Early Triaging Support System for Hospitalized COVID-19 Patients: a Machine-Learning based Severity Prediction Model using Nationwide Multi-Center Real World Data

Sangwon Baek[1], Kyunga Kim[2,3], Myung Jin Chung[1,3]

[1] Medical AI Research Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea
[2]Biomedical Statistics Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea
[3]Department of Digital Health & Department of Data Convergence and Future Medicine SAIHST, Sungkyunkwan University, Seoul, Republic of Korea

Presentation Number: Control Number 6228
Presentation Date: June 18th, 10am-4pm
Korea Phone #: +82-10-8805-3873
Personal Email: baeksw98@gmail.com
Work Email: sw98.baek@sbri.co.kr

SAMSUNG MEDICAL CENTER

Introduction

- During the COVID-19 pandemic, the global healthcare system confronted an urgent threat that resulted 6.9 million casualties as of June 2023
- Overwhelming influx of patients into hospitals has strained medical resources
- Asymptomatic progression and presentation of non-specific symptoms in COVID-19 cases further exacerbated situation
- Many CDSSs for patient triaging have been proposed, but were highly susceptible to risk of biases
- Biases can result in overoptimistic prediction performance with poor calibration, and need to be minimized to develop a valid and generalizable predictive model suitable for clinical implementation
- Aimed to develop and validate a short-term prognostic prediction model; explore its transportability to periods with limited information on new variants; and characterize clinical characteristics of patients.

Methods

Study Setting and Design
- A nationwide, multicentre, prospective cohort study
- All hospitalized COVID-19 patients at 19 main referral hospitals between Jan 5th, 2020 and Aug 29th, 2022

Definition of Variant Dominant Periods
- Segmented study period into three variant dominant periods: alpha-variant dominant period (Jan 5th, 2020~May 1st, 2021); delta-variant dominant period (May 1st, 2021~Nov 24th, 2021); omicron-variant dominant period (Nov 24th, 2021~Aug 24th, 2022)

Definition of COVID-19 Severity
- Declared the COVID-19 severity for patients under one or more of the following conditions during their hospitalization: (1) mechanical ventilation required; (2) ECMO required; (3) admission to ICU; (4) patient's death

Data Collection
- Adhering to standard data collection guideline, we gathered patient data with 32 features from demographic, clinical, laboratory, and radiological findings within the first day of hospitalization.
- The final severity status of each patient was determined at day 15 of his/her hospitalization.
- All data collected in each hospital were de-identified and then uploaded onto the cloud database storage

Model development process
- Among 32 collected features, 27 readily accessible ones with no missing data remained for prediction modeling.
- The detailed implementation of candidate feature subsets identification; model development and evaluation with feature selection; model performance visualization and feature interpretation are shown in Figure 2

Model Transportability to New Variant Dominant Period
- 'all-variants' model was built from the entire development cohort
- 'Omicron-only' model was developed using the patients in the development cohort diagnosed during Omicron-dominant period
- 'Alpha-Omicron' and 'Delta-Omicron' models were built in the same way to develop ensemble model that combined these two models and the Omicron-only model

Statistical Analysis
- Patient characteristics were summarized as median [IQR] and number (%) for continuous and categorical variables
- SMD was calculated as Cohen's D and H formula for continuous and categorical variables
- Youden's index was used to find optimal thresholds that maximize the average of sensitivity and specificity
- Integrated calibration index was derived from the weighted average of the absolute difference between the calibration curve and the perfectly calibrated diagonal line.
- A two-sided p value below 0.05 was set to declare statistical significance
- All statistical analyses were performed using Python (Python Software Foundation, V3.9)

Figure 1. Development and validation cohort generation and partitioning

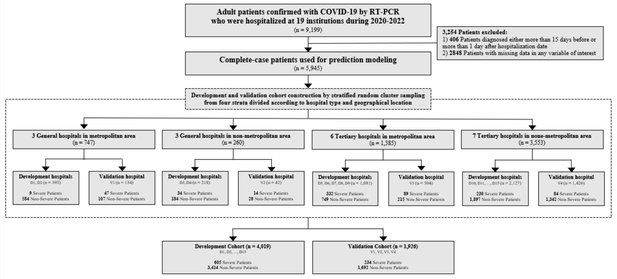
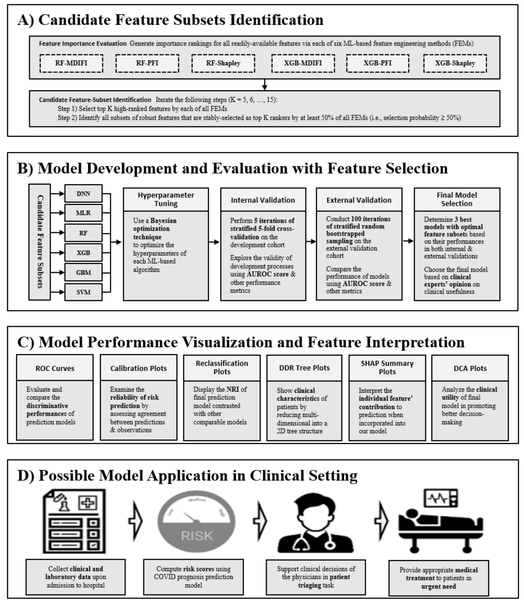


