

1   **A DEEP LEARNING MODEL TO PREDICT THE NEED FOR MECHANICAL VENTILATION USING**  
2   **CHEST X-RAY IMAGES IN HOSPITALIZED COVID-19 PATIENTS**

3   **Anoop Kulkarni<sup>1,2,\*</sup>, Ambarish M. Athavale<sup>3,\*</sup>, Ashima Sahni<sup>4</sup>, Shashvat Sukhal<sup>5</sup>, Abhimanyu**  
4   **Sahni<sup>6</sup>, Mathew Itteera<sup>3</sup>, Sara Zhukovsky<sup>7</sup>, Jane Vernik<sup>3</sup>, Mohan Abraham<sup>3</sup>, Amit Joshi<sup>3</sup>,**  
5   **Amatur Amarah<sup>3</sup>, Juan Ruiz<sup>3</sup>, Peter D. Hart<sup>3</sup>, Hemant Kulkarni<sup>2,8</sup>**

6   <sup>1</sup> Innotomy Consulting, Bengaluru, India

7   <sup>2</sup> Lata Medical Research Foundation, Nagpur, India

8   <sup>3</sup> Division of Nephrology, Department of Medicine, Cook County Health, Chicago, Illinois, USA

9   <sup>4</sup> Division of Pulmonary, Critical Care, Sleep, and Allergy, University of Illinois Hospital and  
10   Health Sciences System, Chicago, Illinois, USA

11   <sup>5</sup> Division of Pulmonary and Critical Care, Department of Medicine, Cook County Health,  
12   Chicago, Illinois, USA

13   <sup>6</sup> Division of Cardiology, Department of Medicine, Cook County Health, Chicago, Illinois, USA

14   <sup>7</sup> Rush University Medical Center, Chicago, IL, USA

15   <sup>8</sup> M&H Research, LLC, San Antonio, Texas, USA

16   \*, these authors contributed equally

17   **Running Title: Deep learning from X-rays of COVID-19 patients**

18   **Word Count: 3,272;**              **Number of Tables: 1;**              **Number of Figures: 4**

19   **Corresponding author:**

20   Hemant Kulkarni, MD  
21   12023 Waterway Rdg,  
22   San Antonio, TX 78249, USA  
23   Phone: +1 (210) 602 5537

24   Email: [hemant.kulkarni@mnhresearch.com](mailto:hemant.kulkarni@mnhresearch.com)

25   **DISCLOSURES:** None of the authors have a conflict of interest to disclose.

26 **ABSTRACT**

27 **Purpose:** Early identification of a potentially deteriorating clinical course in hospitalized COVID-

28 19 patients is critical since there exists a resource-demand gap for the ventilators.

29 **Materials:** We aimed to develop and validate a deep learning-based approach to predict the

30 need for mechanical ventilation as early as at the time of initial radiographic evaluation. We

31 exploited the well-established DenseNet121 deep learning architecture for this purpose on 663

32 X-ray images derived from 528 hospitalized COVID-19 patients. Two Pulmonary and Critical Care

33 experts blindly and independently evaluated the same X-ray images for purpose of validation.

34 **Results:** We found that our deep learning model predicted the need for ventilation with a high

35 accuracy, sensitivity and specificity (90.06%, 86.34% and 84.38%, respectively). This prediction

36 was done approximately three days ahead of the actual intubation event. Our model also

37 outperformed two Pulmonary and Critical Care experts who evaluated the same X-ray images

38 and provided an incremental accuracy of 7.24–13.25%.

39 **Conclusion:** Our deep learning model accurately predicted the need for mechanical ventilation

40 early during hospitalization of COVID-19 patients. Until effective preventive or treatment

41 measures become widely available for COVID-19 patients, prognostic stratification as provided

42 by our model is likely to be highly valuable.

43

44

45 **KEYWORDS:** deep learning, chest radiograph, COVID-19, mechanical ventilation

## 46 INTRODUCTION

47 Coronavirus Disease-19 (COVID-19) is a global pandemic which has caused an estimated 20  
48 million infections and 738,668 deaths within a span of just over seven months. [1] Strikingly,  
49 the cumulative COVID -19 hospitalization rate is 137.6 per 100,000 infections. [2] COVID-19 can  
50 affect major organ systems such as lungs, heart, kidney and brain but the morbidity and  
51 mortality in COVID-19 patients is primarily due to lung infection from a pneumonic process.  
52 Thus, a significant number of COVID-19 patients need supportive care such as intravenous fluid  
53 administration, supplemental oxygen. Further, as many as 32% of hospitalized COVID-19  
54 patients need admission to an intensive care unit [3] and mechanical ventilation [4, 5]. This has  
55 caused a great strain in hospital resources in certain geographic regions with high rates of  
56 COVID-19 infection. For example, at the height of COVID-19 pandemic in Wuhan and New-York,  
57 there were concerns of the health system being overwhelmed from the sheer number of  
58 patients requiring hospitalization. In Wuhan, a temporary COVID-19 facility was built and in  
59 New-York, a United States Navy hospital ship was dispatched to help cope with the number of  
60 patients requiring hospitalization[6, 7]. Since then, COVID-19 has now spread and threatens to  
61 overwhelm hospital systems in Texas, Florida and Arizona with the possibility of a second wave  
62 in the fall and winter coinciding with the annual rise of influenza. In the absence of a vaccine or  
63 effective anti-viral treatment, hospital systems will need to be prepared for an increase in  
64 hospitalization rates, ICU admissions and need for mechanical ventilation.

65

66 It has been observed that many COVID-19 patients experience a worsening of shortness of  
67 breath and need for supplemental oxygen or mechanical ventilation during the second week of

68 the illness[8]. However, not every patient who is hospitalized with COVID-19 infection needs  
69 mechanical ventilation or ICU level of care. Thus, a tool that can effectively predict the need for  
70 mechanical ventilation would ensure a better triage at initial point of contact with healthcare  
71 system and enable better allocation of healthcare resources by avoiding unnecessary  
72 hospitalizations. This was the motivation for the present study.

