Original research

Deep learning model to predict the need for mechanical ventilation using chest X-ray images in hospitalised patients with COVID-19

Anoop R Kulkarni,[1,2] Ambarish M Athavale,[3] Ashima Sahni,[4] Shashvat Sukhal,[5] Abhimanyu Saini,[6] Mathew Itteera,[3] Sara Zhukovsky,[7] Jane Vernik,[3] Mohan Abraham,[3] Amit Joshi,[3] Amatur Amarah,[3] Juan Ruiz,[3] Peter D Hart,[3] Hemant Kulkarni [2,8]

For numbered affiliations see end of article.

Correspondence to
Dr Hemant Kulkarni, M&H Research LLC, San Antonio, TX 78249, USA; hemant.kulkarni@mnhresearch.com

ARK and AMA contributed equally.

Received 12 November 2020
Revised 18 January 2021
Accepted 13 February 2021

Check for updates

© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Kulkarni AR, Athavale AM, Sahni A, *et al*. *BMJ Innov* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjinnov-2020-000593

ABSTRACT

Objectives There exists a wide gap in the availability of mechanical ventilator devices and their acute need in the context of the COVID-19 pandemic. An initial triaging method that accurately identifies the need for mechanical ventilation in hospitalised patients with COVID-19 is needed. We aimed to investigate if a potentially deteriorating clinical course in hospitalised patients with COVID-19 can be detected using all X-ray images taken during hospitalisation.

Methods We exploited the well-established DenseNet121 deep learning architecture for this purpose on 663 X-ray images acquired from 528 hospitalised patients with COVID-19. Two Pulmonary and Critical Care experts blindly and independently evaluated the same X-ray images for the purpose of validation.

Results We found that our deep learning model predicted the need for mechanical ventilation with a high accuracy, sensitivity and specificity (90.06%, 86.34% and 84.38%, respectively). This prediction was done approximately 3 days ahead of the actual intubation event. Our model also outperformed two Pulmonary and Critical Care experts who evaluated the same X-ray images and provided an incremental accuracy of 7.24%–13.25%.

Conclusions Our deep learning model accurately predicted the need for mechanical ventilation early during hospitalisation of patients with COVID-19. Until effective preventive or treatment measures become widely available for patients with COVID-19, prognostic stratification as provided by our model is likely to be highly valuable.

Summary box

What are the new findings?

- Artificial intelligence techniques can be used to accurately predict the need for mechanical ventilation in patients with COVID-19.
- Deep leaning–based prediction of the need for mechanical ventilation outperforms predictions offered by subject experts.

How might it impact on healthcare in the future?

- Triaging of hospital resources for admitted patients with COVID-19 is paramount. Our tool can play an important role in this regard.
- Conceptually, our tool can therefore help in a rational and optimal use of resources for management of patients with COVID-19.

INTRODUCTION

The COVID-19 global pandemic has caused almost 50 million infections and over 1.2 million deaths within a span of just over 10 months.[1] Strikingly, the cumulative COVID-19 hospitalisation rate is 137.6 per 100 000 infections.[2] A significant number of patients with COVID-19 need supportive care such as intravenous fluid administration and supplemental oxygen. Further, as many as 32% of hospitalised patients with COVID-19 need admission to an intensive care unit[3] and respiratory support through mechanical ventilation.[4 5] This has caused a great strain in hospital resources in certain

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.

geographical regions with high rates of COVID-19 infection. For example, at the height of COVID-19 pandemic in Wuhan and New York, there were concerns of the health system being overwhelmed from the sheer number of patients requiring hospitalisation. In Wuhan, a temporary COVID-19 facility was built, and in New York, a United States Navy hospital ship was dispatched to help cope with the number of patients requiring hospitalisation.[6] Since then, COVID-19 has now spread globally and has, in many instances, severely tested healthcare system capacity to handle the sheer number of patients that have continued to flood healthcare facilities. Considering inadequate vaccination resources and suboptimal antiviral treatments, hospital systems will need to be prepared for an increase in hospitalisation rates, ICU admissions and need for mechanical ventilation.

It has been observed that many patients with COVID-19 experience a worsening of shortness of breath and need for supplemental oxygen or mechanical ventilation during the second week of the illness.[7] However, not every patient who is hospitalised with COVID-19 infection needs mechanical ventilation or ICU level of care. Thus, a tool that can effectively predict the potential need for mechanical ventilation would ensure a better triage at initial point of contact with healthcare system and enable better allocation of healthcare resources by avoiding unnecessary hospitalisations. This was the motivation for the present study.

In this context, a deep learning analysis of chest radiograph was able to identify patients with COVID-19 infection with more than 90% accuracy.[8] Also, Wang *et al*[9] were able to stratify patients in high-risk and low-risk groups by a deep learning analysis of lung CT images. In this study, we focused on using the information contained within chest X-ray images to predict the need for mechanical ventilation. A chest radiograph has practical advantages over CT scans in being more readily available especially in resource-challenged scenarios and less risk of equipment contamination. Indeed, a chest radiograph was performed for every patient with COVID-19 evaluated in our hospital emergency room. Here, we present a deep learning analysis of chest radiograph of patients with COVID-19 to predict need for mechanical ventilation.

METHODS

Study participants

The clinical and image data for this study were collected at the John H. Stroger, Jr Hospital of Cook County, Chicago, IL. All patients with COVID-19 who were admitted to the study centre between 15 March 2020 and 31 May 2020 and followed up until the censoring date of 16 June 2020 were included. The study cohort was identified in two: first, all confirmed COVID-19 cases were identified, and second, only the new inpatients were selected from those identified in step 1. COVID-19 positivity was confirmed for all patients using PCR for the RdRp and N genes of the SARS-CoV-2 virus. Clinical data of these patients were collected by chart reviews. The study did not require informed consent from the patients as the data were retrospectively collected after de-identification.

