

# Optimal triage for COVID-19 patients under limited healthcare resources: Development of a parsimonious machine learning prediction model and threshold optimization using discrete-event simulation

Jeong Min Kim, Hwa Kyung Lim, Jae-Hyeon Ahn, Kyoung Hwa Lee, Kwang Suk Lee, Kyo Chul Koo

Submitted to: JMIR Medical Informatics  
on: August 10, 2021

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

## ***Table of Contents***

---

<b>Original Manuscript</b> .....	<b>5</b>
<b>Supplementary Files</b> .....	<b>35</b>
Figures .....	<b>36</b>
Multimedia Appendixes .....	<b>37</b>
Multimedia Appendix 1.....	<b>38</b>



# Optimal triage for COVID-19 patients under limited healthcare resources: Development of a parsimonious machine learning prediction model and threshold optimization using discrete-event simulation

Jeong Min Kim<sup>1\*</sup> BBA; Hwa Kyung Lim<sup>1\*</sup> MSc; Jae-Hyeon Ahn<sup>1</sup> PhD; Kyoung Hwa Lee<sup>2</sup> MD, MMS; Kwang Suk Lee<sup>3\*</sup> MD, MMS; Kyo Chul Koo<sup>3\*</sup> MD, PhD

<sup>1</sup>College of Business, Korea Advanced Institute of Science and Technology Seoul KR

<sup>2</sup>Division of Infectious Disease, Department of Internal Medicine, Yonsei University College of Medicine Seoul KR

<sup>3</sup>Department of Urology, Yonsei University College of Medicine Seoul KR

\* these authors contributed equally

## Corresponding Author:

Kyo Chul Koo MD, PhD

Department of Urology, Yonsei University College of Medicine

211 Eonju-ro, Gangnam-gu

Seoul

KR

## Abstract

**Background:** The coronavirus disease 2019 (COVID-19) pandemic has caused an unprecedented burden on healthcare systems.

**Objective:** To effectively triage COVID-19 patients within situations of limited data availability and to explore optimal thresholds to minimize mortality rates while maintaining the healthcare system capacity.

**Methods:** A nationwide sample of 5601 patients confirmed for COVID-19 up until April 2020 was retrospectively reviewed. XGBoost and logistic regression analysis were used to develop prediction models for the patients' maximum clinical severity during hospitalization, classified according to the WHO Ordinal Scale for Clinical Improvement (OSCI). The recursive feature elimination technique was used to evaluate the extent of the model performance's maintenance when clinical and laboratory variables are eliminated. Using populations based on hypothetical patient influx scenarios, discrete-event simulation was performed to find the optimal threshold within limited resource environments that minimizes mortality rates.

**Results:** The cross-validated area under the receiver operating characteristics (AUROC) of the baseline XGBoost model that utilized all 37 variables was 0.965 for OSCI  $\geq 6$ . Compared to the baseline model's performance, the AUROC of the feature-eliminated model that utilized 17 variables was maintained at 0.963 with statistical insignificance. Our prediction model was provided online for clinical implementation. Optimal thresholds were found to minimize mortality rates in a hypothetical patient influx scenario. The benefit of utilizing an optimal triage threshold was clear, reducing mortality up to 18.1% compared to the conventional Youden Index.

**Conclusions:** Our adaptive triage model and its threshold optimization capability reveal that COVID-19 management can be integrated using both medical and healthcare management sectors to guarantee maximum treatment efficacy.

(JMIR Preprints 10/08/2021:32726)

DOI: <https://doi.org/10.2196/preprints.32726>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

**Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

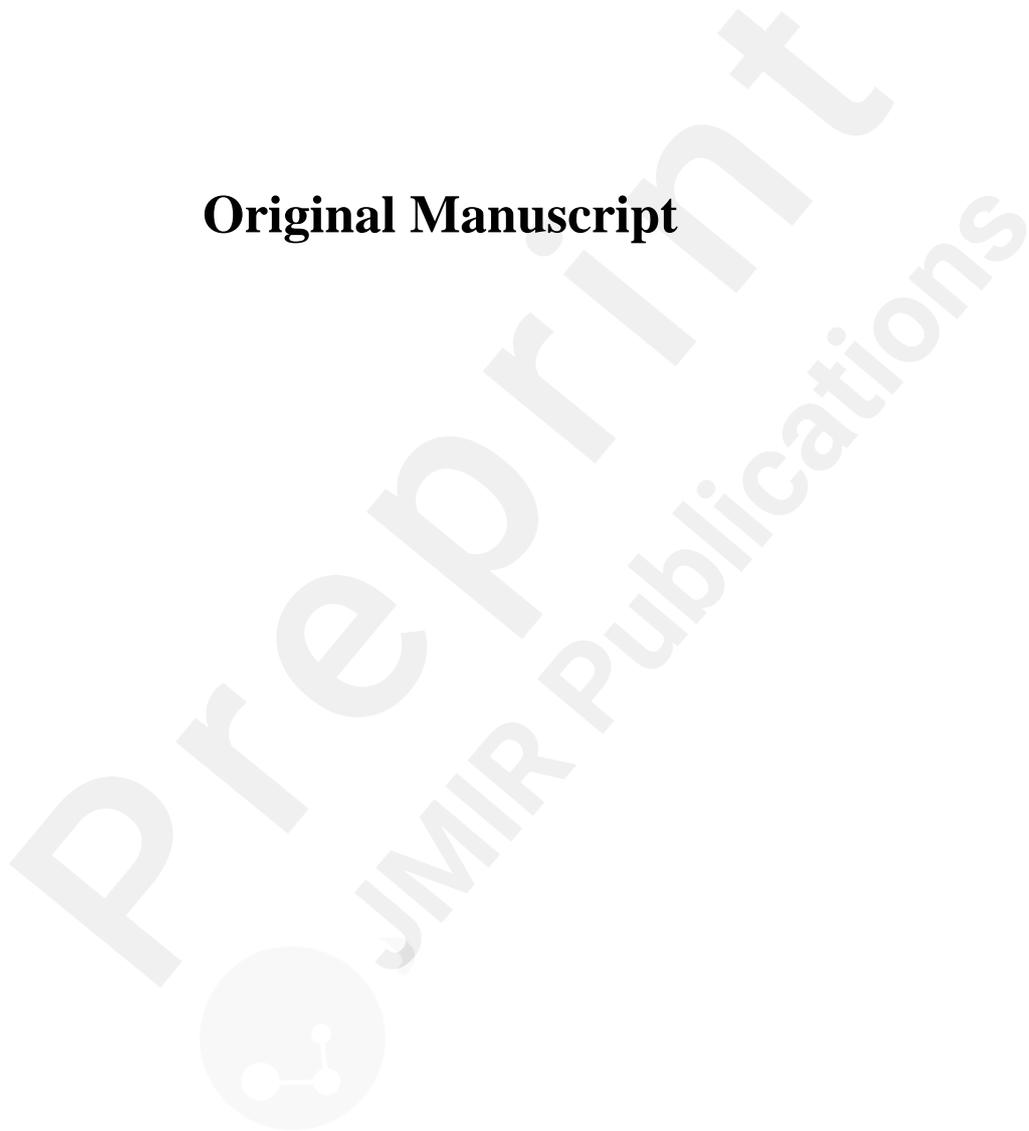
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

**Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.  
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http://www.jmir.org](#), I will be able to access the full text of my manuscript.



**Original Manuscript**



## **Optimal triage for COVID-19 patients under limited healthcare resources: Development of a parsimonious machine learning prediction model and threshold optimization using discrete-event simulation**

**Authors:** <sup>1,†</sup>Jeong Min Kim, BBA, <sup>1,†</sup>Hwa Kyung Lim, MSc, <sup>1</sup>Jae-Hyeon Ahn, PhD, <sup>2</sup>Kyoung Hwa Lee, MD, MMS, <sup>3,\*</sup>Kwang Suk Lee, MD, MMS, <sup>3,\*</sup>Kyo Chul Koo, MD, PhD

<sup>†</sup> Jeong Min Kim and Hwa Kyung Lim equally contributed as first authors

\* Kwang Suk Lee and Kyo Chul Koo equally contributed as corresponding authors

### **Affiliations**

<sup>1</sup> College of Business, Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea

<sup>2</sup> Division of Infectious Disease, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

<sup>3</sup> Department of Urology, Yonsei University College of Medicine, Seoul, Republic of Korea

### **Corresponding author**

Kyo Chul Koo, MD, PhD.