73

74 Deep learning is a branch of artificial intelligence that has shown great promise in diagnosis and  
75 prognosis of various health conditions. Specifically, in the context of COVID-19 disease, a deep  
76 learning analysis of chest radiograph was able to identify patients with COVID-19 infection with  
77 more than 90% accuracy. [9] Also, Wang et al [10] were able to stratify patients in high risk and  
78 low risk groups by a deep learning analysis of lung computed tomography (CT) images. In this  
79 study, we focused on using the information contained within chest X-ray images to predict the  
80 need for mechanical ventilation. A chest radiograph has practical advantages over CT scans in  
81 being more readily available especially in resource-challenged scenarios and less risk of  
82 equipment contamination. Indeed, a chest radiograph was performed for every patient with  
83 COVID-19 evaluated in our hospital emergency room. Here, we present a deep learning analysis  
84 of Chest radiograph of patients with COVID-19 to predict need for mechanical ventilation.

85

86 **MATERIAL AND METHODS**

87 **Study participants**

88 The clinical and image data for this study were collected at the John H. Stroger, Jr Hospital of  
89 Cook County, Chicago, IL. All COVID-19 patients who were admitted to the study center

90 between March 15, 2020 and May 31, 2020 and followed up till the censoring date of June 16,  
91 2020 were included. On the last day, 7 (1.3%) of the patients were still in hospital all of whom  
92 had completed at least 16 days of inpatient follow-up. The study cohort was identified in two:  
93 first, all confirmed COVID-19 cases were identified and second, only the new inpatients were  
94 selected from those identified in step one. COVID-19 positivity was confirmed for all patients  
95 using the polymerase chain reaction for the RdRp and N genes of the SARS-CoV-2 virus. Clinical  
96 data of these patients was collected by chart reviews. The study did not require informed  
97 consent from the patients as the data was retrospectively collected after de-identification. The  
98 study was approved by the Institutional Review Board of the Cook County Health, Chicago, IL.

99

#### 100 **Chest X-ray images**

101 All patients with symptoms suggestive of a possible infection with SARS-CoV-2 underwent  
102 portable antero-posterior chest X-ray assessment at the study center. The protocol followed  
103 was as follows: Ensuring appropriate isolation and distancing practices the X-ray images were  
104 acquired in upright or near-upright posture. Images were saved in dicom and jpg format and  
105 were manually scrubbed to remove all identifiable information.

106

#### 107 **Data preprocessing**

108 The acquired X-ray images were first resized to 224 x 224 pixels and then center cropped as  
109 required for many deep learning networks that use convolutional layers to parse out image  
110 features. To ensure robustness in training and validation of the deep learning network, we  
111 undertook two steps in data preprocessing. First, we augmented each image using a random

112 combination of right- or left-rotation (max 30°), random cropping and random lighting. These  
113 augmentations permitted us to use different variations of the original image for training the  
114 deep learning algorithm thereby reducing the potential overfitting. Second, since the patients  
115 who needed mechanical ventilation in the study dataset represented a minority class, for  
116 training the network we first oversampled the number of ventilated patients so as to achieve a  
117 class balance of ~50% of X-ray images for ventilated and non-ventilated patients in the training  
118 sample. Combination of the first and second steps in data preprocessing yielded a set of 1320  
119 X-ray images from the ventilated patients and 1200 X-ray images from non-ventilated patients.  
120 This set of 2520 images was used for network training.

121

## 122 **Network architecture**

123 The established and validated CheXNeXt deep learning algorithm [11] as well as the PXR  
124 network [12] are based on the DenseNet121 [13] architecture. While the CheXNeXt predicts  
125 one or more of 14 lung pathologies from an X-ray image of the chest, the PXR network scores  
126 an X-ray image for severity of acute respiratory distress syndrome (ARDS). We used the same  
127 backbone for our proposed prognosticator algorithm. The architecture of a DenseNet121  
128 network is shown in Figure 2B. Briefly, the DenseNet121 represents a series of convolutional  
129 operations on the image array (size 224 x 224 pixels) and is characterized by a serial  
130 combination of 4 dense blocks (D1-D4, Figure 2B) interspersed with 3 transitional blocks (T1-T3,  
131 Figure 2B). Each dense block is, in turn, a serial combination of densely connected convolutional  
132 layers such that each succeeding layer receives inputs from all preceding layers. The total  
133 number of hidden layers in a DenseNet121 network are 121 (hence the name) and the output is

134 typically given as a multi-probability array which is subjected to a softmax function to obtain  
135 likely classifications. In our case, since the outcome (need for mechanical ventilation) was  
136 binary, we changed the last layer to a sigmoid function (equivalent to a logit function in logistic  
137 regression) as shown in Figure 2B. We used this modified DenseNet121 network architecture in  
138 our study.

139

#### 140 **Network Training and Validation**

141 We used the Tensorflow 2.2.0 (<https://www.tensorflow.org/>) and Keras 2.3.0-tf framework  
142 (<https://keras.io/>) for model training and evaluation. The Jupyter notebook congaing all the  
143 Python code is available with the authors and will be shared upon receipt of reasonable  
144 request. Training of the model was done on all the layers of DenseNet121 (i.e. no layers were  
145 frozen) with the following pre-specifications: batch size: 32, optimizer: stochastic gradient  
146 descent (SGD), loss function: binary cross-entropy, learning rate: 0.003, epoch  
147 s per cycle length: 4 (with plateaued loss) and cycle length: 10. Model that provided the best  
148 validation accuracy was selected as the final model.