Chest X-ray images

All patients with symptoms suggestive of a possible infection with SARS-CoV-2 underwent portable antero-posterior chest X-ray assessment at the study centre. Chest X-ray images were acquired using the GE Healthcare Optima XR240-amx system rated at 90 kV and 1.5 mA. The protocol followed was as follows: Ensuring appropriate isolation and distancing practices, the X-ray images were acquired in upright or near-upright posture. Images were saved in dicom and jpg format and were manually scrubbed to remove all identifiable information.

Data preprocessing

The acquired X-ray images were first resized to 224×224 pixels and then centre cropped as required for many deep learning networks that use convolutional layers to parse out image features. To ensure robustness in training and validation of the deep learning network, we undertook two steps in data preprocessing. First, we augmented each image using a random combination of right or left rotation (maximum 30°), random cropping and random lighting. These augmentations permitted us to use different variations of the original image for training the deep learning algorithm thereby reducing the potential overfitting. Second, since the patients who needed mechanical ventilation in the study dataset represented a minority class, for training the network we first oversampled the number of ventilated patients so as to achieve a class balance of ∼50% of X-ray images for ventilated and non-ventilated patients in the training sample. Combination of the first and second steps in data preprocessing yielded a set of 1320 X-ray images from the ventilated patients and 1200 X-ray images from non-ventilated patients. This set of 2520 images was used for network training.

Network architecture

The established and validated CheXNeXt deep learning algorithm[10] as well as the PXR network[11] are based on the DenseNet121[12] architecture. While the CheXNeXt predicts one or more of 14 lung pathologies from an X-ray image of the chest, the PXR network scores an X-ray image for severity of acute respiratory distress syndrome (ARDS). We used the same backbone for our proposed prognosticator algorithm. The architecture of a DenseNet121 network is shown in figure 1B. Briefly, the DenseNet121 represents a series of convolutional operations on the image array (size 224×224 pixels) and is characterised by a serial combination of four dense blocks (D1–D4, figure 1B) interspersed with three transitional blocks (T1–T3,

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.

