# A novel model to predict severe COVID-19 and mortality using an artificial intelligence algorithm to interpret chest radiographs and clinical variables

Nicolás Munera[1,6], Esteban Garcia-Gallo[2,6], Álvaro Gonzalez[1], José Zea[1], Yuli V. Fuentes[2,3], Cristian Serrano[2,3], Alejandra Ruiz-Cuartas[2], Alejandro Rodriguez[4] and Luis F. Reyes[2,3,5]

[1]Arkangel AI, Bogotá, Colombia. [2]Department of Infectious Diseases, Universidad de la Sabana, Chía, Colombia. [3]Critical Care Department, Clínica Universidad de La Sabana, Chía, Colombia. [4]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias, Hospital Universitari Joan XXIII and Rovira & Virgili University, Tarragona, Spain. [5]Nuffield School of Medicine, University of Oxford, Oxford, UK. [6]Joint first authors.

Corresponding author: Luis F. Reyes (luis.reyes5@unisabana.edu.co)



Shareable abstract (@ERSpublications)
**In patients with #COVID19, an automated chest radiograph interpretation algorithm, along with clinical variables, is a reliable alternative to identify patients at risk of developing severe COVID-19, who might require admission to the intensive care unit** https://bit.ly/3Kf61TK

**Cite this article as:** Munera N, Garcia-Gallo E, Gonzalez Á, *et al*. A novel model to predict severe COVID-19 and mortality using an artificial intelligence algorithm to interpret chest radiographs and clinical variables. *ERJ Open Res* 2022; 8: 00010-2022 [DOI: 10.1183/23120541.00010-2022].



Received: 7 Jan 2022
Accepted: 19 April 2022

## Abstract

***Background*** Patients with coronavirus disease 2019 (COVID-19) could develop severe disease requiring admission to the intensive care unit (ICU). This article presents a novel method that predicts whether a patient will need admission to the ICU and assesses the risk of in-hospital mortality by training a deep-learning model that combines a set of clinical variables and features in chest radiographs.

***Methods*** This was a prospective diagnostic test study. Patients with confirmed severe acute respiratory syndrome coronavirus 2 infection between March 2020 and January 2021 were included. This study was designed to build predictive models obtained by training convolutional neural networks for chest radiograph images using an artificial intelligence (AI) tool and a random forest analysis to identify critical clinical variables. Then, both architectures were connected and fine-tuned to provide combined models.

***Results*** 2552 patients were included in the clinical cohort. The variables independently associated with ICU admission were age, fraction of inspired oxygen ($F_{iO_2}$) on admission, dyspnoea on admission and obesity. Moreover, the variables associated with hospital mortality were age, $F_{iO_2}$ on admission and dyspnoea. When implementing the AI model to interpret the chest radiographs and the clinical variables identified by random forest, we developed a model that accurately predicts ICU admission (area under the curve (AUC) 0.92±0.04) and hospital mortality (AUC 0.81±0.06) in patients with confirmed COVID-19.

***Conclusions*** This automated chest radiograph interpretation algorithm, along with clinical variables, is a reliable alternative to identify patients at risk of developing severe COVID-19 who might require admission to the ICU.

## Introduction

The disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), better known as coronavirus disease 2019 (COVID-19), has become an international issue due to its social, economic and health impact [1, 2]. Most patients present a mild disease; however, the infection may evolve to pneumonia and critical infection in some cases [1, 3, 4]. Patients can develop complications such as ventilatory failure, coagulopathies, thrombosis (*e.g.* disseminated intravascular coagulation), sepsis, multiple organ dysfunction and death [5, 6]. More than 260 million cases have been confirmed, and <4.5 million people have died. Patients at risk of dying due to COVID-19 are male, older adults and patients with several comorbid conditions [7, 8]. According to the disease's severity and past medical history, the mortality rate associated with the COVID-19 ranges from 2.1% to 55%; therefore, it has become a global public health problem [9–11].

Several stratification strategies have been described to identify patients at higher risk of developing severe COVID-19 and dying. For instance, the CALL score evaluates a patient's comorbidities, age, lymphocyte count and serum levels of lactate dehydrogenase [12, 13]. Another widely used score is the 4C score, which uses a combination of clinical variables and laboratory results to identify patients with severe COVID-19 and a high risk of dying due to COVID-19 [7]. Other studies have assessed chest radiograph abnormalities, age, comorbidities and abnormal laboratory results to identify patients with severe COVID-19 [1, 14]. However, the predictive capacity of these scores is limited, because some of them only include clinical variables or radiological variables, but none of them have a combination of these two. Additionally, the scores that include radiological information require the subjective interpretation of the treating physician, who might not have enough expertise to interpret these images.

The evaluation of diagnostic images is crucial for diagnosing COVID-19 pneumonia regardless of the result of the reverse transcriptase (RT)-PCR, especially in patients with high clinical suspicion [15, 16]. A chest radiograph is the most frequently utilised image to diagnose pneumonia and COVID-19; however, the image reading process is highly variable among observers [17]. Features identified in chest radiographs could be lung consolidations, ground-glass opacity, nodules and reticular-nodular opacities, leaving the diagnostic capability of the test to the subjective physician interpretation [18]. This limitation is fundamental in areas or hospitals where untrained radiologists are available.

Therefore, there is a need for novel approaches that use easy-to-access clinical data and computer-based image interpretation algorithms that allow untrained clinicians to accurately identify patients at higher risk of developing severe COVID-19 or dying. We hypothesise that using artificial intelligence (AI) and advanced statistical models, we could create an algorithm that detects patients at risk of dying using chest radiographs and some easy-to-access clinical data. We conducted a study to test this hypothesis, using a previously developed AI algorithm to interpret chest radiographs and clinical data collected for a prospective multicentre study.