Address: Department of Urology, Yonsei University College of Medicine, 211 Eonju-ro, Gangnam-gu, Seoul, Republic of Korea, 135-720

Tel: +82-2-2019-3470

Fax: +82-2-3462-8887

E-mail: gckoo@yuhs.ac

## Abstract

**Background:** The coronavirus disease 2019 (COVID-19) pandemic has placed an unprecedented burden on healthcare systems.

**Objective:** To effectively triage COVID-19 patients within situations of limited data availability and to explore optimal thresholds to minimize mortality rates while maintaining healthcare system capacity.

**Methods:** A nationwide sample of 5,601 patients confirmed for COVID-19 up until April 2020 was retrospectively reviewed. XGBoost and logistic regression analysis were used to develop prediction models for the maximum clinical severity during hospitalization, classified according to the WHO Ordinal Scale for Clinical Improvement (OSCI). The recursive feature elimination technique was used to evaluate maintenance of the model performance when clinical and laboratory variables were eliminated. Using populations based on hypothetical patient influx scenarios, discrete-event simulation was performed to find an optimal threshold within limited resource environments that minimizes mortality rates.

**Results:** The cross-validated area under the receiver operating characteristics (AUROC) of the baseline XGBoost model that utilized all 37 variables was 0.965 for  $OSCI \geq 6$ . Compared to the

baseline model's performance, the AUROC of the feature-eliminated model that utilized 17 variables was maintained at 0.963 with statistical insignificance. Optimal thresholds were found to minimize mortality rates in a hypothetical patient influx scenario. The benefit of utilizing an optimal triage threshold was clear, reducing mortality up to 18.1%, compared to the conventional Youden Index.

**Conclusions:** Our adaptive triage model and its threshold optimization capability revealed that COVID-19 management can be achieved via the cooperation of both medical and healthcare management sectors for maximum treatment efficacy. The model is available online for clinical implementation.

**Keywords:** COVID-19; decision support techniques; machine learning; prediction; triage

## Introduction

The high incidences of infection, critical illness, and mortality due to coronavirus disease 2019 (COVID-19) have placed unprecedented burdens on international healthcare systems. In response, the World Health Organization (WHO) guidelines have recommended that all countries should prepare for infection surges in their healthcare facilities and implement appropriate triage protocols [1]. Unfortunately, these guidelines fail to provide a one-size-fits-all approach that works for individual regions while accounting for unique outbreak surges.

Numerous prognostic models have been developed to ensure effective triage for COVID-19

patients [2-7]. While these models exhibit modest predictive accuracy, their generalizability has been questioned due to their confinement to single clinical outcome measures and reductions in their discrimination performance when using insufficient data. Most importantly, the classification thresholds of these prediction models, which are crucial for ensuring effective resource utilization by healthcare systems, have been neglected, thereby limiting their practicality. To overcome these models' shortcomings, combining multi-institutional data with advanced prediction models, such as those using machine learning and simulation modeling, is needed.

COVID-19 is associated with significant disruptions to most healthcare infrastructures. Therefore, an adjustable risk stratification model that considers various regions' resource availability, as well as one that identifies patients who will likely require hospitalization and intensive care, will help to reduce these systems' burdens. In this study, we propose an adaptive triage model that takes into account deficits in established healthcare resources due to the COVID-19 pandemic. The main contributions of our study are as follows: First is a powerful and interpretable prediction model using extreme gradient boosting (XGBoost) and Shapley additive explanations (SHAP) that provides accurate prognoses to facilitate preemptive treatments, thereby ensuring improvements in patient survival outcomes. Second is the ability to apply the model using readily available assessment parameters using the recursive feature elimination (RFE) technique, thereby maintaining its reliability in data-limited environments [8,9]. Third is a consideration of resource availability at either a facility or national level relative to varying patient influx volumes by employing the discrete-event simulation (DES) technique.

Our study objectives were three-fold. First, we sought to develop a baseline prediction model with an explanatory feature for triaging COVID-19 patients. Second, based on this model, we aimed to utilize the RFE technique to develop feature-eliminated models that would help ensure efficient resource utilization under limited data availability. Finally, we set out to develop an adaptive triage model using the DES technique to assist in efficient resource utilization under limited

healthcare resources.

## Methods

### Ethics statement

This study was approved by an institutional ethics committee (2020-0883-001) and the Korea Disease Control and Prevention Agency (KDCA) epidemiological survey and analysis committee (20201120\_4a). All study procedures complied with the 1946 Declaration of Helsinki and its 2008 update.

### Patient cohort

Demographic, clinical, laboratory, and disease outcome records of the 5,628 patients, confirmed by a real-time reverse transcription-polymerase chain reaction using nasopharyngeal/oropharyngeal swab or sputum specimens for severe acute respiratory syndrome coronavirus 2 up until April 2020 were retrospectively retrieved. The data were collected and comprehensively managed by the KDCA. Among 10,774 patients consecutively diagnosed with COVID-19 within this time frame, data on 52.2% (5,628/10,774) of the patient population were publicized for research purposes after excluding patients with any missing data. The database did not account for the location of diagnosis within Korea. The database included patients who had been treated and released from quarantine or hospitalization, as well as those who died from COVID-19 sequela. The criteria for patient release included obtaining two consecutive negative results at least 24 hours apart and asymptomatic status. Among 5,628 patients, 27 patients with missing clinical severity data were excluded, resulting in a final development cohort of 5,601 patients.

## Covariates and outcome definitions

Baseline data collected at each patient's diagnosis were used for model development. Demographic data included patient age, sex, systolic and diastolic blood pressure, heart rate, body temperature, and body mass index (BMI). Medical comorbidities included hypertension, diabetes mellitus, heart failure, cardiovascular disease, asthma, chronic kidney disease, chronic obstructive pulmonary disease, chronic liver disease, autoimmune disease, dementia, malignancy, and pregnancy. The clinical findings included a history of fever ( $\geq 37.5^{\circ}\text{C}$ ), cough, sputum production, myalgia, fatigue, sore throat, rhinorrhea, dyspnea, vomiting, nausea, diarrhea, headache, and altered consciousness. Laboratory data included hemoglobin, hematocrit, white blood cell count, %leukocyte, and platelet count. The patients' maximum clinical severity during quarantine or hospitalization was classified according to the WHO Ordinal Scale for Clinical Improvement (OSCI) [10].

## Statistical analysis

### *Model development*

Multivariate logistic regression (LR) and XGBoost were used to select the best-performing prediction model using all available clinical and laboratory data [11]. The models were developed and cross-validated using 5,037 (89.9%) patient data and were then revalidated using a hold-out cohort of 564 (10.1%) patients. Performance metrics were calculated using 10-fold cross-validation to avoid any overfitting. Model development was performed using the *caret* package in R Model. The best-performing model derived from XGBoost was defined as Model 1 and was used as a baseline model for RFE.

### *Variable elimination*

The RFE technique was used to evaluate the extent of the model performance's maintenance when various predictors were eliminated. RFE was performed for two models that incorporated all clinical data with and without laboratory data: Model 1 (clinical data with laboratory data) and Model 2 (clinical data without laboratory data). SHAP was used to rank each variable based on its significance to the models for its desirable properties, including local accuracy, missingness, and consistency [12]. At each RFE iteration, the lowest-ranked feature was eliminated, the model was refitted, and its performance was assessed using 10-fold cross-validation. The feature-eliminated models (Model 3: limited clinical data with laboratory data and Model 4: limited clinical data without laboratory data) were then selected at a point wherein the number of features was minimized while differences in area under receiver operating characteristic (AUROC) curve values remained statistically insignificant. The four classification models were revalidated with the hold-out cohort to avoid any overfitting. Analysis was performed using *caret* and the *SHAPforXGBOOST* package in R.

### *Model interpretation and comparison*

To interpret Model 1, we used SHAP as it provides visible post-hoc interpretability to black-box machine learning models [12]. Patient-specific plots were created by aggregating the SHAP score of each variable for a specific prediction.

Hyperparameters of the XGBoost algorithm were optimized to maximize its AUROC values using a simple grid search with 10-fold cross-validation. Accuracy, AUROC, sensitivity, positive predicted value (PPV), and negative predicted value (NPV) were calculated at a 90% specificity using the *pROC* package in R. Confidence intervals (CI) of the performance measures were then calculated using a stratified bootstrap method with 2,000 replicates.

## Threshold optimization

### Discrete-event simulation and patient influx generation

The DES technique replicates complex behaviors and interactions among individuals, populations, and their environments. Therefore, it has been widely used to form more effective clinical decisions to minimize mortality rates under medical resource constraints [13]. Therefore, we applied DES to identify the optimal threshold within limited medical resource environments that minimizes mortality

rates, as calculated by  $\frac{n(\text{total deaths})}{n(\text{total patients})}$ , using the *simmer* R package.