149

#### 150 **X-ray evaluation by Pulmonary and Critical Care (PCC) Experts**

151 Two experienced experts from the field of Pulmonary and Critical Care (PCC) evaluated all the  
152 X-ray images included in the independent test set (153 images on 118 patients). Both the PCC  
153 experts answered the following question for each X-ray image evaluated: “Based on this X-ray  
154 image, do you think that this COVID-19 patient will need to be mechanically ventilated during  
155 the index hospitalization?” These evaluations were done in a blinded fashion, independent of

156 the knowledge of the prediction by the DL algorithm as well as to other clinical characteristics  
157 like age, sex and comorbidities at the time of admission.

158

159 **Statistical analyses**

160 Descriptive statistics included mean and standard deviation for continuous variables and  
161 proportions for categorical variables. Agreement between PCC Experts' evaluation and the DL  
162 algorithm's prediction with the ground truth was assessed using Cohen's kappa. Performance  
163 metrics for the image classification task were precision (synonymous with positive predictive  
164 value as used in epidemiology), recall (synonymous with sensitivity), accuracy and F1 score  
165 (which was estimated as the harmonic mean of precision and recall). In addition, area under a  
166 receiver operating characteristic curve (AUROC) was estimated for the DL predictions.  
167 Predictive performance of the DL model was assessed at the level of the image as well as at the  
168 level of the patient. To summarize the performance at the level of a patient, we considered the  
169 prediction to be "mechanical ventilation needed" if any of the multiple X-ray images on the  
170 same patient had indicated a high likelihood of ventilation need by the DL model.  
171 Correspondingly, the maximum probability estimated by the DL model for multiple X-rays on a  
172 given patient was considered as the predicted probability of the need for mechanical  
173 ventilation at the level of the patient. Prognostic value of the predictions from the deep  
174 learning model and evaluations from the PCC experts was conducted using Kaplan-Meier plots  
175 and Cox proportional hazards models. Incremental performance attributable to the deep  
176 learning model was estimated using the Harrell's C statistic for survival models [14, 15] and  
177 compared for statistical significance using the likelihood  $\chi^2$  test. All statistical analyses were

178 conducted in Stata 12.0 (Stata Corp, College Station, TX) software package. A global type I error  
179 rate of 0.05 was used to test statistical significance.

180

## 181 **RESULTS**

182

### 183 **Study participants**

184 Data and images for this study come from 528 COVID-19 positive, hospitalized patients and a  
185 total of 663 X-ray images (Figure 1). Of the 528 patients, 79 (~15%) required mechanical  
186 ventilation. Clinical characteristics of the study participants based on the need for mechanical  
187 ventilation are shown in Table 1. Interestingly, none of the socio-demographic and comorbidity  
188 variables were statistically significantly different in patients who received mechanical  
189 ventilation as compared to those who did not. However, in general, those who received  
190 mechanical ventilation were more likely to be aged over 60 years and have hypertension,  
191 obesity, diabetes or chronic kidney disease as a comorbidity. The death rate in those who were  
192 mechanically ventilated was very high (~66%) as compared to those who did not need  
193 mechanical ventilation (~4%) as shown in Table 1.

194

### 195 **Model training results**

196 The results of training of the proposed model are shown in Figure 2E. The loss function  
197 monotonically (except for cycle length 7) decreased in both the training and the validation  
198 subsets and, conversely, the accuracy of prediction increased in a mirror-image fashion in both  
199 subsets. The model achieved convergence quickly. At the end of 10 cycle lengths, the training  
200 set and test set accuracy was very high – almost 100% in the training set and 100% in the

201 validation set. As shown in Figure 2F, the model perfectly predicted the need for mechanical  
202 ventilation in the validation set.

203

204 **Predictive performance of the model in the test set**

205 The predictive performance of the model was assessed at two levels – at the level of X-ray  
206 images ( $n = 153$ ) and at the level of an individual patient ( $n = 118$ ). These results are shown in  
207 Figure 3 (panels A-B for image-level analyses and panels E-F for patient-level analyses). The ROC  
208 curve using mechanical ventilation needed (22 patients, 43 X-ray images) as the ground truth  
209 and the probability estimates from the DL model as predictor (Figure 3A) showed an AUROC of  
210 79.34% at the image-level. The optimum cutoff point on this ROC had a sensitivity and  
211 specificity of 70% and 84%, respectively. The confusion matrix (Figure 3B) showed that the  
212 performance of the DL model was good with a high accuracy (0.7974), good recall, precision  
213 and F1-score (0.6976, 0.6250 and 0.6593) as well as a good Cohen's kappa (0.5158).

214

215 We replicated these analyses at the level of the patient with the maximum predicted  
216 probability (from multiple X-ray images). We observed (Figure 3E) that the AUROC increased to  
217 90.06% with an optimum sensitivity and specificity of (86.34% and 84.38%, respectively).  
218 Comparing these estimates with the corresponding image-level estimates (Figure 3A), we found  
219 that analyses at the level of the patient yielded substantially higher sensitivity without loss of  
220 specificity. The confusion matrix for comparison of the patient-level prediction with ground  
221 truth (Figure 3F) showed a markedly improved predictive performance: accuracy (0.8474),  
222 recall (0.8636), precision (0.5588), F1-score (0.6786) and Cohen's kappa (0.5845).