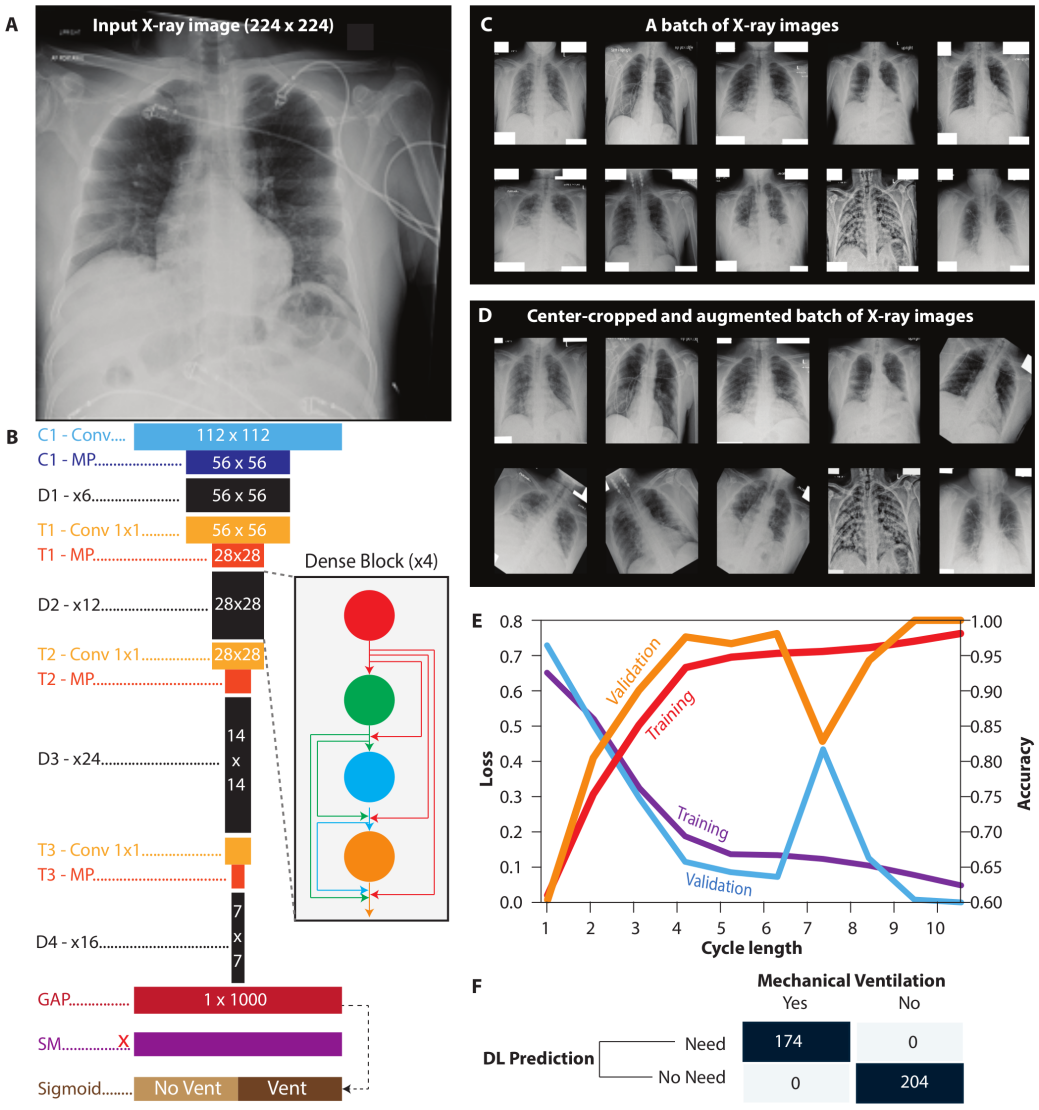

Figure 1 DenseNet121 model, data preprocessing and model training. (A) Example of a preprocessed X-ray image submitted to modelling. (B) The DenseNet121 architecture. Convolutional layers are prefixed with C (cyan), dense blocks with D (black) and transition blocks (orange) with T. GAP, MP, SM and Sigmoid indicate the global average pooling, maxpooling, softmax and binarisation layers within the classifier portion of DenseNet121. Inset shows a dense block with four layers and depicts how each succeeding layer receives inputs from all preceding layers. Shown within each proportionately sized coloured block is the output size in pixels. (C–D) Data preprocessing. Shown in panel C is a batch of resized X-ray images. Panel D shows the same batch after data augmentation that included centre cropping, rotation and horizontal displacement. (E) Training log of DenseNet121 to predict the need for mechanical ventilation. Left axis shows the categorical cross-entropy loss at the end of each cycle length and the right axis shows the estimated accuracy of prediction. Results are shown separately for the training (n=2142) and the validation (n=378) set of X-ray images. (F) Confusion matrix at the end of DenseNet121 training. All the images were correctly classified at this stage.

figure 1B). Each dense block is, in turn, a serial combination of densely connected convolutional layers such that each succeeding layer receives inputs from all preceding layers. The total number of hidden layers in a DenseNet121 network are 121 (hence the name) and the output is typically given as a multi-probability array which is subjected to a softmax function to obtain likely classifications. In our case, since the outcome (need for mechanical ventilation) was binary, we changed the last layer to a sigmoid function (equivalent to a logit function in logistic regression) as shown in figure 1B. We used this modified DenseNet121 network architecture in our study.

Network training and validation

We used the Tensorflow 2.2.0 (https://www.tensorflow.org/) and Keras 2.3.0-tf framework (https://keras.io/) for model training and evaluation. The Jupyter notebook containing all the Python code is available with the authors and will be shared on receipt of reasonable request. Training of the model was done on all the layers of DenseNet121 (ie, no layers were

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.

frozen) with the following pre-specifications: batch size: 32, optimizer: stochastic gradient descent (SGD), loss function: binary cross-entropy, learning rate: 0.003, epochs per cycle length: 4 (with plateaued loss) and cycle length: 10. The model that provided the best validation accuracy was selected as the final model.

X-Ray evaluation by Pulmonary and Critical Care (PCC) experts

Two experienced (3 years beyond Fellowship) experts from the field of PCC evaluated all the X-ray images included in the independent test set (153 images on 118 patients). This evaluation by the PCC experts was done retrospectively, conducted blindly and independently, and was based on clinical gestalt. Both the PCC experts answered the following question for each X-ray image evaluated: "Based on this X-ray image, do you think that this COVID-19 patient will need to be mechanically ventilated during the index hospitalisation?" These evaluations were done in a blinded fashion, independent of the knowledge of the prediction by the DL algorithm as well as to other clinical characteristics like age, sex and comorbidities at the time of admission.

Statistical analyses

Descriptive statistics included mean and SD for continuous variables and proportions for categorical variables. Agreement between PCC experts' evaluation and the DL algorithm's prediction with the ground truth was assessed using Cohen's kappa. Performance metrics for the image classification task were precision (synonymous with positive predictive value as used in epidemiology), recall (synonymous with sensitivity), accuracy and F1 score (which was estimated as the harmonic mean of precision and recall). In addition, area under a receiver operating characteristic curve (AUROC) was estimated for the DL predictions. Predictive performance of the DL model was assessed at the level of the image as well as at the level of the patient. To summarise the performance at the level of a patient, we considered the prediction to be 'mechanical ventilation needed' if any of the multiple X-ray images on the same patient had indicated a high likelihood of ventilation need by the DL model. Correspondingly, the maximum probability estimated by the DL model for multiple X-rays on a given patient was considered as the predicted probability of the need for mechanical ventilation at the level of the patient.

Prognostic value of the predictions from the deep learning model and evaluations from the PCC experts was conducted using Kaplan-Meier plots and Cox proportional-hazards models. Incremental performance attributable to the deep learning model was estimated using Harrell's C statistic for survival models[13 14] and compared for statistical significance using the likelihood χ^2 test. All statistical analyses were conducted in Stata V.12.0 (Stata Corp, College Station, TX) software package. A global type I error rate of 0.05 was used to test statistical significance.

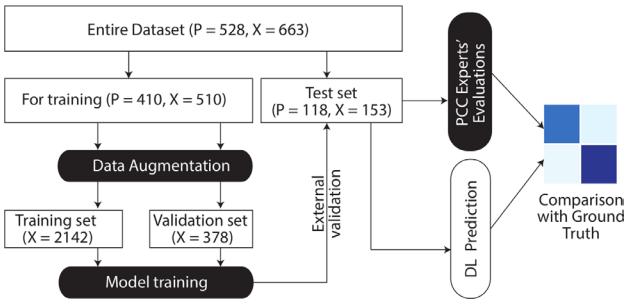

Figure 2 Overall analysis pipeline. P, number of patients; X, number of X-ray images. PCC, Pulmonary and Critical Care.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

RESULTS

Study participants

Data and images for this study come from 528 COVID-19 positive, hospitalised patients and a total of 663 X-ray images (figure 2). On the last day, 7 (1.3%) of the patients were still in hospital all of whom had completed at least 16 days of inpatient follow-up. Of the 528 patients, 79 (~15%) required mechanical ventilation. Clinical characteristics of the study participants based on the need for mechanical ventilation are shown in table 1. None of the sociodemographic and comorbidity variables were statistically significantly different in patients who received mechanical ventilation as compared with those who did not. Also, the average time interval between symptom onset and hospital admission was comparable in patients who did or did not require mechanical ventilation (6–7 days in both groups of patients). However, in general, those who received mechanical ventilation were more likely to be aged over 60 years and have hypertension, obesity, diabetes or chronic kidney disease as a comorbidity. Interestingly, those patients who eventually required mechanical ventilation had been ordered ~2 radiographs on an average as compared with a single radiograph ordered in most patients who did not require mechanical ventilation. The death rate in those who were mechanically ventilated was very high (~66%) as compared with those who did not need mechanical ventilation (~4%) as shown in table 1.