First, we ran a simulation using different COVID-19 historical epidemic patient influx scenarios (H1, H2, H3, and H4) that were observed between February 2020 and February 2021 (Supplementary Fig. 1) [14]. Second, hypothetical patient influx scenarios were created using the susceptible-infectious-recovered (SIR) model for the spread of disease [15]. The total population calculated was fixed at 60,000, considering that the largest historical influx observed in South Korea was H4 (58,654 cumulative patients). We defined initial conditions at time  $t = 0$ ,  $S(0)$ ,  $I(0)$ , and  $R(0)$  and  $I(0)$  and  $R(0)$  were fixed at 6 and 0, respectively. The recovery rate  $\gamma$  was set at 0.05 because the average COVID-19 recovery time was 20.1 days [16]. The transmission rate  $\beta$  ranged between 0.75 and 5 when generating influxes with different  $R_0$  (basic reproduction rate) levels. The number of newly confirmed patients per day was obtained from the SIR modeling data (Supplementary Fig. 2).

### Probability generation

Out-of-fold prediction results of the 10-fold cross-validation were aggregated to generate an empirical probability distribution of the disease severity probability. We used the results of Model 3 because of its high performance and its potential use in instances of limited diagnostic tools. Inverse transformation sampling was performed on the empirical probability distribution function, which was

approximated using Gaussian kernel density estimation and linear interpolation [17]. The process was performed separately for severe and non-severe patients, with sampled probabilities being randomly matched with generated patient influx rates while maintaining the prevalence of severe patients. The prediction probability distribution of the out-of-fold samples and the generated prediction probability distribution are presented in Supplementary Fig. 3.

### **Simulation scenarios**

Patients with a severe disease probability above the threshold are directed to the intensive care unit (ICU), with admission to this unit then being dependent on its current capacity. Rejected patients are then directed to the general ward along with those who have a severe disease probability below the threshold. The probability of severe disease patients dying while in the ICU was 0.507, while it was 0.990 for those outside of the ICU [18]. We assumed that non-severe patients would survive regardless of ICU admission. Patient deaths were categorized as follows: resource-independent deaths, wherein severe patients expired despite ICU care (Type I); resource-dependent deaths in which severe patients died due to ICU unavailability (Type II); and threshold-dependent deaths wherein severe patients expired after being incorrectly classified as “non-severe” and were subsequently directed to the general ward (Type III).

The maximum capacity of the ICU was established as 504 beds based on the number of isolation beds under negative pressure [14]. To estimate the distribution of length of stay, we used a previously suggested gamma distribution with a shape parameter of 1.5488 and a rate parameter of 0.1331 for those who expired and with a shape parameter of 0.8904 and a rate parameter of 0.0477 for those who survived to approximate the median and interquartile range [18,19]. Simulations were repeated 20 times for each influx scenario to ensure robustness.

## Results

### Patient characteristics

Descriptive characteristics of the training and hold-out cohorts are provided in Table 1. A total of 5,330 (95.2%) patients exhibited non-severe disease symptoms with an OSCI value <6, while 271 (4.8%) exhibited severe disease symptoms with an OSCI value  $\geq 6$ .

**Table 1. Demographics and clinical characteristics.**

Variables	Total cohort (n = 5601)		Train cohort (n = 5037)		Holdout cohort (n = 564)		P
	M (SD) / n (%)	Missing (%)	M (SD) / n (%)	Missing (%)	M (SD) / n (%)	Missing (%)	
Age (years)		0.0%		0.0%		0.0%	0.407
0 - 9	66 (1.2%)		61 (1.2%)		5 (0.9%)		
10 - 19	205 (3.7%)		185 (3.7%)		20 (3.6%)		
20 - 29	1110 (19.8%)		988 (19.6%)		122 (21.6%)		
30 - 39	564 (10.1%)		513 (10.2%)		51 (9.0%)		
40 - 49	739 (13.2%)		652 (12.9%)		87 (15.4%)		
50 - 59	1141 (20.4%)		1039 (20.6%)		102 (18.1%)		
60 - 69	907 (16.2%)		809 (16.1%)		98 (17.4%)		
70 - 79	545 (9.7%)		495 (9.8%)		50 (8.9%)		
$\geq 80$	324 (5.8%)		295 (5.9%)		29 (5.1%)		
Sex (male)	2310 (41.2%)	0.0%	2073 (41.2%)	0.0%	237 (42.0%)	0.0%	0.726
Pregnancy	19 (0.3%)	0.4%	17 (0.3%)	0.3%	2 (0.4%)	0.5%	1.000
Pregnancy weeks	0.05 (1.1)	0.4%	0.06 (1.1)	0.4%	0.03 (0.5)	0.5%	0.399
Body mass index		21.4%		21.5%		20.9%	0.653
< 18.5	259 (4.6%)		236 (4.7%)		23 (4.1%)		
18.5 - 22.9	1854 (33.1%)		1666 (33.1%)		188 (33.3%)		
23.0 - 24.9	1035 (18.5%)		929 (18.4%)		106 (18.8%)		
25.0 - 29.9	1045 (18.7%)		938 (18.6%)		107 (19.0%)		
$\geq 30$	207 (3.7%)		185 (3.7%)		22 (3.9%)		
SBP		2.5%		2.5%		2.7%	0.600
< 120	1306 (23.3%)		1177 (23.4%)		129 (22.9%)		
120 - 129	1138 (20.3%)		1012 (20.1%)		126 (22.3%)		
130 - 139	1084 (19.4%)		977 (19.4%)		107 (19.0%)		
140 - 159	1418 (25.3%)		1281 (25.4%)		137 (24.3%)		
$\geq 160$	513 (9.2%)		463 (9.2%)		50 (8.9%)		
DBP		2.5%		2.5%		2.7%	0.036
< 80	2102 (37.5%)		1878 (37.3%)		224 (39.7%)		
80 - 89	1797 (32.1%)		1601 (31.8%)		196 (34.8%)		
90 - 99	1056 (18.9%)		971 (19.3%)		85 (15.1%)		
$\geq 100$	504 (9.0%)		460 (9.1%)		44 (7.8%)		
Heart rate	85.8 (15.1)	2.3%	85.8 (15.0)	2.3%	86.3 (15.4)	2.5%	0.474
Body temperature	36.9 (0.6)	0.7%	36.9 (0.6)	0.8%	37.0 (0.6)	0.7%	0.861
Fever	1302 (23.3%)	0.1%	1168 (23.2%)	0.1%	134 (23.8%)	0.0%	0.801
Cough	2331 (41.6%)	0.1%	2103 (41.8%)	0.1%	228 (40.4%)	0.0%	0.575
Sputum	1611 (28.8%)	0.1%	1460 (29.0%)	0.1%	151 (26.8%)	0.0%	0.293
Sore throat	872 (15.6%)	0.1%	779 (15.5%)	0.1%	93 (16.5%)	0.0%	0.565
Rhinorrhea	617 (11%)	0.1%	560 (11.1%)	0.1%	57 (10.1%)	0.0%	0.511
Myalgia	920 (16.4%)	0.1%	820 (16.3%)	0.1%	100 (17.7%)	0.0%	0.411
Fatigue	233 (4.2%)	0.1%	207 (4.1%)	0.1%	26 (4.6%)	0.0%	0.650
Shortness of breath	665 (11.9%)	0.1%	608 (12.1%)	0.1%	57 (10.1%)	0.0%	0.194
Headache	963 (17.2%)	0.1%	873 (17.3%)	0.1%	90 (16.0%)	0.0%	0.446
Altered consciousness	35 (0.6%)	0.1%	31 (0.6%)	0.1%	4 (0.7%)	0.0%	0.776
Vomiting	244 (4.4%)	0.1%	210 (4.2%)	0.1%	34 (6.0%)	0.0%	0.052
Diarrhea	516 (9.2%)	0.1%	457 (9.1%)	0.1%	59 (10.5%)	0.0%	0.315
Diabetes mellitus	688 (12.3%)	0.1%	620 (12.3%)	0.1%	68 (12.1%)	0.0%	0.916
Hypertension	1198 (21.4%)	0.1%	1087 (21.6%)	0.1%	111 (19.7%)	0.0%	0.323
Heart failure	59 (1.1%)	0.1%	52 (1.0%)	0.1%	7 (1.2%)	0.0%	0.808
Cardiovascular disease	179 (3.2%)	0.3%	156 (3.1%)	0.3%	23 (4.1%)	0.4%	0.259
Asthma	128 (2.3%)	0.1%	118 (2.3%)	0.1%	10 (1.8%)	0.0%	0.478

