “Register” button in the lowest part of the screen (Fig. 3b). If the user’s preferred option is the second, they will only need to provide their name, surname, email address, and password. An email will be automatically sent with a link to verify the account.

**Third Step:** Finally, a drop-down menu will allow the user to select their specialty or professional category (Fig. 3c, d). Available options are shown on Table 2.

### Training and Obtaining the Diagnostic Accuracy Values

Once the authentication is done and the profile category is selected, a new screen with CXR will be open. Besides the CXR, patient’s age and gender are also provided (Fig. 4a).

On the left side of the screen, the user can choose in a drop-down menu one diagnosis among the three possible categories described before: *non-pathological*, *pathological compatible with COVID-19*, and *pathological non-compatible with COVID-19* (Fig. 4b). After selecting the desired option, by pressing the “Next” button, the answer will be registered and a new case will be charged. At any time, the user can finish the training session by pressing the “End Test” button (Fig. 4). Each time the user takes a test, a new set of training images is selected.

A virtual magnifying glass is also implemented. Users simply need to press on the image to summon the magnifying glass, and they will see a zoomed image within

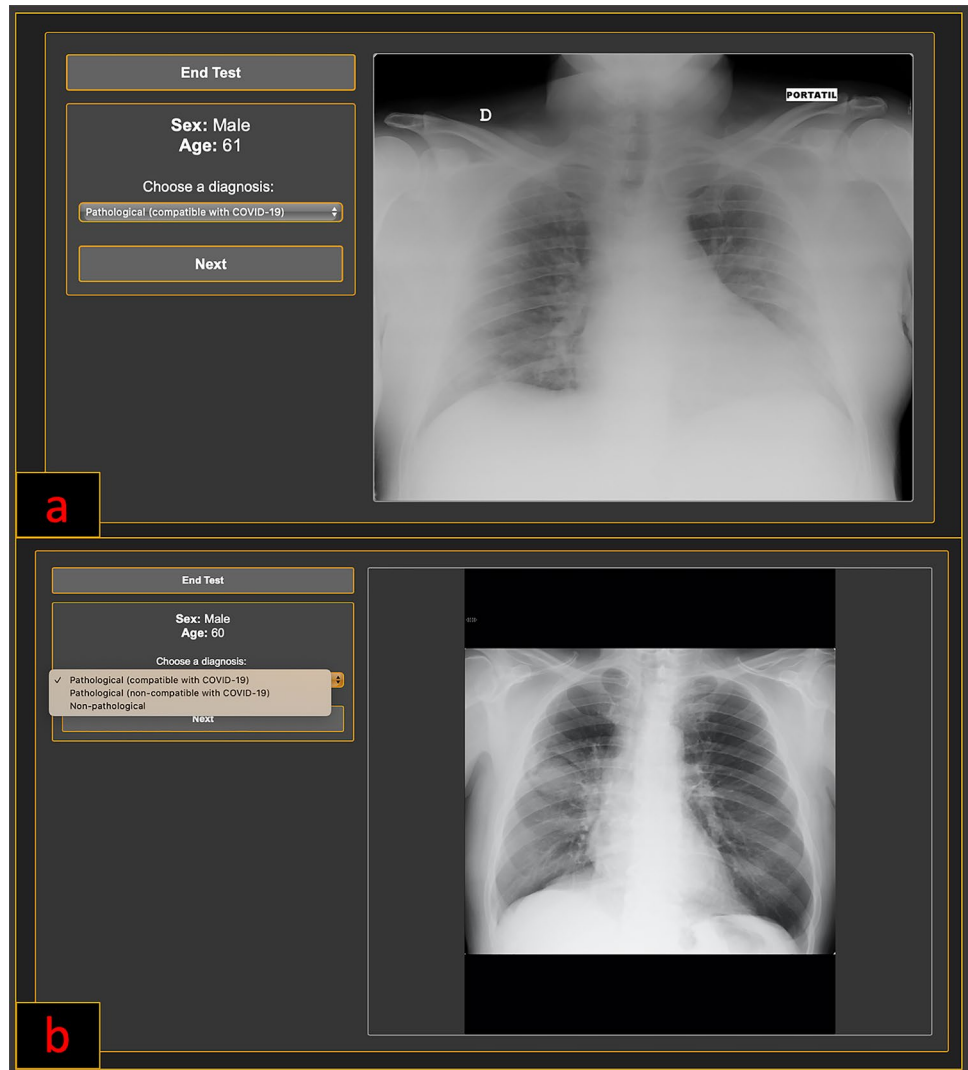
its radius, without disturbing the rest of the page (Fig. 5). To remove it, they only need to press on the image again.

The “End Test” button links to the results interface, where the user’s sensitivity, specificity, predictive values and global accuracy are provided after each training (Figs. 6 and 7). If the users scroll down on this interface, those questions that they answered incorrectly will be displayed, indicating the given response and the correct answer, so they will learn from their mistakes. *Pathological non compatible with COVID-19* images also include a reference to the real pathology of the patient.

**Table 2** Specialties and professional categories available to select on the drop-down menu

Specialty/category
Abdominal radiologist
Neuroradiologist
Breast radiologist
Musculoskeletal radiologist
General radiologist
Interventional radiologist
Pediatric radiologist
Thoracic radiologist
Emergency radiologists
Radiology resident
Resident of other specialty
Medical student
Pulmonology physician
Internal medicine physician
Emergency physician
Intensive care physician
Other kind of physician
Diagnostic Imaging Technician (radiographer)
Other category

**Fig. 4** **a** Training interface and **b** the diagnoses available on the drop-down menu, as viewed from a desktop device



