

# Predicting intensive care need for COVID-19 patients using deep learning on chest radiography

Hui Li<sup>†</sup>,<sup>\*,†</sup> Karen Drukker<sup>†</sup>, Qiyuan Hu<sup>†</sup>, Heather M. Whitney<sup>†</sup>,  
Jordan D. Fuhrman, and Maryellen L. Giger<sup>†</sup>

The University of Chicago, Department of Radiology, Chicago, Illinois, United States

---

**ABSTRACT. Purpose:** Image-based prediction of coronavirus disease 2019 (COVID-19) severity and resource needs can be an important means to address the COVID-19 pandemic. In this study, we propose an artificial intelligence/machine learning (AI/ML) COVID-19 prognosis method to predict patients' needs for intensive care by analyzing chest X-ray radiography (CXR) images using deep learning.

**Approach:** The dataset consisted of 8357 CXR exams from 5046 COVID-19–positive patients as confirmed by reverse transcription polymerase chain reaction (RT-PCR) tests for the SARS-CoV-2 virus with a training/validation/test split of 64%/16%/20% on a by patient level. Our model involved a DenseNet121 network with a sequential transfer learning technique employed to train on a sequence of gradually more specific and complex tasks: (1) fine-tuning a model pretrained on ImageNet using a previously established CXR dataset with a broad spectrum of pathologies; (2) refining on another established dataset to detect pneumonia; and (3) fine-tuning using our in-house training/validation datasets to predict patients' needs for intensive care within 24, 48, 72, and 96 h following the CXR exams. The classification performances were evaluated on our independent test set (CXR exams of 1048 patients) using the area under the receiver operating characteristic curve (AUC) as the figure of merit in the task of distinguishing between those COVID-19–positive patients who required intensive care following the imaging exam and those who did not.

**Results:** Our proposed AI/ML model achieved an AUC (95% confidence interval) of 0.78 (0.74, 0.81) when predicting the need for intensive care 24 h in advance, and at least 0.76 (0.73, 0.80) for 48 h or more in advance using predictions based on the AI prognostic marker derived from CXR images.

**Conclusions:** This AI/ML prediction model for patients' needs for intensive care has the potential to support both clinical decision-making and resource management.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.10.4.044504](https://doi.org/10.1117/1.JMI.10.4.044504)]

**Keywords:** COVID-19; artificial intelligence; machine learning; chest X-ray radiography

Paper 23001GR received Jan. 1, 2023; revised Jul. 12, 2023; accepted Aug. 1, 2023; published Aug. 21, 2023.

---

<sup>†</sup>Hui Li and Karen Drukker are co-first authors

\*Address all correspondence to Hui Li, [huli@uchicago.edu](mailto:huli@uchicago.edu)

## 1 Introduction

The coronavirus disease 2019 (COVID-19) is an ongoing pandemic caused by severe acute respiratory syndrome coronavirus 2, which was first reported in late 2019. As of June 28, 2023, there have been 767,518,723 confirmed cases of COVID-19, including 6,947,192 deaths.<sup>1</sup> The reverse transcription polymerase chain reaction (RT-PCR) is the reference standard currently used for COVID-19 disease diagnosis. In addition, clinical assessment<sup>2</sup> and multimodality medical imaging<sup>3</sup> are also used in disease diagnosis and patient management.

Artificial intelligence/machine learning (AI/ML), including deep learning, has been applied in medical imaging and radiation therapy for several decades.<sup>4-8</sup> Accordingly, various studies have been reported using AI/ML on medical imaging for COVID-19 disease. AI/ML algorithms have been developed to differentiate COVID-19 pneumonia from non-COVID-19 pneumonia when RT-PCR is not readily available.<sup>9-12</sup> Various AI/ML methods have been developed to assess the severity/extent of disease<sup>13-16</sup> and predict the prognosis of the disease,<sup>17</sup> as well as for patient management in therapeutic treatment planning and monitoring patients' response.<sup>13,18</sup> Image-based studies of long-term COVID-19 effects on other organs, including the heart and brain, are also underway.<sup>19</sup>

Accurate prognosis prediction for COVID-19 patients is crucial not only for implementing appropriate treatment for individual patients, but also for optimizing medical resource allocation during the pandemic. Chest X-ray radiography (CXR) is recommended for triaging at patient presentation and disease monitoring due to its ease of use, relatively low cost, wide availability, and portability.<sup>3,20,21</sup> Characteristics such as bilateral lower lobe consolidations, ground glass opacities, peripheral air space opacities, and diffuse air space disease on CXR have been related to COVID-19.<sup>22,23</sup> However, the non-specificity of these features to COVID-19 and the shortage of radiological expertise in some resource-strained healthcare systems during a pandemic make precise image assessments challenging.

There are various studies in intensive care unit (ICU) requirement prediction for COVID-19 patients using AI/ML.<sup>24-32</sup> Those predicting models are based on clinical data, laboratory test results, comorbidity data, genetic data, and imaging data. Heo et al.<sup>24</sup> performed the logistic regression analysis to predict ICU admission status using clinical, radiological, and laboratory variables. An area under the curve (AUC) value of 0.880 was obtained from an integer-based scoring system using seven selected features. Asteris et al.<sup>26</sup> developed an artificial neural network (ANN) model based on complement-related genetic variants, age, and gender to predict ICU admission. They reported an accuracy of 89.47% in predicting COVID-19 severity using a sample of 133 patients with the developed ANN model. Chierigato et al.<sup>27</sup> built a hybrid ML/deep learning model for ICU prediction using CT images and clinical data from 558 patients with high sensitivity and specificity as well as SHapley Additive exPlanations (SHAP) values for each individual feature corresponding to the importance of each feature in the prediction model to increase the interpretability of the model.