## Material and methods

### *Study design*

This is a prospective diagnostic test study. Clinical data were collected in the LIVEN COVID-19 study, a voluntary registry created by the Latin American Intensive Care Network [19, 20]. Variables were compiled by the attending physicians, who reviewed medical records and diagnostic testing data for patients admitted to 22 hospitals across eight Latin-American countries, with SARS-CoV-2 infection confirmed by RT-PCR between March 2020 and January 2021. This study aimed to determine the risk factors associated with the development of severe COVID-19 and death. In this sense, three models were trained for the evaluation. The first model assessed predictions using only chest radiographs; the second model used clinical variables to predict outcomes; finally, the third model used both images and clinical data to identify patients at risk of developing severe COVID-19 or dying during hospital admission. This study was approved by the institutional review board of the Clinica Universidad de La Sabana (TSICCM CUS0012). This study was a secondary analysis of a dataset collected prospectively, so informed consent was waived.

### *Data collection*

Data included sociodemographic variables, comorbid conditions, symptoms, vital signs on hospital admission and treatments received during the hospitalisation. Obesity was determined by treating physicians when the patients had body mass index >30 $kg \cdot m^{-2}$. Additionally, chest radiograph images collected on hospital admission were reported in some patients and were used in our models. Physiological variables and laboratory results were gathered during the first 24 h of hospital admission. All data were collected in the Research Electronic Data Capture platform (REDCap, version 8.11.11; Vanderbilt University, Nashville, TN, USA) [21] hosted at the Universidad de La Sabana (Chía, Colombia). Clinical variables were pre-processed before training the proposed classifier. Incomplete clinical information was considered as a general exclusion criterion. Subjects without a chest radiograph were excluded from the image-based model and were included only in the clinical cohort (figure 1).

### *Model construction*

Transfer learning was used to train two hybrid architectures. These architectures were designed to extract features from images and clinical data to predict intensive care unit (ICU) admission and hospital mortality. A hold-out scheme was used whenever each cohort was assessed (clinical or images), specifically for the images cohort; 70% was reserved for training (these data were used to fit the model); 12% for validation (these data were used to provide an unbiased evaluation of the model fit on the training dataset while tuning model hyperparameters and stopping the training early); finally, the remaining 18%

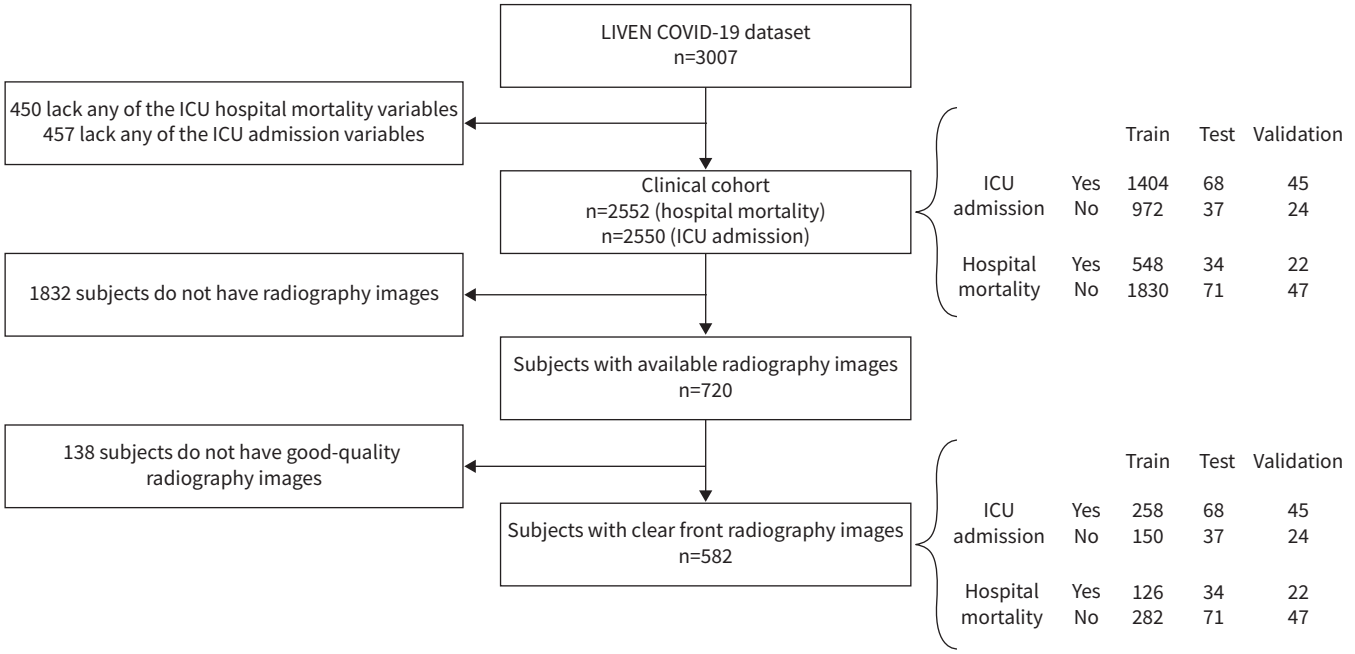


FIGURE 1 Cohorts for outcome assessments for the Latin American Intensive Care Network (LIVEN) coronavirus disease 2019 (COVID-19) dataset. Exclusion criteria are presented, and splits for the clinical cohort and images cohort are specified. ICU: intensive care unit.

was used for testing (these data were used to provide an unbiased evaluation of the final model). Furthermore, to keep all experimentation under the same testing conditions, the subjects selected for validation and testing in the images cohort were also used as such for the clinical cohort, which, due to different sample sizes, results in a differently proportioned split whenever clinical models were assessed, to 92.6% training, 2.9% validation and 4.5% testing.

The images model was a fine-tuned model pre-trained with ImageNet weights (figure 2a). It uses "Hippocrates", a tool that tests for five different backbones (MobileNet, InceptionV3, DenseNet121, Xception), a range of neurons number that goes from 32 to 256 in the last fully connected layer, several values for dropout weights ranging from 0.3 to 0.7 and multiple top-layer weights for classification. As a result, one loss optimisation was carried out numerous times per backbone and hyperparameters setup, and the model yielding the best performance was selected.