223

224 **Independent evaluation by PCC Experts**

225 Independent evaluations by the two PCC Experts are shown as confusion matrices for image-  
226 level analyses (Figure 3C-D) and patient-level analyses (Figure 3G-H). Performance of both PCC  
227 Experts was relatively lower as compared to the DL model at the image-level as well as at the  
228 patient-level. At the image-level the performance characteristics of PCC Expert 1 were: accuracy  
229 (0.7451), recall (0.6744), precision (0.5370), F1-score (0.5979) and Cohen's kappa (0.4148).  
230 Similarly, the performance characteristics of PCC Expert 2 at the level of images were: accuracy  
231 (0.7059), recall (0.8140), precision (0.4861), F1-score (0.6087) and Cohen's kappa (0.3962). Like  
232 the DL model performance, the performance characteristics improved when the analyses were  
233 done at the level of patients (Figure 3G-H). For example, the performance characteristics of PCC  
234 Expert 1 at the level of patient were: accuracy (0.8051), recall (0.8636), precision (0.4872), F1-  
235 score (0.6230) and Cohen's kappa (0.5049). For PCC Expert 2 the performance characteristics  
236 were: accuracy (0.7119), recall (0.9090), precision (0.3846), F1-score (0.5405) and Cohen's  
237 kappa (0.3774). Despite these improved estimates at the level of the patient and with the  
238 exception of recall for PCC Expert 2, all the performance characteristics of both the PCC Experts  
239 were either on par or below the corresponding estimates for the DL model.

240

241 **Incremental predictive performance of DL model**

242 To assess the incremental predictive performance of the DL model we conducted survival  
243 analyses with time to mechanical ventilation as the outcome of interest. These results are  
244 shown in Figure 4. Kaplan-Meier plots (Figure 4A-C) showed that patients predicted to need

245 mechanical ventilation by the DL model or the PCC Experts rapidly progressed to mechanical  
246 ventilation (red curves in Figure 4A-C). However, the relative hazards of progressing to  
247 mechanical ventilation were highest for the DL model (15.3) as compared to those of PCC  
248 Experts 1 (11.3) and PCC Expert 2 (9.9) indirectly implying better prognostic stratification by the  
249 DL model. To directly assess the incremental value of the DL model over the stratification done  
250 by the PCC Experts, we conducted pairwise comparisons of a series of Cox proportional hazards  
251 models using the Harrell's C-statistic. This statistic was estimated to be 0.7454 for stratification  
252 offered by PCC Expert 1 but increased to 0.8331 (improvement 0.0877,  $p<0.001$ ) upon addition  
253 of DL model prediction as a covariate in the Cox Model (compare models 1 and 1A in Figure  
254 4D). Similarly, the addition of DL model prediction to the stratification offered by PCC Expert 2  
255 improved Harrell's C-statistic from 0.6921 to 0.8246 (improvement 0.1325,  $p<0.001$ ). Lastly,  
256 when stratifications offered by both the PCC Experts were simultaneously used as covariates,  
257 the Harrel's C-statistic was estimated to be 0.7685 which increased to 0.8382 upon addition of  
258 the DL model prediction as a covariate (improvement 0.0724,  $p = 0.001$ ). Together, the results  
259 in Figure 4 demonstrate that the DL model significantly and incrementally contributed to an  
260 improved prediction of the need for and time to mechanical ventilation over and beyond the  
261 predictions obtained from two PCC Experts.

262

263 **Time gained by early prediction using the DL model**

264 Lastly, we examined the time gained by using the DL model predictions for the need of  
265 mechanical ventilation. These analyses were done at the level of the patient with start point  
266 defined as the time at which the DL model first predicted the need for mechanical ventilation.

267 Using this strategy, we observed that the median time to mechanical ventilation was 2.98 (95%  
268 CI 1.63 – 4.32) days. Thus, the DL model developed in this study antedated mechanical  
269 ventilation early at the time of or during index hospitalization.

270  
271 **DISCUSSION**

272 We have developed a novel, X-ray image-based, deep learning model to predict the need for  
273 mechanical ventilation early during hospitalization of COVID-19 patients. Our model was  
274 accurate (90% at the level of the patient), externally validated in an independent test set and  
275 provided improved prediction as compared to the prognostic performance of stratification  
276 provided by two PCC Experts. Considering the urgent need for effective rationalization of health  
277 care resources for COVID-19 patients, especially the ventilators, we believe that our DL model  
278 can have an important role in critical care of COVID-19 patients. This anticipation is contingent  
279 upon the observation that our DL model was able to predict the need for mechanical  
280 ventilation approximately 5 days ahead of the actual intubation event.

281  
282 Previously, the CheXNeXt deep learning system [11] has been used to predict 14 pathologies  
283 based on chest X-rays. Also, recently Li et al [16] have developed a deep-learning Siamese  
284 network to predict the Radiographic Assessment of Lung Edema (RALE) scores [17] used to  
285 quantify severity of ARDS in COVID-19 patients. These landmark studies have proffered  
286 definitive directions for the potential use of chest X-rays in clinical care of critical patients.  
287 However, a direct application of these systems to predict actionable outcomes like the need for  
288 mechanical ventilation is currently lacking. Using cytokine / chemokine data on hospitalized  
289 COVID-19 patients, Donlan et al [12] have shown that circulating concentration of interleukin-

290 13 (IL-13) can predict the need for mechanical ventilation. Since IL-13 can contribute to  
291 pulmonary eosinophilia and tissue remodeling, it is conceivable that radiographically detectable  
292 texture alterations may accompany these cytokine profiles. This hypothesis is, in part,  
293 supported by investigations in COVID-19 negative patients.[18, 19] Whether such a correlation  
294 exists within the context of COVID-19 is currently unknown. Considered in totality, these  
295 previous studies provide a possible biological explanation to why chest radiographs can predict  
296 the need for mechanical ventilation in immediate future.