Model training results

The results of training of the proposed model are shown in figure 1E. The loss function monotonically decreased (except for cycle length 7) in both the training and the validation subsets and, conversely, the accuracy of prediction increased in a mirror-image fashion in both subsets. The model achieved convergence quickly. At the end of

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.

Table 1 Baseline characteristics of study participants (n=528)

Characteristic*	MV needed (n=79)	MV not needed (n=449)	P value
Sociodemographic characteristics			
Age (years)*	57.18 (13.87)	53.99 (13.81)	0.059
Age >60 years	36 (46.57)	163 (36.30)	0.117
Males	51 (64.56)	307 (68.37)	0.503
Hispanic/Latino ethnicity	44 (55.70)	259 (57.68)	0.742
Black/African-American race	31 (39.24)	155 (34.52)	0.418
Body mass index (kg/m^2)*	32.14 (7.54)	31.21 (10.34)	0.451
Symptom onset → hospital admission (days)*	6.42 (7.86)	7.34 (6.94)	0.314
No of X-ray images per patient*	1.96 (0.64)	1.14 (0.47)	<0.001
Comorbidities			
Hypertension	36 (46.75)	170 (37.95)	0.144
Obesity	42 (53.16)	195 (43.43)	0.109
Diabetes	40 (51.95)	183 (40.85)	0.069
Coronary artery disease	5 (6.49)	38 (8.48)	0.659†
Chronic kidney disease	9 (11.69)	27 (6.03)	0.085†
Asthma	3 (3.90)	36 (8.04)	0.246†
Chronic liver disease	7 (9.09)	23 (5.13)	0.182†
Congestive heart failure	2 (2.60)	23 (5.13)	0.560†
COPD	4 (5.19)	18 (4.02)	0.548†
ESRD	5 (6.49)	16 (3.57)	0.214†
HIV/AIDS	3 (3.90)	14 (3.13)	0.726†
Atrial fibrillation	1 (1.30)	20 (4.46)	0.340†
Ever smoker	19 (24.05)	87 (19.38)	0.339
Outcomes			
Death	52 (65.82)	17 (3.79)	<0.001

Cells indicate the number (percentage) for categorical variables and mean (SD) for continuous variables indicated by a dagger (†).
*Cells indicate mean (SD) for the continuous variables; all other cells indicate number (percentage).
†Fisher's exact test.
COPD, chronic obstructive pulmonary disease; ESRD, end-stage renal disease; MV, mechanical ventilation.

10 cycle lengths, the training set and test set accuracy was very high—almost 100% in the training set and 100% in the validation set. As shown in figure 1F, the model perfectly predicted the need for mechanical ventilation in the validation set.

Predictive performance of the model in the test set

The predictive performance of the model was assessed at two levels—at the level of X-ray images (n=153) and at the level of an individual patient (n=118). These results are shown in figure 3 (panels A–B for image-level analyses and panels E–F for patient-level analyses). The ROC curve using mechanical ventilation (22 patients, 43 X-ray images) as the ground truth and the probability estimates from the DL model as predictor (figure 3A) showed an AUROC of 79.34% at the image level. The optimum cut-off point on this ROC had a sensitivity and specificity of 70% and 84%, respectively. The confusion matrix (figure 3B) showed that the performance of the DL model was good with a high accuracy (0.7974), good recall, precision and F1 score (0.6976, 0.6250 and 0.6593) as well as a good Cohen's kappa (0.5158).

We replicated these analyses at the level of the patient with the maximum predicted probability (from multiple X-ray images). We observed (figure 3E) that the AUROC increased to 90.06% with an optimum sensitivity and specificity of 86.34% and 84.38%, respectively. Comparing these estimates with the corresponding image-level estimates (figure 3A), we found that analyses at the level of the patient yielded substantially higher sensitivity without loss of specificity. The confusion matrix for comparison of the patient-level prediction with ground truth (figure 3F) showed a markedly improved predictive performance: accuracy (0.8474), recall (0.8636), precision (0.5588), F1 score (0.6786) and Cohen's kappa (0.5845).

Independent evaluation by PCC experts

Independent evaluations by the two PCC experts are shown as confusion matrices for image-level analyses (figure 3C,D) and patient-level analyses (figure 3G,H). Performance of both PCC experts was relatively lower as compared with the DL model at the image level as well as at the patient level. At the image level, the performance characteristics of PCC expert 1 were

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.