Figure 2. Machine learning-based Pipeline for COVID prognosis prediction



RF-MDIFI=random forest based mean decrease in Gini Impurity feature importance method. RF-PFI=random forest based permutation feature importance method. RF-Shapley=random forest based Shapley method. XGB-BFI=eXtreme gradient Boosting based built-in feature importance method. XGB-PFI=eXtreme gradient Boosting based permutation feature importance method. XGB-Shapley=eXtreme gradient Boosting based Shapley method. RF=random forest. MLR=multivariable logistic regression. DNN=deep neural network. XGB=eXtreme gradient Boosting, GBM=gradient boosting machine. SVM=support vector machine. AUROC=area under receiver operating characteristic curve. ROC=receiver operating characteristic curve. NRI=net-reclassification improvement. DCA=decision curve analysis. SHAP=Shapley Additive exPlanations.

Results

Table 1. Characteristics of patients in development and validation cohorts

	Total Cohort (N: 5,945)	Development Cohort (N: 4,019)	Validation Cohort (N: 1,926)	ASMD
Patient Characteristics				
Age (years)	60 (40-70)	60 (45-70)	59 (35-65)	0·333
Male Sex (%)	2,887 (48·6%)	2,130 (48·8%)	757 (47·9%)	0·018
Comorbidities (%)				
Hypertension	2,314 (35·6%)	1,622 (37·2%)	692 (31·1%)	0·127
Diabetes Mellitus	1,249 (21·9%)	978 (22·4%)	271 (17·2%)	0·132
CVD	508 (8·7%)	389 (8·9%)	119 (7·5%)	0·059
Cancer	477 (8·0%)	369 (8·5%)	108 (6·8%)	0·061
Others	2,242 (37·7%)	1,813 (41·5%)	429 (27·2%)	0·304
Clinical Symptoms (%)				
Fever	2,366 (39·8%)	1,678 (38·4%)	688 (43·5%)	0·104
Cough	2,823 (44·1%)	1,852 (42·4%)	771 (48·8%)	0·128
Sputum	1,502 (25·3%)	1,038 (23·8%)	464 (29·4%)	0·127
Dyspnea	1,318 (22·2%)	1,042 (23·9%)	276 (17·5%)	0·159
Myalgia	1,388 (23·3%)	938 (21·5%)	450 (28·5%)	0·162
Sore throat	1,142 (19·2%)	763 (17·5%)	379 (24·0%)	0·181
Loss of Sensor	330 (5·6%)	212 (4·9%)	118 (7·5%)	0·109
GI symptom	472 (7·9%)	295 (6·8%)	177 (11·2%)	0·157
Vital Sign				
BT (°C)	36·6 (36·3-37·2)	36·6 (36·3-37·2)	36·5 (36·3-37·2)	0·031
SBP (mmHg)	129 (116-141)	129 (116-140)	129 (117-141)	0·021
DBP (mmHg)	80 (70-87)	80 (70-86)	80 (72-90)	0·200
PR (counts)	84 (74-95)	84 (74-95)	86 (76-97)	0·128
RR (counts)	20 (18-20)	20 (18-20)	20 (18-20)	0·191
SpO2 (%)	97 (96-98)	97 (96-98)	97 (96-98)	0·021
Blood biochemistry				
WBC (10^3/μL)	5·2 (4·1-6·9)	5·3 (4·1-7·1)	5·1 (4·0-6·6)	0·143
ANC (10^3/μL)	3·5 (2·4-5·4)	3·7 (2·5-5·8)	3·0 (2·1-4·3)	0·319
ALC (10^3/μL)	1·2 (0·8-1·7)	1·3 (0·8-1·8)	1·1 (0·8-1·5)	0·105
PLT (10^3/μL)	200 (157-248)	198 (154-248)	203 (166-248)	0·068
CRP (mg/dL)	1·2 (0·3-5·2)	1·3 (0·3-6·0)	0·7 (0·2-2·9)	0·077
LDH (U/L)	316 (221-443)	287 (212-428)	369 (289-476)	0·161