297

298 The results of our study should be considered in the light of some limitations. First, this was a  
299 retrospective, observational evaluation and the confounding and bias implicit in such an  
300 investigation will remain a limitation. Second, the data for this study were derived from a single  
301 center and the generalizability of this approach to other settings needs to be established in  
302 further studies. Third, we restricted our model to the use of chest radiographs only. However,  
303 additional clinical parameters at the time of hospital admission such as respiratory rate,  
304 oxygenation status (e.g. the ROX index[20]) and altered mental status [21] along with socio-  
305 demographic characteristics, comorbidity profile and laboratory investigations can potentially  
306 further improve the prediction. Future studies need to evaluate these possibilities, but our  
307 focus was to use an objective measure such as a chest radiograph and provide a tool to the  
308 critical care provider with a reasonable expectation of the future course of disease in a given  
309 patient.

310

311 **CONCLUSIONS**

312 Till the time effective preventive and management options for COVID-19 patients become  
313 widely available, consorted efforts that reduce the risks to the patient and thus the burden on  
314 healthcare system are needed. To that end, our study demonstrates a proof-of-principle that  
315 early prediction of the need for ventilation using chest X-ray images acquired early during  
316 hospitalization can accurately predict the need for mechanical ventilation in COVID-19 patients.  
317 We believe that such a tool has value in effectively triaging COVID-19 patients at the time of  
318 initial healthcare contact.

319 REFERENCES

- 320 [1] COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns  
321 Hopkins University (JHU), 2020 [accessed April 25.2020].
- 322 [2] Coronavirus disease 2019, <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>; 2020 [accessed August 11.2020].
- 324 [3] Abate SM, Ahmed Ali S, Mantfardo B, Basu B. Rate of Intensive Care Unit admission and  
325 outcomes among patients with coronavirus: A systematic review and Meta-analysis.  
326 PloS one 2020;15(7):e0235653.
- 327 [4] Current hospital capacity estimates - snapshot,  
328 <https://www.cdc.gov/nhsn/covid19/report-patient-impact.html>; 2020 [accessed August  
329 11.2020].
- 330 [5] Duca A, Memaj I, Zanardi F, Preti C, Alesi A, Della Bella L, et al. Severity of respiratory  
331 failure and outcome of patients needing a ventilatory support in the Emergency  
332 Department during Italian novel coronavirus SARS-CoV2 outbreak: Preliminary data on  
333 the role of Helmet CPAP and Non-Invasive Positive Pressure Ventilation.  
334 EClinical Medicine 2020;24:100419.
- 335 [6] Zhu W, Wang Y, Xiao K, Zhang H, Tian Y, Clifford SP, et al. Establishing and Managing a  
336 Temporary Coronavirus Disease 2019 Specialty Hospital in Wuhan, China.  
337 Anesthesiology: The Journal of the American Society of Anesthesiologists  
338 2020;132(6):1339-45.
- 339 [7] USNS Mercy, USNS Comfort receiving patients in LA, New York City, [accessed August  
340 11.2020].

- 341 [8] Interim clinical guidance for management of patients with confirmed coronavirus  
342 disease, 2020 [accessed August 11.2020].
- 343 [9] Khan AI, Shah JL, Bhat MM. Coronet: A deep neural network for detection and diagnosis  
344 of COVID-19 from chest x-ray images. Computer Methods and Programs in Biomedicine  
345 2020;105581.
- 346 [10] Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for  
347 COVID-19 diagnostic and prognostic analysis. European Respiratory Journal 2020.
- 348 [11] Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest  
349 radiograph diagnosis: A retrospective comparison of the CheXNeXT algorithm to  
350 practicing radiologists. PLoS medicine 2018;15(11):e1002686.
- 351 [12] Donlan AN, Young M, Petri WA, Abhyankar M. IL-13 Predicts the Need for Mechanical  
352 Ventilation in COVID-19 Patients. medRxiv : the preprint server for health sciences  
353 2020;2020.06.18.20134353.
- 354 [13] Artacho Ruiz R, Artacho Jurado B, Caballero Gueto F, Cano Yuste A, Durban Garcia I,  
355 Garcia Delgado F, et al. Predictors of success of high-flow nasal cannula in the treatment  
356 of acute hypoxemic respiratory failure. Medicina intensiva 2019.
- 357 [14] Alexander M, Wolfe R, Ball D, Conron M, Stirling RG, Solomon B, et al. Lung cancer  
358 prognostic index: a risk score to predict overall survival after the diagnosis of non-small-  
359 cell lung cancer. British journal of cancer 2017;117(5):744-51.
- 360 [15] Peters M, van der Voort van Zyp JRN, Moerland MA, Hoekstra CJ, van de Pol S,  
361 Westendorp H, et al. Development and internal validation of a multivariable prediction

- 362 model for biochemical failure after whole-gland salvage iodine-125 prostate  
363 brachytherapy for recurrent prostate cancer. Brachytherapy 2016;15(3):296-305.
- 364 [16] Li MD, Arun NT, Gidwani M, Chang K, Deng F, Little BP, et al. Automated assessment of  
365 COVID-19 pulmonary disease severity on chest radiographs using convolutional Siamese  
366 neural networks. medRxiv : the preprint server for health sciences 2020.
- 367 [17] Warren MA, Zhao Z, Koyama T, Bastarache JA, Shaver CM, Semler MW, et al. Severity  
368 scoring of lung oedema on the chest radiograph is associated with clinical outcomes in  
369 ARDS. Thorax 2018;73(9):840-6.
- 370 [18] Svenningsen S, Haider E, Boylan C, Mukherjee M, Eddy RL, Capaldi DP, et al. CT and  
371 Functional MRI to Evaluate Airway Mucus in Severe Asthma. Chest 2019;155(6):1178-89.
- 372 [19] Kim YH, Kim KW, Lee KE, Lee MJ, Kim SK, Kim SH, et al. Transforming growth factor-beta  
373 1 in humidifier disinfectant-associated children's interstitial lung disease. Pediatric  
374 pulmonology 2016;51(2):173-82.
- 375 [20] Roca O, Caralt B, Messika J, Samper M, Sztrymf B, Hernandez G, et al. An Index  
376 Combining Respiratory Rate and Oxygenation to Predict Outcome of Nasal High-Flow  
377 Therapy. American journal of respiratory and critical care medicine 2019;199(11):1368-  
378 76.
- 379 [21] Lee JY, Kim HA, Huh K, Hyun M, Rhee JY, Jang S, et al. Risk Factors for Mortality and  
380 Respiratory Support in Elderly Patients Hospitalized with COVID-19 in Korea. Journal of  
381 Korean medical science 2020;35(23):e223.
- 382  
383
- 384