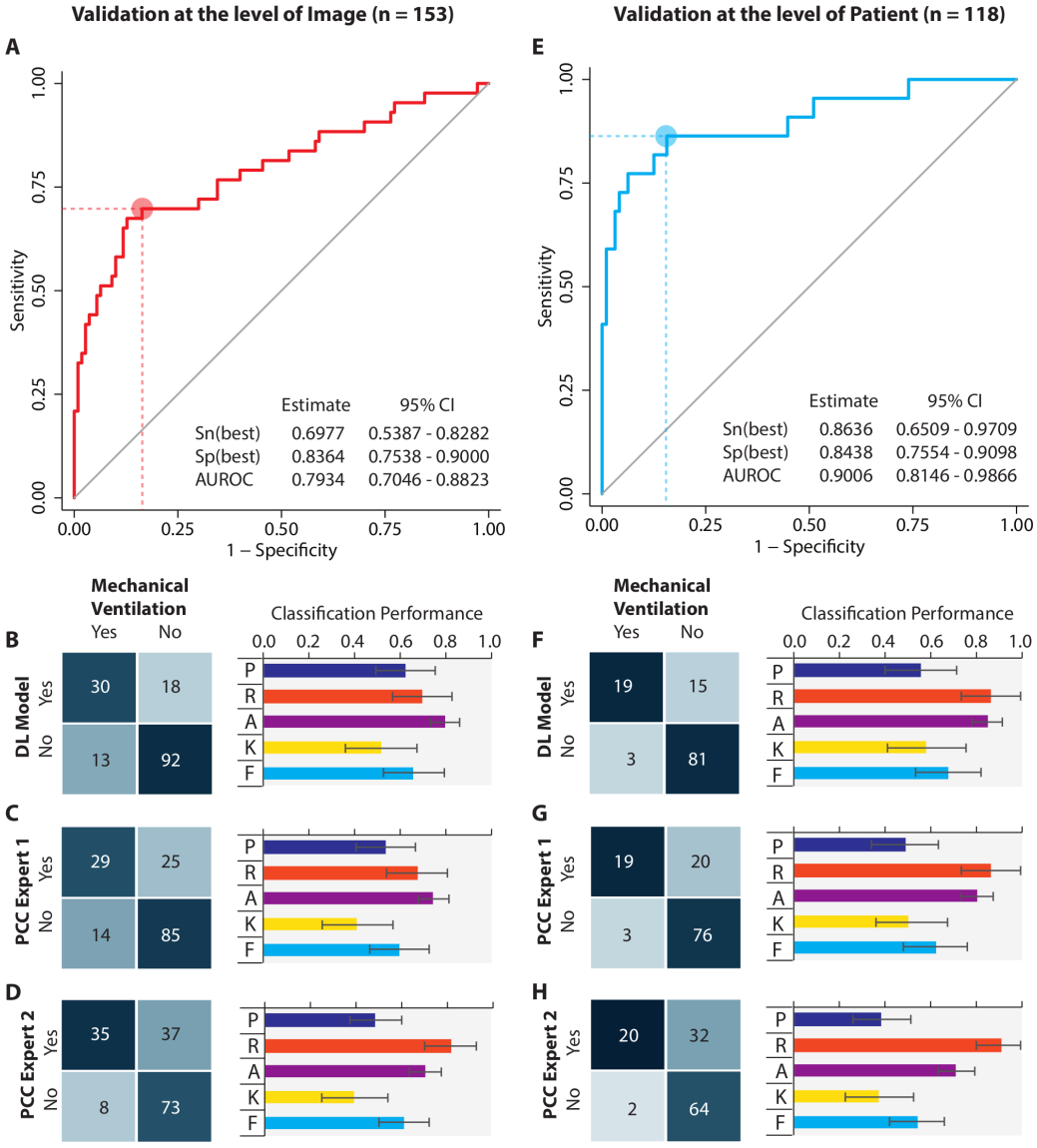

Figure 3 Prediction for the need of mechanical ventilation. Analyses were done at the level of X-ray image (A–D) and at the level of each patient (E–H). Panels A and E show the predictive accuracy as AUROC. The optimum cut-off was chosen as the point on ROC closest to the upper left corner of the plot and is indicated by a colour-coded circle. The sensitivity (dashed perpendicular to y-axis) and specificity (inverse of the dashed perpendicular to the x-axis) at the optimal cut-off is shown as Sn(best) and Sp(best), respectively. AUROC, area under the receiver operating characteristic curve. (B–D) Each panel shows the confusion matrix on the left side and five performance metrics in a bar chart on the right side. The bars and error bars show the point and 95% CI for each indicated and (colour-coded) performance metric. The metrics shown in the plot are P, precision; R, recall; A, accuracy; K, Cohen's kappa; and F, F1 score. (F–H) These panels respectively correspond to B–D but the results are shown at the level of the patient. Panels B–D and panels F–H use the same horizontal scale. PCC, Pulmonary and Critical Care.

accuracy (0.7451), recall (0.6744), precision (0.5370), F1 score (0.5979) and Cohen's kappa (0.4148). Similarly, the performance characteristics of PCC expert 2 at the level of images were accuracy (0.7059), recall (0.8140), precision (0.4861), F1 score (0.6087) and Cohen's kappa (0.3962). Like the DL model performance, the performance characteristics improved when the analyses were done at the level of patients (figure 3G,H). For example, the performance characteristics of PCC expert 1 at the level of patient were accuracy (0.8051), recall (0.8636), precision (0.4872), F1 score (0.6230) and Cohen's kappa (0.5049). For PCC expert 2, the performance characteristics were accuracy (0.7119), recall (0.9090), precision (0.3846), F1 score (0.5405) and Cohen's kappa (0.3774). Despite these improved estimates at the level of the patient and with the exception of recall for PCC expert 2, all the performance characteristics of both the PCC experts were either on par or below the corresponding estimates for the DL model.

Incremental predictive performance of DL model

To assess the incremental predictive performance of the DL model, we conducted survival analyses with time

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.