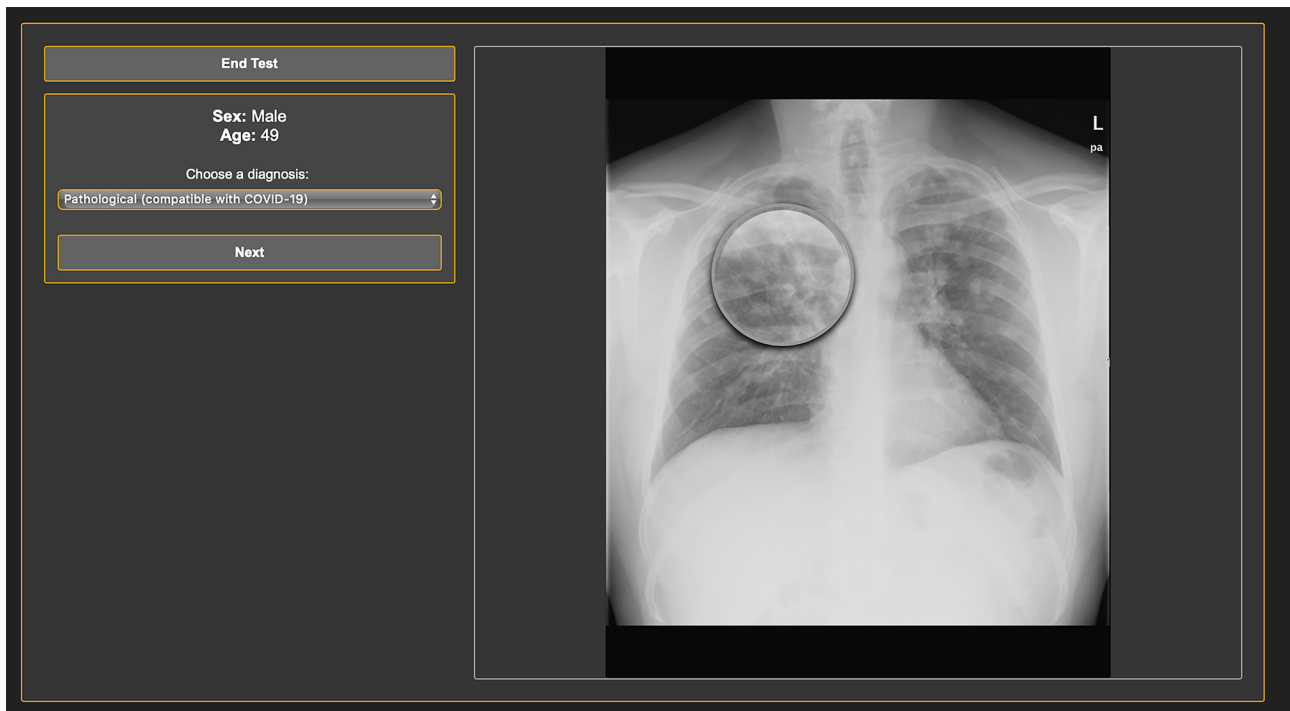
## Data Collection and Analysis of the User's Results

The *COVID-19 TRAINING* application allows the user to evaluate themselves through the estimation of their diagnostic accuracy values: sensitivity, specificity, positive predictive value, negative predictive value, and global accuracy. Diagnostic accuracy values are calculated for COVID-19 diagnosis. The application registers the answers of the users and classifies them into the following four conventional categories [29, 30]: (1) true positive (TP): the user classifies properly a CXR which belongs to the category of *pathological compatible with COVID-19*; (2) false positive (FP): the user classifies improperly a CXR which does not belong to the category of *pathological compatible with COVID-19*, in this category; (3) true negative (TN): the user classifies properly a CXR which does not belong to the category of *pathological compatible with COVID-19* in another category; (4) false negative (FN): the user

classifies improperly a CXR which belongs to the category of *pathological compatible with COVID-19*, in another category.

The sensitivity represents the user's ability to determine the COVID-19 cases correctly. It accounts for the proportion of true positives in patient-cases:  $TP/(TP + FN)$  [29]. The specificity shows the user's capacity to rule out COVID-19 correctly. To estimate specificity, the proportion of true negatives in healthy cases should be calculated:  $TN/(TN + FP)$  [29].

Positive predictive value (PPV) defines the probability of having COVID-19 when the user classifies the CXR into the category *pathological compatible with COVID-19*. Therefore, it represents the proportion of COVID-19 patients within the patients with positive CXR for COVID-19 according to the user's criteria:  $TP/(TP + FP)$  [30]. By contrast, negative predictive value (NPV) describes the probability of not having COVID-19 when the user does not



**Fig. 5** Virtual magnifying glass, as viewed from a desktop device

classify the CXR into *pathological compatible with COVID-19* category. It is defined as the proportion of subjects without COVID-19 within the patients with a negative CXR according to the user's criteria:  $TN/(TN + FN)$  [30].

Finally, global accuracy depicts the ability of the user to differentiate COVID-19 patients and non-COVID-19 patients. It is the proportion of true positives and true negatives in all evaluated cases:  $(TP + TN)/(TP + TN + FP + FN)$  [29].

The information gathered by the application includes TP, TN, FP, FN, sensitivity, specificity, PPV, NPV, and global accuracy for each user and each time they take the test. The specialty or professional category selected after the authentication and the answers given to each CXR are also stored. Nevertheless, the specialty or category chosen before starting the training have only been recorded since the launch of the last version on the 24th of April. Beta test did not record it.

Engagement with the application is tracked using a customized data capture system and validated with Google Analytics [31]. This program is an effective resource for measuring the diffusion and understanding the geodemographics of users [32].