Some preparation and pre-processing for training models with images had to be done before trying backbone learning and convolutional neural network training. Firstly, only frontal views of posterior–anterior and anterior–posterior chest radiograph images with well-defined anatomical structures were selected; as a result, images with strong artefacts or heavily blurred structures were discarded from the study. Secondly, as there were variations in contrast, grey-level intensity and capture methods, a pre-processing algorithm was used to take all images to the same dynamic range and remove elements that were not part of the image.

After pre-processing, backbone learning was performed by letting each backbone and parameter configuration learn during a fixed number of iterations. Each network processed the whole dataset five times (five epochs). Then, the setup that yielded the best performance was used to train over 20 epochs with an early stop to avoid overfitting during the process. The combined model was used to exploit both clinical and image information in the classification process. A custom architecture was proposed for this model by connecting single sigmoid outputs of both models to a single neuron that predicts the likelihood of any given class (figure 2c). For this case, 002C feature extraction weights for both separate models were frozen, since the only weights that could be learned were related to the contribution of clinical and image information for the output prediction and its respective bias. The combined models were trained for 150 epochs with an early-stop call-back that prevented the model from overfitting by monitoring validation loss decay. The binary cross-entropy and stochastic gradient descent (SGD) with a learning rate of 0.01 were used as cost function and optimiser.

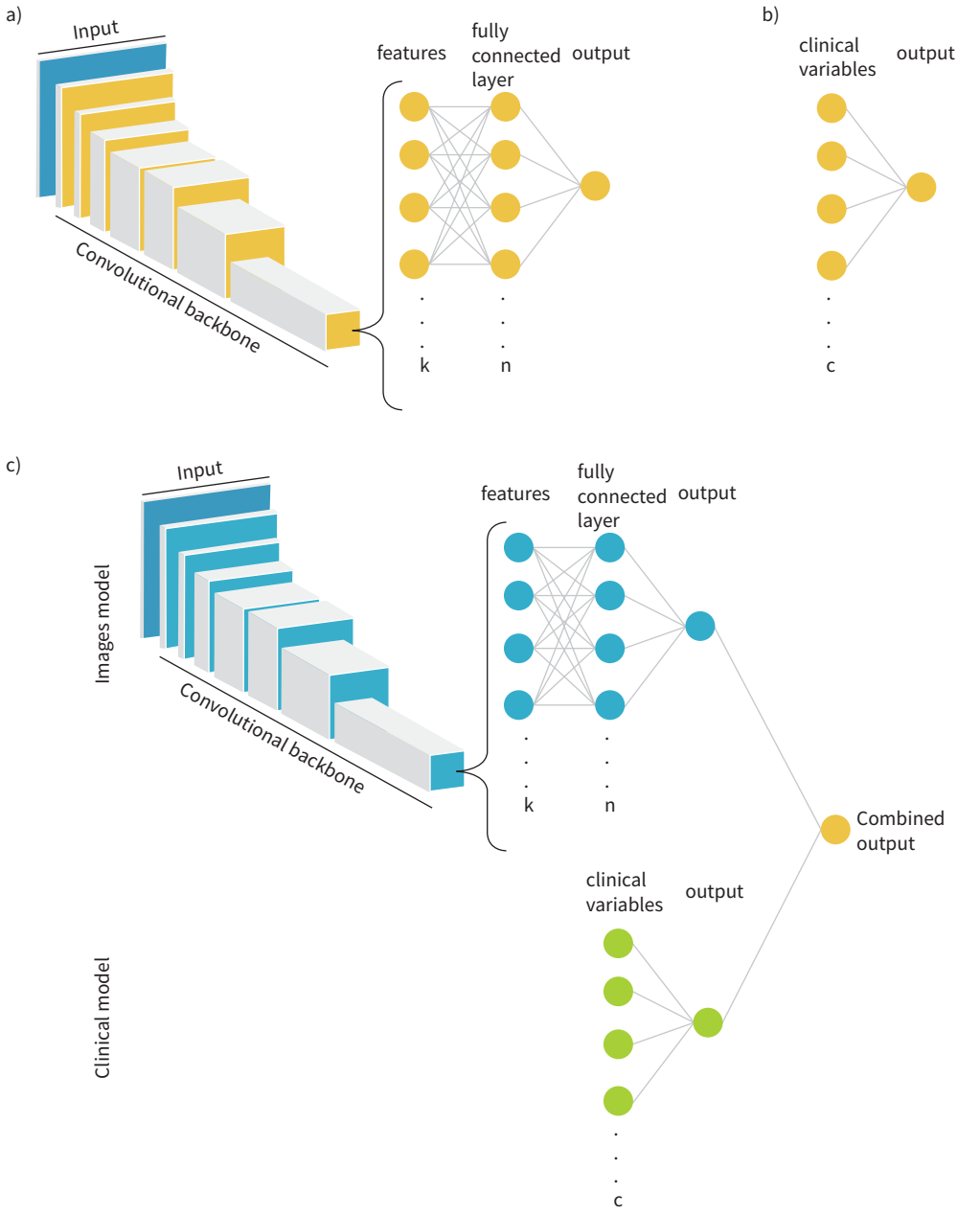


**FIGURE 2** Convolutional neural network model construction. a) Proposed approach for obtaining a model from images by backbone learning; b) proposed perceptron model to use clinical data for outcome assessment; c) proposed combination of a) and b).

### *Statistical analysis*

To predict the probability of ICU admission or hospital mortality, a random forest model was used. The random forest model is an ensemble-learning model that uses multiple decision trees as its base models. In the end, a majority voting system is implemented to synthesise the results of all the base models. Additionally, a logistic regression model was designed to select the clinical variables and laboratory results that best predicted the outcomes. Sociodemographic and physiological data selected by the random forest model were included as independent variables in the multivariate analysis. Some variables were included for biological plausibility. Odds ratios were obtained based on the exponentials of the final logistic regression model coefficients.

A predictive clinical model was built as a simple perceptron model for the AI prototype (figure 2b). All the selected clinical variables in the logistic regression model were connected to a single neuron output

layer with a sigmoid activation function. Then, the model was trained for 150 epochs with an early-stop call-back that prevents it from overfitting by monitoring validation loss decay. For this training, the binary cross-entropy was defined as the cost function, and SGD with a learning rate of 0.001 was used as the optimiser.

