Original Paper

# Deep Learning With Chest Radiographs for Making Prognoses in Patients With COVID-19: Retrospective Cohort Study

Hyun Woo Lee[1,2*], MD; Hyun Jun Yang[2*], MS; Hyungjin Kim[2,3], MD; Ue-Hwan Kim[4], PhD; Dong Hyun Kim[2,5], MD; Soon Ho Yoon[2,3], MD; Soo-Youn Ham[6], MD; Bo Da Nam[7], MD; Kum Ju Chae[8], MD; Dabee Lee[9], MD; Jin Young Yoo[10], MD; So Hyeon Bak[11], MD; Jin Young Kim[12], MD; Jin Hwan Kim[13], MD; Ki Beom Kim[14], MD; Jung Im Jung[15], MD; Jae-Kwang Lim[16], MD; Jong Eun Lee[17], MD; Myung Jin Chung[18], MD; Young Kyung Lee[19], MD; Young Seon Kim[20], MD; Sang Min Lee[21], MD; Woocheol Kwon[22], MD; Chang Min Park[2,3], MD; Yun-Hyeon Kim[17], MD; Yeon Joo Jeong[23], MD; Kwang Nam Jin[2,5], MD; Jin Mo Goo[2,3], MD

[1]Division of Respiratory and Critical Care, Department of Internal Medicine, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Republic of Korea

[2]College of Medicine, Seoul National University, Seoul, Republic of Korea

[3]Department of Radiology, Seoul National University Medical Research Center, Seoul, Republic of Korea

[4]AI Graduate School, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea

[5]Department of Radiology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Republic of Korea

[6]Department of Radiology, Kangbuk Samsung Hospital, Seoul, Republic of Korea

[7]Department of Radiology, Soonchunhyang University Seoul Hospital, Soonchunhyang University College of Medicine, Seoul, Republic of Korea

[8]Department of Radiology, Research Institute of Clinical Medicine, Jeonbuk National University-Biomedical Research Institute of Jeonbuk National University Hospital, Jeonju, Republic of Korea

[9]Department of Radiology, Dankook University Hospital, Cheonan, Republic of Korea

[10]Department of Radiology, Chungbuk National University Hospital, Cheongju, Republic of Korea

[11]Department of Radiology, Kangwon National University Hospital, Kangwon National University School of Medicine, Chuncheon, Republic of Korea

[12]Department of Radiology, Keimyung University Dongsan Hospital, Keimyung University School of Medicine, Daegu, Republic of Korea

[13]Department of Radiology, Chungnam National University Hospital, College of Medicine, Daejeon, Republic of Korea

[14]Department of Radiology, Daegu Fatima Hospital, Daegu, Republic of Korea

[15]Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[16]Department of Radiology, Kyungpook National University Hospital, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

[17]Department of Radiology, Chonnam National University Hospital, Gwangju, Republic of Korea

[18]Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[19]Department of Radiology, Seoul Medical Center, Seoul, Republic of Korea

[20]Department of Radiology, Yeungnam University Hospital, Yeungnam University College of Medicine, Daegu, Republic of Korea

[21]Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

[22]Department of Radiology, Ewha Womans University Seoul Hospital, Seoul, Republic of Korea

[23]Department of Radiology, Pusan National University Hospital, Pusan National University School of Medicine and Biomedical Research Institute, Busan, Republic of Korea

[*]these authors contributed equally

**Corresponding Author:**
Kwang Nam Jin, MD
Department of Radiology
Seoul Metropolitan Government-Seoul National University Boramae Medical Center
20, Boramae-ro 5-gil, Dongjak-gu
Seoul, 07061
Republic of Korea
Phone: 82 2 870 2536
Email: wlsrhkdska@gmail.com

## *Abstract*

**Background:**   An artificial intelligence (AI) model using chest radiography (CXR) may provide good performance in making prognoses for COVID-19.

**Objective:**   We aimed to develop and validate a prediction model using CXR based on an AI model and clinical variables to predict clinical outcomes in patients with COVID-19.

**Methods:**   This retrospective longitudinal study included patients hospitalized for COVID-19 at multiple COVID-19 medical centers between February 2020 and October 2020. Patients at Boramae Medical Center were randomly classified into training, validation, and internal testing sets (at a ratio of 8:1:1, respectively). An AI model using initial CXR images as input, a logistic regression model using clinical information, and a combined model using the output of the AI model (as CXR score) and clinical information were developed and trained to predict hospital length of stay (LOS) ≤2 weeks, need for oxygen supplementation, and acute respiratory distress syndrome (ARDS). The models were externally validated in the Korean Imaging Cohort of COVID-19 data set for discrimination and calibration.

**Results:**   The AI model using CXR and the logistic regression model using clinical variables were suboptimal to predict hospital LOS ≤2 weeks or the need for oxygen supplementation but performed acceptably in the prediction of ARDS (AI model area under the curve [AUC] 0.782, 95% CI 0.720-0.845; logistic regression model AUC 0.878, 95% CI 0.838-0.919). The combined model performed better in predicting the need for oxygen supplementation (AUC 0.704, 95% CI 0.646-0.762) and ARDS (AUC 0.890, 95% CI 0.853-0.928) compared to the CXR score alone. Both the AI and combined models showed good calibration for predicting ARDS (*P*=.079 and *P*=.859).

**Conclusions:**   The combined prediction model, comprising the CXR score and clinical information, was externally validated as having acceptable performance in predicting severe illness and excellent performance in predicting ARDS in patients with COVID-19.

## Introduction

SARS-CoV-2 infection causes COVID-19 pneumonia of varying severity. As of March 18, 2022, the global cumulative number of confirmed COVID-19 cases was more than 464.8 million, with >6 million deaths [1]. The occurrence of new variants of SARS-CoV-2 makes appropriate medical resource allocation, based on COVID-19 severity, challenging. A reliable prediction model for COVID-19 pneumonia would help in screening patients at a high risk of progression to severe disease or respiratory failure in a timely manner [2]. While many COVID-19 prediction models have been suggested, most have not been sufficiently validated [3].

Chest radiography (CXR) is not recommended for confirmation of diagnosis or assessment of COVID-19 severity [4]. Most studies have developed prognosis prediction models using clinical information and chest computed tomography (CT) scans [5,6]. However, with the high volume of patients during the pandemic, CXR is used more widely than chest CT because of its rapid speed, better portability, and lower cost [7,8]. Additionally, the role of CXR has been reexamined using advanced deep learning (DL) techniques. A DL model could make prognoses for COVID-19 more accurately than conventional severity score systems [9]. CXR information can increase the accuracy of severity assessment or risk stratification. An artificial intelligence (AI) model using CXR alone performed well in predicting COVID-19 severity [10]. A recent study showed that an AI model with CXR had the potential to predict mortality more accurately [11]. An external validation study reported that predictive modeling with CXR and clinical information improved prognoses compared to clinical information alone or radiologist-derived severity scores [12].