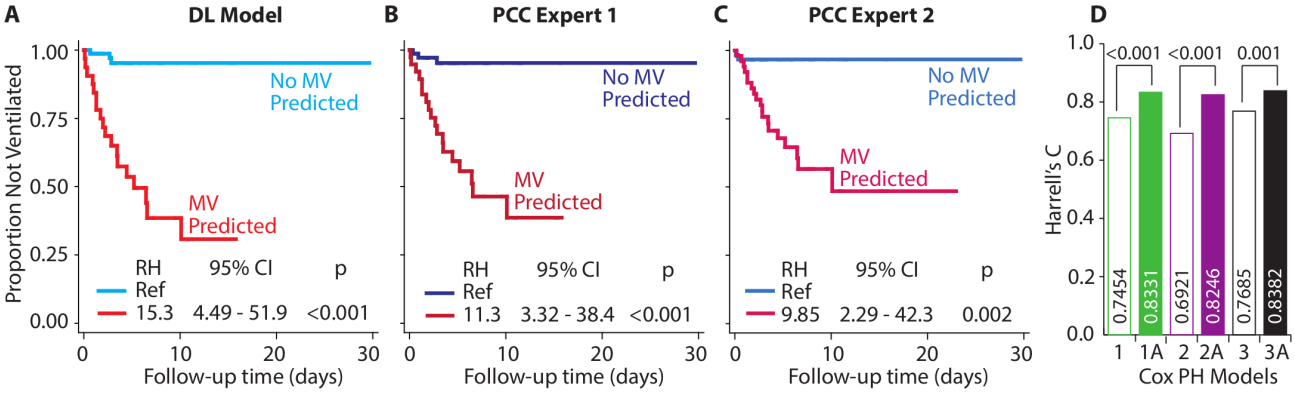

Figure 4 Incremental prognostic value of the DL model as compared with the PCC experts' evaluation. (A–C) Kaplan-Meier plots for time to mechanical ventilation since the time of first X-ray image. For patients with multiple X-ray images, the time was left-censored at the first image indicating the need of mechanical ventilation. Panels A–C indicate classifications based on the DL model (A), PCC expert 1 (B) and PCC expert 2 (C), respectively. Relative hazards (RH) and 95% CIs were estimated using Cox proportional-hazards (PH) models. Since different patients were classified as needing mechanical ventilation (MV) by the DL model and the PCC experts, different shades of red (for MV needed) and blue (for MV not needed) are used. (D) Incremental value of DL model to prognosticate patients. Models 1 and 1A compare the prediction from a Cox PH model that used only PCC expert 1 (model 1) vs that from a Cox PH model that used PCC expert 1 and the DL model as covariates. Models 2 and 2A correspondingly compare models with only PCC expert 2 and PCC expert 2 with DL model as covariates. Models 3 and 3A compare models with PCC experts 1 and 2 as covariates and both PCC experts with DL model, respectively. Bars indicate Harrell's C statistic for the indicated model. The statistical significance for the difference was tested using likelihood χ^2 test and is shown at the top of the bars depicting the indicated paired comparisons. PCC, Pulmonary and Critical Care.

to mechanical ventilation as the outcome of interest. These results are shown in figure 4. Kaplan-Meier plots (figure 4A–C) showed that patients predicted to need mechanical ventilation by the DL model or the PCC experts rapidly progressed to mechanical ventilation (red curves in figure 4A–C). However, the relative hazards of progressing to mechanical ventilation were highest for the DL model (15.3) as compared with those of PCC experts 1 (11.3) and PCC expert 2 (9.9) indirectly implying better prognostic stratification by the DL model. To directly assess the incremental value of the DL model over the stratification done by the PCC experts, we conducted pairwise comparisons of a series of Cox proportional-hazards models using Harrell's C-statistic. This statistic was estimated to be 0.7454 for stratification offered by PCC expert 1 but increased to 0.8331 (improvement 0.0877, p<0.001) on addition of DL model prediction as a covariate in the Cox model (compare models 1 and 1A in figure 4D). Similarly, the addition of DL model prediction to the stratification offered by PCC expert 2 improved Harrell's C-statistic from 0.6921 to 0.8246 (improvement 0.1325, p<0.001). Lastly, when stratifications offered by both the PCC experts were simultaneously used as covariates, Harrel's C-statistic was estimated to be 0.7685 which increased to 0.8382 on addition of the DL model prediction as a covariate (improvement 0.0724, p=0.001). Together, the results in figure 4 demonstrate that the DL model significantly and incrementally contributed to an improved prediction of the need for and time to mechanical ventilation over and beyond the predictions obtained from two PCC experts.

Time gained by early prediction using the DL model

Lastly, we examined the time gained by using the DL model predictions for the need of mechanical ventilation. These analyses were done at the level of the patient with start point defined as the time at which the DL model first predicted the need for mechanical ventilation. Using this strategy, we observed that the median time to mechanical ventilation was 2.98 (95% CI 1.63 to 4.32) days. Thus, the DL model developed in this study predicted mechanical ventilation early at the time of or during index hospitalisation.

DISCUSSION

We have developed a novel, X-ray image-based, deep learning model to predict the need for mechanical ventilation early during hospitalisation of patients with COVID-19. Our model was accurate (90% at the level of the patient), externally validated in an independent test set and provided improved prediction as compared with the prognostic performance of stratification provided by two PCC experts. Considering the urgent need for effective rationalisation of healthcare resources for patients with COVID-19, especially the ventilators, we believe that our DL model can have an important role in critical care of patients with COVID-19. This anticipation is contingent on the observation that our DL model was able to predict the need for mechanical ventilation approximately 3 days on average ahead of the actual intubation event. It needs to be mentioned, however, that we do not anticipate that clinicians will use our DL model to make clinical decisions—rather, the DL model is intended to be used

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.

as an aid for patient triaging and informed resource allocation by hospitals.

We conducted the analyses both at the level of the images and at the level of the individual patient. This distinction is important to understand. The training of the model was indeed done using all available images, but to translate the DL predictions to clinical situations, we needed to determine whether a selected patient (rather than an X-ray image) would need mechanical ventilation. To that end, when any of the X-ray images for a patient was classified by the DL algorithm as high risk (that is likely needing mechanical ventilation), we considered that the patient should be classified as in need of mechanical ventilation. Operationally, this strategy of patient-level classification can be expected to improve the sensitivity of prediction for two reasons. First, the average per-patient X-rays ordered was higher for the patients who required mechanical ventilation as compared with those who did not (table 1). At the level of the patient, therefore, combining information from multiple X-rays is likely to predict the need for mechanical ventilation in those who eventually required mechanical ventilation. Second, the option to use *any* indicative X-ray as diagnostic for a patient will likely lower the diagnostic threshold making the algorithm more sensitive. A comparison of the ROC curves shown in figure 3A,E demonstrates that there indeed was a substantial improvement in sensitivity (from 70% to 86%) without loss of specificity when those analyses were done at the level of the patient rather than at the level of each X-ray image.