Training a deep learning model from scratch in the medical imaging field is a challenging task since it requires large well-curated medical imaging datasets with annotations provided by medical professionals. Due to the nature of medical imaging datasets, most with necessary human-delineated annotations are small in size. Therefore, a technique called "transfer learning" has emerged to bridge this gap and has been applied in medical imaging analysis.<sup>33</sup> In these situations, deep learning models pretrained on nonmedical image datasets or medical image datasets from either a different imaging modality or same imaging modality but for different clinical tasks are fine-tuned with a relatively small medical imaging dataset for clinical decision-making tasks.<sup>33-39</sup> For example, Antropova et al.<sup>34</sup> applied transfer learning on three different imaging modalities to extract deep features and fused them with human engineered radiomic features for the diagnostic classification of breast tumors, with results demonstrating statistically significant improved classification performance as compared to previous developed computer-aided diagnosis methods. Huang et al.<sup>35</sup> applied deep transfer learning to identify possible disease on CXR images for multilabel classification task with improved prediction capacities. Samala et al.<sup>36</sup> performed a multi-stage transfer learning for the classification of malignant and benign masses in digital breast tomosynthesis images and reported improved classification performance.

The purpose of our study was to develop an AI/ML COVID-19 prognosis method to predict patients' need for intensive care by analyzing CXR images of COVID-19-positive patients using deep learning with a sequential transfer learning strategy.

## 2 Materials and Methods

### 2.1 Dataset

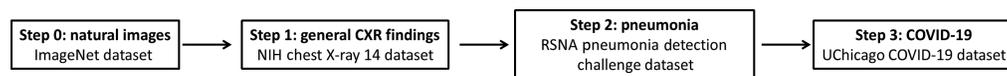
A limited deidentified dataset was retrospectively collected from our institution under a Health Insurance Portability and Accountability Act (HIPAA)-compliant, Institutional Review Board-approved protocol during the COVID-19 outbreak, consisting of CXR exams acquired between Feb 27, 2020 and January 21, 2022. From patients who underwent the RT-PCR test for the SARS-CoV-2 virus, CXR exams and clinical data were collected after the initial RT-PCR tests. The clinical data used in this study were last updated on March 13, 2022. In this study, intensive care is defined as intubation (invasive mechanical ventilation) and/or ICU admission. We assumed that all patients who needed intensive care were admitted without delay during this study period. Chest radiographs of two groups of COVID-19-positive patients were included in this study. One group consisted of COVID-19-positive patients who needed intubation or ICU support. The other group consisted of COVID-19-positive patients who were not admitted to ICU and did not need intubation following their COVID-19 diagnosis. The intubation or ICU admission information was extracted from patients' clinical information and radiology reports. The ICU admission or intubation time was compared with the imaging exam time to determine the time elapsed between imaging and any potential subsequent intubation or ICU admission event. For example, if the CXR exam was obtained within the 24 h prior to ICU admission or intubation, then the ICU admission status for 24, 48, 72, and 96 h would all be true; if the CXR exam was obtained less than 48 h but more than 24 h prior to the intubation/ICU admission event, then the 24-h status would be false, while the 48, 72, and 96 statuses would be true. For a patient without an intubation or ICU admission event, all statuses would be false. Only images acquired after a positive RT-PCR were included, and images obtained after ICU admission or intubation were excluded. Ultimately, the dataset for this study consisted of 8357 CXR images from 5046 COVID-19-positive patients. Patient demographics are summarized in Table 1. Patients were largely unvaccinated, with only 16% having received one or more vaccinations against COVID-19 at the time of imaging.

### 2.2 Classifier Training

The DenseNet121 architecture was chosen for this study because of its success in the diagnosis of various diseases on CXR in previous publications.<sup>40–42</sup> Instead of presenting the model with a random mixture of CXR examples to learn to detect COVID-19, a sequential transfer learning technique was employed to train the model on a sequence of gradually more specific and complex tasks to mimic the human learning process.<sup>43</sup> First, a model pretrained on ImageNet<sup>44</sup> with 1.2 million natural images was fine-tuned on the National Institutes of Health (NIH) ChestX-ray14 dataset to detect 14 common disease types.<sup>44,45</sup> Then, the model was fine-tuned on the Radiological Society of North America Pneumonia Detection Challenge dataset, which has a high pneumonia prevalence, ~24%, to detect evidence of pneumonia.<sup>46</sup> The data for this pneumonia detection challenge can be accessed through the challenge website.<sup>46</sup> The ground truth was provided by the radiologists at the Society for Thoracic Radiology by labeling pneumonia cases. Finally, the model was fine-tuned again on the training set of our COVID-19 dataset and then ultimately evaluated on the independent held-out test set in the task of intensive care prediction for COVID-19 patients, as conducted in our previous preliminary study.<sup>47</sup> For the preprocessing, the images were down sampled to  $256 \times 256$  pixels and gray-scale normalized. Images were randomly augmented by horizontal flipping, rotation of up to 8 deg and shifting by up to 10% of the image size. The model was trained with weighted cross-entropy loss function, Adam optimizer, and a batch size of 64 with an initial learning rate of 0.0001. Step decay on learning rate and early stopping were employed. The details regarding this cascade model training approach can be found elsewhere.<sup>10,47</sup> The sequential transfer learning diagram for predicting ICU admission of COVID-19 patients is shown in Fig. 1. The dataset was randomly split at the patient level into 64% for training, 16% for validation, and 20% for testing using stratified