We aimed to create an AI model using CXR and clinical variables to predict early recovery, severe illness, and acute respiratory distress syndrome (ARDS) in patients with COVID-19 and validate these models in an external cohort.

## Methods

Our study adhered to the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines [13].

### Study Design and Eligibility Criteria

This retrospective longitudinal study included hospitalized patients with COVID-19 in an isolation ward who received negative pressure ventilation from February 2020 to October 2020. COVID-19 was diagnosed by confirmatory quantitative reverse transcription–polymerase chain reaction using upper or lower respiratory tract samples. A protocol was used for systematic questionnaires and anthropometric measurements to ascertain demographic information, initial symptoms, and comorbidities of the patients. Within 24 hours of hospitalization, the patients routinely underwent blood tests and anteroposterior-view CXR. Medical decisions regarding treatment and discharge to home were made by each physician based on guidance from the Korea Disease Control and Prevention Agency. We excluded patients with missing clinical information or when there were technical difficulties in reading CXR images due to compromised software compatibility.

For model training, validation, and internal testing, we used CXR images and clinical information from patients who were admitted at Boramae Medical Center (BMC), Seoul, Korea. For external testing, we used the Korean Imaging Cohort of COVID-19 (KICC-19) data set, which collects imaging data and clinical information from patients with COVID-19 at 17 medical centers in Korea. Details on the profiles of the KICC-19 data set were published in an earlier report [14].

## Variables and Measurements

Data were collected on baseline characteristics, including age, sex, BMI, smoking status, and comorbidities. A protocolized questionnaire was used at the BMC to identify the symptoms of patients with COVID-19, including abnormal smell or taste, myalgia, sore throat, cough, sputum, chest discomfort, dyspnea, fever or chills, rhinorrhea, and nausea or diarrhea. In the KICC-19 data set, we extracted information on symptoms such as cough, dyspnea, and fever. We obtained laboratory test results, including white blood cell counts and lymphocyte percentage, as well as C-reactive protein (CRP), procalcitonin, troponin-I, and lactate dehydrogenase levels. Information on treatment and disease severity was also acquired. Information was obtained on clinical outcomes, including hospital length of stay (LOS) and oxygen supplementation, as well as the use of a high-flow nasal cannula (HFNC), mechanical ventilator (MV), or extracorporeal membrane oxygenation (ECMO). ARDS was operationally defined as a medical condition needing an HFNC, MV, or ECMO.

We obtained 26,684 CXR images with information on the location and extent of pneumonia provided by the Radiological Society of North America (RSNA) pneumonia-detection challenge. We extracted the initial CXR images from the electronic medical records of patients hospitalized for COVID-19 at the BMC and registered them in the KICC-19.
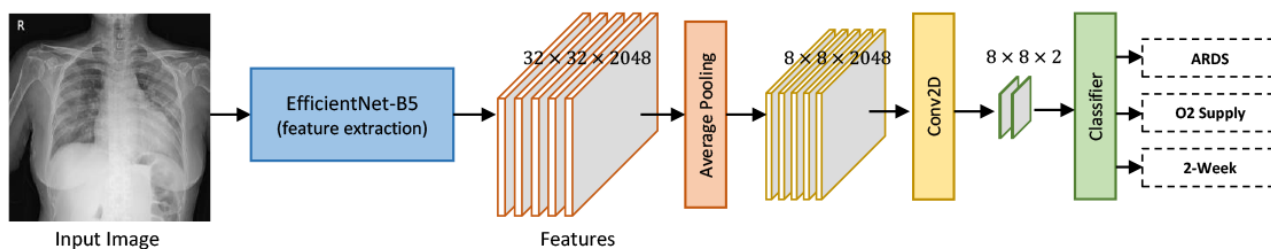
## Study Outcomes

The primary outcome was the performance of the AI model in predicting clinical outcomes such as hospital LOS ≤2 weeks, need for oxygen supplementation, and development of ARDS based on CXR images. The output value from this model was defined as the CXR score. The secondary outcome was whether the performance of the prediction model using clinical information could be improved by combining it with the CXR score.

## Development of the AI Model

The DL model was implemented using the open-source PyTorch library (version 1.7.0+cu101) with the CUDA/cuDNN (versions 10.1 and 7.6.3, respectively) computing frameworks on a single graphics processing unit (Geforce RTX 3090; NVIDIA). The overall data flow and proposed model architecture is summarized in Figure 1. We developed our model in two stages, to ensure robust performance: (1) backbone training and (2) model training. First, we trained the backbone of our model with a data set (n=26,684) including information on the location and extent of pneumonia from the RSNA pneumonia-detection challenge to boost performance by learning robust features from a large quantity of data. The EfficientNet B5 architecture was used as the backbone architecture due to its computational efficiency and performance compared to those of other convolutional neural network architectures, such as ResNet and DenseNet. For backbone training, we attached a region proposal network, a region-of-interest pooling layer, and a classifier network to the backbone, which was pretrained on ImageNet, to configure the faster region-based convolutional neural network (RCNN) architecture (Multimedia Appendix 1, Figure S1). We trained the faster-RCNN model with the Adam optimizer under the following settings: learning rate of $1\times10^{-6}$, learning decay rate of 0.8, learning rate decay step size of 4, and batch size of 3. We selected the model with the minimum validation loss, which was a combination of classification and bounding box regression loss. After the backbone training stage, the backbone learns to extract pertinent features for detecting pneumonia from posteroanterior or anteroposterior CXR images.

Next, we developed and trained the model for the classification of clinical outcomes. The backbone was followed by 3 branches: one each for hospital LOS ≤2 weeks, oxygen supplementation, and development of ARDS. Each branch consisted of an average pooling layer, a 2D convolution layer, a clinical data channel, and a fully connected layer. The outputs of the model were 3 probabilities between 0 and 1: one each for hospital LOS ≤2 weeks, oxygen supplementation, and development of ARDS; for each output, a probability >0.5 indicated a positive prediction. We initialized the weights of the models using the weights from the previous backbone training stage. We fixed the weights of the backbone during the model training stage to keep the feature extractor untouched. To train the proposed model, we used a configuration of the Adam optimizer with a learning rate of $1\times10^{-4}$ and batch size of 12. We selected the model with minimum validation loss, that is, classification loss.

**Figure 1.** Illustration of the data flow model for artificial intelligence–assisted prediction. Our deep learning model was developed in two stages to ensure robust performance: (1) backbone training and (2) model training. ARDS: acute respiratory distress syndrome. Conv2D: Convolution 2D.