COPD	40 (0.7%)	0.1%	38 (0.8%)	0.1%	2 (0.4%)	0.0%	0.428
Chronic kidney disease	55 (1.0%)	0.1%	48 (1.0%)	0.1%	7 (1.2%)	0.0%	0.665
Malignancy	145 (2.6%)	0.1%	134 (2.7%)	0.1%	11 (2.0%)	0.0%	0.386
Chronic liver disease	83 (1.6%)	5.8%	75 (1.6%)	5.7%	8 (1.5%)	6.7%	1.000
Autoimmune disease	38 (0.7%)	5.9%	32 (0.7%)	5.8%	6 (1.1%)	6.9%	0.365
Dementia	224 (4.2%)	5.9%	203 (4.3%)	5.8%	21 (3.7%)	6.7%	0.811
Hemoglobin	13.3 (1.8)	27.2%	13.29 (1.8)	26.7%	13.2 (1.8)	31.6%	0.412
Hematocrit	39.2 (5.0)	27.2%	39.3 (4.9)	26.7%	39.1 (5.2)	31.7%	0.558
Lymphocyte	29.2 (11.7)	27.6%	29.3 (11.7)	27.1%	28.2 (11.0)	32.1%	0.088
Platelet	236697 (82897)	27.1%	236776 (82534)	26.7%	235943 (86395)	31.4%	0.856
White blood cell	6126 (2824)	27.1%	6121 (2841)	26.7%	6167 (2666)	31.4%	0.748
WHO OSCI $\geq 6$	271 (4.8%)	0.0%	242 (4.8%)	0.0%	29 (5.1%)	0.0%	0.802

Data are presented as a mean (standard deviation) or number (%). Differences between groups were analyzed using Welch's *t*-test for continuous variables, the Mann-Whitney U test for ordinal variables, the Chi-square test for categorical variables with frequencies above 5, and Fisher's exact test for categorical variables with frequencies below 5. Two-sided *P*-values are reported. COPD = chronic obstructive pulmonary disease, DBP = diastolic blood pressure, M = mean, SBP = systolic blood pressure, SD = standard deviation, WHO OSCI = World Health Organization Ordinal Scale for Clinical Improvement.

## Model performance

The cross-validated AUROC values of the XGBoost and LR models were 0.965 (95% CI: 0.958–0.972) and 0.938 (95% CI: 0.911–0.959), respectively (*P*=.039). We chose the XGBoost model as our baseline Model 1 since it outperformed the LR across all performance measures. Regarding AUROC, we also examined XGBoost's outperformance across four different severity endpoints (Supplementary Table 1). An online clinical decision-support system based on Model 3 is provided for clinical implementation at <http://covid19severity.duckdns.org>.

## Model interpretability

According to SHAP, age and lymphocyte count were the most important risk factors for predicting disease severity of OSCI  $\geq 6$  (Fig. 1). Patient age, lymphocyte count, platelet count, BMI, hematocrit, and heart rate all exhibited non-linear influences in predicting disease severity (Fig. 2). In addition to the overall impact of each feature on the model's output, SHAP provides patient-specific influences of each variable on their predicted disease severity (Supplementary Fig. 4).

## Predictive performance under limited data availability

An AUROC of 0.965 (95% CI: 0.958–0.972) was obtained with Model 1, which included all 37 variables. Notably, a reduction in its performance was found to be insignificant when 20 variables were eliminated, resulting in Model 3 (Supplementary Fig. 5; Supplementary Table 2). Model 1 achieved both a sensitivity and specificity greater than 90%. Model 3 achieved a sensitivity of 88% and a PPV of 31% at the specificity level of 90%. Model 3 still outperformed the LR model regarding all performance measures.

Model 2 obtained an AUROC of 0.946 (95% CI: 0.936–0.956), which included 32 variables. The reduction in performance was found to be insignificant when 21 variables were eliminated, resulting in Model 4 (Supplementary Fig. 6; Supplementary Table 2). Models 2 and 4 achieved sensitivities of 84% and 81% at a fixed specificity level of 90%, respectively (Table 2). Significant differences in AUROCs were observed when laboratory variables were excluded in these models, which implied that the laboratory variables had a solid discriminative power (all  $P \leq .01$ ).

**Table 2. Comparison of model performances.**

Model	Number of variables	AUROC	Specificity	Sensitivity	Accuracy	PPV	NPV
1	37	0.965 (0.958-0.972)	0.900 (0.892-0.909)	0.905 (0.868-0.942)	0.900 (0.892-0.908)	0.314 (0.295-0.335)	0.995 (0.993-0.997)
2	32	0.946 (0.936-0.956)	0.900 (0.891-0.908)	0.839 (0.793-0.884)	0.897 (0.888-0.905)	0.297 (0.276-0.319)	0.991 (0.988-0.994)
3	17	0.963 (0.955-0.971)	0.900 (0.892-0.908)	0.884 (0.839-0.921)	0.899 (0.891-0.907)	0.309 (0.289-0.329)	0.994 (0.991-0.996)
4	11	0.942 (0.931-0.953)	0.901 (0.892-0.909)	0.810 (0.756-0.860)	0.896 (0.888-0.904)	0.291 (0.270-0.313)	0.989 (0.987-0.992)

AUROC = area under the receiver operating characteristic, NPV = negative predictive value, PPV = positive predictive value

The AUROCs of Models 1 and 2 for the held-out cohorts were 0.958 (95% CI: 0.924–0.991) and 0.943 (95% CI: 0.901–0.985), respectively, which were both indifferent from the cross-validation results ( $P=.663$  and  $P=.889$ , respectively). The AUROCs of Models 3 and 4 for the held-out cohorts

were 0.949 (95% CI: 0.906–0.990) and 0.941 (95% CI: 0.903–0.978), respectively, and were also indifferent from the cross-validation results ( $P=.541$  and  $P=.950$ , respectively). The indifferences between the cross-validation and hold-out results revealed that all models had a degree of generalizability to unseen data (Supplementary Table 3). Detailed results and the selected variables used at each step of the RFE are presented in Supplementary Tables 2 and 4.

### **Optimal triage under limited resource availability**

The overall DES workflow is illustrated in Fig. 3. Mortality rates were minimized at thresholds of 0.1, 0.01, 0.04, and 0.24 for H1, H2, H3, and H4, respectively (Supplementary Fig. 7). The mortality rates showed a convex shape in accordance with these thresholds (Supplementary Fig. 8).

We can infer that as the death rate increases, the threshold should be raised when a large increase is accompanied. While the association between mortality rates and triage thresholds across various patient influx scenarios is inferable through an analysis of historical influx data, it is impractical to draw general conclusions from this information. For example, looking at Supplementary Fig. 7, an upward trend in the optimal threshold and optimized mortality rate occurred when comparing H2, H3, and H4, wherein there was a clear increase in the patient influx volume. However, it is difficult to infer this information when comparing H1 with H3 or H4 because of differences in their multi-dimensional characteristics, including duration, maximum daily patients, and cumulative patients. To further support our results, we performed additional simulations using patient flow data that were generated using the SIR model with varying  $R_0$ s.

The DES using hypothetical patient influxes revealed that the optimal threshold ranged from 0.02 to 0.66, while the respective minimized mortality rates ranged from 0.017 (1.7%) to 0.042 (4.2%) (Supplementary Fig. 9). The optimal threshold values and minimized mortality rates for each  $R_0$  showed that a larger  $R_0$  value tends to result in increases in both of these variables. The optimal

threshold is increased along with the R0 values to increase precision for severe patients while fully utilizing the ICU. The optimized mortality rates were increased due to an increased proportion of deaths outside the ICU resulting from a larger volume of patient influx. The benefits of utilizing an optimal triage threshold were clear when compared with the conventional Youden Index (J-index) as a benchmark value, which was 0.013. Decreased mortality rates (=

$\frac{J\text{-index mortality rate} - \text{optimized mortality rate}}{J\text{-index mortality rate}}$ ) were notably large in a magnitude ranging from

6.1% to 18.1% (Fig. 4). Detailed data are listed in Table 3.

**Table 3.** Optimized threshold and its benefits on mortality outcomes according to patient influx settings.

Influx	Optimal threshold	Optimized mortality rate	Decreased mortality rate
H1	0.1	0.022	0.298
H2	0.01	0.015	0.047
H3	0.04	0.019	0.146
H4	0.24	0.031	0.209
R0 = 1.5	0.02	0.017	0.061
R0 = 2	0.16	0.025	0.179
R0 = 4	0.39	0.032	0.181
R0 = 6	0.43	0.041	0.068
R0 = 8	0.62	0.042	0.071
R0 = 10	0.66	0.042	0.069

Decreased mortality rate = (J index mortality rate - optimized mortality rate)/J index mortality rate, H = historical epidemic patient influx scenario, R0 = basic reproduction rate.