385 **Table 1. Baseline characteristics of study participants (n=528).** Cells indicate the number (N)

386 and proportion (%)

387

Characteristic†	MV Needed (N = 79)	MV not needed (N = 449)	P
<b>Socio-demographic characteristics</b>			
Age >60y	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
<b>Comorbidities</b>			
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
<b>Outcomes</b>			
Death	52 (65.82)	17 (3.79)	<0.001

388 \*, Fisher's exact test

389 **FIGURE LEGENDS:**

390 **Figure 1: Overall analysis pipeline.** P, number of patients; X, number of X-ray images

391

392 **Figure 2: Densenet121 model, data preprocessing and model training.** (A) Example of a

393 preprocessed X-ray image submitted to modeling. (B) The DenseNet121 architecture.

394 Convolutional layers are prefixed with C (cyan), dense blocks with D (black) and transition

395 blocks (orange) with T. GAP, MP, SM and Sigmoid indicate the global average pooling,

396 softmax and binarization layers within the classifier portion of DenseNet121. Inset

397 shows a dense block with 4 layers and depicts how each succeeding layer receives inputs from

398 all preceding layers. Shown within each proportionately sized colored block is the output size in

399 pixels. (C-D) Data preprocessing. Shown in panel C is a batch of resized X-ray images. Panel D

400 shows the same batch after data augmentation that included center cropping, rotation and

401 horizontal displacement. (E) Training log of DenseNet121 to predict the need for mechanical

402 ventilation. Left axis shows the categorical cross-entropy loss at the end of each cycle length

403 and the right axis shows the estimated accuracy of prediction. Results are shown separately for

404 the training ( $n = 2142$ ) and the validation ( $n=378$ ) set of X-ray images. (F) Confusion matrix at

405 the end of DenseNet121 training. All the images were correctly classified at this stage.

406

407 **Figure 3: Prediction for the need of mechanical ventilation.** Analyses were done at the level of

408 X-ray image (A-D) and at the level of each patient (E-H). Panels A and E show the predictive

409 accuracy as AUROC. The optimum cutoff was chosen as the point on ROC closest to the upper

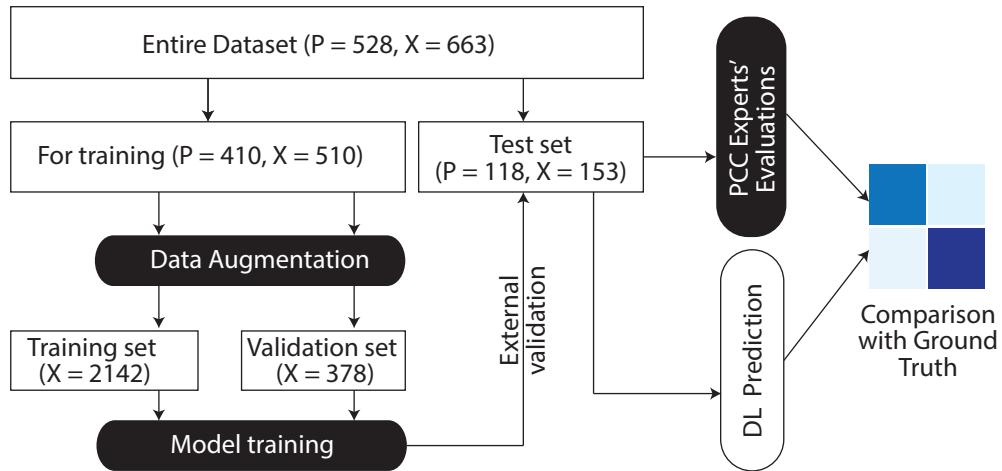
410 left corner of the plot and is indicated by a color-coded circle. The sensitivity (dashed