Statistical analysis was performed with IBM SPSS Statistics [33]. Averages for diagnostic accuracy values

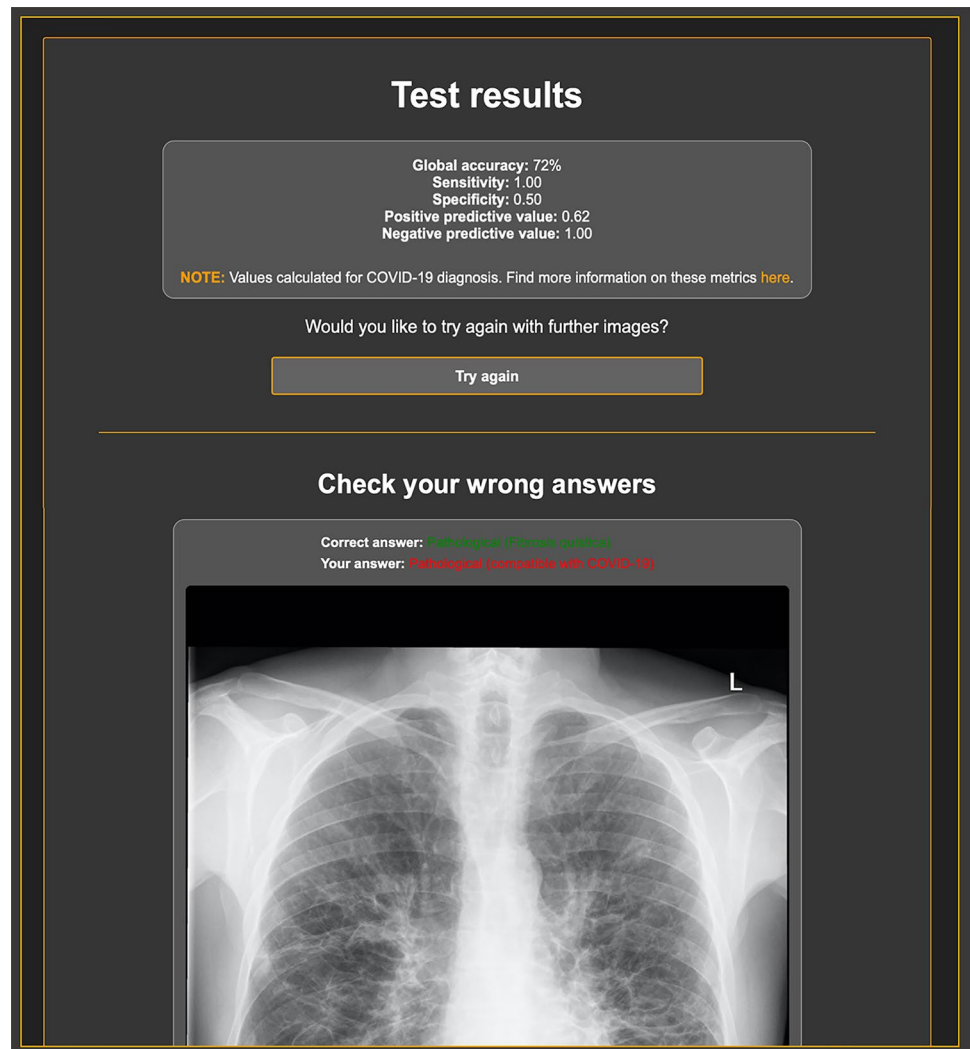
were computed following two different approaches: firstly, population diagnostic values were calculated from the totals of TP, TN, FP, and FN collected; differently, the means for the diagnostic values obtained by the users were also estimated. The test used to assess differences in the performances between two groups was chi-squared test. To indicate a statistically significant difference, a  $p < 0.05$  was considered.

## Results

After the beta launch on the 11th of April, the application had 431 users within the first 3 days and 704 within the first week, with a total number of answers of 23,130. This version had users in more than 20 countries, according to Google Analytics reports. Figure 8 shows a map with users' locations.

The application registered 16,360 (70.73%) correct answers (TP + TN) and 6770 (29.27%) incorrect answers (FP + FN). Consequently, the odds of answering correctly was 2.42. Positive likelihood ratio was 2.26, 95% CI [2.20, 2.33], and negative likelihood ratio was 0.39, 95% CI [0.38, 0.40]. Each user sent a mean of 32.86 (SD 39.21) answers, ranging from 1 to 434 (Figs. 9 and 10): an average of 23.24 (SD 28.09) of them were correct, and an average of 9.62 (SD 12.66) were incorrect.

**Fig. 6** Screen of results, as viewed from a desktop device



Overall, this first version collected the following answers: 9026 (39.02%) TP; 7334 (31.71%) TN; 3545 (15.33%) FP; and 3225 (13.94%) FN. The mean of TP, TN, FP, and FN per user were 12.82 (SD 15.91), 10.42 (SD 12.88), 4.58 (SD 6.31), and 5.04 (SD 7.65), respectively (Table 3).

The global population diagnostic accuracy values registered were the following: sensitivity 0.74, 95% CI [0.73, 0.74]; specificity 0.67, 95% CI [0.67, 0.68]; PPV 0.72, 95% CI [0.71, 0.72]; NPV 0.69, 95% CI [0.69, 0.70]; and accuracy 0.71, 95% CI [0.70, 0.71] (Table 4). Differently, the means for the diagnostic values obtained by the users were as follows: sensitivity 0.72, 95% CI [0.71, 0.74]; specificity 0.64, 95% CI [0.62, 0.66]; PPV