## Statistical Analyses

Demographic information, symptoms, laboratory test results, treatments, and study outcomes were compared among the training, validation, internal testing, and external testing sets using the Student 2-tailed *t* test or Mann-Whitney *U* test for continuous variables and the Pearson chi-square test or Fisher exact test for categorical variables. Univariate and multivariate logistic regression analyses were performed using the stepwise selection method. For 3 different prediction models for each clinical outcome (model 1, using the CXR score derived from the DL model; model 2, using clinical information derived from the multivariable regression model; and model 3, using both the CXR scores and clinical information from the multivariable regression model), performance was evaluated using sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and the area under the receiver operating characteristic curve (AUROC). We considered an AUROC <0.7 suboptimal performance, 0.7 to 0.79 acceptable, 0.8 to 0.89 excellent, and ≥0.9 outstanding [15]. A comparison of the predictive performance between the 2 different models used the DeLong test [16] or the bootstrap test [17]. Statistical significance was set at *P*<.05. Calibration of the CXR score–based models (models 1 and 3) was evaluated by plotting the observed versus predicted probabilities and using the *P* value for the Spiegelhalter statistic [18,19]. Statistical significance in the Spiegelhalter *z* test indicates poor calibration. All the statistical analyses were performed using R (version 4.1.0; R Core Team).

## Ethics Approval

The Institutional Review Board Committee of the Boramae Medical Center (BMC) approved the study protocol and waived the need for informed consent for access to the electronic medical records (30-2020-307).

# Results

## Patient Characteristics

We used CXR images and clinical information from hospitalized patients with COVID-19 for model training (n=589), validation (n=75), and internal testing (n=75); we used patients with COVID-19 (n=467) registered in the KICC-19 for external testing. The median interval between symptom onset and CXR was 3 (IQR 1-6) days. The baseline characteristics of the combined total of 1206 patients are summarized in Table 1. The mean age was 53.4 years; 52.3% (n=631) were female; and 9.4% (n=113) were every-day smokers. Comorbidities included hypertension (n=317, 26.3%), diabetes mellitus (n=188, 15.6%), cancer (n=74, 6.1%), cardiovascular disease (n=73, 6.1%), chronic lung disease (n=50, 4.1%), chronic liver disease (n=26, 2.2%), and chronic kidney disease (n=20, 1.7%). The baseline characteristics of the patients in the training, validation, and internal testing sets are described in Multimedia Appendix 1, Table S1.

The clinical features of the 1206 patients are presented in Table 2. Cough, fever, and dyspnea were present in 53% (n=639), 51.7% (n=624), and 17.7% (n=213) of patients, respectively. At baseline assessment, mean white blood cell (WBC) count was 5024 cells/μL, with 29.4% lymphocytes. The median CRP and procalcitonin levels were 0.42 mg/dL and 0.03 ng/mL, respectively. Treatment for COVID-19 was remdesivir in 5.8% (n=70) and corticosteroids in 9.5% (n=115) of patients. Eligible patients were hospitalized for a median of 15 (IQR 11-24) days. HFNC, MV, and ECMO were used in 5.4% (n=65), 5% (n=60), and 1.5% (n=18) of patients, respectively. The clinical features of patients included in the training, validation, and internal testing sets are described in Multimedia Appendix 1, Table S2.

**Table 1.** Baseline characteristics of the patients diagnosed with COVID-19 in different data sets. *P* values were estimated using the Student 2-tailed *t* test or the Mann-Whitney *U* test for continuous variables and the Pearson chi-square test or Fisher exact test for categorical variables.

| Characteristics | Total (n=1206) | Boramae Medical Center (n=739) | Korean Imaging Cohort of COVID-19 (n=467) | *P* value |
|---|---|---|---|---|
| Age, mean (SD) | 53.4 (18.4) | 54.5 (17.6) | 51.8 (19.4) | .01 |
| Female, n (%) | 631 (52.3) | 376 (50.9) | 255 (54.6) | .23 |
| Every-day smoker, n (%) | 113 (9.4) | 84 (11.3) | 29 (6.2) | .004 |
| **Comorbidities, n (%)** | | | | |
| Hypertension | 317 (26.3) | 213 (28.8) | 104 (22.3) | .01 |
| Diabetes mellitus | 188 (15.6) | 118 (16.0) | 70 (15) | .71 |
| Cancer | 74 (6.1) | 51 (6.9) | 23 (4.9) | .20 |
| Cardiovascular disease | 73 (6.1) | 50 (6.8) | 23 (4.9) | .24 |
| Chronic lung disease | 50 (4.1) | 38 (5.1) | 12 (2.6) | .04 |
| Chronic liver disease | 26 (2.2) | 24 (3.2) | 2 (0.4) | .002 |
| Chronic kidney disease | 20 (1.7) | 13 (1.8) | 7 (1.5) | .91 |

**Table 2.** Clinical manifestations of the patients diagnosed with COVID-19 in different data sets. *P* values were estimated using the Student 2-tailed *t* test or the Mann-Whitney *U* test for continuous variables and the Pearson chi-square test or Fisher exact test for categorical variables.

| Clinical manifestations | Total (n=1206) | Boramae Medical Center (n=739) | Korean Imaging Cohort of COVID-19 (n=467) | *P* value |
|---|---|---|---|---|
| **Symptoms, n (%)** | | | | |
| Cough | 639 (53) | 417 (56.4) | 222 (47.5) | .003 |
| Fever | 624 (51.7) | 376 (50.9) | 248 (53.1) | .49 |
| Dyspnea | 213 (17.7) | 128 (17.3) | 85 (18.2) | .79 |
| **Laboratory tests** | | | | |
| White blood cells (cells/μL), mean (SD) | 5024 (1941) | 4984 (1932) | 6007 (1926) | .005 |
| Lymphocytes (%), mean (SD) | 29.4 (10.9) | 29.5 (11.4) | 29.2 (9.0) | .70 |
| C-reactive protein (mg/dL), median (IQR) | 0.42 (0.12 to 1.83) | 0.65 (0.17 to 3.14) | 0.30 (0.10 to 0.38) | <.001 |
| Procalcitonin (ng/mL), median (IQR) | 0.03 (0.01 to –0.04) | 0.03 (0.02 to –0.05) | 0.02 (0.00 to –0.04) | .02 |
| **Treatment, n (%)** | | | | |
| Remdesivir | 70 (5.8) | 58 (7.8) | 12 (2.6) | <.001 |
| Corticosteroid | 115 (9.5) | 74 (10) | 41 (8.8) | .55 |
| Length of stay (days), median (IQR) | 15 (11 to 24) | 13 (10 to 19) | 22 (15 to 30) | <.001 |
| Length of stay ≤2 weeks, n (%) | 562 (46.6) | 451 (61) | 111 (23.8) | <.001 |
| Oxygen supplementation, n (%) | 222 (18.4) | 160 (21.7) | 62 (13.3) | <.001 |
| High-flow nasal cannula, n (%) | 65 (5.4) | 55 (7.4) | 10 (2.1) | <.001 |
| Mechanical ventilator, n (%) | 60 (5) | 31 (4.2) | 29 (6.2) | .15 |
| Extracorporeal membrane oxygenator, n (%) | 18 (1.5) | 7 (0.9) | 11 (2.4) | .08 |