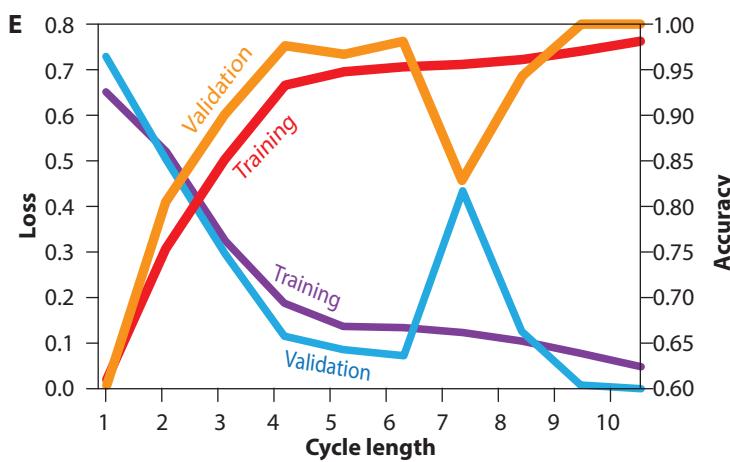
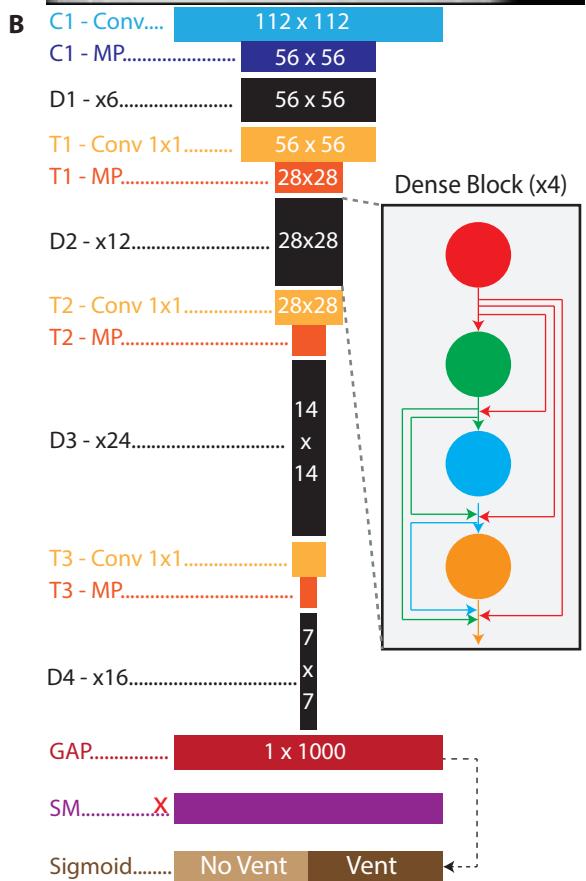
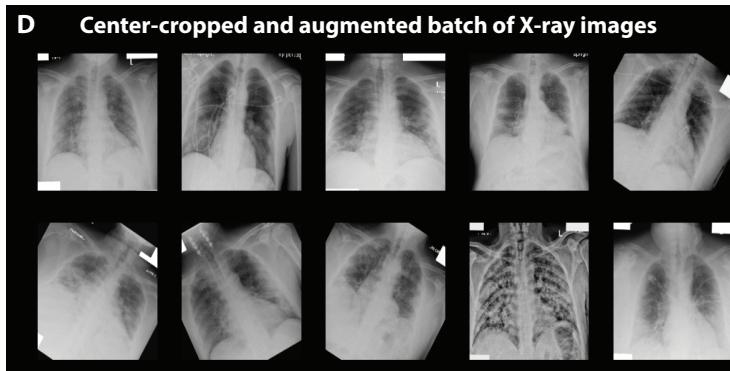
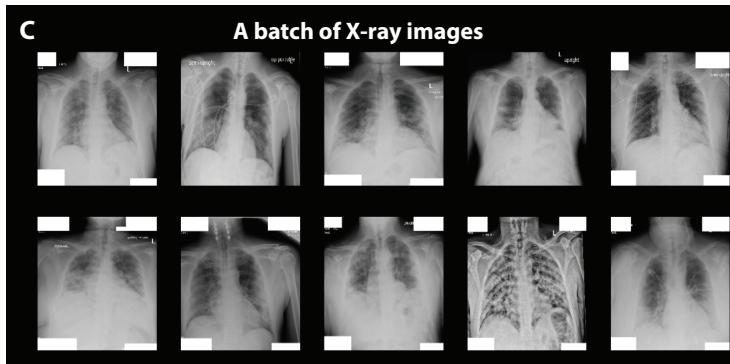
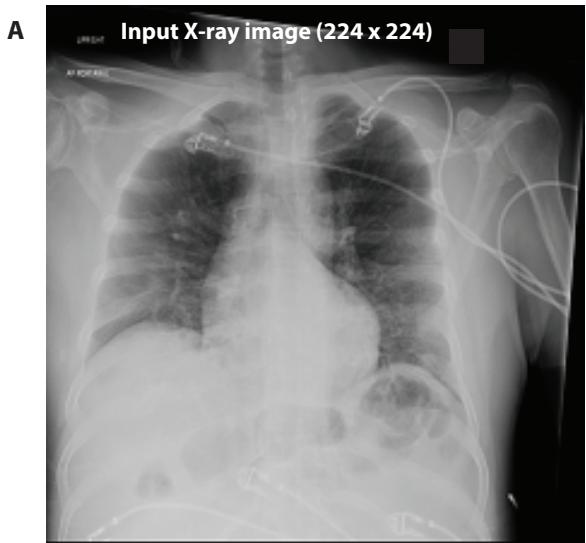
411 perpendicular to y axis) and specificity (inverse of the dashed perpendicular to the x axis) at the  
412 optimal cutoff is shown as Sn(best) and Sp(best), respectively. AUROC, area under the receiver  
413 operating characteristic curve; CI, confidence interval. (B-D) Each panel shows the confusion  
414 matrix on the left side and five performance metrics in a bar chart on the right side. The bars  
415 and error bars show the point and 95% confidence interval for each indicated and (color-coded)  
416 performance metric. The metrics shown in the plot are: P, precision; R, recall; A, accuracy; K,  
417 Cohen's kappa; and F, F1 score. (F-H) These panels respectively correspond to B-D but the  
418 results are shown at the level of the patient. Panels B-D and panels F-H use the same horizontal  
419 scale.

420

421 **Figure 4. Incremental prognostic value of the DL model as compared to the PCC Experts'**  
422 **evaluation.** (A-C) Kaplan-Meier plots for time to mechanical ventilation since the time of first X-  
423 ray image. For patients with multiple X-ray images the time was left-censored at the first image  
424 indicating the need of mechanical ventilation. Panels A-C indicate classifications based on the  
425 DL model (A), PCC Expert 1 (B) and PCC Expert 2 (C), respectively. Relative hazards (RH) and 95%  
426 confidence intervals (CI) were estimated using Cox proportional hazards (PH) models. Since  
427 different patients were classified as needing mechanical ventilation (MV) by the DL model and  
428 the PCC Experts, different shades of red (for MV needed) and blue (for MV not needed) are  
429 used. (D) Incremental value of DL model to prognosticate patients. Models 1 and 1A compare  
430 the prediction from a Cox PH model that used only PCC Expert 1 (Model 1) versus that from a  
431 Cox PH model that used PCC Expert 1 and the DL model as covariates. Models 2 and 2A,  
432 correspondingly compare models with only PCC Expert 2 and PCC Expert 2 with DL model as

433 covariates. Models 3 and 3A compare models with PCC Experts 1 and 2 as covariates and both  
434 PCC Experts with DL model, respectively. Bars indicate Harrell's C statistic for the indicated  
435 model. The statistical significance for the difference was tested using likelihood  $\chi^2$  test and is  
436 shown at the top of the bars depicting the indicated paired comparisons.