A statistical analysis using bootstrapping was performed to validate each model. First, the testing set was sampled with replacement to obtain 250 samples, each with a sample size equal to the 50% of the size of the original set. Then, each sample was assessed by using the area under the receiver operating characteristic curve (AUC-ROC), obtaining a metric population (AUC population). Finally, statistics such as mean, standard deviation and 95% confidence intervals were also computed for the AUC population. As a result, the bootstrapped test of the AUC population calculated per model was used to compare the performance of the proposed models. Then, a distribution comparison of the three AUC populations under a t-test to establish statistically significant differences across the proposed models were performed. Additional evaluation measurements like sensitivity, specificity and accuracy metrics were also computed over all the testing sets for each model. All statistical analyses were performed using SciPy 1.7.1 in Python 3.8 and R Studio version 1.3.1056.

## Results

3007 patients were registered in the LIVEN COVID-19 study. After excluding patients with no ICU hospital mortality data or the clinical variables needed on clinical admission, 2550 patients were included in the clinical cohort for the ICU admission predictive model, and 2552 for the hospital mortality analysis. 59.5% (1517 out of 2550) of the patients required ICU admission and were distributed in the models as follows: 92.6% (1404 out of 1517) underwent training, 4.5% (68 out of 1517) underwent testing and 2.9% (45 out of 1517) underwent validation. Of all the patients included in the clinical cohort, 23.7% (604 out of 2552) died during hospital admission. Figure 1 presents how these patients were distributed in the models.

23.9% (720 out of 3007) of the overall cohort had chest radiograph images available; however, 80.8% (582 out of 720) had clear frontal images. Of the images cohort, 31.3% (182 out of 582) of patients died, and of these, 69.2% (126 out of 182) underwent training, 18.7% (34 out of 182) underwent testing and 12.1% (22 out of 182) underwent validation. 63.7% (371 out of 582) required ICU admission and were distributed as presented in figure 1.

The variables independently associated with ICU admission were age (OR 1.62, 95% CI 1.43–1.83; p<0.001), fraction of inspired oxygen ($F_{iO_2}$) on admission (OR 4.10, 95% CI 3.55–4.73; p<0.001), systolic pressure on admission (OR 1.20, 95% CI 1.05–1.38; p=0.007), diastolic pressure on admission (OR 0.80, 95% CI 0.70–0.93; p=0.003), oxygen saturation ($S_{O_2}$) (OR 0.84, 95% CI 0.76–0.94; p=0.002), Glasgow Coma Scale score on admission (OR 0.60, 95% CI 0.53–0.69; p=0.007), male sex (OR 1.42, 95% CI 1.28–1.59; p<0.001), dyspnoea on admission (OR 1.42, 95% CI 1.28–1.58; p<0.001), obesity (OR 1.42, 95% CI 1.28–1.58; p<0.001), arterial hypertension (OR 1.17, 95% CI 1.05–1.32; p=0.005) and diabetes mellitus (OR 1.22, 95% CI 1.10–1.36; p<0.001). Vomiting/nausea, chronic kidney disease, conjunctivitis and skin ulcers were not relevant for this final model (table 1).

The variables associated with hospital mortality were age (OR 1.68, 95% CI 1.51–1.87; p<0.001), $F_{iO_2}$ on admission (OR 4.32, 95% CI 3.75–4.97; p<0.001), systolic blood pressure on admission (OR 1.20, 95% CI 1.05–1.38; p=0.007), diastolic blood pressure on admission (OR 0.80, 95% CI 0.70–0.93; p=0.003), $S_{O_2}$ (OR 0.82, 95% CI 0.74–0.91; p<0.001), Glasgow Coma Scale score on admission (OR 0.61, 95% CI 0.54–0.69; p<0.001), male sex (OR 1.44, 95% CI 1.29–1.60; p<0.001), dyspnoea on admission (OR 1.50, 95% CI 1.35–1.66; p<0.001), obesity (OR 1.43, 95% CI 1.28–1.59; p<0.001), chronic kidney disease (OR 1.20, 95% CI 1.08–1.33; p<0.001) and arterial hypertension (OR 1.21, 95% CI 1.08–1.35; p=0.001). Diabetes mellitus on admission was not relevant for this final model (table 1).

### *ICU admission models*

ROC curves for ICU admission are presented in figure 3a, c, e. This assessment yielded a performance of 0.88±0.05 for the images-based model, 0.90±0.04 for the clinical model and 0.92±0.04 for the combined model. Furthermore, additional metrics such as sensitivity and specificity are provided for each model in table 2. All possible combinations of each AUC for the three models showed statistical differences (p<0.0001). Visualisation of ROC populations and mean curves for each model are displayed in figure 4a, b.

### *Hospital mortality*

The ROC curves for hospital mortality are presented in figure 3b, d, f. Additionally, metrics such as sensitivity and specificity are provided for each model in table 2. This assessment yielded performances of

**TABLE 1** Clinical variables selected by logistic regression models to assess intensive care unit (ICU) admission and hospital mortality prediction

| | ICU admission | | Hospital mortality | |
|---|---|---|---|---|
| | Logistic regression variables OR (95% CI) | p-value | Logistic regression variables OR (95% CI) | p-value |
| Age | 1.62 (1.43–1.83) | <0.001 | 1.68 (1.51–1.87) | <0.001 |
| $F_{iO_2}$ | 4.10 (3.55–4.73) | <0.001 | 4.32 (3.75–4.97) | <0.001 |
| Systolic blood pressure | 1.20 (1.05–1.38) | 0.007 | 1.20 (1.05–1.38) | 0.007 |
| Diastolic blood pressure | 0.80 (0.70–0.93) | 0.003 | 0.80 (0.70–0.93) | 0.003 |
| $S_{O_2}$ | 0.84 (0.76–0.94) | 0.002 | 0.82 (0.74–0.91) | <0.001 |
| Glasgow Coma Scale | 0.60 (0.53–0.69) | <0.001 | 0.61 (0.54–0.69) | <0.001 |
| Sex | 1.42 (1.28–1.59) | <0.001 | 1.44 (1.29–1.60) | <0.001 |
| Dyspnoea | 1.42 (1.28–1.58) | <0.001 | 1.50 (1.35–1.66) | <0.001 |
| Obesity | 1.42 (1.28–1.58) | <0.001 | 1.43 (1.28–1.59) | <0.001 |
| Vomiting/nausea[#] | | | 1.00 (0.90–1.11) | 0.88 |
| Abdominal pain | 1.08 (0.98–1.20) | 0.11 | 1.05 (0.94–1.16) | 0.34 |
| Chronic kidney disease[#] | | | 1.20 (1.08–1.33) | <0.001 |
| Conjunctivitis[#] | | | 1.00 (0.90–1.12) | 0.91 |
| Arterial hypertension | 1.17 (1.05–1.32) | 0.005 | 1.21 (1.08–1.35) | 0.001 |
| Skin ulcers[#] | | | 0.95 (0.86–1.06) | 0.43 |
| Diabetes mellitus[#,¶] | 1.22 (1.10–1.36) | <0.001 | | |