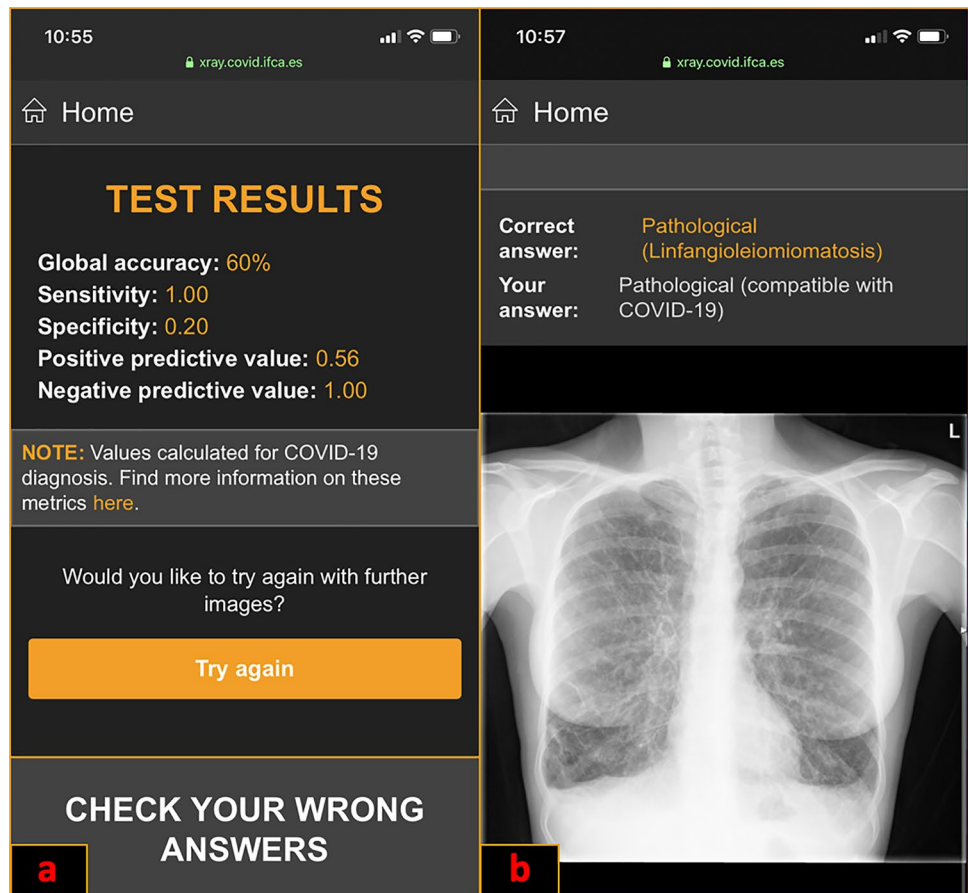
0.70, 95% CI [0.69, 0.72]; NPV 0.68, 95% CI [0.66, 0.69]; and accuracy 0.68, 95% CI [0.67, 0.69] (Table 5).

The hardest CXR obtained an odds of answering correctly of 0.32 (Table 6). On the other side, the easiest CXR odds of answering correctly was 18.70 (Table 7). Both CXR belonged to the category *pathological compatible with COVID-19* (Figs. 11a and 12).

As would be expected, users who achieved a sensitivity equal or higher to the average sensitivity (0.72), also registered less specificity ( $p < 0.0001$ ). Similarly, the user's with higher specificity, those who achieved the average specificity (0.64) or more, decrease their sensitivity ( $p < 0.0001$ ). In addition, the users who sent more answers than the 50th percentile (68



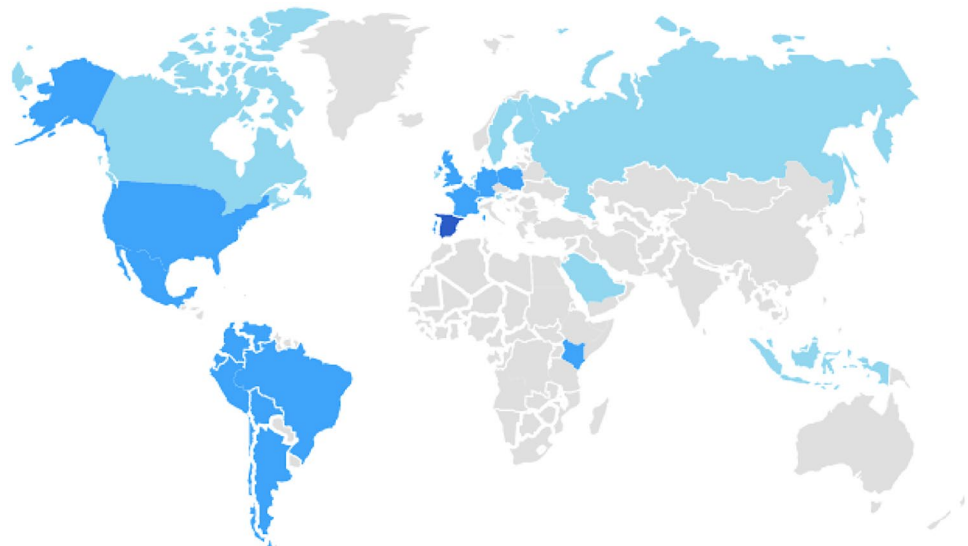
**Fig. 7** Screen of results, as viewed from a mobile device



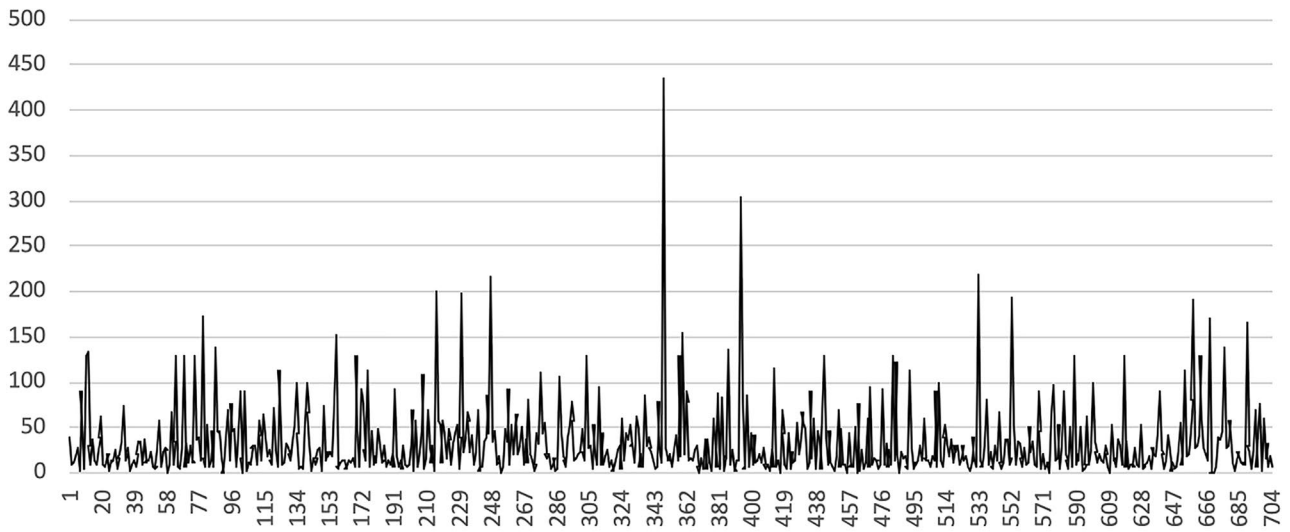
answers) achieved better final average global accuracy ( $p < 0.0001$ ); excluding users who sent less than 10 answers from the analysis, results were also statistically significant ( $p = 0.0002$ ) (Table 8).