**Table 1** Patient demographics of the COVID-19 dataset. Age is reported in years as mean  $\pm$  standard deviation.

Dataset	Entire dataset		Training set		Validation set		Test set	
By patients	Number of patients	Age (years)						
Number of patients	5046	54.5 $\pm$ 19.1	3181 (63.0%)	54.3 $\pm$ 19.0	817 (16.2%)	55.7 $\pm$ 19.2	1048 (20.8%)	54.2 $\pm$ 19.3
<b>Sex</b>								
Female	2833 (56.1%)	54.7 $\pm$ 19.7	1780 (56.0%)	54.6 $\pm$ 19.6	453 (55.4%)	55.1 $\pm$ 19.7	600 (57.3%)	54.4 $\pm$ 19.9
Male	2213 (43.9%)	54.4 $\pm$ 18.3	1401 (44.0%)	54.0 $\pm$ 18.2	364 (44.6%)	56.5 $\pm$ 18.5	448 (42.7%)	54.0 $\pm$ 18.5
<b>Race</b>								
American Indian or Alaska Native	9 (0.2%)		4 (0.1%)		0 (0.0%)		5 (0.5%)	
Asian/Mideast Indian	44 (0.9%)		28 (0.9%)		10 (1.2%)		6 (0.6%)	
Black/African-American	4241 (84.0%)		2687 (84.5%)		666 (81.5%)		888 (84.7%)	
More than one race	198 (3.9%)		120 (3.8%)		37 (4.5%)		41 (3.9%)	
Native Hawaiian/other Pacific Islander	4 (0.1%)		2 (0.1%)		0 (0.0%)		2 (0.2%)	
White	464 (9.2%)		278 (8.7%)		92 (11.3%)		94 (9.0%)	
Unknown/patient declined	86 (1.7%)		62 (1.9%)		12 (1.5%)		12 (1.1%)	
<b>Ethnicity</b>								
Hispanic or Latino	271 (5.4%)		166 (5.2%)		45 (5.5%)		60 (5.7%)	
Not Hispanic or Latino	4701 (93.1%)		2965 (93.2%)		759 (92.9%)		977 (93.2%)	
Unknown/patient declined	74 (1.5%)		50 (1.6%)		13 (1.6%)		11 (1.1%)	

**Fig. 1** Flow chart of sequential transfer learning diagram for ICU admission prediction of COVID-19 patients.

sampling, holding the class prevalence for the least frequent outcome, i.e., intubation or ICU admission within 24 h, constant across all subsets. Dataset statistics and the prevalence of cases that required intensive care within 24, 48, 72, and 96 h after chest radiography exams are summarized in Table 2.

### 2.3 Performance Evaluation

Performance was evaluated for the task of predicting the need for intensive care within 24, 48, 72, and 96 h after each CXR exam in the test set (1048 patients, 1672 CXR exams). Here, the classification performance for each label was evaluated using receiver operating characteristic (ROC) analysis with area under the proper binormal ROC curve (AUC) as the figure of merit.<sup>48,49</sup> The 95% confidence intervals (CIs) of the AUC values were calculated by bootstrapping the posterior probabilities of the test set (5000 bootstrap samples).<sup>50</sup> The statistical difference between the

**Table 2** Dataset statistics and the prevalence of cases that required intensive care within 24, 48, 72, and 96 h after chest radiography exams. The number of patients and images in each subset are listed.

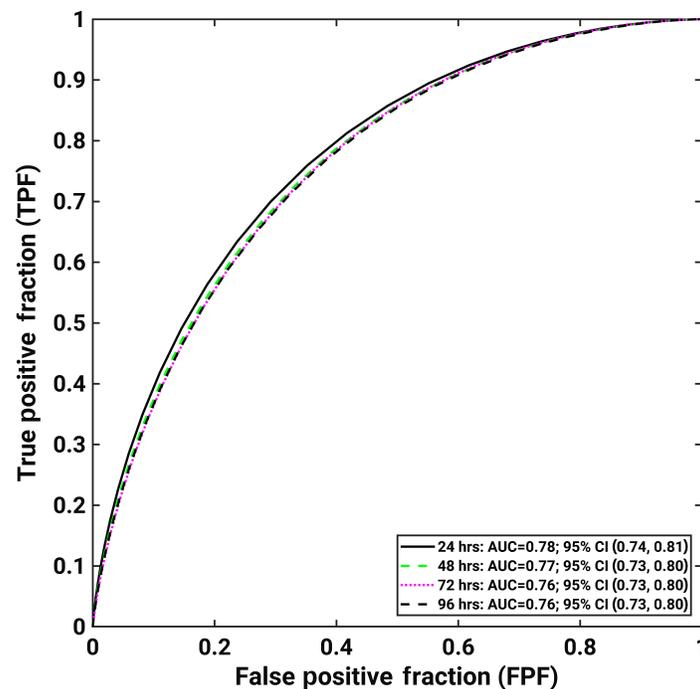
Entire dataset		Overall	Training	Validation	Test
<b>Total</b>	Patient	5046	3181 (63.0%)	817 (16.2%)	1048 (20.8%)
	Image	8357	5347 (64.0%)	1338 (16.0%)	1672 (20.0%)
<b>ICU cases</b>		<b>Overall</b>	<b>Training</b>	<b>Validation</b>	<b>Test</b>
<b>24 h</b>	Patient	730 (14.5%)	468 (14.7%)	115 (14.1%)	147 (14.0%)
	Image	979 (11.7%)	626 (11.7%)	157 (11.7%)	196 (11.7%)
<b>48 h</b>	Patient	790 (15.7%)	505 (15.9%)	125 (15.3%)	160 (15.3%)
	Image	1104 (13.2%)	718 (13.4%)	172 (12.9%)	214 (12.8%)
<b>72 h</b>	Patient	801 (15.9%)	512 (16.1%)	126 (15.4%)	163 (15.6%)
	Image	1174 (14.0%)	772 (14.4%)	179 (13.4%)	223 (13.3%)
<b>96 h</b>	Patient	809 (16.0%)	519 (16.3%)	126 (15.4%)	164 (15.6%)
	Image	1222 (14.6%)	808 (15.1%)	185 (13.8%)	229 (13.7%)

AUC values for different models was computed using ROCKIT software.<sup>51</sup> Gradient-weighted class activation mapping (Grad-CAM) was generated to provide a visual explanation of the model's classification.<sup>52</sup> The second performance evaluation was performed by patient and involved the first CXR exam of each patient only (1048 patients, 1048 CXR exams). Here, time-to-event analysis<sup>53,54</sup> was performed based on the AI/ML output for the task of predicting the need for intensive care within 96 h after the initial CXR exam. The median of the intensive care risk score (the AI/ML output) was used to divide the patient cohort into “high risk” and “low risk” subsets, and the corresponding hazard ratio was calculated. The third analysis involved post-hoc stepwise fitting of a linear regression model using the intensive care risk score, patient age, sex, race, ethnicity, and immunization status as initial variables to investigate whether variables other than the AI/ML output, i.e., the ICU/intubation risk score, were important for determining the patient prognosis within our test cohort. All reported performances pertain to the independent test set (1048 patients).