## Performance of the Prediction Model

The AI model was trained, validated, and internally tested using the CXR images of the hospitalized patients with COVID-19. The performance of each model in the internal testing set is described in Multimedia Appendix 1, Figure S2. The probability of each prespecified outcome (CXR score) was calculated in the external testing set (ie, KICC-19); the AUROCs were 0.602 (95% CI 0.540-0.664) for hospital LOS ≤2 weeks, 0.647 (95% CI 0.586-0.708) for oxygen supplementation, and 0.782 (95% CI 0.720-0.845) for development of ARDS. Representative heat maps visually explain how DL preferentially recognized pneumonic lesions in the CXR images (Figure 2). The heat maps used the feature maps from the clinical data channel of the model weighted by the output probabilities.

We identified clinical variables that were significantly associated with clinical outcomes in hospitalized patients with COVID-19 (Multimedia Appendix 1, Tables S3-4; Table 3). Patients with hypertension, chronic liver disease, low lymphocyte count, or corticosteroid treatment were less likely to be discharged from the hospital within 2 weeks. Patients who needed oxygen supplementation were older; had hypertension, diabetes, or dyspnea; and had a higher level of inflammatory markers, including CRP and procalcitonin. ARDS was more common in those who were older, had dyspnea, or had higher procalcitonin levels. The performance of the prediction models using significant clinical variables was evaluated in the external testing set, with AUROCs of 0.618 (95% CI 0.558-0.678) for hospital LOS ≤2 weeks, 0.567 (95% CI 0.501-0.632) for oxygen supplementation, and 0.878 (95% CI 0.835-0.920) for development of ARDS.

**Figure 2.** Representative cases in the test set database. (A) Chest radiograph of a 65-year-old woman who survived for 32 days after hospitalization. She had no cardiopulmonary comorbidities. She required oxygen supplementation but did not meet the operational definition of acute respiratory distress syndrome. The radiograph shows multiple consolidations and ground-glass opacities in both lung fields. The heat map mainly distinguishes the focal consolidative opacities of both lung fields. The image demonstrates red areas not only in the right lower and left upper lung fields but also around both shoulder joints, because lung segmentation was not applied in our model. The combined model, using chest radiography scores and clinical information, predicted a 40.9% chance of hospital length of stay ≤2 weeks, 74.5% chance of oxygen supplementation, and 33% chance of developing acute respiratory distress syndrome. (B) Chest radiograph of a 93-year-old man who died after 18 days of hospitalization. This patient had a previous history of heart disease. He required oxygen supplementation and met the operational definition for acute respiratory distress syndrome. The radiograph shows diffuse ground-glass opacities in both lung fields. The heat map mainly distinguishes the bilateral ground-glass opacities of both lung fields. The combined model, using chest radiography scores and clinical information, predicted a 57.8% chance of hospital length of stay ≤2 weeks, 96.4% chance of oxygen supplementation, and 99.1% chance of acute respiratory distress syndrome.
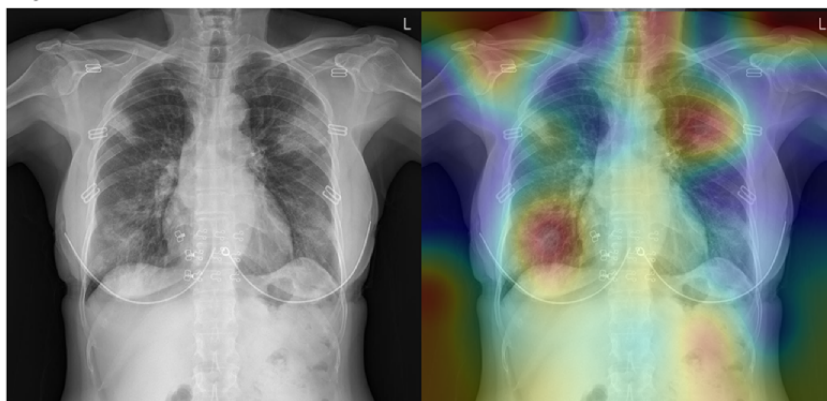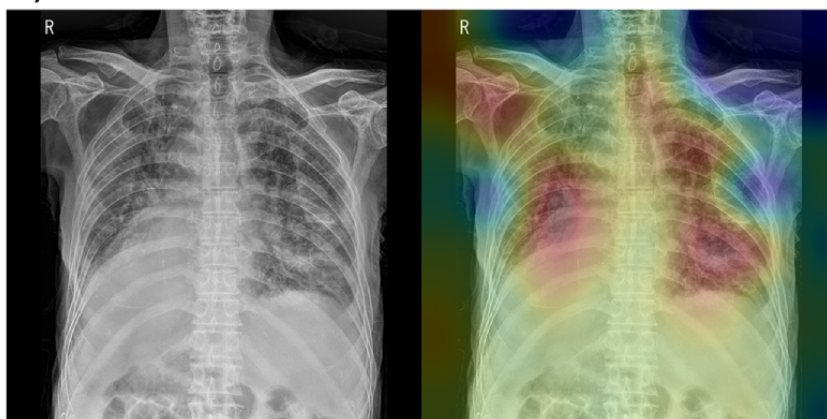
**Table 3.** Univariable and multivariable logistic regression model for each clinical outcome in COVID-19 patients hospitalized at Boramae Medical Center.