Beta version did not ask for the user's category or specialty, so no data about it was recorded before the final version that was launched on the 24th of April.

**Fig. 8** Countries of the application users. The first version of the application extended to more than 20 countries as the map shows in blue color: a darker color means more visits



### Number of answers sent by each user



**Fig. 9** A frequency distribution graph showing the number of answers sent by each user. The x-axis represents the users and the y-axis represents the number of responses

### Discussion

Several medical educational applications have been developed [33–35], but only a few of them are focused on the diagnosis training, and even less on COVID-19 diagnosis. Currently, the society is facing a health situation which has no precedents [36]. In this context, the role of the medical community is vital [37, 38] and the medical education of those who are taking part of the solution is essential [18, 39].

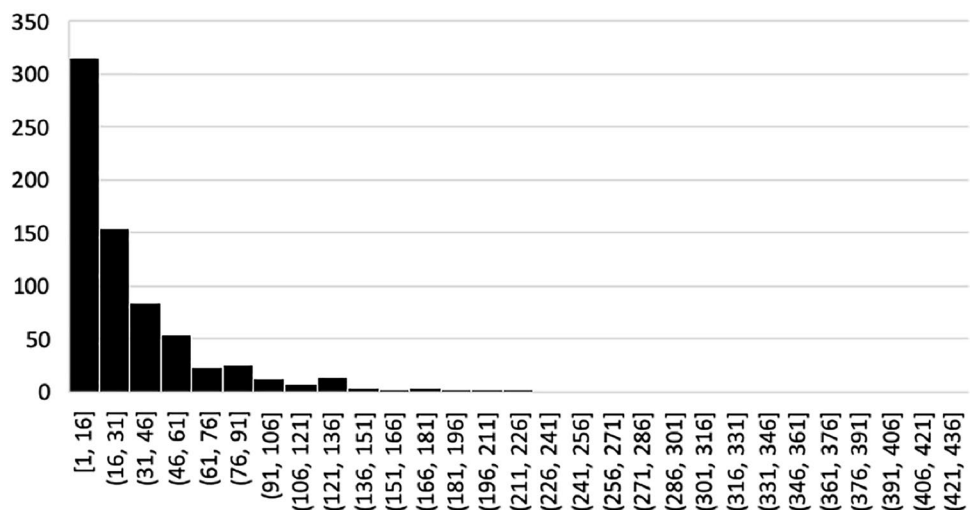
The application *COVID-19 TRAINING* was developed to help professionals to acquire the required competencies to diagnose COVID-19 on CXR. This application brings a new manner to ease the learning of the necessary skills to successfully evaluate COVID-19 on CXR. According to

us, this is a different educational technique that could set a precedent for developing further applications on the training on the diagnosis of other pathologies.

Furthermore, we have introduced an original evaluation method based on the estimation of the users’ diagnostic accuracy values for COVID-19. In our opinion, this mechanism of assessment could prove useful for medical professionals since it provides them with information about their sensitivity, specificity, predictive values, and global accuracy. Armed with that knowledge, they will recognize and improve their weaknesses.

The results obtained by the first version of the application showed that those who achieved better sensitivity also decrease their specificity. Similarly, those who registered better specificity also had lower sensitivity. These results

**Fig. 10** An histogram depicting the number of answers collected by the first version of the application. The x-axis shows intervals of 16 responses width and the y-axis shows the frequency of the intervals



**Table 3** Basic statistics for the user's answers

Answers	<i>n</i> (%)	Mean ± SD (per user)	Range (per user)
TP	9026 (39.02)	12.82 ± 15.91	0–136
FN	3225 (13.94)	4.58 ± 6.31	0–91
TN	7334 (31.71)	10.42 ± 12.88	0–88
FP	3545 (15.33)	5.04 ± 7.65	0–123
Correct responses	16,360 (70.73)	23.24 ± 28.09	0–220
Incorrect responses	6770 (29.27)	9.62 ± 12.66	0–214
Total responses	23,130 (100)	32.86 ± 39.21	1–434

TP true positive, FN false negative, TN true negative, FP false positive

evidence that usually, increasing sensitivity entails decreasing specificity, and vice versa. Our application can help the users to find an optimal balance between their sensitivity and specificity and consequently, to achieve their best possible global accuracy.

In addition, the analysis and display of the failed answers is also a key feature of this tool. At the end of the test, users can review their incorrect responses so they will become aware about their mistakes and they will recognize their weak points. This feature can be also useful in research to identify the cases missed by numerous users. Later on, the examination of these problematic images may give some clues to find solutions to improvable areas in the medical practice not detected before.

Following this approach, we have analyzed the hardest and the easiest CXR trying to understand why users found difficult or obvious the respective cases.

The most difficult CXR was a COVID-19 case that had subtle and ill-defined ground-glass opacities in the right lung. The most evident findings on this image were in the right upper lobe, nearly overlapped with the first costochondral junction (Fig. 11). This first costochondral junction is a well-recognized pitfall on CXR and it sometimes mimics a rounded opacity [40]. Perhaps, in the situations where a COVID-19 is suspected and a prominence on either the first costochondral or first costosternal or sternoclavicular junctions is seen on CXR, further evaluation should be performed.

On the other side, the easiest CXR was also a COVID-19 patient, but in this case, the image illustrated evident peripheral bilateral opacities, which are one of the most typical radiological finding of this disease [7, 13] (Fig. 12).

Overall, this application seems to be helpful. From the beginning, it had a dizzyingly fast diffusion with hundreds of users registered from more than 20 different countries within the first days and more than 20.000 answers sent. Additionally, many medical societies spread and shared this first beta version on their webpages.