$F_{iO_2}$: fraction of inspired oxygen; $S_{O_2}$: oxygen saturation. [#]: missing values were variables not selected for the predictive model analysis due to statistical significance or biological plausibility; [¶]: noncomplicated diabetes.

0.75±0.07 for the images-based model, 0.81±0.06 for the clinical model and 0.81±0.06 for the combined model. Sensitivity performance was 71%, 75% and 75% for the images, clinical and combined models, respectively. Similarly, specificity metrics were 76%, 71% and 74%, respectively, for the three proposed models. Additionally, positive predictive values were 59%, 57% and 58%, respectively, and negative predictive values were 84%, 84% and 85%, respectively. No statistically significant differences were found in the AUC comparison between clinical and combined models (p=0.13). However, when the AUC of the imaging model was compared with the combined model (p<0.0001) and the AUC of the clinical model with the images-based model (p<0.0001), statistically significant differences were found (figure 4c, d).

## Discussion

This study presents algorithms for predicting whether COVID-19 patients may require ICU admission or are likely to die during hospitalisation by using an automatised method to interpret chest radiographs, clinical variables and a combination of both. We found that models constructed with chest radiograph images (interpreted by an AI algorithm) and clinical data presented good discriminatory performance regarding ICU admission and hospital mortality. Notably, the models using clinical data and the AI algorithm combined had an excellent discriminatory power to identify patients at risk of developing severe COVID-19. Nevertheless, predicting hospital mortality by combining chest radiography features and clinical information was not statistically significant. Finally, the chest radiograph images model alone had the lower predictive potential for both outcomes.

Different predictive models for the COVID-19 illness progression have been developed throughout the pandemic. Routinely measured clinical variables have been used as essential predictors of severity. Zhao *et al.* [22], in a retrospective study of 4997 COVID-19 patients, showed that the presence of shortness of breath, elevated heart rate, elevated respiratory rate and decreased pulse oxygen saturation was significantly associated with a higher proportion of patients admitted to the ICU. Additionally, other authors have described that diagnostic image analysis provides consistent information of pulmonary involvement and complements clinical prediction in COVID-19 patients. In another study, Jiao *et al.* [23] developed a multicentre retrospective study of 1834 patients with COVID-19, reporting that when chest radiographs were added to clinical data, the ROC curve increased from 0.82 (95% CI 0.79–0.82) to 0.84 (0.81–0.85), with p<0.0001 for severity prediction. On top of that, Soda *et al.* [24] designed a hybrid approach model using clinical data associated with chest radiograph images of 820 COVID-19 patients finding the best performance for critical infection prediction when using both inputs. Our combined model also demonstrated that the clinical information provides consistent performance that improves the classification metrics when complemented with an AI image features extraction algorithm. These results suggest a complementary role between imaging, demographics, routine laboratories involving lung function and others to determine whether a patient is likely to require ICU admission.