### 3 Results

The ROC curves for predicting COVID-19 patients' potential need for intensive care in 24, 48, 72, and 96 h in advance are shown in Fig. 2. We achieved an AUC (95% CI) of 0.78 (0.74, 0.81) when predicting ICU admission 24 h in advance, while also achieving promising performances in predictions made more in advance: 0.77 (0.73, 0.80), 0.76 (0.73, 0.80), and 0.76 (0.73, 0.80) when predicting ICU admission 48, 72, and 96 h in advance, respectively.

Figure 3(a) shows two examples, each with the original CXR image and the Grad-CAM heatmap from the last batch normalization layer of the model overlaid on the CXR image. The top row in Fig. 3(a) is from a COVID-19–positive patient who was admitted to ICU within 4 h following image acquisition. The bottom row in Fig. 3(a) is from a COVID-19–positive patient who did not receive intensive care within the 96 h after the CXR image was acquired, most likely due to a mild assessment of the likelihood of receiving intensive care. The predictions for intensive care within 24, 48, 72, and 96 h after CXR images agreed with the clinical assessment with both patients. The highlighted areas from the Grad-CAM heatmaps showed the abnormalities in the lungs indicating those areas of lung that had the most impact on the classification score, i.e., on the probability of COVID-19–positive patient to be admitted into ICU. This elevated Grad-CAM signal in the COVID-19 patient could be an indication of pneumonia and may be associated with the extent of ground glass/hazy opacities and consolidation of the lung



**Fig. 2** ROC curves for classification tasks requiring intensive care or not within 24, 48, 72, and 96 h from image acquisition. The legend gives the AUC with 95% CI for each task.

area. Figure 3(b) shows two examples, the top row is a false positive example and the bottom one is a false negative example.

The time-to-event analysis demonstrated that the “high risk” subset of patients (the half of the cohort with a risk score larger than/equal to the median score) had a significantly increased risk of the need for intensive care than the “low risk” subset (the half of the cohort with a risk score lower than the median score, Table 3, Fig. 4). The hazard ratio was 0.22 [95% CI (0.16, 0.30);  $p$ -value < 0.0001].

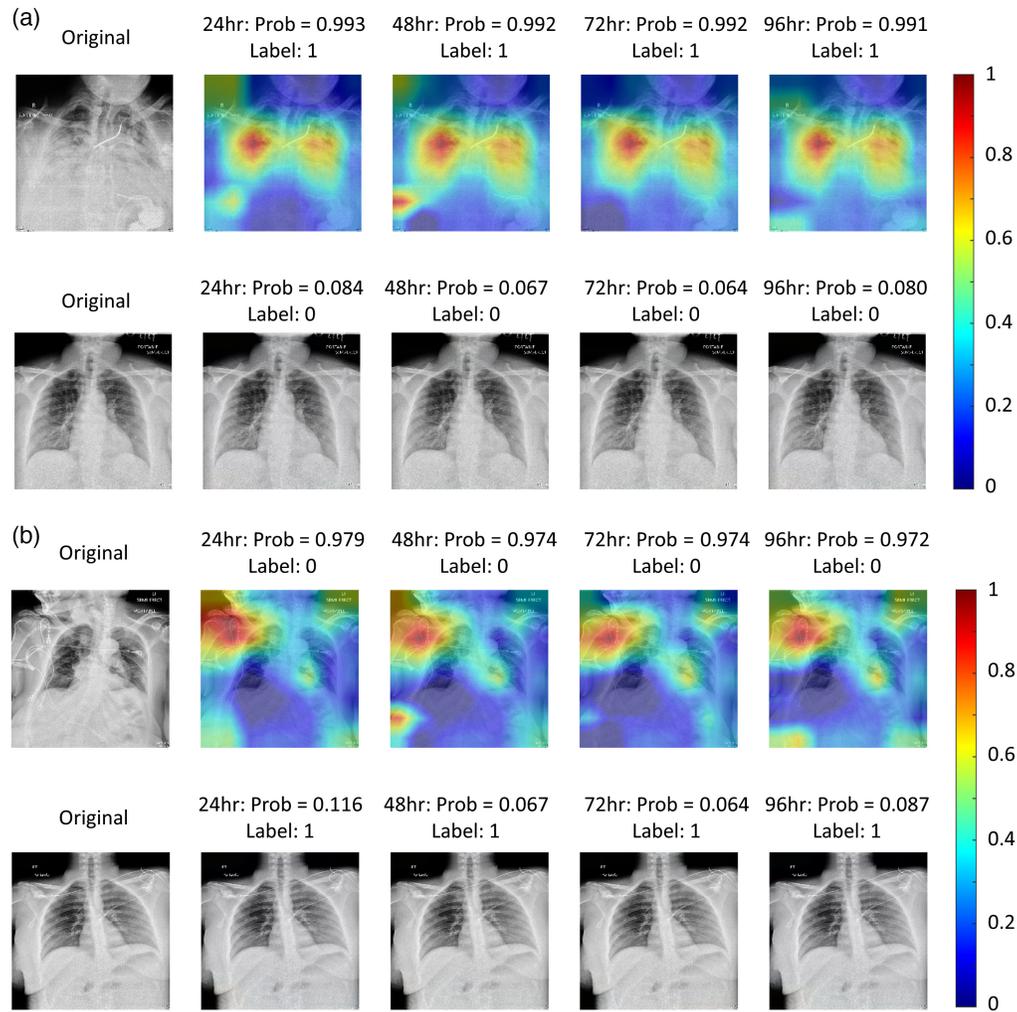
In the stepwise fitting of a linear regression model using the intensive care risk score, patient age, sex, race, ethnicity, and immunization status as initial variables, the intensive risk score was selected first ( $p$ -value < 0.0001), and patient sex was selected second ( $p$ -value = 0.020), with the women being at a slightly lower risk of needing intensive care than the men in our cohort. All other variables failed to reach statistical significance and were not selected.