Results showed that the users who sent more answers achieved better global accuracy. In addition, only 6 users (0.9%) exceeded the number of cases contained in our dataset, so most of them did not repeat answers. Users who sent less than 10 answers were excluded

from the analysis to avoid potential biases. These facts could indicate that the application improves the users' diagnostic skills.

Finally, we also found that medical educational applications may be used in research as a new method to collect relevant information. Since the application was officially launched on the 24th of April, data about the users' specialty or professional category have been recorded. Our intention is to use this data to analyze the difference in the diagnostic values between the users belonging to different specialties or categories and to try to estimate the real utility of the CXR on the evaluation of COVID-19.

**Table 4** Population diagnostic accuracy values

Statistic	Value	95% CI
Sensitivity	.74	[.73, .74]
Specificity	.67	[.67, .68]
Positive predictive value	.72	[.71, .72]
Negative predictive value	.69	[.69, .70]
Global accuracy	.71	[.70, .71]

Values computed from the total answers collected

CI confidence interval

**Table 5** Average of user's diagnostic accuracy values

Statistic	Mean ± SD	95% CI of the mean
Sensitivity	.72 ± .21	[.71, .74]
Specificity	.64 ± .25	[.62, .66]
Positive predictive value	.70 ± .21	[.69, .72]
Negative predictive value	.68 ± .22	[.66, .69]
Global accuracy	.68 ± .16	[.67, .69]

Values estimated from the diagnostic values achieved by each user

CI confidence interval

**Table 6** The top 10 hardest chest X-rays

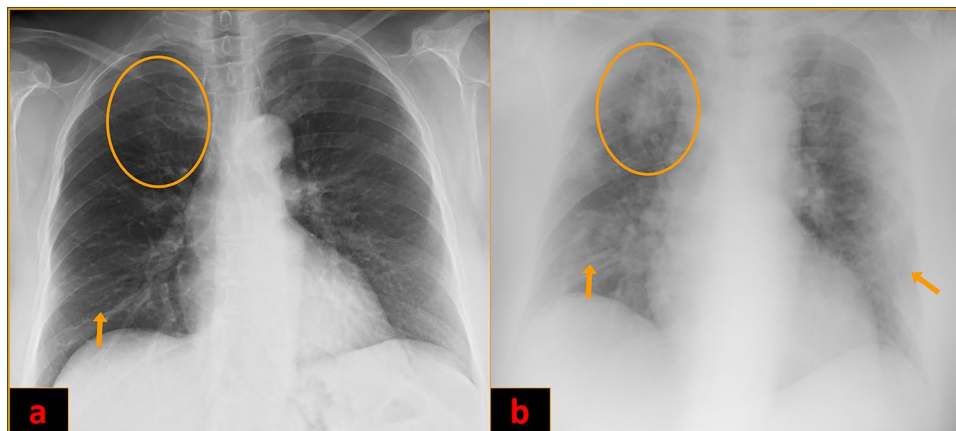
Position	Image	TP	FN	TN	FP	Odds of answering correctly
1	F06	47	149	0	0	.32
2	F020_2	0	0	48	117	.41
3	M053	53	128	0	0	.41
4	M034	0	0	57	130	.44
5	M06	0	0	56	125	.45
6	M030_2	52	116	0	0	.45
7	M09	66	121	0	0	.55
8	F09	66	118	0	0	.56
9	M032_2	0	0	62	98	.63
10	M031	70	104	0	0	.67

*TP* true positive, *FN* false negative, *TN* true negative, *FP* false positive

**Table 7** The top 10 easiest chest X-rays

Position	Image	TP	FN	TN	FP	Odds of answering correctly
1	M027	187	10	0	0	18.70
2	F024	0	0	185	10	18.50
3	M027_2	182	10	0	0	18.20
4	F022	0	0	163	10	16.30
5	F014	175	11	0	0	15.91
6	M019	158	11	0	0	14.30
7	M018	181	13	0	0	13.92
8	F06_3	173	13	0	0	13.31
9	M02_2	151	12	0	0	12.58
10	M016	163	14	0	0	11.64

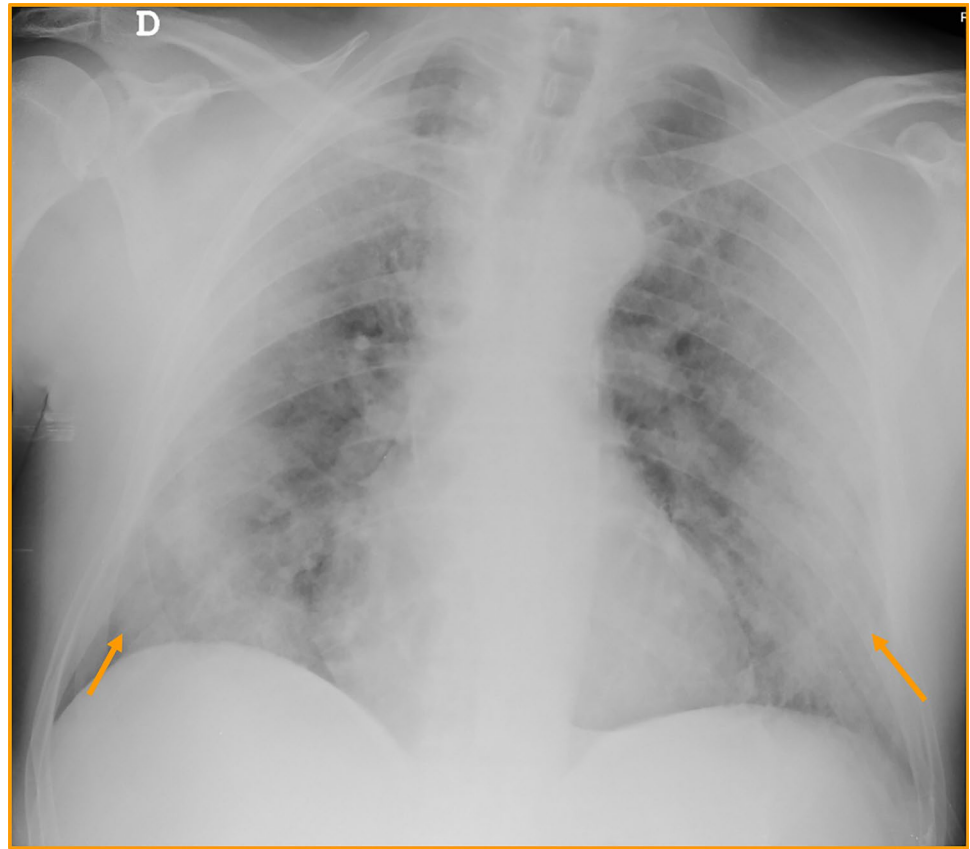
*TP* true positive, *FN* false negative, *TN* true negative, *FP* false positive