We observed a convex relationship for mortality rates in accordance with the thresholds in Fig. 5. The mortality rate was minimized at a point where type I deaths, which had the lowest  $P_{death}$  (50.7%), was maximized in proportion to total death. For example, when R0 was 1.5, the proportion of type I deaths was maximized at the optimal threshold, accounting for 66.4% of all deaths. However, a threshold that is too low leads to inadequate capacity exhaustion with misclassified non-severe patients. Consequently, the resulting limited capacity for actual severe patients then

decreases the proportion of type I deaths and increases those of type II deaths. Conversely, a threshold that is too high would result in unnecessary rejection for severe patients, which then decreases the proportion of type I deaths and increases those of type III deaths.

In situations of excessively high  $R_0$  values and increased ICU demands, increasing the triage threshold to reject more patients will still deplete the ICU capacity. Therefore, adjusting the threshold will mostly result in trade-offs between the number of threshold- and capacity-dependent rejections, limiting the influence of threshold adjustment on minimizing patient mortality. In situations of sufficiently low  $R_0$  values, the effect of threshold optimization is reduced along with its necessity. Nonetheless, the large reduction in mortality rates among the remaining influxes highlights the substantial benefits of optimizing the patient triage threshold under resource constraints.

## Discussion

A distinctive feature of our Model 1 is its high discriminative power with an AUROC that exceeded 0.97 in both cross-validation and hold-out settings. Previous prediction models for determining the clinical deterioration of COVID-19 patients have reported predictive accuracies ranging from 0.77 to 0.91 [2-5]. Additionally, these models require specific diagnostic data, including laboratory data, peripheral oxygen saturation, or radiographic findings to maintain their predictive accuracies. Moreover, to what extent the performance abilities of these models are maintained during the partial absence of data has not been studied. Given this unmet clinical need, we have developed Model 1. In addition, we confirmed that our feature-eliminated models maintained an adequate discriminative power even in the partial absence of data. The advantages of our feature-eliminated models include not only their increased generalizability to unseen data, but also their applicability within scenarios wherein there is limited medical data. We have uploaded

Model 3 online to be implemented in clinical practice. Given the acute exacerbation of pneumonia in COVID-19 patients, our model can also be used to re-evaluate hospitalized patients in the short-term, so that individuals whose clinical manifestations are likely to worsen can be identified as early as possible [20].

A noteworthy feature of our model is its ability to discriminate between patient-specific factors contributing to disease exacerbation and their individual contributions using SHAP values. Current COVID-19 treatment guidelines provide recommendations based on the average-risk patient under limited available insights into their disease stage [10]. These recommendations provide a one-size-fits-all approach to all patients, which is problematic for those with more complex or atypical disease presentations. Our model obviates the need for arbitrary patient risk groupings and is, therefore, useful in maximizing survival odds based on individual risk stratification. Furthermore, our models can be integrated into electronic medical record systems, which utilize coding algorithms, as a notification system that helps in the early identification of disease exacerbation risk factors.

The validity of our model is supported by the high consistency between the results of its interpretation using SHAP and previously reported prognosticators of COVID-19 severity [21-26]. We noted that old age, followed by lymphopenia and thrombocytopenia, exhibited the highest Shapley values for disease exacerbation. We presume that age interacts with relevant features in older adults, including poor functional performance and increased frailty, which are associated with adverse outcomes and increased mortality among patients with respiratory syndromes [27]. Our findings also support literature indicating that lymphopenia plays an important role in COVID-19 exacerbation [23-26]. Lymphopenia is characterized by the lowering of lymphocytes due to injured alveolar epithelial cells and is commonly observed in COVID-19 patients [28]. Consistent with previous studies, thrombocytopenia was also found to be associated with adverse COVID-19 outcomes [24,29]. It has been suggested that a reduction or morphological alternation in the pulmonary capillary bed exerts pathological platelet defragmentation because the lungs are a platelet

release site with mature megakaryocytes [30]. Our prediction model supports the notion that early identification of COVID-19 infection, before a hematological crisis occurs, is necessary for ensuring a better prognosis.

There is no existing study that has examined COVID-19 severity prediction models in an attempt to provide an explicit solution for the delivery of optimal triage using threshold modification that accounts for limited resource availability. We employed DES in our Model 3 to examine discrimination thresholds that are usable in an adaptive manner across various patient influx scenarios and the related healthcare resource availability. Our simulations revealed that applying the optimal thresholds of both historical and generated patient influxes will minimize the mortality rate of each patient influx scenario. Our hypothesis is supported by the significant differences found in mortality rates between the J-index and our optimized thresholds when applied to the expected patient influx volumes. This observation supports the potential usability of our model to substantially reduce COVID-19 mortality rates through appropriate and effective adjusting of triage thresholds.

## Limitations

One limitation of our study was its incorporation of a single, national cohort of Asian ethnicity with a relatively small sample size, which impacts the generalizability of our findings. An external validation using a more multi-ethnic population is thus needed to determine if a similar discrimination performance occurs among other ethnic groups. However, to ensure our model's robustness, we implemented 10-fold cross-validation with additional confirmation using a hold-out cohort. Second, the triage threshold was evaluated using a simulation. Simulations do not yield concrete answers, nor are they able to assess all kinds of potential situations [31]. Third, the applicability of utilizing SHAP values to discriminate patient-specific contributing factors for disease exacerbation has not been prospectively validated. Whether early identification of disease

exacerbation risk factors and their individual contributions would turn into a better prognosis would need to be validated after the implementation of our online system into clinical practice. Lastly, clinical data, including self-reported measurements, may not be objectively interpreted, and models utilizing these parameters should be interpreted cautiously.

## Conclusions

We developed and validated a robust prediction model, with an explanatory feature, that offers an effective means of enhancing the efficiency of COVID-19 triage. We further proposed an adaptive triage model that utilizes both patient influx volume and the capacity of a healthcare system to minimize mortality rates within the scope of their resource limits. Our model has the potential for effective application because it is available online for patients and providers in both inpatient and outpatient settings. Overall, our results imply that COVID-19 treatment plans need to integrate both medical and healthcare management expertise to guarantee maximum efficacy.

## Acknowledgments

We would like to thank the Korea Disease Control and Prevention Agency (KDCA) and the Health Information Managers in the participating hospitals for their assistance in collecting the medical records utilized in this study. This study was supported through the Infection Prevention Strategy Development Program of Korea (HW20C2103).

## Conflicts of interest

All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare that all authors had financial support from the Infection Prevention

Strategy Development Program of Korea for the submitted work; that there were no financial relationships with any organizations that might have an interest in the submitted work within the previous 3 years; and that there were no other relationships or activities that could appear to have influenced the submitted work.

## Abbreviations

AUROC = area under the receiver operating characteristic

BMI = body mass index

CI = confidence interval

COVID-19 = coronavirus disease 2019

DES = discrete-event simulation

ICU = intensive care unit

LR = logistic regression

NPV = negative predicted value

OSCI = Ordinal Scale for Clinical Improvement

PPV = positive predicted value

RFE = recursive feature elimination

R0 = basic reproduction rate

SHAP = Shapley additive explanations

SIR = susceptible–infectious–recovered

WHO = World Health Organization

XGBoost = extreme gradient boosting

## Contributions

J.M.K., H.K.L., and K.C.K. contributed towards the conception and writing of the original draft; J.M.K., H.K.L., K.H.L., and K.S.L. contributed towards the acquisition, analysis, and interpretation of the study data; J.H.A. and K.C.K. contributed towards the review, editing, and supervision of this study. All authors have read and approved the submitted version of this manuscript.

## Funding

This study was supported through the Infection Prevention Strategy Development Program of Korea (HW20C2103). The funding source had no role in the study design, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to submit the article for publication.

## Code availability

The code used to develop and evaluate this study's models is available at <https://github.com/minkim88/Optimal-Triage-COVID-19>

## Multimedia Appendix 1

**Supplementary Table 1.** Performance of the models according to the World Health Organization Ordinal Scale for Clinical Improvement. AUROC = area under the receiver operating characteristic, CSS = clinical severity score, LR = logistic regression, NPV = negative predictive value, PPV = positive predictive value, WHO OSCI = World Health Organization Ordinal Scale for Clinical Improvement, XGB = extreme gradient boosting.