| Clinical outcomes | Unadjusted odds ratio (95% CI) | P value | Adjusted odds ratio (95% CI) | P value |
|---|---|---|---|---|
| **Hospital length of stay ≤2 weeks** | | | | |
| Age | 0.99 (0.98 to 1.00) | .01 | 1.00 (0.99 to 1.01) | .64 |
| Female | 1.30 (0.96 to 1.74) | .09 | 1.18 (0.87 to 1.61) | .29 |
| Every-day smoker | 0.84 (0.53 to 1.32) | .44 | N/A[a] | N/A |
| Hypertension | 0.57 (0.41 to 0.78) | <.001 | 0.68 (0.47 to 0.98) | .04 |
| Diabetes mellitus | 0.61 (0.41 to 0.91) | .02 | 0.79 (0.52 to 1.22) | .29 |
| Cancer | 0.91 (0.51 to 1.62) | .74 | N/A | N/A |
| Cardiovascular disease | 0.80 (0.45 to 1.43) | .46 | N/A | N/A |
| Chronic lung disease | 0.98 (0.50 to 1.91) | .95 | N/A | N/A |
| Chronic kidney disease | 0.28 (0.08 to 0.91) | .04 | 0.35 (0.11 to 1.20) | .01 |
| Chronic liver disease | 0.37 (0.16 to 0.86) | .02 | 0.40 (0.17 to 0.93) | .03 |
| Cough | 1.03 (0.77 to 1.39) | .84 | N/A | N/A |
| Fever | 1.03 (0.77 to 1.39) | .84 | N/A | N/A |
| Dyspnea | 1.17 (0.79 to 1.74) | .43 | N/A | N/A |
| White blood cells, 1000/µL | 1.04 (0.96 to 1.12) | .36 | N/A | N/A |
| Lymphocyte, % | 1.02 (1.01 to 1.04) | .001 | 1.02 (1.00 to 1.03) | .03 |
| C-reactive protein >0.5 mg/dL | 1.14 (0.84 to 1.53) | .40 | N/A | N/A |
| Procalcitonin >0.05 ng/mL | 0.82 (0.54 to 1.25) | .36 | N/A | N/A |
| **Oxygen supplementation** | | | | |
| Age | 1.08 (1.07 to 1.10) | <.001 | 1.06 (1.04 to 1.08) | <.001 |
| Female | 0.67 (0.47 to 0.96) | .03 | 0.91 (0.53 to 1.56) | .73 |
| Every-day smoker | 0.45 (0.23 to 0.90) | .02 | 0.63 (0.24 to 1.69) | .36 |
| Hypertension | 3.85 (2.67 to 5.55) | <.001 | 1.90 (1.12 to 3.24) | .02 |
| Diabetes mellitus | 4.91 (3.23 to 7.46) | <.001 | 2.23 (1.24 to 4.01) | .008 |
| Cancer | 0.99 (0.50 to 1.98) | .98 | N/A | N/A |
| Cardiovascular disease | 2.86 (1.58 to 5.17) | <.001 | 1.24 (0.51 to 3.00) | .64 |
| Chronic lung disease | 2.50 (1.27 to 4.91) | .008 | 1.17 (0.42 to 3.27) | .77 |
| Chronic kidney disease | 3.18 (1.05 to 9.59) | .04 | 2.44 (0.48 to 12.48) | .28 |
| Chronic liver disease | N/A | N/A | N/A | N/A |
| Cough | 2.35 (1.61 to 3.44) | <.001 | 1.26 (0.74 to 2.14) | .40 |
| Fever | 3.13 (2.13 to 4.58) | <.001 | 1.55 (0.91 to 2.62) | .11 |
| Dyspnea | 11.04 (7.20 to 16.94) | <.001 | 8.93 (4.85 to 16.43) | <.001 |
| White blood cells, 1000 cells/µL | 1.17 (1.07 to 1.27) | <.001 | 0.94 (0.82 to 1.08) | .38 |
| Lymphocyte, % | 0.92 (0.91 to 0.94) | <.001 | 0.98 (0.95 to 1.01) | .11 |
| C-reactive protein >0.5 mg/dL | 9.52 (6.05 to 14.98) | <.001 | 3.04 (1.68 to 5.48) | <.001 |
| Procalcitonin >0.05 ng/mL | 12.01 (7.59 to 19.01) | <.001 | 3.50 (1.87 to 6.57) | <.001 |
| **Development of acute respiratory distress syndrome** | | | | |
| Age | 1.08 (1.06 to 1.11) | <.001 | 1.07 (1.03 to 1.12) | <.001 |
| Female | 0.44 (0.25 to 0.78) | .005 | 0.56 (0.21 to 1.52) | .26 |
| Every-day smoker | 0.72 (0.28 to 1.85) | .49 | N/A | N/A |
| Hypertension | 4.29 (2.47 to 7.46) | <.001 | 1.46 (0.56 to 3.81) | .44 |

| Clinical outcomes | Unadjusted odds ratio (95% CI) | *P* value | Adjusted odds ratio (95% CI) | *P* value |
|---|---|---|---|---|
| Diabetes mellitus | 4.40 (2.50 to –7.75) | <.001 | 0.80 (0.31 to 2.10) | .65 |
| Cancer | 1.63 (0.66 to 3.99) | .29 | N/A | N/A |
| Cardiovascular disease | 3.85 (1.85 to 8.00) | <.001 | 1.42 (0.42 to 4.81) | .57 |
| Chronic lung disease | 2.34 (0.93 to 5.84) | .07 | 0.81 (0.20 to 3.31) | 0770 |
| Chronic kidney disease | 2.17 (0.47 to 10.04) | .32 | N/A | N/A |
| Chronic liver disease | 1.07 (0.24 to 4.66) | .93 | N/A | N/A |
| Cough | 2.36 (1.29 to 4.33) | .006 | 0.52 (0.18 to 1.53) | .24 |
| Fever | 5.18 (2.58 to 10.41) | <.001 | 1.55 (0.55 to 4.40) | 4.111 |
| Dyspnea | 18.13 (9.77 to 33.66) | <.001 | 12.76 (4.48 to 36.36) | <.001 |
| White blood cells, 1000 cells/μL | 1.28 (1.15 to 1.42) | <.001 | 1.04 (0.87 to 1.25) | .66 |
| Lymphocyte, % | 0.87 (0.84 to 0.90) | <.001 | 0.96 (0.91 to 1.01) | .09 |
| C-reactive protein >0.5 mg/dL | 14.78 (5.83 to 37.44) | <.001 | 2.24 (0.56 to 8.98) | .26 |
| Procalcitonin >0.05 ng/mL | 24.84 (13.23 to 46.63) | <.001 | 8.07 (2.96 to 22.00) | <.001 |

[a]N/A: not applicable.

## Comparison Between Different Prediction Models

The sensitivity, specificity, PPV, NPV, and accuracy of each model are summarized in Multimedia Appendix 1, Table S5. A comparison of the performance of the different prediction models for each clinical outcome is shown in Figure 3. We found no significant difference in the performance of the 3 prediction models in predicting hospital LOS ≤2 weeks (Multimedia Appendix 1, Figure S3). Model 3 showed an AUROC of 0.704 (95% CI 0.646-0.762) in predicting oxygen supplementation, which was significantly superior to models 1 and 2 (*P*<.001 and *P*=.02, respectively; Multimedia Appendix 1, Figure S4). Model 2 showed better performance in predicting ARDS than model 1 (*P*=.01) (Multimedia Appendix 1, Figure S5). Model 3 showed an AUROC of 0.890 (95% CI 0.853-0.928) for ARDS, which was significantly superior to model 1 (*P*=.004).