**Fig. 11 a** The most difficult chest X-ray. It is a 66-year-old female with a chest X-ray belonging to the category *pathological compatible with COVID-19*. On this image, subtle ground-glass opacities on the right lung are visible (arrow); the most evident opacity is in the right

upper lobe (circle). **b** On a subsequent chest X-ray performed to the same patient after three days, this ground-glass opacity in the upper lobe became even clearer (circle). In addition, new opacities appeared in the lower lobes on this chest X-ray (arrows)

**Fig. 12** The easiest chest X-ray. It is a 69-year-old male diagnosed of COVID-19. The image shows extensive, multiple and bilateral opacities (arrows) indicating a severe form of COVID-19 pneumonia



**Table 8** Comparison of the results between groups and  $p$  values for chi-squared tests

Statistic	Group 1	Group 2	Proportion difference	95% CI of difference	$\chi^2$ value	$p$ value
Sensitivity $[TP/(TP + FN)]^a$	.76 [3504/4604]	.72 [5479/7591]	-.04	[-.06, -.02]	22.814	<.0001
Specificity $[TN/(TN + FP)]^b$	.72 [3192/4425]	.64 [4130/6431]	-.08	[-.10, -.06]	74.795	<.0001
Global accuracy $[(TP + TN)/(TP + TN + FP + FN)]^c$	.70 [10430/14956]	.73 [5930/8174]	.03	[.02, .04]	20.147	<.0001
Global accuracy $(TP + TN)/(TP + TN + FP + FN)^d$	.70 [9815/13988]	.73 [5930/8174]	.02	[.01, .04]	14.203	.0002

CI confidence interval, TP true positive, FN false negative, TN true negative, FP false positive

<sup>a</sup>Group 1: users who achieved a specificity lesser than the mean (.64); group 2: users who achieved a specificity equal to or higher than the mean (.64)

<sup>b</sup>Group 1: users who registered a lower sensitivity than the mean (.72); group 2: users who registered a sensitivity equal to or higher than the mean (.72)

<sup>c</sup>Group 1: users who sent a number of answers equal to the 50th percentile (68 answers) or less; group 2: users who sent more than 68 answers

<sup>d</sup>Group 1: users who sent a number of answers between 10 and the 50th percentile (68 answers); group 2: users who sent more than 68 answers

## Conclusion

In present COVID-19 pandemic, the medical education of the professionals involved in patients care is vital and *COVID-19 TRAINING* brings a different solution to help them in this purpose. Applications focused on the training on diagnosis could provide a new original manner of evaluation for medical professionals. The assessment of users by estimating their diagnostic accuracy values make them aware of their weak points. In addition, this kind of application also collects valuable information that can be used for research purposes.

## Appendix 1

A link to a video navigating the application: <https://api.cloud.ifca.es:8080/swift/v1/covid19/VIDEO%20APP.mov>.

**Acknowledgments** We would like to acknowledge all the professionals who have battled in an exemplary manner to safeguard the health and life of all citizens worldwide since the beginning of this pandemic.