**Supplementary Table 2.** AUROCs at each step of the RFE and *P* values of differences in AUROC values for Models 1 and 2. For Model 1, pregnancy and pregnancy week variables showed SHAP values of 0 and were eliminated at the beginning of the RFE. As Model 3 was derived from Model 1 by RFE, Model 3 corresponded to the model with the top 17 variables in the Model 1 column (i.e., 20 variables were removed from Model 1). Likewise, Model 4 corresponded to the model with the top 11 variables in the Model 2 column. AUROC = area under the receiver operating characteristic, RFE = recursive feature elimination.

**Supplementary Table 3.** Performance of the models in the hold-out cohort. AUROC = area under the receiver operating characteristic, NPV = negative predictive value, PPV = positive predictive value.

**Supplementary Table 4.** Order of feature importance for each model.

**Supplementary Figure 1.** Epidemic incidence curves of historical patient influxes of COVID-19 in South Korea.

**Supplementary Figure 2.** SIR simulated patient influx. SIR = susceptible-infectious-recovered,  $R_0$  = basic reproduction rate.

**Supplementary Figure 3.** Prediction probability distribution graph of patients in the out-of-fold samples.

**Supplementary Figure 4.** Patient-specific SHAP plots. The top 10 features, which were ranked for each patient by averaging their SHAP values, depicting the model's predictive behaviors. In this specific example, the SHAP plot for an actual severe patient shows that their age, lymphocyte count, and experiencing shortness of breath significantly affected their prediction output in a positive direction. However, their white blood cell count and hematocrit, as well as their hemoglobin and BMI, contributed in a negative direction. On the other hand, the SHAP plot for an actual non-severe patient shows that their age, lymphocyte count, platelet counts, and hemoglobin significantly affected their prediction output in the negative direction.

**Supplementary Figure 5.** Changes to the model's performance after applying recursive feature elimination (RFE). The outcome of the entire RFE was presented in terms of its AUROC values. RFE was performed on Model 1, which consisted of 37 variables, including five laboratory variables. Model 1 during the first iteration was developed using 37 variables, while Model 3 was developed using 17 variables. In Model 1, pregnancy and pregnancy week variables showed SHAP values of 0 and were eliminated at the first iteration.

**Supplementary Figure 6.** Changes to the model's performance after applying RFE. The outcome of the entire RFE is presented in terms of AUROC values. RFE was performed on Model 2, which consisted of 32 variables, in which five laboratory variables were excluded. Model 2 at the first

iteration was developed based on 32 variables, while Model 4 was developed using 11 variables.

**Supplementary Figure 7.** Optimized results of the patient triage simulations for the historical influx. Decreased mortality rate = (J index mortality rate - optimized mortality rate) / J index mortality rate.

**Supplementary Figure 8.** Mortality rates of the historical patient influx scenarios according to each threshold and at each threshold across different scenarios. Black dots represent the optimized mortality rate and the respective optimal threshold for each patient influx scenario.

**Supplementary Figure 9.** Mortality rates of the hypothetical patient influx scenarios according to each threshold and at each threshold across different scenarios. Black dots represent the optimized mortality rate and the respective optimal threshold for each patient influx scenario.

## References

1. World Health Organization. Critical preparedness, readiness and response actions for COVID-19: interim guidance, 7 March 2020. World Health Organization. <https://apps.who.int/iris/handle/10665/331422> (2020).
2. Gupta RK, Harrison EM, Ho A, et al. Development and validation of the ISARIC 4C Deterioration model for adults hospitalised with COVID-19: a prospective cohort study. *Lancet Respir Med*

2021;9:349–359. doi: 10.1016/S2213-2600(20)30559-2.

3. Mejia-Vilet JM, Córdova-Sánchez BM, Fernández-Camargo DA, Méndez-Pérez RA, Morales-Buenrostro LE, Hernández-Gilsoul T. A risk score to predict admission to the intensive care unit in patients with Covid-19: the ABC-GOALS score. *Salud Publica Mex* 2020;63:1–11. doi: 10.21149/11684.

4. Ji D, Zhang D, Xu J, et al. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL Score. *Clin Infect Dis* 2020;71:1393–1399. doi: 10.1093/cid/ciaa414.

5. Knight SR, Ho A, Pius R, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. *BMJ* 2020;370:m3339. doi: 10.1136/bmj.m3339.

6. Gao Y, Cai G, Fang W, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat Commun* 2020;11:5033. doi: 10.1038/s41467-020-18684-2.

7. Bolourani S, Brenner M, Wang P, et al. A machine learning prediction model of respiratory failure within 48 hours of patient admission for COVID-19: model development and validation. *J Med Internet Res* 2021;23:e24246. doi: 10.2196/24246.

8. Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 2002;46:389–422.

9. Karnon J, Stahl J, Brennan A, et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--4. *Value Health* 2012;15:821–827. doi: 10.1016/j.jval.2012.04.013.

10. World Health Organization. WHO R&D Blueprint novel Coronavirus COVID-19 Therapeutic Trial Synopsis. February 18, 2020, Geneva, Switzerland. [https://www.who.int/blueprint/priority-diseases/key-action/COVID-19\\_Treatment\\_Trial\\_Design\\_Master\\_Protocol\\_synopsis\\_Final\\_18022020.pdf](https://www.who.int/blueprint/priority-diseases/key-action/COVID-19_Treatment_Trial_Design_Master_Protocol_synopsis_Final_18022020.pdf)

11. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: ACM; 2016:785–794. <https://doi.org/10.1145/2939672.2939785>.
12. Lundberg SM, Lee S. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA;2017:4768–4777.
13. Pidd M. *Computer simulation in management science*. 5th ed. Chichester: John Wiley and Sons Ltd;2004:328p.
14. Korea Disease Control and Prevention Agency. Infectious Disease Portal. <http://www.kdca.go.kr/npt/biz/npp/portal/nppIssueIcdMain.do>
15. Hethcote HW. The mathematics of infectious diseases. *SIAM review* 42.4;2000:599–653. <https://doi.org/10.1137/S0036144500371907>.
16. Lee Y, Hong C, Kim D, Lee T, Lee J. Clinical course of asymptomatic and mildly symptomatic patients with coronavirus disease admitted to community treatment centers, South Korea. *Emerg Infect Dis* 2020;26:2346–2352. doi: 10.3201/eid2610.201620.
17. Caflisch R. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica* 1998;7:1–49.
18. Wood RM, Pratt AC, Kenward C, et al. The value of triage during periods of intense COVID-19 demand: simulation modeling study. *Med Decis Making* 2021;41:393–407. doi: 10.1177/0272989X21994035.
19. ICNARC. ICNARC report on COVID-19 in critical care. 2020. <https://www.icnarc.org/Our-Audit/Audits/Cmp/Reports>.
20. Haimovich AD, Ravindra NG, Stoytchev S, et al. Development and validation of the quick COVID-19 Severity Index: a prognostic tool for early clinical decompensation. *Ann Emerg Med* 2020;76:442–453. doi: 10.1016/j.annemergmed.2020.07.022.
21. Ye J, Zhang X, Zhu F, Tang Y. Application of a prediction model with laboratory indexes in the

risk stratification of patients with COVID-19. *Exp Ther Med* 2021;21:182. doi: 10.3892/etm.2021.9613.

22. Hu L, Chen S, Fu Y, et al. Risk factors associated with clinical outcomes in 323 coronavirus disease 2019 (COVID-19) hospitalized patients in Wuhan, China. *Clin Infect Dis* 2020;71:2089–2098. doi: 10.1093/cid/ciaa539.

23. Lassau N, Ammari S, Chouzenoux E, et al. Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. *Nat Commun* 2021;12:634. doi: 10.1038/s41467-020-20657-4.

24. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. *Clin Chim Acta* 2020;506:145–148. doi: 10.1016/j.cca.2020.03.022.

25. Henry BM, Lippi G. Chronic kidney disease is associated with severe coronavirus disease 2019 (COVID-19) infection. *Int Urol Nephrol* 2020;52:1193–1194. doi: 10.1007/s11255-020-02451-9.

26. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med* 2020;46:846–848. doi: 10.1007/s00134-020-05991-x.

27. Wu G, Yang P, Xie Y, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. *Eur Respir J* 2020;56:2001104. doi: 10.1183/13993003.01104-2020.

28. Chan JF, Yuan S, Kok K, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020;395:514–523. doi: 10.1016/S0140-6736(20)30154-9.

29. Goodall JW, Reed TAN, Ardissino M, et al. Risk factors for severe disease in patients admitted with COVID-19 to a hospital in London, England: a retrospective cohort study. *Epidemiol Infect* 2020;148:e251. doi: 10.1017/S0950268820002472.