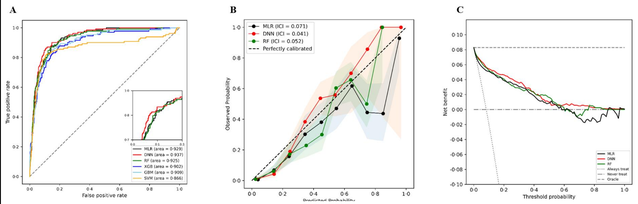
Data are median (IQR) or n (%). ASMD=absolute standardized mean difference. CVD=cardio-vascular disease. GI= gastrointestinal. BT=body temperature. SBP=systolic blood pressure. DBP=diastolic blood pressure. PR=pulse rate. RR=respiratory rate. SpO2=saturation of peripheral oxygen. WBC= white blood cell. ANC=absolute neutrophil count. ALC=absolute lymphocyte count. PLT= platelet count. CRP=c-reactive protein. LDH=lactate dehydrogenase.

Table 2. Discriminative performance in the external validation

	Predictive Measures								
Model	**AUROC**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**LRP**	**LRN**	**DOR**	**Cut-Off**
DNN	**0·937 (0·935-0·938)**	0·883 (0·878-0·889)	0·881 (0·877-0·886)	41·19 (40·26-42·12)	98·70 (98·74-98·85)	7·68 (7·43-7·94)	0·13 (0·13-0·14)	**61·18 (57·61-64·75)**	0·191
MLR	0·929 (0·927-0·930)	0·890 (0·893-0·906)	0·842 (0·836-0·848)	34·97 (34·13-35·82)	98·91 (98·84-98·98)	5·88 (5·70-6·07)	0·12 (0·11-0·13)	53·10 (50·31-55·90)	0·170
RF	0·925 (0·923-0·927)	0·885 (0·879-0·891)	0·853 (0·848-0·858)	36·17 (35·27-37·06)	98·77 (98·71-98·84)	6·19 (6·00-6·39)	0·13 (0·13-0·14)	49·68 (46·22-53·15)	0·177
XGB	0·900 (0·898-0·903)	0·849 (0·842-0·856)	0·838 (0·832-0·844)	33·14 (32·23-34·05)	98·36 (98·29-98·44)	5·44 (5·23-5·64)	0·18 (0·17-0·19)	31·10 (29·73-32·48)	0·111
GBM	0·907 (0·904-0·910)	0·859 (0·853-0·866)	0·850 (0·844-0·855)	34·99 (34·11-35·87)	98·50 (98·43-98·57)	5·89 (5·70-6·08)	0·17 (0·16-0·17)	37·35 (35·39-39·32)	0·015
SVM	0·862 (0·858-0·866)	0·824 (0·817-0·831)	0·892 (0·889-0·896)	41·68 (40·88-42·47)	98·21 (98·14-98·28)	7·82 (7·60-8·04)	0·20 (0·19-0·20)	41·06 (39·13-43·00)	0·102

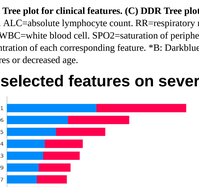
All results, 95% CIs in parentheses, were computed from bootstrapping external validation cohorts 100 times with replacements. Youden's Index was used to determine optimal cut-off point. DNN=deep neural network. MLR=multivariable logistic regression. RF=random forest. XGB=eXtreme gradient boosting. GBM=gradient boosting machine. SVM=support vector machine. AUROC=area under receiver operating characteristic curve. PPV=positive predictive value. NPV=negative predictive value. LRP=likelihood ratio positive. LRN=likelihood ratio negative. DOR=diagnostic odds ratio.

Figure 3. Performance comparisons of prediction models during external validation



(A) Receiver operating characteristic curve showing the performance of the prediction models. (B) Calibration plot displaying the correlation between predicted and observed risks of the prediction models by segmenting the cohort into 10 bins. (C) Decision curve analysis plot demonstrating the potential net benefit of applying the prediction models in clinical applications. AUROC=area under receiver operating characteristic curve. MLR=multivariable logistic regression. DNN=deep neural network. RF=random forest. XGB=eXtreme gradient boosting. GBM=gradient boosting machine. SVM=support vector machine.