## 4 Discussion

In this work, we present a deep learning method that can predict the need for intensive care of COVID-19–positive patients using CXR images, where intensive care is defined as intubation and/or ICU care, i.e., a prognostic marker of COVID severity.

Note here, without fine-tuning, the AUCs of 0.72 (0.68, 0.76), 0.70 (0.67, 0.74), 0.70 (0.66, 0.73), and 0.70 (0.66, 0.73) were obtained when predicting ICU admission 24, 48, 72, and 96 h in advance, respectively. We observed statistically significant improved performance in predicting ICU admission between two schemes, with fine-tuning and without fine-tuning, 0.78 versus 0.72 [95% CI of  $\Delta$ AUC (0.0200, 0.0941),  $p$  = 0.0025]; 0.77 versus 0.70 [for 95% CI of  $\Delta$ AUC (0.0305, 0.1038),  $p$  = 0.0003]; 0.76 versus 0.70 [95% CI of  $\Delta$ AUC (0.0305, 0.1011),  $p$  = 0.0003]; and 0.76 versus 0.70 [95% CI of  $\Delta$ AUC (0.0322, 0.1027),  $p$  = 0.0002] for predicting ICU admission 24, 48, 72, and 96 h in advance, respectively. These results indicated that this sequential transfer learning strategy may be useful on improving the model performance.

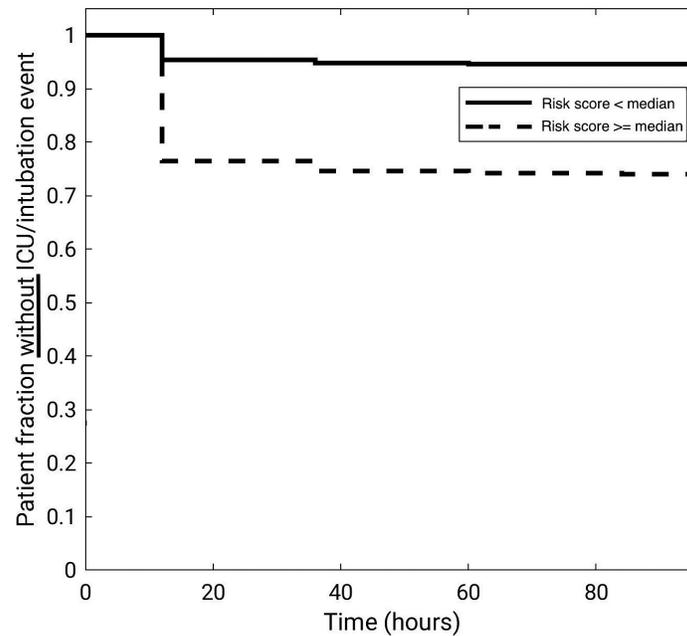
A similar study by Shamout et al.<sup>56</sup> predicted patient deterioration and achieved an AUC of 0.786 (0.745, 0.830) when using both clinical variables and imaging data and 0.738 (0.695, 0.785) when using CXR image data alone. Although a direct quantitative comparison with the existing approaches was not feasible due to the differences in the task definition and datasets,



**Fig. 3** Example CXRs overlaid with their Grad-CAM heatmaps for prediction of the need for intensive care within 24, 48, 72, and 96 h, respectively, for instances (a) in which the AI/ML prediction was correct and (b) in which the output was incorrect. The probability is the model output for the likelihood of receiving intensive care scaled to 50% prevalence.<sup>55</sup> The term “label” in the figure reflects the “ground truth” for the intensive care requirement: 1 for ICU admission/intubation and 0 for no ICU admission/intubation. (a) The patient in the top example was admitted into ICU within 4 h after image acquisition (true positive example). The patient in the bottom example did not require intensive care within 96 h after image acquisition (true-negative example). (b) The top is a false positive example and the bottom is a false negative example.

**Table 3** The number of ICU admission/intubation events within the different time windows for the “high risk” and “low risk” patient subsets of the test set, i.e., for those patients receiving a risk score smaller than, or larger/equal to, the median score of the test cohort in its entirety.

Time window (h)	ICU admission/intubation events		
	“High risk” cohort (N = 524)	“Low risk” cohort (N = 524)	Entire cohort (N = 1048)
0 to 24	123 (23.5%)	24 (4.6%)	147 (14.0%)
24 to 48	10 (1.9%)	3 (0.6%)	13 (1.2%)
48 to 72	2 (0.4%)	1 (0.2%)	3 (0.3%)
72 to 96	1 (0.2%)	0 (0%)	1 (0.1%)
0 to 96	136 (26.0%)	28 (5.3%)	164 (15.6%)



**Fig. 4** Time-to-event analysis for the need of intensive care within the 96-h time window after each patient's first CXR exam. The time progression of the data is plotted at the midpoint of the time interval. For example, the patient fraction without ICU/intubation event at 24 h post-imaging is plotted at 12 h.

interestingly, our study, using imaging data alone, yielded a similar AUC to Shamout's results incorporating both clinical and imaging data. Li et al.<sup>16</sup> also reported a COVID-19 pulmonary disease severity model using CXR and achieved an AUC of 0.80 (95% CI 0.75 to 0.85) in prediction of subsequent incubation or death within 3 days of hospital admission. Others have investigated ICU admission prediction based on clinical characteristics alone. Zhao et al.<sup>32</sup> built a prediction model for ICU admission based on clinical characteristics of COVID-19 patients. That risk score model yielded an AUC of 0.74 (0.63, 0.85) for predicting ICU admission. A similar study by Li et al.<sup>30</sup> using only clinical variables achieved an AUC of 0.780 (0.760, 0.785) in ICU admission prediction with deep learning model, interestingly, our study achieved comparable performance using image data alone.