Previously, the CheXNeXT system[10] has been used to predict 14 pathologies based on chest X-rays but not in the context of COVID-19 infection. There have been several studies[15–19] to detect or diagnose COVID-19 based on chest radiographs, but deep learning attempts to prognosticate COVID-19 disease course have been few and far between. Cohen *et al*[20] developed an algorithm to predict patients at high risk of mortality; Zhu *et al*[21] have developed a deep learning method to stage disease severity, the CheXNeXt deep learning; and, recently, Li *et al*[22] have developed a deep-learning Siamese network to predict the Radiographic Assessment of Lung Edema (RALE) scores[23] used to quantify severity of ARDS in patients with COVID-19. These landmark studies have proffered definitive directions for the potential use of chest X-rays in clinical care of critical patients. However, a direct application of these systems to predict actionable outcomes like the need for mechanical ventilation is currently lacking. Existing studies have tended to focus on different aspects of chest X-ray images for COVID-19 prognostication. For example, the primary purpose of RALE score is quantification of pulmonary oedema[23] while the algorithms developed by Zhu *et al*[21] and Cohen *et al*[20] focus on identification of ground-glass opacities and the geographical extent. Considering the novelty of COVID-19 pathophysiology, we aimed to include all possible and detectable abnormalities and therefore used the CheXNeXT system that can accurately identify 14 different pathologies. In this context, it is noteworthy that using cytokine/chemokine data on hospitalised patients with COVID-19, Donlan *et al*[11] have shown that circulating concentration of interleukin-13 (IL-13) can predict the need for mechanical ventilation. Since IL-13 can contribute to pulmonary eosinophilia and tissue remodelling, it is thus possible that radiographically detectable texture alterations that are neither captured by the RALE score nor the CheXNeXT system may accompany these cytokine profiles. This hypothesis is, in part, supported by investigations in COVID-19 negative patients.[24 25] Whether such a correlation exists within the context of COVID-19 is currently unknown. In totality, these previous studies provide a possible biological explanation as to why chest radiographs can predict the need for mechanical ventilation in immediate future.

The results of our study should be considered in the light of some limitations. First, this was a retrospective, observational evaluation and the confounding and bias implicit in such an investigation will remain a limitation. Second, the data for this study were derived from a single centre and the generalisability of this approach to other settings needs to be established in further studies. Third, we restricted our model to the use of chest radiographs only. However, additional clinical parameters at the time of hospital admission such as respiratory rate, oxygenation status (eg, the ROX index)[26] and altered mental status[27] along with sociodemographic characteristics, comorbidity profile and laboratory investigations can potentially further improve the prediction. Future studies need to evaluate these possibilities, but our focus was to use an objective measure such as a chest radiograph and provide a tool to the critical care provider with a reasonable expectation of the future course of disease in a given patient. Fourth, we used Harrel's C statistic as a measure of the predictive accuracy of the Cox regression models. In the setting of a binary outcome, Harrell's C statistic behaves statistically similar to the AUROC curve. This behaviour of the statistic is however influenced by censoring in the context of survival analyses. This is generally considered as a limitation of Harrell's C statistic for Cox models.[28] Alternative methods like Uno's C statistic[28] and time-dependent receiver operating characteristic curves[29] are available for the survival analysis framework. Notwithstanding these methodological nuances, since comparison of two models on the same set of patients with the same censoring characteristics was undertaken in our study, a comparison in Harrell's C statistics did provide us with an estimate of the improved prediction by the DL model over that of the PCC experts. Fifth, the evaluation of the X-rays by the PCC experts was done retrospective solely for the purpose of this research and not as a part of routine patient care. Also, currently there are no clinical protocols in place to predict the need for mechanical ventilation based on chest radiography. Therefore, whether the DL

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.

model supersedes the PCC experts in terms of the time gained in advance cannot be answered based on this study. Future studies need to address that question specifically.

Until effective preventive and management options for patients with COVID-19 become widely available, consorted efforts that reduce the risks to the patient and thus the burden on healthcare system are needed. To that end, our study demonstrates a proof-of-principle that chest X-ray images acquired early during hospitalisation can accurately predict the need for mechanical ventilation in patients with COVID-19. Such a tool can be valuable in effectively triaging patients with COVID-19 at the time of initial healthcare contact.

Author affiliations
[1]Innotomy Consulting, Bengaluru, India
[2]Lata Medical Research Foundation, Nagpur, India
[3]Department of Medicine, Division of Nephrology, Cook County Hospital, Chicago, Illinois, USA
[4]Division of Pulmonary, Critical Care, Sleep, and Allergy, University of Illinois Hospital and Health Sciences System, Chicago, Illinois, USA
[5]Department of Medicine, Division of Pulmonary and Critical Care, Cook County Hospital, Chicago, Illinois, USA
[6]Department of Medicine, Division of Cardiology, Cook County Hospital, Chicago, Illinois, USA
[7]Rush Medical College, Rush University Medical Center, Chicago, Illinois, USA
[8]M&H Research LLC, San Antonio, Texas, USA

Twitter Anoop R Kulkarni @DrAnoopKulkarni

Contributors AMA and HK conceptualised the study; ARK and HK conceptualised the deep learning solution; ARK wrote Python scripts, trained and tested the model; AsS and SS provided expertise in pulmonary and critical care; AbS, MI, SZ, JV, MA, AJ, AA and JR participated in data collection along with AMA; HK conducted statistical analyses; AMA, ARK and HK wrote the first draft of the manuscript; all authors reviewed and approved the final manuscript. AMA and HK are the guarantors of this manuscript and take responsibility for data integrity, analytical accuracy and draft writing.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the Institutional Review Board of the Cook County Health, Chicago, IL (Approval No. 20-038 x).