Figure 4. Performance of prediction models on omicron-variant cases



Used final prediction model (Deep neural network-based model with 11 selected features) to develop each model. Ensemble model was developed by combining the prediction probabilities of the three models trained using Alpha&Omicron, Delta&Omicron, and Omicron development cohorts; All variants model was developed by using whole development cohort; Omicron only model was developed by using only Omicron development cohort. Patient size (N): whole development cohort, 4,365 (Alpha development cohort, 2,622; Delta development cohort, 1,406; Omicron development cohort, 337); whole validation cohort, 1580 (Alpha validation cohort, 840; Delta validation cohort, 628; Omicron validation cohort, 112).

Figure 5. Visual representation for clinical and laboratory characteristics of patients

(A) DDR Tree plot for laboratory features. (B) DDR Tree plot for clinical features. (C) DDR Tree plot for severity probability. DDR=discriminative dimensionality reduction. LDH=lactate dehydrogenase. ALC=absolute lymphocyte count. RR=respiratory rate. DM=diabetes mellitus. CRP=c-reactive protein. ANC=absolute neutrophil count. PLT=platelet counts. WBC=white blood cell. SPO2=saturation of peripheral oxygen. *A&C: Darkred color indicates high concentration and lightgreen color indicates low concentration of each corresponding feature. *B: Darkblue color indicates presence of clinical features or increased age, lightgreen color indicates absence of clinical features or decreased age.

Figure 6. Average impact of selected features on severity prediction

Used random forest-based model with 11 selected features to construct the summary plot above. Used TreeExplainer to compute Shapley Additive exPlanations (SHAP) values. CRP=c-reactive protein. LDH=lactate dehydrogenase. ALC=absolute lymphocyte count. RR=respiratory rate. ANC=absolute neutrophil count. WBC=white blood cell. PLT=platelet counts. SPO2=saturation of peripheral oxygen. DM=diabetes mellitus. *Specific SHAP values are shown on the right of each feature names in the y-label.

Discussion

- We developed a novel clinical decision support system for COVID-19 severity that is generalizable across temporally and geographically diverse population

Engineering strength
- Model exhibits robust prediction and high calibration
- Model maintains stable prediction without data- nor algorithmic-level modifications
- Model can be easily and promptly accessible with minimal feature usage
- Model retains transportable prediction despite the limited sample availability.

Clinical strength
- Patient characterization was performed through displaying distribution of clinical and laboratory features in two-dimensional tree structure
- Feature-level contributions on prediction was computed using shapley values
- Visualizations served to provide additional index to be used as a reference for the clinicians.

Limitations
- The studied subjects were limited to hospitalized COVID-19 patients in South Korea that validation with international cohort is needed
- COVID variants in our analysis was determined by the predominant period that there may exist misclassification for certain patients
- Vaccination status was not accounted during the analysis due to high missing rate
- Model transportability against future variant cases is unknown since our cohort comprised patients up to omicron-variant period.

Conclusion
- Although worldwide healthcare system currently progress towards endemic, COVID-19 still remains a health emergency as concerns for similar future outbreaks continue to exist.
- Appropriate preparation by learning lessons from the current pandemic is essential to prevent next global future pandemics
- Despite the operational and clinical benefit our model provided through facilitating early patient triage, we acknowledge its limited scope and influence in thoroughly addressing COVID-19 management.
- To minimize unwanted victims, the demand for more precision approach persists that much more extensive data explorations (i.e. lab & genomic) seem necessary.
- We present our work not as an ultimate solution, but as a catalyst for further progress towards an enhanced healthcare system

Acknowledgements

Authors sincerely thanks all studied patients across the 19 participated hospitals. We particularly thank healthcare practitioners for assistance with data collection and extraction. This research was supported by a grant of the Information and Communications Promotion Fund (ICT promotion fund) through the National IT Industry Promotion Agency (NIPA), funded by the Ministry of Science and ICT (MSIT), South Korea.