**Funding** The authors received no financial support for this work.

## Declarations

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al: A Novel Coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382(8): 727–733, 2020. <https://doi.org/10.1056/NEJMoa2001017>
- Raoult D, Zumla A, Locatelli F, Ippolito G, Kroemer G: Coronavirus infections: Epidemiological, clinical and immunological features and hypotheses. *Cell Stress* 4(4): 66–75, 2020. <https://doi.org/10.15698/cst2020.04.216>
- WHO. WHO's Coronavirus disease (COVID-19) outbreak situation dashboard. (n.d.) Available at <https://www.who.int>. Accessed 6 May 2020.
- Cucinotta D, Vanelli M: WHO Declares COVID-19 a Pandemic. *Acta Biomed* 91(1): 157–160, 2020. <https://doi.org/10.23750/abm.v91i1.9397>
- Wu Z, McGoogan JM: Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 323(13): 1239–1242, 2020. <https://doi.org/10.1001/jama.2020.2648>
- Xiea M, Chen Q: Insight into 2019 novel coronavirus — an updated interim review and lessons from SARS-CoV and MERS-CoV. *Int J Infect Dis* 94: 119–124, 2020. <https://doi.org/10.1016/j.ijid.2020.03.071>
- Cao Y, Liu X, Xiong L, Cai K: Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2: A systematic review and meta-analysis. *J Med Virol*, <https://doi.org/10.1002/jmv.25822>, April 3, 2020.
- Lechien JR, Chiesa-Estomba CM, De Siati DR, Horoi M, Le Bon SD, Rodríguez A, 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 Otorhinolaryngol*, <https://doi.org/10.1007/s00405-020-05965-1>, April 6, 2020.
- Udugama B, Kadhiresan P, Kozłowski HN, Malekjahani A, Osborne M, Li VYC, et al: Diagnosing COVID-19: The Disease and Tools for Detection. *ACS Nano* 14(4): 3822–3835, 2020. <https://doi.org/10.1021/acsnano.0c02624>
- Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al: Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology*, <https://doi.org/10.1148/radiol.2020200432>, February 19, 2020.
- He JL, Luo L, Luo ZD, Lyu JX, Ng MY, Shen XP, et al: Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside Wuhan, China. *Resp Med* 168:105980, 2020. <https://doi.org/10.1016/j.rmed.2020.105980>
- Li Y, Yao L, Li J, Chen L, Song Y, Cai Z, et al: Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19. *J Med Virol*, <https://doi.org/10.1002/jmv.25786>, March 26, 2020.
- Wong HYF, Lam HYS, Fong AHT, Leung ST, Chin TWY, Lo CSY, et al: Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology*, <https://doi.org/10.1148/radiol.2020201160>, March 27, 2020.
- Jacobi A, Chung M, Bernheim A, Eber C: Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. *Clin Imaging*, 64:35–42, 2020. <https://doi.org/10.1016/j.clinimag.2020.04.001>
- Borghesi A, Maroldi R: COVID-19 outbreak in Italy: Experimental chest X-ray scoring system for quantifying and monitoring disease progression. *Radiol Med* 125(5):509–513, 2020. <https://doi.org/10.1007/s11547-020-01200-3>, 2020.
- Kooraki S, Hosseiny M, Myers L, Gholamrezanezhad A: Coronavirus (COVID-19) Outbreak: What the Department of Radiology Should Know. *J Am Coll Radiol* 17(4): 447–451, 2020. <https://doi.org/10.1016/j.jacr.2020.02.008>
- Hwang EJ, Nam JG, Lim WH, Park SJ, Jeong YS, Kang JH, et al: Deep learning for chest radiograph diagnosis in the Emergency Department. *Radiology* 293(3):573–580, 2019. <https://doi.org/10.1148/radiol.2019191225>
- van der Gijp A, Ravesloot CJ, Jarodzka H, van der Schaaf MF, van der Schaaf IC, van Schaik JPI, et al: How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. *Adv Health Sci Educ Theory Pract* 22: 765–787, 2017. <https://doi.org/10.1007/s10459-016-9698-1>
- Yoon JS, Boutis K, Pecaric MR, Fefferman NR, Ericsson KA, Pusic MV: A think-aloud study to inform the design of radiograph interpretation practice. *Adv Health Sci Educ Theory Pract*, <https://doi.org/10.1007/s10459-020-09963-0>, March 5, 2020.
- American College of Radiology. ACR practice parameter for radiologist coverage of imaging performed in hospital emergency departments. Available at <https://www.acr.org/-/media/ACR/Files/Practice-Parameters/HospER.pdf?la=en>. Accessed 19 May 2020.
- Eng J, Mysko WK, Weller GE, et al: Interpretation of Emergency Department radiographs: a comparison of emergency medicine physicians with radiologists, residents with faculty, and film with digital display. *AJR Am J Roentgenol* 175(5):1233–1238, 2000. <https://doi.org/10.2214/ajr.175.5.1751233>
- Gatt ME, Spectre G, Paltiel O, Hiller N, Stalnikowicz R: Chest radiographs in the emergency department: is the radiologist really

- necessary?. *Postgrad Med J* 79(930):214–217, 2003. <https://doi.org/10.1136/pmj.79.930.214>
23. Petinaux B, Bhat R, Boniface K, Aristizabal J: Accuracy of radiographic readings in the emergency department. *Am J Emerg Med* 29(1):18–25, 2011. <https://doi.org/10.1016/j.ajem.2009.07.011>
  24. Chung JH, Duszak R Jr, Hemingway J, Hughes DR, Rosenkrantz AB: Increasing Utilization of Chest Imaging in US Emergency Departments From 1994 to 2015. *J Am Coll Radiol* 16(5):674–682, 2019. <https://doi.org/10.1016/j.jacr.2018.11.011>
  25. Sellers A, Hillman BJ, Wintermark M: Survey of after-hours coverage of emergency department imaging studies by US academic radiology departments. *J Am Coll Radiol* 11(7):725–730, 2014. <https://doi.org/10.1016/j.jacr.2013.11.015>
  26. Yi Y, Lagniton PNP, Ye S, Li E, Xu RH: COVID-19: What has been learned and to be learned about the novel coronavirus disease. *Int J Biol Sci* 16(10):1753–1766, 2020. <https://doi.org/10.7150/ijbs.45134>
  27. Python 3 Programming Language, RRID:SCR\_008394.
  28. Anonymized
  29. Baratloo A, Hosseini M, Negida A, El Ashal G: Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emerg (Tehran)* 3(2):48–49, 2015.
  30. Šimundić AM: Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC*. 2009;19(4):203–211, 2009.
  31. Google Analytics: <http://google.com/analytics>.
  32. McGuckin C, Crowley N: Using Google Analytics to evaluate the impact of the CyberTraining project. *Cyberpsychol Behav Soc Netw* 15(11): 625–9, 2012. <https://doi.org/10.1089/cyber.2011.0460>
  33. SPSS, RRID:SCR\_002865; IBM, New York, USA.
  34. Wood LE, Picard MM, Kovacs MD: App Review: The Radiology Assistant 2.0. *J Digit Imaging* 31: 383–386, 2018. <https://doi.org/10.1007/s10278-018-0070-2>.
  35. Mosa AS, Yoo I, Sheets L: A systematic review of healthcare applications for smartphones. *BMC Med Inform Decis Mak* 12:67, 2012. <https://doi.org/10.1186/1472-6947-12-67>
  36. WHO. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. Available at <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>. Accessed 5 April 2020.
  37. Adams JG, Walls RM: Supporting the Health Care Workforce During the COVID-19 Global Epidemic. *JAMA* 323(15):1439–1440, 2020. <https://doi.org/10.1001/jama.2020.3972>
  38. DeWitt DE: Fighting COVID-19: Enabling Graduating Students to Start Internship Early at Their Own Medical School. *Ann Intern Med*, <https://doi.org/10.7326/M20-1262>, April 7, 2020.
  39. Lei P, Huang Z, Liu G, Wang P, Song W, Mao J, et al: Clinical and computed tomographic (CT) images characteristics in the patients with COVID-19 infection: What should radiologists need to know? *J X-ray Sci Technol*, <https://doi.org/10.3233/XST-200670>, April 7, 2020.
  40. Haramati LB, Haramati N: Pulmonary pseudonodules on computed tomography: a common pitfall caused by degenerative arthritis. *J Thorac Imaging* 11(4):283–285, 1996. <https://doi.org/10.1097/00005382-199623000-00007>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.