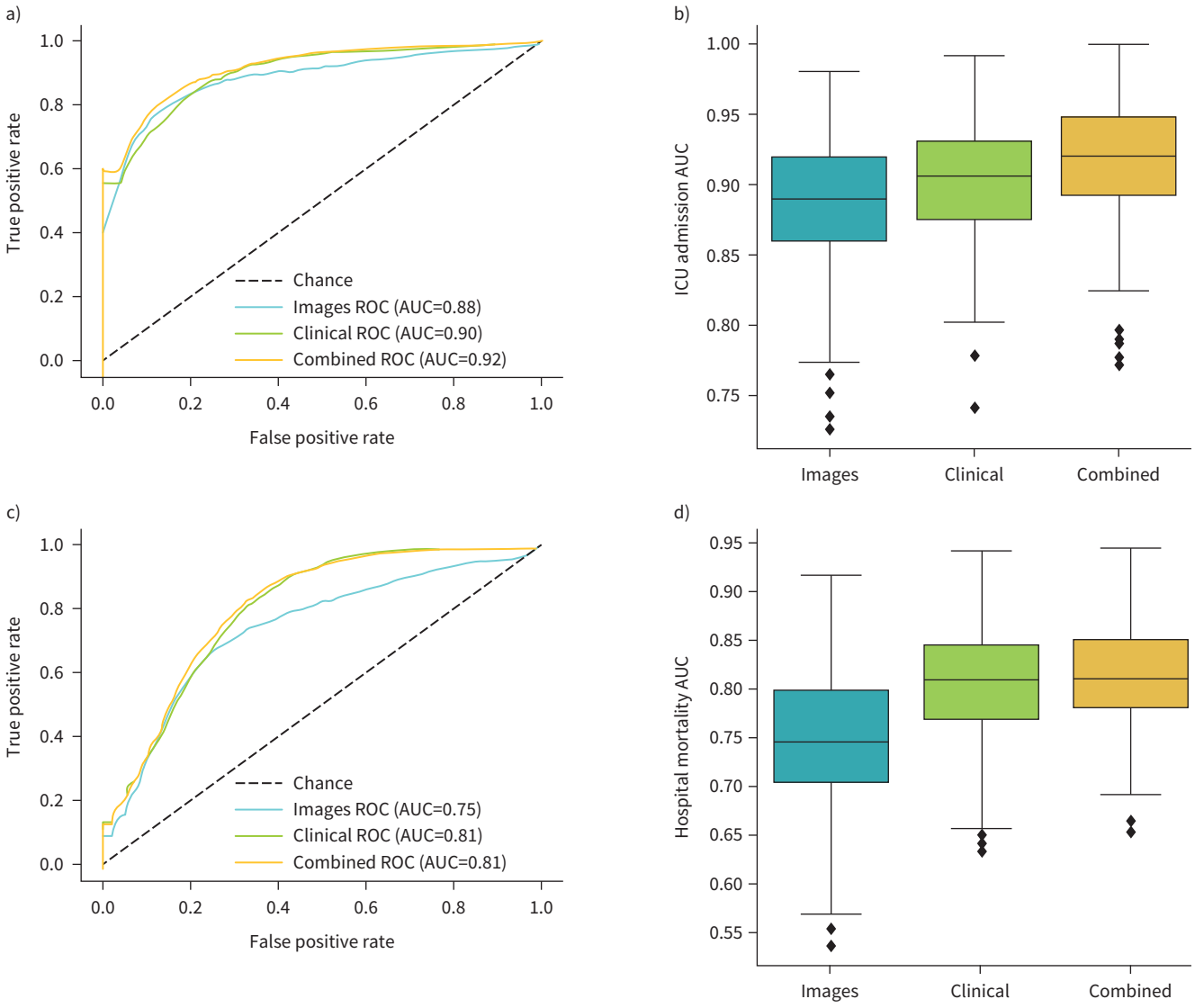


**FIGURE 3** Receiver operating characteristic (ROC) curves of a) intensive care unit (ICU) admission and c) hospital mortality assessment and b, d) statistical comparison of models per outcome assessment. AUC: area under the curve.

**TABLE 2** Performance metrics for intensive care unit (ICU) admission and hospital mortality model assessment

| | Sensitivity | Specificity | PPV | NPV | AUC | |
|---|---|---|---|---|---|---|
| | | | | | Mean±SD | 95% CI |
| **ICU admission** | | | | | | |
| Images | 0.85 | 0.81 | 0.89 | 0.75 | 0.88±0.05 | 0.8788–0.8911 |
| Clinical | 0.87 | 0.78 | 0.88 | 0.76 | 0.90±0.04 | 0.8956–0.9059 |
| Combined | 0.91 | 0.78 | 0.89 | 0.83 | 0.92±0.04 | 0.9113–0.9218 |
| **Hospital mortality** | | | | | | |
| Images | 0.71 | 0.76 | 0.59 | 0.84 | 0.75±0.07 | 0.7368–0.7546 |
| Clinical | 0.71 | 0.75 | 0.57 | 0.84 | 0.81±0.06 | 0.7981–0.8132 |
| Combined | 0.74 | 0.75 | 0.58 | 0.85 | 0.81±0.06 | 0.8066–0.8205 |

PPV: positive predictive value; NPV: negative predictive value; AUC: area under the curve.