References

1. *WHO Coronavirus (COVID-19) Dashboard.*
2. Knight, S.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**: p. m3339.
3. Gallo Marin, B., et al., *Predictors of COVID-19 severity: A literature review.* Rev Med Virol, 2021. **31**(1): p. 1-10.
4. Miller, J.L., et al., *Prediction models for severe manifestations and mortality due to COVID-19: A systematic review.* Acad Emerg Med, 2022. **29**(2): p. 206-216.
5. Malik, P., et al., *Biomarkers and outcomes of COVID-19 hospitalisations: systematic review and meta-analysis.* BMJ Evid Based Med, 2021. **26**(3): p. 107-108.
6. Soltan, A.A.S., et al., *Real-world evaluation of rapid and laboratory-free COVID-19 triage for emergency care: external validation and pilot deployment of artificial intelligence driven screening.* Lancet Digit Health, 2022. **4**(4): p. e266-e278.
7. Wynants, L., et al., *Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal.* BMJ, 2020. **369**: p. m1328.
8. Gilbert, A. and A. Ghuysen, *Triage in the time of COVID-19.* Lancet Digit Health, 2022. **4**(4): p. e210-e211.
9. Lee, D.W., et al., *Genomic epidemiology of SARS- CoV-2 Omicron variants in the Republic of Korea.* Sci Rep, 2022. **12**(1): p. 22414.
10. Marhsall, J., et al., *A minimal common outcome measure set for COVID-19 clinical research.* Lancet Infect Dis, 2020. **20**(8): p. e192-e197.
11. Bagherzadeh-Khiabani, F., et al., *A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results.* J Clin Epidemiol, 2016. **71**: p. 76-85.
12. Austin, P.C. and E.W. Steyerberg, *The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models.* Stat Med, 2019. **38**(21): p. 4051-4065.
13. Mienye, I.D. and Y. Sun, *Performance analysis of cost-sensitive learning methods with application to imbalanced medical data.* Informatics in Medicine Unlocked, 2021. **25**.
14. Wu, G., 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**(2).
15. Lodato, I., et al., *Prognostic Model of COVID-19 Severity and Survival among Hospitalized Patients Using Machine Learning Techniques.* Diagnostics (Basel), 2022. **12**(11).
16. Bhatia, S., et al., *Severity and mortality prediction models to triage Indian COVID-19 patients.* PLOS Digit Health, 2022. **1**(3): p. e0000020.
17. Gao, Y., et al., *Machine learning based early warning system enables accurate mortality risk prediction for COVID-19.* Nat Commun, 2020. **11**(1): p. 5033.
18. Johnson, J.M. and T.M. Khoshgoftaar, *Survey on deep learning with class imbalance.* Journal of Big Data, 2019. **6**(1).
19. Kim, J., et al., *Optimal Triage for COVID-19 Patients Under Limited Health Care Resources With a Parsimonious Machine Learning Prediction Model and Threshold Optimization Using Discrete-Event Simulation: Development Study.* JMIR Med Inform, 2021. **9**(11): p. e32726.
20. Yadaw, A.S., et al., *Clinical features of COVID-19 mortality: development and validation of a clinical prediction model.* Lancet Digit Health, 2020. **2**(10): p. e516-e525.
21. Abiodun, E.O., et al., *A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities.* Neural Comput Appl, 2021. **33**(22): p. 15091-15118.
22. Alakija, A., *Leveraging lessons from the COVID-19 pandemic to strengthen low-income and middle-income country preparedness for future global health threats.* Lancet Infect Dis, 2023.
23. Esper, F.P., et al., *Alpha to Omicron: Disease Severity and Clinical Outcomes of Major SARS-CoV-2 Variants.* J Infect Dis, 2022.
24. Nair, A.T.N., et al., *Heterogeneity in phenotype, disease progression and drug response in type 2 diabetes.* Nat Med, 2022. **28**(5): p. 982-988.
25. Park, M.S., et al., *Machine Learning-Based COVID-19 Patients Triage Algorithm Using Patient-Generated Health Data from Nationwide Multicenter Database.* Infect Dis Ther, 2022. **11**(2): p. 787-805.
26. Talukder, A., et al., *Association Between Symptoms and Severity of Disease in Hospitalised Novel Coronavirus (COVID-19) Patients: A Systematic Review and Meta-Analysis.* J Multidiscip Healthc, 2022. **15**: p. 1101-1110.
27. Loyer, C., et al., *Impairment of neutrophil functions and homeostasis in COVID-19 patients: association with disease severity.* Crit Care, 2022. **26**(1): p. 155.
28. Zhu, B., et al., *Correlation between white blood cell count at admission and mortality in COVID-19 patients: a retrospective study.* BMC Infect Dis, 2021. **21**(1): p. 574.
29. Abdullah, T.A.A., M.S.M. Zahid, and W. Ali, *A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions.* Symmetry, 2021. **13**(12).
30. Vickers, A.J., B. Van Calster, and E.W. Steyerberg, *Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests.* BMJ, 2016. **352**: p. i6.