The potential clinical utility of our CXR imaging-based ICU admission/intubation risk score is further emphasized by both the presented time-to-event analysis and the fitted linear regression model. In the former, patients deemed to be "high risk" by our AI/ML model were almost five times as likely to require intensive care compared to those deemed "low risk." In the latter, the linear regression model included only patient sex as contributing significantly to the prediction of the need for intensive care and coming second after the AI/ML predicted risk score. It should be noted; however, that for different patient cohorts, demographical characteristics may play a larger role since our institution serves a population with a demographic distribution that is different from those of the US census<sup>57</sup> or CDC.<sup>58</sup>

The majority of previous publications using imaging data of COVID-19 patients focus on diagnosis rather than prognosis.<sup>12,59–65</sup> While early and rapid diagnosis is crucial for highly infectious diseases, such as COVID-19, laboratory testing ability has largely advanced so that timely diagnosis by imaging is a lesser concern. Prognostic tasks are challenging but have substantial benefits including accurately triaging patients and forecasting demands on related hospitalization resources. An imaging-based model that can predict intensive care needs could potentially help to alleviate these challenges. We expect our CXR-based AI model could supplement prior AI studies, which only incorporated clinical variables, such as vital signs and laboratory tests or CT images,<sup>56,66–68</sup> in the prognosis of COVID-19 patients.

Some cases were classified as false positive or false negative by the model and there are some factors that could have contributed to this. First, the influence from irrelevant regions on CXR images on the prediction of ICU admission status may contribute to the false positive cases.

Incorporating lung region segmentation and cropping in the model could reduce false positive cases. Second, CXR images are a primary imaging modality for assessing the COVID-19 disease progression and pulmonary disease is the main complication associated with COVID-19 patients. However, some COVID-19 patients may have other non-pulmonary related comorbidities contributing to their deteriorating health and their ICU admission. These could cause a false negative prediction by the model. By incorporating both imaging and non-imaging data, including clinical variables and lab test results in the model could reduce the false negative and improve the model performance.

Our study has some limitations, which will be addressed in future work. First, we will expand the database to include more images as well as images from other institutions, so that we can assess the robustness of our approach. While we had access to patient demographics, clinical variables were not readily available. Thus, we will gather clinical variables as part of future investigations. AI/ML models combining imaging data with clinical variables to predicting ICU admission will be explored. We will also investigate the role of temporal analysis, taking advantage of previous and follow-up CXR exams of COVID-19 patients to evaluate disease progression. Finally, we did not compare the performance of our AI/ML model to clinician performance in predicting ICU admission from CXR. A reader study will be conducted to gather clinicians' performance on this ICU prediction task and compare with the proposed model to assess the potential clinical benefit of our model.

In summary, a deep learning CXR-based model was developed to predict patients' risk of requiring intensive care for COVID-19 at 24, 48, 72, and 96 h post-imaging. Overall, our findings show the promise of AI-assisted medical image analysis in COVID-19 prognostic task, which bear the potential to play an important role in supporting clinical decision-making especially in situations of limited resources. Our proposed model may be potentially useful for efficient patient triage and for low resourced regions that need to prioritize care, knowing who to treat immediately during a pandemic. This work has the potential to support both clinical decision-making and resource management.

---

## Disclosures

M.L.G. is a stockholder in R2 technology/Hologic and QView; receives royalties from Hologic, GE Medical Systems, MEDIAN Technologies, Riverain Medical, and Mitsubishi and Toshiba; and is a cofounder of and equity holder in Quantitative Insights (now Qlarity Imaging). K.D. and H.L. receive royalties from UCTech. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest that would reasonably appear to be directly and significantly affected by the research activities.

## Data and Code Availability Statement

The data used for this manuscript, including CXR are not publicly available due to patient privacy and data sharing agreements.

## Acknowledgments

The authors are grateful to Feng Li, MD, PhD, for the scientific discussion. This work was partially supported by an award from the C3.AI Digital Transformation Institute, the National Institute of Biomedical Imaging and Bioengineering (NIBIB) COVID-19 (Contract No. 75N92020D00021), and the National Institutes of Health (NIH) Shared Instrument Grant (S10 OD025081).

## References

1. World Health Organization, "WHO Coronavirus (COVID-19) dashboard," <https://covid19.who.int/> (accessed 28 June 2023).
2. M. A. Lake, "What we know so far: COVID-19 current clinical knowledge and research," *Clin. Med.* **20**(2), 124–127 (2020).
3. E. A. Akl et al., "Use of chest imaging in the diagnosis and management of COVID-19: a WHO rapid advice guide," *Radiology* **298**(2), E63–E69 (2020).
4. B. Sahiner et al., "Deep learning in medical imaging and radiation therapy," *Med. Phys.* **46**(1), e1–e36 (2019).
5. W. L. Bi et al., "Artificial intelligence in cancer imaging: clinical challenges and applications," *CA Cancer J. Clin.* **69**(2), 127–157 (2019).