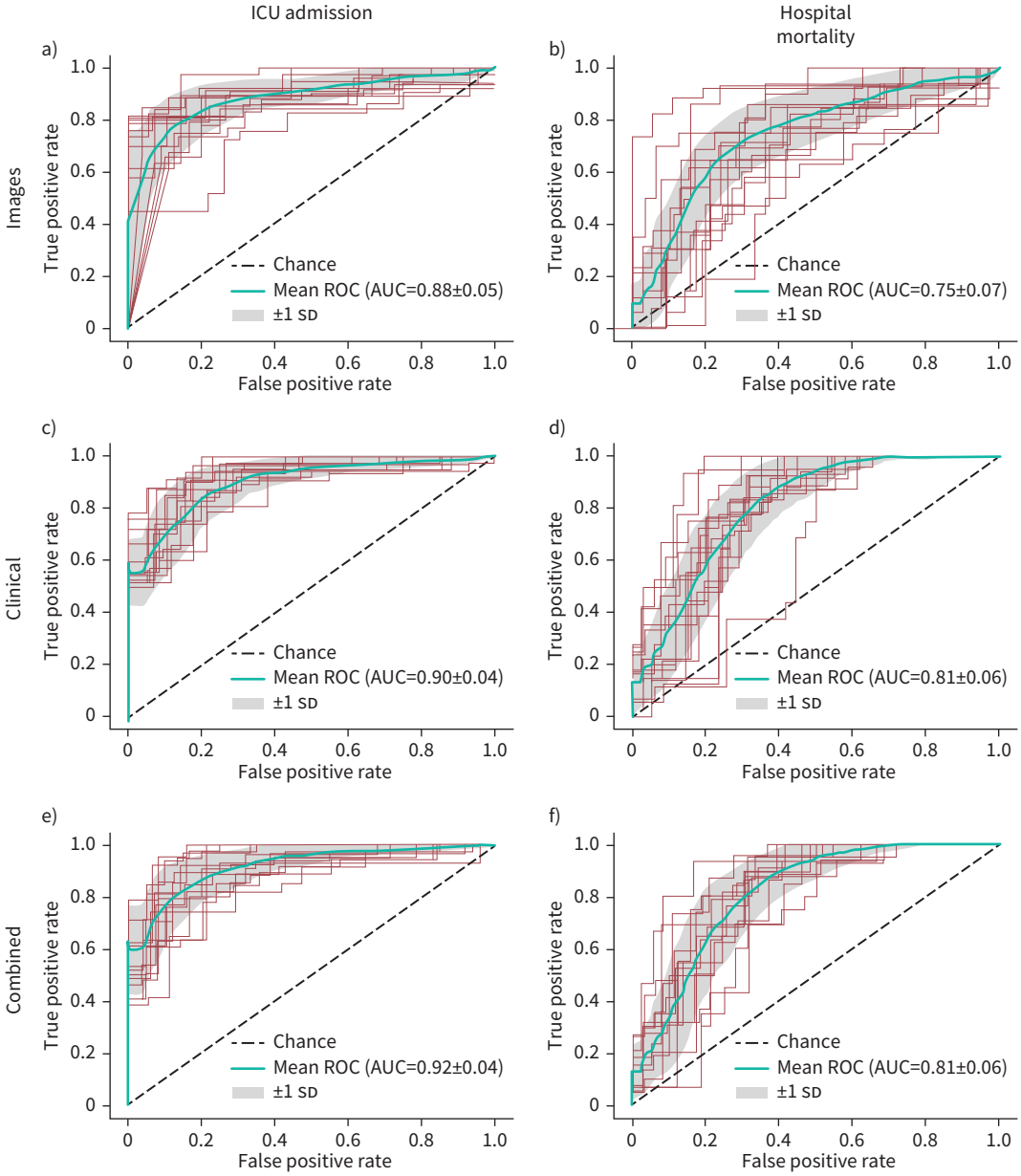


**FIGURE 4** Receiver operating characteristic (ROC) curves of a, c, e) intensive care unit (ICU) admission and b, d, f) hospital mortality assessment using proposed models. AUC: area under the curve.

In the case of prediction of hospital mortality, Balbi *et al.* [25] found that arterial oxygen tension/$F_{iO_2}$ ratio was associated with higher mortality rate (OR 0.99, 95% CI 0.98–1.00; p= 0.002), as well as the presence of cardiovascular disease (OR 3.21, 95% CI 1.28–8.39; p<0.014) and age (OR 1.16, 95% CI 1.11–1.22; p<0.001), with statistical differences, despite having only 340 COVID-19 patients in their retrospective study. Similarly, Ryan *et al.* [26] developed a model by a methodology of "boosted" decision trees including variables such as age, heart rate, respiratory rate, peripheral oxygen saturation, temperature, systolic blood pressure, diastolic blood pressure, white blood cell counts, platelets, lactate, creatinine and bilirubin levels, reporting an AUC-ROC of 0.86 to predict 48-h mortality in COVID-19 patients. Moreover, the model showed a better performance compared with the AUC-ROC of current 48-h mortality scores (quick Sepsis Related Organ Failure Assessment 0.792, Modified Early Warning Score 0.724 and CURB-65 (confusion, urea >7 mmol·L$^{-1}$, respiratory rate ⩾30 breaths·min$^{-1}$, blood pressure <90 mmHg systolic and/or 60 mmHg diastolic, age ⩾65 years) 0.802) [26]. Nevertheless, this study had a community hospital dataset of only 114 COVID-19 patients and did not include information about diagnostic images.

Our clinical model included similar variables and had better discriminatory power than previous studies. Importantly, in our study, we used the random forest analysis which is more robust to identify the variables associated with the outcomes. However, our model using automatised chest radiograph interpretation and clinical data had a modest prediction power for mortality.

Different imaging findings have been associated with COVID-19 disease. Balbi *et al.* [25] described how ground-glass opacities with consolidation (69%) were the most common chest radiograph finding evaluated in COVID-19 patients, with an almost perfect inter-rater agreement related with the parenchymal opacity ($\kappa$=0.90), Brixia score (intraclass coefficient (ICC) 0.91) and percentage of lung involvement (ICC 0.95). Nevertheless, the chest radiograph characteristics and the risk of mortality or risk for ICU admission were not assessed in the multivariate analysis. Likewise, Au-Yong *et al.* [27], in their retrospective cohort study of 751 patients with COVID-19, demonstrated that a higher percentage of chest radiograph opacity is related to lower survival: 50–75% opacity had a median 7.6 escalation-free survival days (95% CI 5.4–23.7 days) and 76–100% opacity had 2.6 days (95% CI 1.5–16.6 days) ($p<0.001$). Despite this, our results suggest that an automatised model of interpretation of chest radiograph characteristics alone did not achieve the best performance for predicting COVID-19 severity or in-hospital mortality compared to models that combined clinical data. Thus, we believe that using our automated algorithm to interpret chest radiographs and some clinical characteristics might be extremely useful to identify patients at risk of developing severe COVID-19 and those at risk of dying due to this infection. Determining these high-risk patients might be critical when chest radiographs are interpreted by untrained personnel.

There are some limitations of this study that are important to consider. First, a few chest radiograph images had inconsistent quality and noisy data which may have affected the performance of the automatised algorithm for reading the images in clinical practice. However, the images could be pre-processed and fixed to mitigate this issue. Second, the number of patients with images available was small. Therefore, the sample size of both images and combined cohorts was less than the clinical information sample. The algorithm's performance might have been affected, causing it to fail to make robust predictions. However, this is one of the few studies that have included radiological findings in predictor models, which is a strength of our algorithms and allows us to generate new hypotheses about using artificial intelligence in medical practice. Indeed, the model's performance with a chest radiograph could improve the prediction capacity when combined with clinical variables. Third, although deep neural networks have exhibited superior performance in various tasks, interpretability is always the Achilles' heel of deep neural networks. At present, deep neural networks obtain high discrimination power at the cost of low interpretability of their black-box representations. We believe that high model interpretability may help people break several bottlenecks of deep learning, *e.g.* learning from very few annotations, learning *via* human–computer communications at the semantic level and semantically debugging network representations. Likewise, the applicability of the models will depend to a great extent on the local health systems and the willingness of the clinicians to request the images and the corresponding laboratories. Future follow-up studies will add tremendous value to the current evidence by testing this model, specifically those based on radiograph images.

To summarise, our study presents evidence that our automatised algorithm to interpret chest radiographs along with some clinical data might be an instrumental tool to identify patients at higher risk of developing severe COVID-19. Notably, our model does not require the physician's interpretation of the images; it only requires an image, and the AI system interprets the variables and makes an automated analysis. Predicting ICU admission using images, clinical information or a combination of both yielded consistent results across all three experiments. The combined model is the best to identify patients at risk of severe COVID-19. Additionally, chest radiograph images demonstrated better predictive power in the case of ICU admission compared with mortality prediction, and their utility is improved when it is complemented with clinical information. Future work involves clinical trials with ICU admission predictors that evaluate our models' external validation and improve clinical outcomes of COVID-19 patients.

Provenance: Submitted article, peer reviewed.