Data availability statement Data and code are available from the author on reasonable request.

This article is made freely available for use in accordance with BMJ's website terms and conditions for the duration of the covid-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

ORCID iD
Hemant Kulkarni http://orcid.org/0000-0001-6950-3341

REFERENCES

1 COVID-19 Dashboard by the center for systems science and engineering (CSSE) at Johns Hopkins University (JHU), 2020. Available: https://coronavirus.jhu.edu/map.html
2 CDC. Coronavirus disease 2019, 2020. Available: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html
3 Abate SM, Ahmed Ali S, Mantfardo B, *et al*. Rate of intensive care unit admission and outcomes among patients with coronavirus: a systematic review and meta-analysis. *PLoS One* 2020;15:e0235653.
4 CDC. Current hospital capacity estimates – snapshot, 2020. Available: https://www.cdc.gov/nhsn/covid19/report-patient-impact.html
5 Duca A, Memaj I, Zanardi F, *et al*. Severity of respiratory failure and outcome of patients needing a ventilatory support in the emergency department during Italian novel coronavirus SARS-CoV2 outbreak: preliminary data on the role of helmet CPAP and non-invasive positive pressure ventilation. *EClinicalMedicine* 2020;24:100419.
6 Zhu W, Wang Y, Xiao K, *et al*. Establishing and managing a temporary coronavirus disease 2019 specialty hospital in Wuhan, China. *Anesthesiology* 2020;132:1339–45.
7 CDC. Interim clinical guidance for management of patients with confirmed coronavirus disease, 2020. Available: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html
8 Khan AI, Shah JL, Bhat MM. CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. *Comput Methods Programs Biomed* 2020;196:105581.
9 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. doi:10.1183/13993003.00775-2020. [Epub ahead of print: 06 Aug 2020].
10 Rajpurkar P, Irvin J, Ball RL, *et al*. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* 2018;15:e1002686.
11 Donlan AN, Young M, Petri WA. IL-13 predicts the need for mechanical ventilation in COVID-19 patients. *medRxiv*.
12 Artacho Ruiz R, Artacho Jurado B, Caballero Güeto F, *et al*. Predictors of success of high-flow nasal cannula in the treatment of acute hypoxemic respiratory failure. *Med Intensiva* 2019. doi:10.1016/j.medin.2019.07.012. [Epub ahead of print: 24 Aug 2019].
13 Alexander M, Wolfe R, Ball D, *et al*. Lung cancer prognostic index: a risk score to predict overall survival after the diagnosis of non-small-cell lung cancer. *Br J Cancer* 2017;117:744–51.
14 Peters M, van der Voort van Zyp JRN, Moerland MA, *et al*. Development and internal validation of a multivariable prediction model for biochemical failure after whole-gland salvage iodine-125 prostate brachytherapy for recurrent prostate cancer. *Brachytherapy* 2016;15:296–305.
15 Keles A, Keles MB, Keles A. COV19-CNNet and COV19-ResNet: diagnostic inference Engines for early detection of COVID-19. *Cognit Comput* 2021:1–11.
16 Hussain L, Nguyen T, Li H, *et al*. Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection. *Biomed Eng Online* 2020;19:88.
17 Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci Rep* 2020;10:19549.
18 Wang M, Xia C, Huang L, *et al*. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. *Lancet Digit Health* 2020;2:e506–15.

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.

19 Jain G, Mittal D, Thakur D, *et al*. A deep learning approach to detect Covid-19 coronavirus with X-ray images. *Biocybern Biomed Eng* 2020;40:1391–405.
20 Cohen JP, Dao L, Roth K, *et al*. Predicting COVID-19 pneumonia severity on chest X-ray with deep learning. *Cureus* 2020;12:e9448.
21 Zhu J, Shen B, Abbasi A, *et al*. Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. *PLoS One* 2020;15:e0236621.
22 Li MD, Arun NT, Gidwani M, *et al*. Automated assessment of COVID-19 pulmonary disease severity on chest radiographs using convolutional Siamese neural networks. *medRxiv* 2020. doi:10.1101/2020.05.20.20108159. [Epub ahead of print: 26 May 2020].
23 Warren MA, Zhao Z, Koyama T, *et al*. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. *Thorax* 2018;73:840–6.
24 Svenningsen S, Haider E, Boylan C, *et al*. CT and Functional MRI to Evaluate Airway Mucus in Severe Asthma. *Chest* 2019;155:1178–89.
25 Kim YH, Kim KW, Lee KE, *et al*. Transforming growth factor-beta 1 in humidifier disinfectant-associated children's interstitial lung disease. *Pediatr Pulmonol* 2016;51:173–82.
26 Roca O, Caralt B, Messika J, *et al*. An index combining respiratory rate and oxygenation to predict outcome of nasal high-flow therapy. *Am J Respir Crit Care Med* 2019;199:1368–76.
27 Lee JY, Kim HA, Huh K, *et al*. Risk factors for mortality and respiratory support in elderly patients hospitalized with COVID-19 in Korea. *J Korean Med Sci* 2020;35:e223.
28 Uno H, Cai T, Pencina MJ, *et al*. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011;30:1105–17.
29 Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med Res Methodol* 2017;17:53.

BMJ Innov: first published as 10.1136/bmjinnov-2020-000593 on 2 March 2021. Downloaded from http://innovations.bmj.com/ on March 8, 2021 at India:BMJ-PG Sponsored. Protected by copyright.