6. M. L. Giger, "Machine learning in medical imaging," *J. Am. Coll. Radiol.* **15**(3), 512–520 (2018).
7. G. Currie et al., "Machine learning and deep learning in medical imaging: intelligent imaging," *J. Med. Imaging Radiat. Sci.* **50**(4), 477–487 (2019).
8. I. El Naqa et al., "Artificial intelligence: reshaping the practice of radiological sciences in the 21st century," *BJR* **93**(1106), 20190855 (2020).
9. R. Zhang et al., "Diagnosis of coronavirus disease 2019 pneumonia by using chest radiography: value of artificial intelligence," *Radiology* **298**(2), E88–E97 (2021).
10. Q. Hu, K. Drukker, and M. L. Giger, "Role of standard and soft tissue chest radiography images in deep-learning-based early diagnosis of COVID-19," *J. Med. Imaging* **8**(S1), 014503 (2021).
11. H. X. Bai et al., "Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT," *Radiology* **296**(3), E156–E165 (2020).
12. X. Mei et al., "Artificial intelligence-enabled rapid diagnosis of patients with COVID-19," *Nat. Med.* **26**(8), 1224–1228 (2020).
13. V. V. Danilov et al., "Automatic scoring of COVID-19 severity in x-ray imaging based on a novel deep learning workflow," *Sci. Rep.* **12**(1), 12791 (2022).
14. Z. Tang et al., "Severity assessment of COVID-19 using CT image features and laboratory indices," *Phys. Med. Biol.* **66**(3), 035015 (2021).
15. D. Assaf et al., "Utilization of machine-learning models to accurately predict the risk for critical COVID-19," *Intern. Emerg. Med.* **15**(8), 1435–1443 (2020).
16. M. D. Li et al., "Automated assessment and tracking of COVID-19 pulmonary disease severity on chest radiographs using convolutional Siamese neural networks," *Radiol. Artif. Intell.* **2**(4), e200079 (2020).
17. W. Zhao et al., "Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study," *Am. J. Roentgenol.* **214**(5), 1072–1077 (2020).
18. J. D. Fuhrman et al., "Cascaded deep transfer learning on thoracic CT in COVID-19 patients treated with steroids," *J. Med. Imaging* **8**(S1), 014501 (2021).
19. A. Mahammedi et al., "Brain and lung imaging correlation in patients with COVID-19: could the severity of lung disease reflect the prevalence of acute abnormalities on neuroimaging? A global multicenter observational study," *AJNR Am. J. Neuroradiol.* **42**(6), 1008–1016 (2021).
20. American College of Radiology, "ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection," <https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection> (accessed 17 June 2020).
21. G. D. Rubin et al., "The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society," *Chest* **158**(1), 106–116 (2020).
22. A. Jacobi et al., "Portable chest x-ray in coronavirus disease-19 (COVID-19): a pictorial review," *Clin. Imaging* **64**, 35–42 (2020).
23. M.-Y. Ng et al., "Imaging profile of the COVID-19 infection: radiologic findings and literature review," *Radiol. Cardiothorac. Imaging* **2**(1), e200034 (2020).
24. J. Heo et al., "Prediction of patients requiring intensive care for COVID-19: development and validation of an integer-based score using data from Centers for Disease Control and Prevention of South Korea," *J. Intensive Care* **9**(1), 16 (2021).
25. S. Saadatmand et al., "Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID-19 patients," *Ann. Oper. Res.* 1–29 (2022).
26. P. G. Asteris et al., "Genetic prediction of ICU hospitalization and mortality in COVID-19 patients using artificial neural networks," *J. Cell. Mol. Med.* **26**(5), 1445–1455 (2022).
27. M. Chierigato et al., "A hybrid machine learning/deep learning COVID-19 severity predictive model from CT images and clinical data," *Sci. Rep.* **12**(1), 4329 (2022).
28. S. Subudhi et al., "Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19," *NPJ Digit. Med.* **4**(1), 87 (2021).
29. D. Patel et al., "Machine learning based predictors for COVID-19 disease severity," *Sci. Rep.* **11**(1), 4673 (2021).
30. X. Li et al., "Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables," *PeerJ* **8**, e10337 (2020).
31. F.-Y. Cheng et al., "Using machine learning to predict ICU transfer in hospitalized COVID-19 patients," *JCM* **9**(6), 1668 (2020).
32. Z. Zhao et al., "Prediction model and risk scores of ICU admission and mortality in COVID-19," *PLoS ONE* **15**(7), e0236618 (2020).
33. H.-C. Shin et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imaging* **35**(5), 1285–1298 (2016).
34. N. Antropova, B. Q. Huynh, and M. L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," *Med. Phys.* **44**(10), 5162–5171 (2017).