This article has supplementary material available from openres.ersjournals.com

Conflict of interest: All authors have no conflict of interest.

## References

1 Bloom CI, Drake TM, Docherty AB, *et al.* Risk of adverse outcomes in patients with underlying respiratory conditions admitted to hospital with COVID-19: a national, multicentre prospective cohort study using the ISARIC WHO Clinical Characterisation Protocol UK. *Lancet Respir Med* 2021; 9: 699–711.

2 Doğanay F, Ak R. Performance of the CURB-65, ISARIC-4C and COVID-GRAM scores in terms of severity for COVID-19 patients. *Int J Clin Pract* 2021; 75: e14759.
3 ISARIC Clinical Characterisation Group. COVID-19 symptoms at hospital admission vary with age and sex: results from the ISARIC prospective multinational observational study. *Infection* 2021; 49: 889–905.
4 Reyes LF, Murthy S, Garcia-Gallo E, *et al*. Clinical characteristics, risk factors and outcomes in patients with severe COVID-19 registered in the International Severe Acute Respiratory and Emerging Infection Consortium WHO clinical characterisation protocol: a prospective, multinational, multicentre, observational study. *ERJ Open Res* 2022; 8: 00552-2021.
5 Docherty AB, Harrison EM, Green CA, *et al.* Features of 20 133 UK patients in hospital with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020; 369: m1985.
6 Rello J, James A, Reyes LF. Post-acute COVID-19 Syndrome (PACS): a public health emergency. *Anaesth Crit Care Pain Med* 2021; 40: 100882.
7 Ali R, Qayyum F, Ahmed N, *et al.* ISARIC 4c mortality score as a predictor of in-hospital mortality in COVID-19 patients admitted in Ayub teaching hospital during first wave of the pandemic. *J Ayub Med Coll Abbottabad* 2021; 33: 20–25.
8 Rodríguez A, Moreno G, Gómez J, *et al.* Severe infection due to the SARS-CoV-2 coronavirus: experience of a tertiary hospital with COVID-19 patients during the 2020 pandemic. *Med Intensiva* 2020; 44: 525–533.
9 Bastard P, Gervais A, Le Voyer T, *et al.* Autoantibodies neutralizing type I IFNs are present in ∼4% of uninfected individuals over 70 years old and account for ∼20% of COVID-19 deaths. *Sci Immunol* 2021; 6: eabl4340.
10 Richardson S, Hirsch JS, Narasimhan M, *et al.* Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA* 2020; 323: 2052–2059.
11 Ospina-Tascón GA, Calderón-Tapia LE, García AF, *et al.* Effect of high-flow oxygen therapy *vs* conventional oxygen therapy on invasive mechanical ventilation and clinical recovery in patients with severe COVID-19: a randomized clinical trial. *JAMA* 2021; 326: 2161–2171.
12 COVID-ICU Group on behalf of the REVA Network and the COVID-ICU Investigators. Clinical characteristics and day-90 outcomes of 4244 critically ill adults with COVID-19: a prospective cohort study. *Intensive Care Med* 2021; 47: 60–73.
13 Paules CI, Fauci AS. COVID-19: the therapeutic landscape. *Med* 2021; 2: 493–497.
14 Rodríguez A, Ruiz-Botella M, Martín-Loeches I, *et al.* Deploying unsupervised clustering analysis to derive clinical phenotypes and risk factors associated with mortality risk in 2022 critically ill patients with COVID-19 in Spain. *Crit Care* 2021; 25: 63.
15 Coopersmith CM, Antonelli M, Bauer SR, *et al.* The Surviving Sepsis Campaign: research priorities for coronavirus disease 2019 in critical illness. *Crit Care Med* 2021; 49: 598–622.
16 Chalmers JD, Crichton ML, Goeminne PC, *et al.* Management of hospitalised adults with coronavirus disease 2019 (COVID-19): a European Respiratory Society living guideline. *Eur Respir J* 2021; 57: 2100048.
17 Cozzi D, Albanesi M, Cavigli E, *et al.* Chest X-ray in new coronavirus disease 2019 (COVID-19) infection: findings and correlation with clinical outcome. *Radiol Med* 2020; 125: 730–737.
18 Wang S, Zha Y, Li W, *et al.* A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J* 2020; 56: 2000775.
19 Reyes LF, Bastidas A, Narváez PO, *et al*. Clinical characteristics, systemic complications, and in-hospital outcomes for patients with COVID-19 in Latin America. LIVEN-Covid-19 study: a prospective, multicenter, multinational, cohort study. *PLoS One* 2022; 17: e0265529.
20 Reyes LF, Rodriguez A, Bastidas A, *et al*. Dexamethasone as risk-factor for ICU-acquired respiratory tract infections in severe COVID-19. *J Crit Care* 2022; 69: 154014.
21 Harris PA, Taylor R, Thielke R, *et al.* Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; 42: 377–381.
22 Zhao Z, Chen A, Hou W, *et al.* Prediction model and risk scores of ICU admission and mortality in COVID-19. *PLoS One* 2020; 15: e0236618.
23 Jiao Z, Choi JW, Halsey K, *et al.* Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study. *Lancet Digit Health* 2021; 3: e286–e294.
24 Soda P, D'Amico NC, Tessadori J, *et al.* AIforCOVID: predicting the clinical outcomes in patients with COVID-19 applying AI to chest-X-rays. An Italian multicentre study. *Med Image Anal* 2021; 74: 102216.
25 Balbi M, Caroli A, Corsi A, *et al.* Chest X-ray for predicting mortality and the need for ventilatory support in COVID-19 patients presenting to the emergency department. *Eur Radiol* 2021; 31: 1999–2012.
26 Ryan L, Lam C, Mataraso S, *et al.* Mortality prediction model for the triage of COVID-19, pneumonia, and mechanically ventilated ICU patients: a retrospective study. *Ann Med Surg* 2020; 59: 207–216.
27 Au-Yong I, Higashi Y, Giannotti E, *et al.* Chest radiograph scoring alone or combined with other risk scores for predicting outcomes in COVID-19: a UK study. *Radiology* 2021; 301: 210986.