30. Yang M, Ng MH, Li CK. Thrombocytopenia in patients with severe acute respiratory syndrome (review). *Hematology* 2005;10:101–105. doi: 10.1080/10245330400026170.
31. Tulsian PC, Pandey V. *Quantitative Techniques: Theory & Problems* (New Delhi: Pearson Education, Singapore, 2002).

### Figure legends

**Figure 1.** Relationships between each feature and SHAP values. Summary plot in which each dot-point represents the SHAP value of a patient in the dataset used to construct the developed model. The dots are plotted for every feature used to fit the baseline model, excluding two features (pregnancy and number of weeks pregnant) that were not selected for the developed model. The SHAP values are displayed in rank order, based on their feature importance, along the y-axis as calculated by averaging the absolute SHAP values of each dot. A point's location on the x-axis shows

its impact on the predictive output of the model. Purple indicates a relatively high feature value, while yellow represents a relatively low feature value. Grey-colored dots represent missing values.

**Figure 2.** Relationships between each feature and the SHAP values. Dependence plots for each of the top nine important features, including patient age, lymphocyte count, platelet count, BMI, hematocrit, shortness of breath, sex, body temperature, and heart rate. Each scatter plot shows the impact of each feature on the predictions made by the study model. The x-axis represents the variables' values, and the y-axis represents their SHAP values. The inflection points indicate the non-linear impact of a feature on the model's prediction. BMI = body mass index, SHAP = Shapley additive explanations.

**Figure 3.** Simulation workflow. Diagram showing how medical resources can be allocated among COVID-19 patients according to the machine learning-based triage system. Patients with a prediction probability exceeding a certain threshold are first triaged to an ICU that is currently under its total capacity. Conversely, patients are directed to a general ward if the ICU's capacity is full or if their severity prediction probability is lower than the threshold. Type I deaths represent those occurring in the ICU. Types II and III deaths represent those of patients who have been directed to the general ward due to ICU unavailability or because they were found to have a disease severity probability lower than the threshold, respectively. We used simulations to obtain the optimal threshold wherein

the mortality rate,  $\frac{n(\text{total deaths})}{n(\text{total patients})} = \frac{n(\text{type I death} + \text{type II death} + \text{type III death})}{n(\text{total patients})}$ , is minimized.

COVID-19 = coronavirus disease 2019, ICU = intensive care unit.

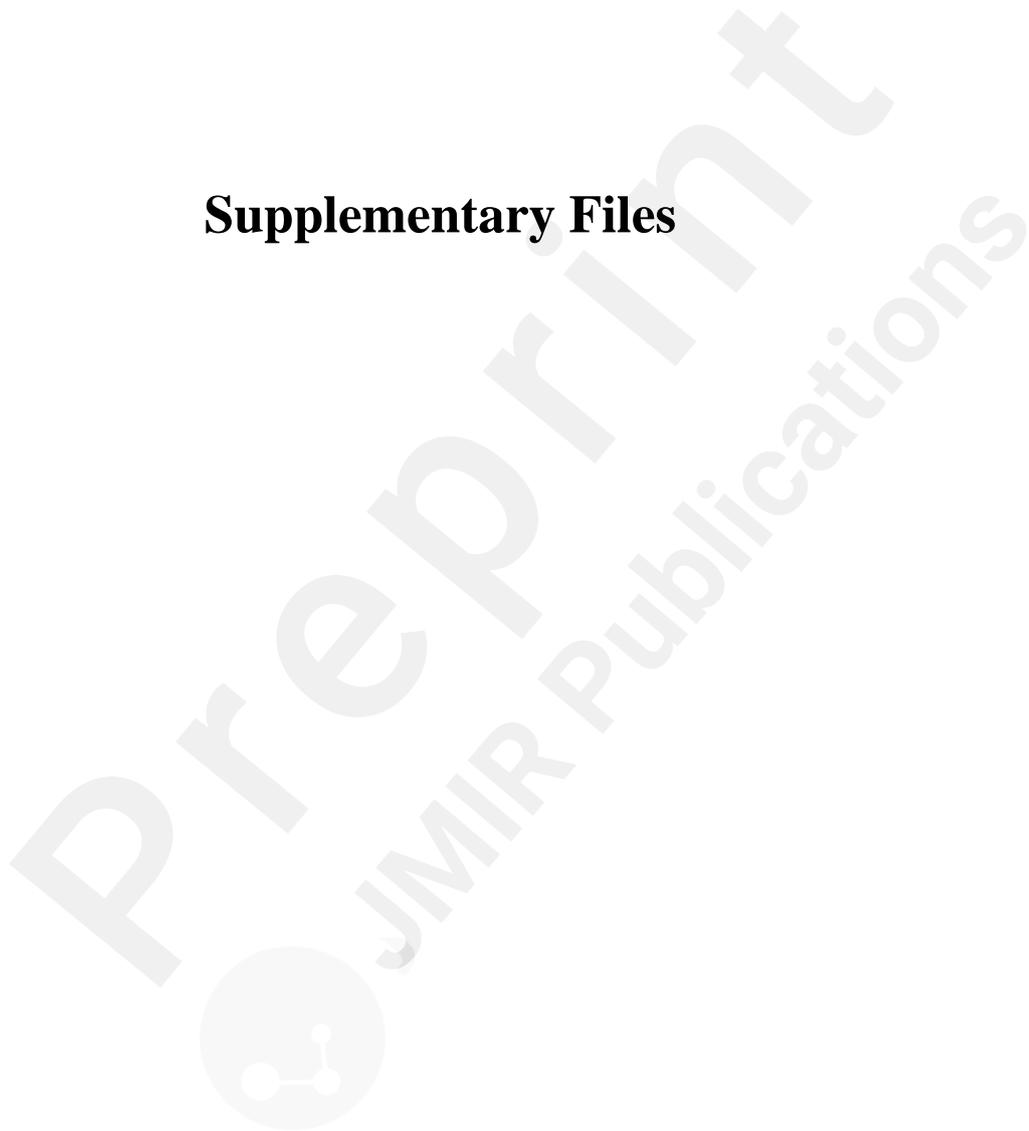
**Figure 4.** Optimized results of the patient triage simulations for hypothetical influx. Decreased mortality rate = (J index mortality rate - optimized mortality rate) / J index mortality rate.

**Figure 5.** Mortality rates in hypothetical patient influxes are decomposed by death subtype at each threshold. The x-axis represents the threshold, and the y-axis represents the stacked proportion of

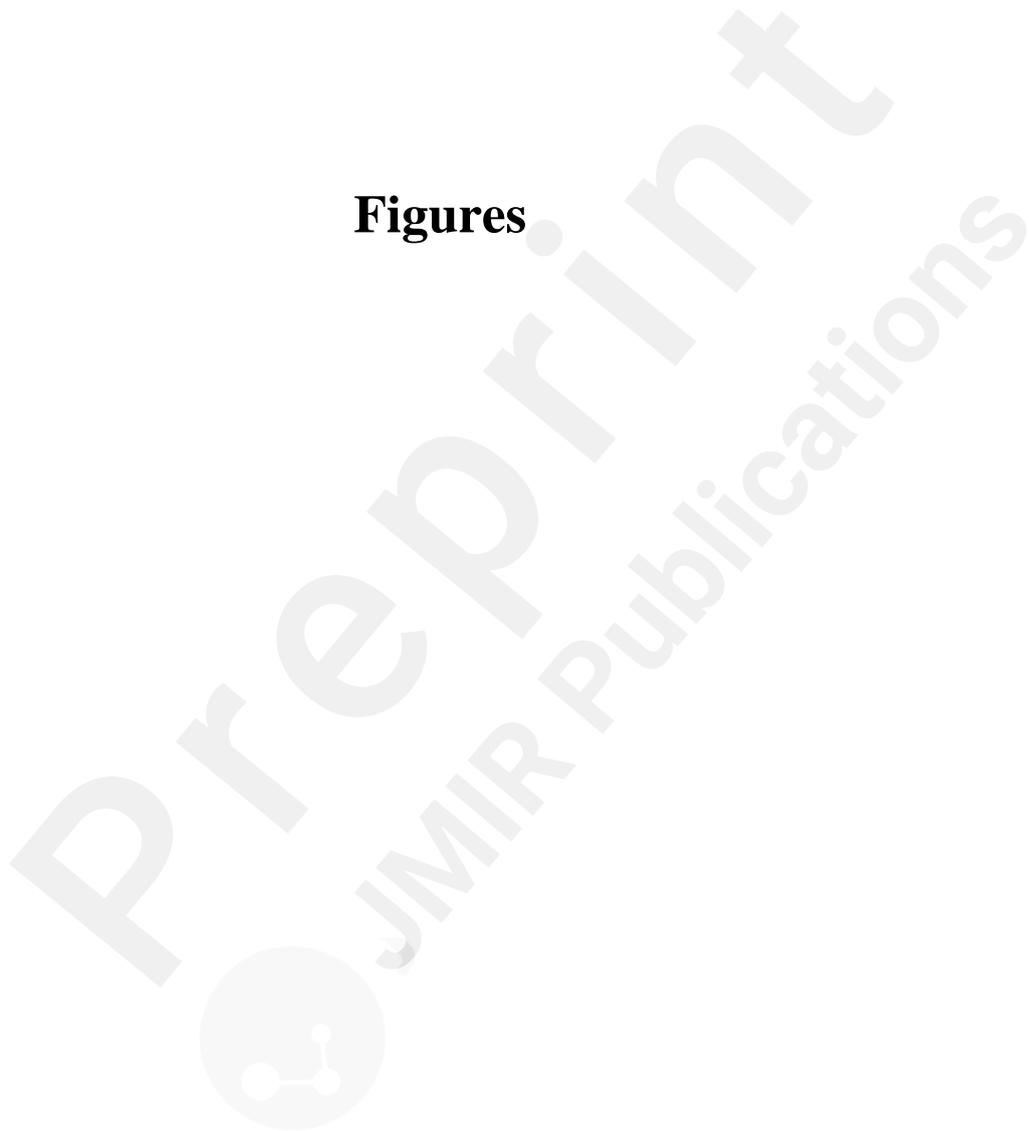
each death subtype to the total number of patients, calculated as  $\frac{n(\text{death subtype})}{n(\text{total patients})}$ , at each threshold.