**Figure 3.** Externally validated performance of the artificial intelligence model with chest radiography score, logistic regression model with clinical information, and the combined prediction model. (A) Hospital LOS ≤2 weeks. (B) Oxygen supplementation. (C) Development of ARDS. ARDS: acute respiratory distress syndrome; AUC: area under the curve; CXR: chest radiography; LOS: length of stay.
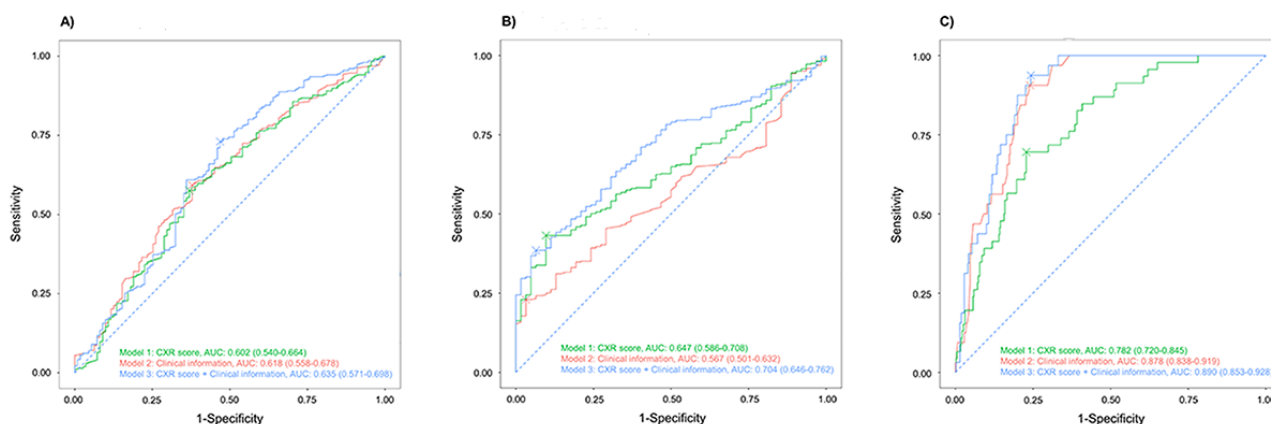


## Calibration of the Prediction Model

The calibration of models 1 and 3 in the internal and external test data sets is described in Table 4 and Multimedia Appendix 1, Figures S6 and S7. The Spiegelhalter *z* test showed good calibration of model 3 for hospital LOS ≤2 weeks, oxygen supplementation, and ARDS in the internal test set. Model 3 showed appropriate calibration only for ARDS in the external test data set (*P*=.86).

**Table 4.** Calibration of the prediction probability of the deep learning–based model using internal and external test data sets. *P* values were calculated with the Spiegelhalter *z* test; significant values indicate inappropriate calibration. Probability 1 and probability 2 were the prediction values of the deep learning–based model without and with symptom probability, respectively.

| Data set and outcome | *P* value (Spiegelhalter *z* test) |
|---|---|
| **Internal test set (n=75)** | |
|   **Hospital length of stay ≤2 weeks** | |
|     Probability 1 | .72 |
|     Probability 2 | .57 |
|   **Oxygen supplementation** | |
|     Probability 1 | .91 |
|     Probability 2 | .80 |
|   **Development of acute respiratory distress syndrome** | |
|     Probability 1 | .002 |
|     Probability 2 | .81 |
| **External test set (n=467)** | |
|   **Hospital length of stay ≤2 weeks** | |
|     Probability 1 | <.001 |
|     Probability 2 | <.001 |
|   **Oxygen supplementation** | |
|     Probability 1 | <.001 |
|     Probability 2 | <.001 |
|   **Development of acute respiratory distress syndrome** | |
|     Probability 1 | .08 |
|     Probability 2 | .86 |

## Discussion

### Principal Findings

We developed and externally validated an AI model to predict prespecified clinical outcomes based on DL using CXR. The performance of the AI model using CXR and the logistic regression model using clinical information were suboptimal for predicting hospital LOS ≤2 weeks or oxygen supplementation; there were no differences between the 2 models. The combined model, with both CXR score and clinical information, performed better in predicting oxygen supplementation. The AI prediction model for ARDS using the CXR score performed acceptably but was inferior to the model using clinical information. The combined model did not perform better in predicting ARDS than clinical information alone. CXR score calibration was appropriate for ARDS in the external test data set but not for hospital LOS ≤2 weeks or oxygen supplementation, suggesting that the CXR score may be important in identifying COVID-19 patients at high risk of progression to severe illness or ARDS. However, it is desirable to refrain from making prognoses for patients with COVID-19 based on CXR alone, considering that the predictive performance of the AI model using CXR was inferior to that of the model combining CXR score and clinical information.

CXR images in patients with COVID-19 pneumonia show various features, including diffuse ground-glass opacities, patchy reticular or nodular opacities, and consolidation [20]. The radiological features of CXR are related to the COVID-19 prognosis [21,22]. The CXR severity scoring system, which is based on radiological interpretation, is significantly associated with the prognosis of patients with COVID-19 [23]. To automate the quantification of the extent and opacity of lung lesions and the consequent assessment of radiological severity, DL models using CXR have been evaluated in COVID-19 pneumonia [24]. Recently, a DL model showed acceptable performance for predicting COVID-19 pneumonia based on CXR [12]. In our study, the CXR score derived from an AI predictive model showed superior performance for oxygen demand and comparable performance for ARDS compared to clinical information. With the application of DL techniques, CXR may need to be reconsidered as a beneficial tool for making COVID-19 pneumonia prognoses.

Hospital LOS is a clinical indicator of disease severity and time to recovery in patients with COVID-19. Prolonged hospital LOS has been associated with specific demographic characteristics and underlying comorbidities [25]. In COVID-19 patients with pneumonic infiltration in CXR, the time to negative conversion is prolonged [26]. However, none of our prediction models using CXR or clinical information were suitable for predicting whether a patient could be discharged within 2 weeks. This is consistent with a previous study reporting that radiological progression in chest CT and hospital LOS were not correlated [27]. Therefore, evidence to support an AI model

using only baseline CXR to predict early recovery from COVID-19 or hospital LOS ≤2 weeks is insufficient.