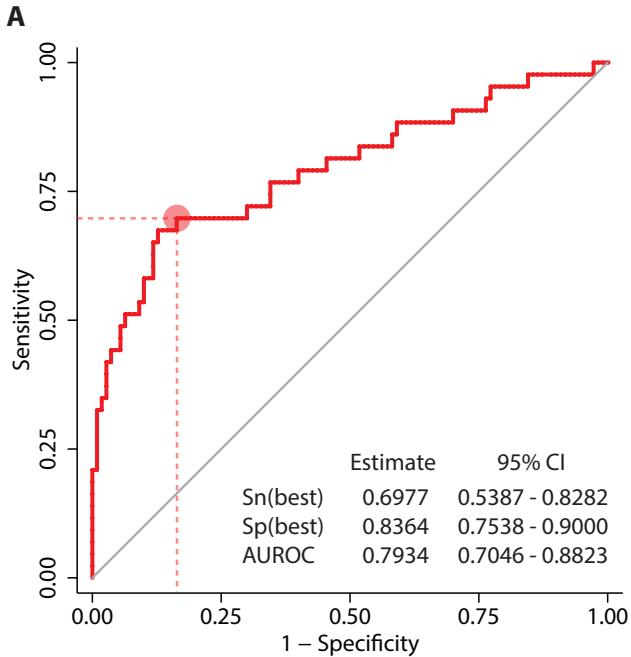




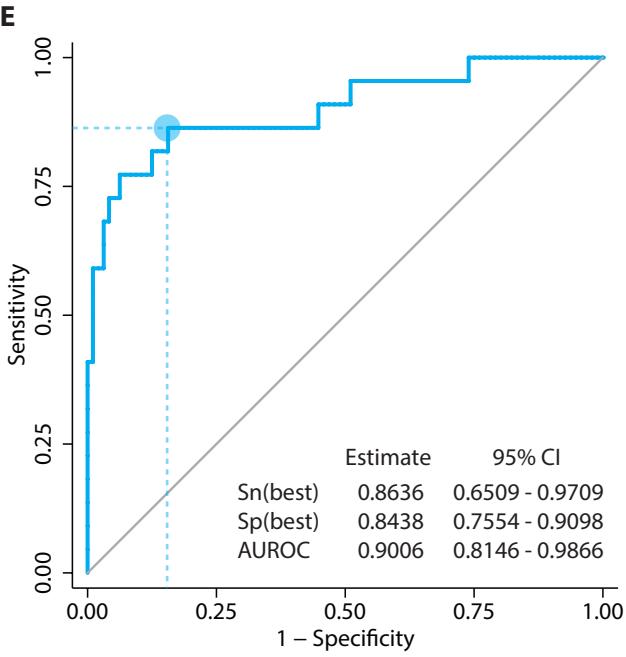
**F**

DL Prediction	Mechanical Ventilation	
	Yes	No
Need	174	0
No Need	0	204

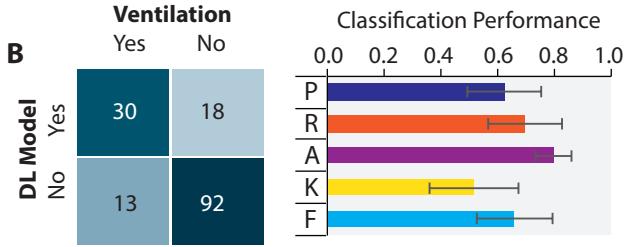
### Validation at the level of Image (n = 153)



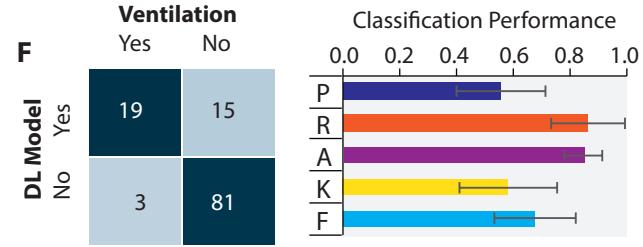
### Validation at the level of Patient (n = 118)



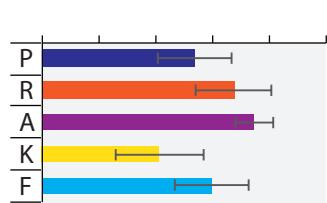
#### Mechanical Ventilation



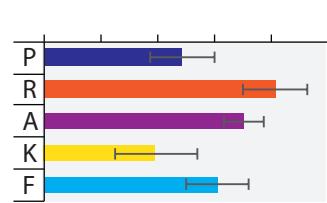
#### Mechanical Ventilation



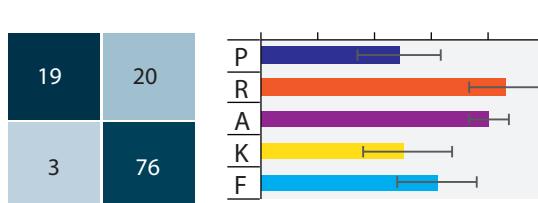
#### PCC Expert 1



#### PCC Expert 2



#### G



#### H