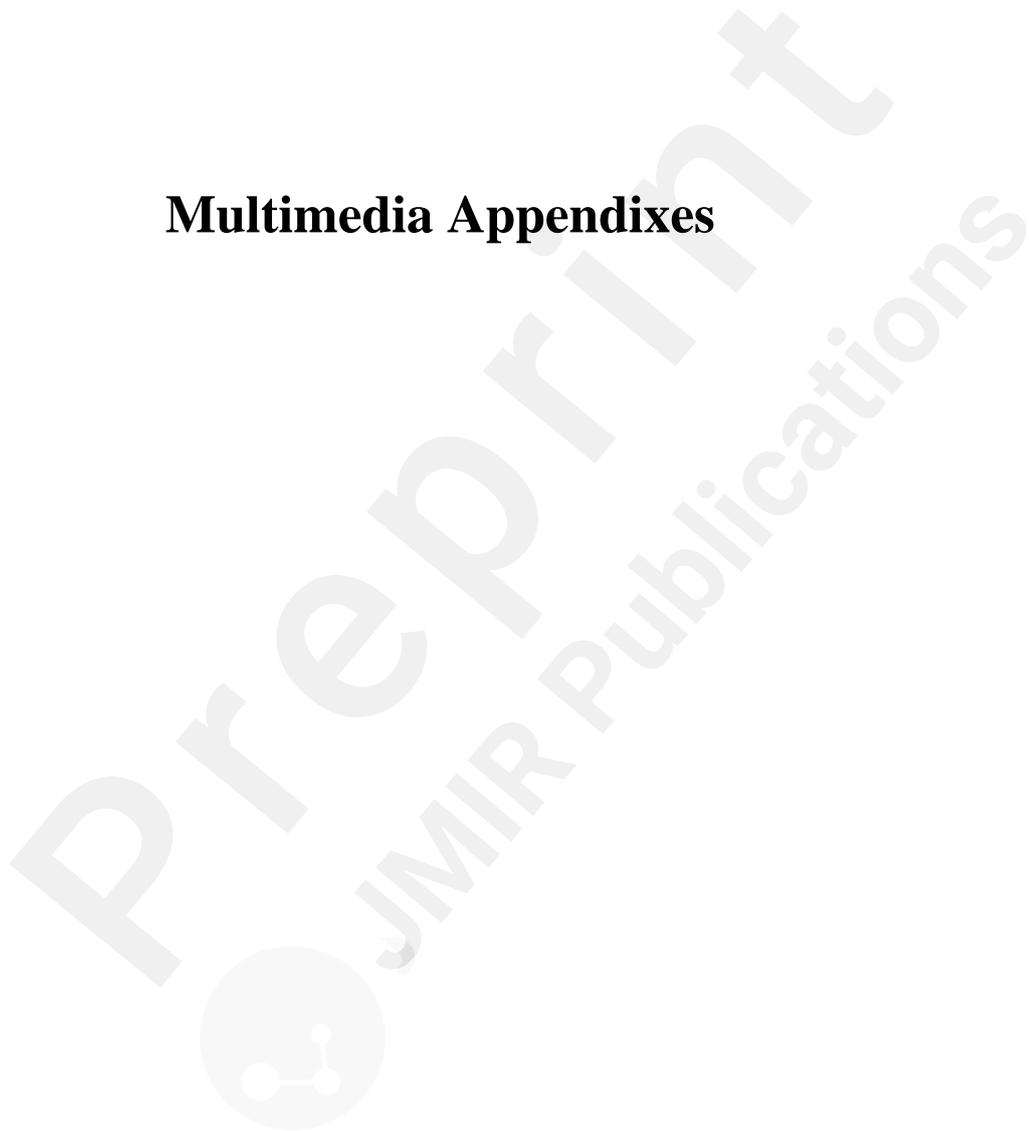
## Supplementary Files



## Figures



## Multimedia Appendixes



Supplementary Table 1. Performance of the models according to the World Health Organization Ordinal Scale for Clinical Improvement. Supplementary Table 2. AUROCs at each step of the RFE and P values of differences in AUROC values for Models 1 and 2. For Model 1, pregnancy and pregnancy week variables showed SHAP values of 0 and were eliminated at the beginning of the RFE. As Model 3 was derived from Model 1 by RFE, Model 3 corresponded to the model with the top 17 variables in the Model 1 column (i.e., 20 variables were removed from Model 1). Likewise, Model 4 corresponded to the model with the top 11 variables in the Model 2 column. Supplementary Table 3. Performance of the models in the hold-out cohort. Supplementary Table 4. Order of feature importance for each model. Supplementary Figure 1. Epidemic incidence curves of historical patient influxes of COVID-19 in South Korea. Supplementary Figure 2. SIR simulated patient influx. SIR = susceptible-infectious-recovered, R0 = basic reproduction rate. Supplementary Figure 3. Prediction probability distribution graph of patients in the out-of-fold samples. Supplementary Figure 4. Patient-specific SHAP plots. The top 10 features, which were ranked for each patient by averaging their SHAP values, depicting the model's predictive behaviors. In this specific example, the SHAP plot for an actual severe patient shows that their age, lymphocyte count, and experiencing shortness of breath significantly affected their prediction output in a positive direction. However, their white blood cell count and hematocrit, as well as their hemoglobin and BMI, contributed in a negative direction. On the other hand, the SHAP plot for an actual non-severe patient shows that their age, lymphocyte count, platelet counts, and hemoglobin significantly affected their prediction output in the negative direction. Supplementary Figure 5. Changes to the model's performance after applying recursive feature elimination (RFE). The outcome of the entire RFE was presented in terms of its AUROC values. RFE was performed on Model 1, which consisted of 37 variables, including five laboratory variables. Model 1 during the first iteration was developed using 37 variables, while Model 3 was developed using 17 variables. In Model 1, pregnancy and pregnancy week variables showed SHAP values of 0 and were eliminated at the first iteration. Supplementary Figure 6. Changes to the model's performance after applying RFE. The outcome of the entire RFE is presented in terms of AUROC values. RFE was performed on Model 2, which consisted of 32 variables, in which five laboratory variables were excluded. Model 2 at the first iteration was developed based on 32 variables, while Model 4 was developed using 11 variables. Supplementary Figure 7. Optimized results of the patient triage simulations for the historical influx. Decreased mortality rate = (J index mortality rate - optimized mortality rate) / J index mortality rate. Supplementary Figure 8. Mortality rates of the historical patient influx scenarios according to each threshold and at each threshold across different scenarios. Black dots represent the optimized mortality rate and the respective optimal threshold for each patient influx scenario. Supplementary Figure 9. Mortality rates of the hypothetical patient influx scenarios according to each threshold and at each threshold across different scenarios. Black dots represent the optimized mortality rate and the respective optimal threshold for each patient influx scenario.

URL: <http://asset.jmir.pub/assets/bcfa378fd23d1625d460e69f201bea16.docx>