Severe illness in COVID-19 is defined as a condition with $SpO_2$ ≤94% in room air, including supplemental oxygen demand [28]. Progression to severe COVID-19 increases mortality risk [29]. Early prognosis and intervention may improve mortality risk in patients with COVID-19 [30,31]. Early administration of dexamethasone is associated with less progression to severe COVID-19 [32]. Therefore, many studies have attempted to predict severe COVID-19 using all available medical information, including CXR, but previous prediction models are not sufficiently validated [3]. Our prognostic model was externally validated to verify its performance in predicting oxygen supplementation need in patients with COVID-19. The AI prediction models using CXR and logistic regression with clinical information were suboptimal for predicting oxygen supplementation in patients with COVID-19. However, the predictive performance for oxygen supplementation improved to an acceptable level after clinical information was combined with the CXR score. Our results suggest that clinical and radiological information are complementary in predicting the need for oxygen supplementation.

Because of high mortality and morbidity, predicting COVID-19–associated ARDS is important [33]. CXR and clinical information, when applied to a time-dependent DL model to predict MV use, showed good predictive performance [12]. Recently, the Berlin definition of ARDS was expanded to include patients treated with HFNC [34]. Therefore, our study operationally defined ARDS as an event in which HFNC, MV, or ECMO were administered. Our AI model using CXR score showed acceptable performance; the combined model using CXR score and clinical information showed excellent performance in predicting the use of HFNC, MV, and ECMO. Further clinical trials are needed to ascertain whether early detection of patients at high ARDS risk can improve outcomes through early treatment for severe COVID-19.

## Limitations

Our study had some limitations. First, the interval between symptom onset and CXR imaging varied. The natural course of COVID-19 suggests that radiological abnormalities would have been ground-glass opacities at the beginning, progressing to consolidative lesions. Therefore, if the interval between symptom onset and CXR imaging was too short, our AI model might have underestimated the severity of prognoses among patients with COVID-19. Second, the performance of our prediction model may have changed according to vaccination history. As the vaccination rate increases, the progression to severe illness or ARDS decreases. Therefore, the PPV of our AI model may decrease in vaccinated patients. Third, most of the included patients were diagnosed with COVID-19 before results on the efficacy of dexamethasone or antiviral agents were reported [35,36]. Therefore, it is necessary to validate the performance of the AI prediction model under recently introduced standard treatments.

## Conclusions

A prediction model combining CXR score and clinical information was externally validated as having acceptable performance in predicting progression to severe illness and excellent performance in predicting the use of HFNC or MV in patients with COVID-19. We hypothesize that making prognoses with AI models using CXR could be applied for patients with COVID-19 in different settings.

## Conflicts of Interest

HK received consulting fees from Radisen; holds stock and stock options in MEDICALIP. Outside this study, SHY works as a chief medical officer in the MEDICAL IP.

## Multimedia Appendix 1

Supplementary tables and figures.
[DOCX File , 2536 KB-Multimedia Appendix 1]

## References

1. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int [accessed 2023-01-30]
2. Chua F, Vancheeswaran R, Draper A, Vaghela T, Knight M, Mogal R, et al. Early prognostication of COVID-19 to guide hospitalisation versus outpatient monitoring using a point-of-test risk prediction score. Thorax 2021 Jul;76(7):696-703 [FREE Full text] [doi: 10.1136/thoraxjnl-2020-216425] [Medline: 33692174]

3.   Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]

4.   Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection. American College of Radiology. URL: https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection [accessed 2023-01-30]

5.   Wang L, Zhang Y, Wang D, Tong X, Liu T, Zhang S, et al. Artificial intelligence for COVID-19: a systematic review. Front Med (Lausanne) 2021;8:704256 [FREE Full text] [doi: 10.3389/fmed.2021.704256] [Medline: 34660623]

6.   Sankaranarayanan S, Balan J, Walsh JR, Wu Y, Minnich S, Piazza A, et al. COVID-19 mortality prediction from deep learning in a large multistate electronic health record and laboratory information. J Med Internet Res 2021 Sep 28;23(9):e30157 [FREE Full text] [doi: 10.2196/30157] [Medline: 34449401]

7.   Wang YXJ, Liu W, Yang M, Chen W. The role of CT for Covid-19 patient's management remains poorly defined. Ann Transl Med 2020 Mar;8(4):145 [FREE Full text] [doi: 10.21037/atm.2020.02.71] [Medline: 32175437]

8.   Sverzellati N, Ryerson CJ, Milanese G, Renzoni EA, Volpi A, Spagnolo P, et al. Chest radiography or computed tomography for COVID-19 pneumonia? Comparative study in a simulated triage setting. Eur Respir J 2021 Sep;58(3):2004188 [FREE Full text] [doi: 10.1183/13993003.04188-2020] [Medline: 33574070]

9.   Zhu JS, Ge P, Jiang C, Zhang Y, Li X, Zhao Z, et al. Deep-learning artificial intelligence analysis of clinical variables predicts mortality in COVID-19 patients. J Am Coll Emerg Physicians Open 2020 Dec;1(6):1364-1373 [FREE Full text] [doi: 10.1002/emp2.12205] [Medline: 32838390]

10.  Kim CK, Choi JW, Jiao Z, Wang D, Wu J, Yi TY, et al. An automated COVID-19 triage pipeline using artificial intelligence based on chest radiographs and clinical data. NPJ Digit Med 2022 Jan 14;5(1):5 [FREE Full text] [doi: 10.1038/s41746-021-00546-w] [Medline: 35031687]

11.  Chamberlin JH, Aquino G, Nance S, Wortham A, Leaphart N, Paladugu N, et al. Automated diagnosis and prognosis of COVID-19 pneumonia from initial ER chest X-rays using deep learning. BMC Infect Dis 2022 Jul 21;22(1):637 [FREE Full text] [doi: 10.1186/s12879-022-07617-7] [Medline: 35864468]

12.  Jiao Z, Choi JW, Halsey K, Tran TML, Hsieh B, Wang D, et al. Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study. Lancet Digit Health 2021 May;3(5):e286-e294 [FREE Full text] [doi: 10.1016/S2589-7500(21)00039-X] [Medline: 33773969]

13.  Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 07;350:g7594 [FREE Full text] [doi: 10.1136/bmj.g7594] [Medline: 25569120]

14.  Yoon SH, Ham SY, Nam BD, Chae KJ, Lee D, Yoo JY, et al. Establishment of a Nationwide Korean Imaging Cohort of Coronavirus Disease 2019. J Korean Med Sci 2020 Nov 30;35(46):e413 [FREE Full text] [doi: 10.3346/jkms.2020.35.e413] [Medline: 33258333]

15.  Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010 Sep;5(9):1315-1316 [FREE Full text] [doi: 10.1097/JTO.0b013e3181ec173d] [Medline: 20736804]

16.  DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988 Sep;44(3):837-845. [Medline: 3203132]

17.  Liu H, Li G, Cumberland W, Wu T. Testing statistical significance of the area under a receiving operating characteristics curve for repeated measures design with bootstrapping. J Data Sci 2005;3(3):257-278 [FREE Full text] [doi: 10.6339/jds.2005.03(3).206]