35. G.-H. Huang et al., “Deep transfer learning for the multilabel classification of chest x-ray images,” *Diagnostics* **12**(6), 1457 (2022).
36. R. K. Samala et al., “Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets,” *IEEE Trans. Med. Imaging* **38**(3), 686–696 (2019).
37. B. Q. Huynh, H. Li, and M. L. Giger, “Digital mammographic tumor classification using transfer learning from deep convolutional neural networks,” *J. Med. Imaging* **3**(3), 034501 (2016).
38. Y. Bar et al., “Deep learning with non-medical training used for chest pathology identification,” *Proc. SPIE* **9414**, 94140V (2015).
39. G. Ayana et al., “A novel multistage transfer learning for ultrasound breast cancer image classification,” *Diagnostics* **12**(1), 135 (2022).
40. G. Huang et al., “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 4700–4708 (2017).
41. P. Rajpurkar et al., “ChexNet: radiologist-level pneumonia detection on chest x-rays with deep learning,” arXiv:1711.05225 (2017).
42. P. Rajpurkar et al., “Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists,” *PLoS Med.* **15**(11), e1002686 (2018).
43. Y. Bengio et al., “Curriculum learning,” in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, pp. 41–48 (2009).
44. J. Deng et al., “ImageNet: a large-scale hierarchical image database,” in *IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 248–255 (2009).
45. X. Wang et al., “ChestX-Ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, pp. 2097–2106 (2017).
46. “RSNA Pneumonia Detection Challenge,” (2018). <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data> (accessed 18 June 2020).
47. Q. Hu, K. Drukker, and M. L. Giger, “Predicting the need for intensive care for COVID-19 patients using deep learning on chest radiography,” in *The 34th Annu. Conf. Neural Inf. Process. Syst. (NeurIPS), Med. Imaging Meets NeurIPS Workshop* (2020).
48. C. E. Metz, B. A. Herman, and J. H. Shen, “Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data,” *Stat. Med.* **17**(9), 1033–1053 (1998).
49. C. E. Metz and X. Pan, “‘Proper’ binormal ROC curves: theory and maximum-likelihood estimation,” *J. Math. Psychol.* **43**(1), 1–33 (1999).
50. B. Efron, “Better bootstrap confidence intervals,” *J. Am. Stat. Assoc.* **82**(397), 171–185 (1987).
51. “ROC software,” <http://metz-roc.uchicago.edu/MetzROC/software>
52. R. R. Selvaraju et al., “GRAD-CAM: visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 618–626 (2017).
53. E. L. Kaplan and P. Meier, “Nonparametric estimation from incomplete observations,” *J. Am. Stat. Assoc.* **53**(282), 457–481 (1958).
54. M. K. Goel, P. Khanna, and J. Kishore, “Understanding survival analysis: Kaplan-Meier estimate,” *Int. J. Ayurveda Res.* **1**(4), 274–278 (2010).
55. K. Horsch, M. L. Giger, and C. E. Metz, “Prevalence scaling,” *Acad. Radiol.* **15**(11), 1446–1457 (2008).
56. F. E. Shamout et al., “An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department,” *NPJ Digit. Med.* **4**(1), 80 (2021).
57. US Census Bureau, “Census.gov,” <https://www.census.gov/> (accessed 28 June 2023).
58. Centers for Disease Control and Prevention, “COVID-19 case surveillance public use data,” <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf> (accessed 28 June 2023).
59. L. Li et al., “Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy,” *Radiology* **296**(2), E65–E71 (2020).
60. S. Minaee et al., “Deep-covid: predicting COVID-19 from chest x-ray images using deep transfer learning,” arXiv:2004.09363 (2020).
61. L. Wang and A. Wong, “Covid-Net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images,” arXiv:2003.09871 (2020).
62. K. Murphy et al., “COVID-19 on the chest radiograph: a multi-reader evaluation of an AI system,” *Radiology* **296**(3), E166–E172 (2020).
63. A. I. Khan, J. L. Shah, and M. M. Bhat, “CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images,” *Comput. Methods Programs Biomed.* **196**, 105581 (2020).
64. T. Ozturk et al., “Automated detection of COVID-19 cases using deep neural networks with x-ray images,” *Comput. Biol. Med.* **121**, 103792 (2020).
65. H. Panwar et al., “Application of deep learning for fast detection of COVID-19 in x-rays using nCOVnet,” *Chaos Solitons Fractals* **138**, 109944 (2020).
66. S. Wang et al., “A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis,” *Eur. Respir. J.* **56**(2), 2000775 (2020).

67. K. Zhang et al., “Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography,” *Cell* **181**(6), 1423–1433.e11 (2020).
68. S. Debnath et al., “Machine learning to assist clinical decision-making during the COVID-19 pandemic,” *Bioelectron. Med.* **6**(1), 1–8 (2020).

**Hui Li**, PhD, is a research associate professor of radiology at the University of Chicago. He has been working on quantitative imaging analysis on medical images for over 20 years. His research interests include risk assessment, diagnosis, prognosis, response to therapy, understanding the relationship between radiomics and genomics, and their future roles in precision medicine with both conventional and deep learning approaches.

**Karen Drukker** received her PhD in physics from the University of Amsterdam. She is a research associate professor of radiology at the University of Chicago, where she has been involved in medical imaging research for 20+ years. Her research interests include machine learning applications in the detection, diagnosis, and prognosis of disease, focusing on rigorous training/testing protocols, generalizability, performance evaluation, and bias and fairness of AI. She is a fellow of SPIE and the American Association of Physicists in Medicine (AAPM).

**Qiyuan Hu** is a machine learning scientist at Tempus Labs. She received her PhD in medical physics from the University of Chicago in 2021 and BA degrees in physics and mathematics from Carleton College. Her research interests include machine learning methodologies for medical image analysis. She was a student member of SPIE and an officer of the University of Chicago SPIE Student Chapter.

**Heather M. Whitney**, PhD, is working as a research assistant professor of radiology at the University of Chicago. Her experience in quantitative medical imaging has ranged from polymer gel dosimetry to radiation damping in nuclear magnetic resonance to radiomics. She is interested in investigating the effects of the physical basis of imaging on radiomics, the repeatability and robustness of radiomics, the development of methods for task-based distribution, and bias and diversity of medical imaging datasets.

**Jordan D. Fuhrman** is a staff scientist in the Department of Radiology at the University of Chicago whose research interests primarily lie in investigating computer-aided diagnosis/AI algorithms for medical image evaluation. His work includes the assessment of lung screening CT scans, head CT scans presented in the neurocritical care unit, and multi-modal COVID-19 modeling. He is a member of AAPM and SPIE, and through AAPM, is affiliated with the Medical Imaging and Data Resource Center.

**Maryellen L. Giger** is the A.N. Pritzker Distinguished Service Professor at the University of Chicago. Her research involves computer-aided diagnosis/machine learning in medical imaging for cancer and now COVID-19 and is contact PI on the NIBIB-funded Medical Imaging and Data Resource Center ([midrc.org](http://midrc.org)), which has published more than 100,000 medical imaging studies for use by AI investigators. She is a member of the National Academy of Engineering, a recipient of the AAPM Coolidge Gold Medal, SPIE Harrison H. Barrett Award, and RSNA Outstanding Researcher Award, and is a Fellow of AAPM, AIMBE, SPIE, and IEEE.