18.  Tammemägi MC, Ten Haaf K, Toumazis I, Kong CY, Han SS, Jeon J, et al. Development and validation of a multivariable lung cancer risk prediction model that includes low-dose computed tomography screening results: a secondary analysis of data from the National Lung Screening Trial. JAMA Netw Open 2019 Mar 01;2(3):e190204 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0204] [Medline: 30821827]

19.  Walsh CG, Sharman K, Hripcsak G. Beyond discrimination: A comparison of calibration methods and clinical usefulness of predictive models of readmission risk. J Biomed Inform 2017 Dec;76:9-18 [FREE Full text] [doi: 10.1016/j.jbi.2017.10.008] [Medline: 29079501]

20.  Cozzi D, Albanesi M, Cavigli E, Moroni C, Bindi A, Luvarà S, et al. Chest X-ray in new Coronavirus Disease 2019 (COVID-19) infection: findings and correlation with clinical outcome. Radiol Med 2020 Aug;125(8):730-737 [FREE Full text] [doi: 10.1007/s11547-020-01232-9] [Medline: 32519256]

21.  Maroldi R, Rondi P, Agazzi GM, Ravanelli M, Borghesi A, Farina D. Which role for chest x-ray score in predicting the outcome in COVID-19 pneumonia? Eur Radiol 2021 Jun;31(6):4016-4022 [FREE Full text] [doi: 10.1007/s00330-020-07504-2] [Medline: 33263159]

22.  Ong SWX, Hui TCH, Lee YS, Haja Mohideen SM, Young BE, Tan CH, et al. High-risk chest radiographic features associated with COVID-19 disease severity. PLoS One 2021;16(1):e0245518. [doi: 10.1371/journal.pone.0245518] [Medline: 33444415]

23.  Reeves RA, Pomeranz C, Gomella AA, Gulati A, Metra B, Hage AN, et al. Performance of a severity score on admission chest radiography in predicting clinical outcomes in hospitalized patients with coronavirus disease (COVID-19). AJR Am J Roentgenol 2021 Sep;217(3):623-632. [doi: 10.2214/AJR.20.24801] [Medline: 33112201]

24.  Cohen JP, Dao L, Roth K, Morrison P, Bengio Y, Abbasi AF, et al. Predicting COVID-19 pneumonia severity on chest x-ray with deep learning. Cureus 2020 Jul 28;12(7):e9448 [FREE Full text] [doi: 10.7759/cureus.9448] [Medline: 32864270]

25.  Guo A, Lu J, Tan H, Kuang Z, Luo Y, Yang T, et al. Risk factors on admission associated with hospital length of stay in patients with COVID-19: a retrospective cohort study. Sci Rep 2021 Mar 31;11(1):7310 [FREE Full text] [doi: 10.1038/s41598-021-86853-4] [Medline: 33790365]

26.  Lee HW, Yoon S, Lee J, Park TY, Kim DK, Chung HS, et al. Clinical implication and risk factor of pneumonia development in mild coronavirus disease 2019 patients. Korean J Intern Med 2021 Jan;36(1):1-10 [FREE Full text] [doi: 10.3904/kjim.2020.329] [Medline: 32972121]

27.  Feng Z, Yu Q, Yao S, Luo L, Zhou W, Mao X, et al. Early prediction of disease progression in COVID-19 pneumonia patients with chest CT and clinical characteristics. Nat Commun 2020 Oct 02;11(1):4968 [FREE Full text] [doi: 10.1038/s41467-020-18786-x] [Medline: 33009413]

28.  IDSA Guidelines on the Treatment and Management of Patients with COVID-19. Infectious Diseases Society of America. URL: https://www.idsociety.org/COVID19guidelines [accessed 2023-01-30]

29.  Berlin DA, Gulick RM, Martinez FJ. Severe Covid-19. N Engl J Med 2020 Dec 17;383(25):2451-2460. [doi: 10.1056/NEJMcp2009575] [Medline: 32412710]

30.  Sun Q, Qiu H, Huang M, Yang Y. Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. Ann Intensive Care 2020 Mar 18;10(1):33 [FREE Full text] [doi: 10.1186/s13613-020-00650-2] [Medline: 32189136]

31.  Goyal DK, Mansab F, Iqbal A, Bhatti S. Early intervention likely improves mortality in COVID-19 infection. Clin Med (Lond) 2020 May 01;20(3):248-250 [FREE Full text] [doi: 10.7861/clinmed.2020-0214] [Medline: 32357975]

32.  Lee HW, Park J, Lee J, Park TY, Heo EY. The effect of the timing of dexamethasone administration in patients with COVID-19 pneumonia. Tuberc Respir Dis (Seoul) 2021 Jul;84(3):217-225 [FREE Full text] [doi: 10.4046/trd.2021.0009] [Medline: 34078038]

33.  Hasan SS, Capstick T, Ahmed R, Kow CS, Mazhar F, Merchant HA, et al. Mortality in COVID-19 patients with acute respiratory distress syndrome and corticosteroids use: a systematic review and meta-analysis. Expert Rev Respir Med 2020 Nov;14(11):1149-1163 [FREE Full text] [doi: 10.1080/17476348.2020.1804365] [Medline: 32734777]

34.  Matthay MA, Thompson BT, Ware LB. The Berlin definition of acute respiratory distress syndrome: should patients receiving high-flow nasal oxygen be included? Lancet Respir Med 2021 Aug;9(8):933-936 [FREE Full text] [doi: 10.1016/S2213-2600(21)00105-3] [Medline: 33915103]

35.  RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in hospitalized patients with Covid-19. N Engl J Med 2021 Mar 25;384(8):693-704 [FREE Full text] [doi: 10.1056/NEJMoa2021436] [Medline: 32678530]

36.  Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, ACTT-1 Study Group Members. Remdesivir for the treatment of Covid-19 - final report. N Engl J Med 2020 Nov 05;383(19):1813-1826 [FREE Full text] [doi: 10.1056/NEJMoa2007764] [Medline: 32445440]

## Abbreviations

**AI:** artificial intelligence
**ARDS:** acute respiratory distress syndrome
**AUROC:** area under the receiver operating characteristic curve
**BMC:** Boramae Medical Center
**CRP:** C-reactive protein
**CT:** computed tomography
**CXR:** chest radiography
**DL:** deep learning
**ECMO:** extracorporeal membrane oxygenator
**HFNC:** high-flow nasal cannula
**KICC-19:** Korean Imaging Cohort of COVID-19
**LOS:** length of stay
**MV:** mechanical ventilator
**NPV:** negative predictive value
**PPV:** positive predictive value
**RCNN:** region-based convolutional neural
**RSNA:** Radiological Society of North America

XSL•FO

**RenderX**