Development of a machine learning-based clinical decision support system to predict clinical deterioration in patients visiting the emergency department

Arom Choi, M.D. [1,3],; Soyeon Choi, M.D. [1]; Kyungsoo Chung, M.D., Ph.D.[2,3]; Hyun Soo Chung, M.D., Ph.D. [1]; Ji Hoon Kim, M.D., MPH, Ph.D. [1,3]

1 Department of Emergency Medicine, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
2 Division of pulmonary and critical care medicine, Department of Internal Medicine, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
3 Institute for Innovation in Digital Healthcare, Yonsei University, Seodaemun-gu, 50 Yonsei-ro, Seoul 03722, Republic of Korea

Background.
The emergency care domain is particularly suitable for the challenge of adopting machine-learning (ML) based clinical decision support system (CDSS) contributing to patient safety and improve clinical outcomes, because of the need for rapid clinical decision-making by physicians. Accordingly, attempts at developing ML-based CDSS that enable efficient prediction for clinical practice have been reported in the setting of emergency departments (EDs). Workflows are delayed in frequently crowded EDs, making patients requiring time-critical interventions vulnerable to worse outcomes [14-16]. Therefore, it is crucial for CDSS to be able to assist the ED physicians who make time-critical decisions and interventions. Therefore, we aimed to develop a practical ML-based CDSS for ED practice utilizing available information according to the decision-making framework of physicians and to validate its clinical usefulness as a supportive tool in the ED.

Methods.
We conducted a retrospective, observational study using data from a Level 1 ED of a tertiary teaching hospital in South Korea from June 2015 to December 2019. We extracted 27 fixed baseline and 93 observational features using data on vital signs, mental status, laboratory results, and electrocardiograms during emergency department stay. we derived 10 important predictors that were highly associated with the outcome upon its occurrence. Next, we developed a basic model to predict the occurrence of the outcome at the current time point (0 h) by weighting 10 important predictors and 69 subfactors. Based on this model, a total of 25 predictive models were developed that brought the input features up to the previous 1 h, 2 h, 3 h, and 6 h before (lagging), and predicted the occurrence of outcomes 1 h, 2 h, 3 h, and 6 h later (leading) from the time of prediction. Outcomes included intubation, admission to the intensive care unit, inotrope or vasopressor administration, and in-hospital cardiac arrest. eXtreme gradient boosting algorithm was used to learn and predict each outcome. Specificity, sensitivity, precision, F1 score, area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC) were assessed.

Results.
We analyzed 303,345 patients with 4,787,121 input data, resampled into 24,148,958 1-h units. The models displayed a discriminative ability to predict outcomes (AUROC >0.9), and the model with lagging 6 and leading 0 displayed the highest value. The AUROC curve of in-hospital cardiac arrest had the smallest change, with increased lagging for all outcomes. With inotropic use, intubation, and intensive care admission, the range of AUROC curve change with the leading 6 was the highest according to different amounts of previous information (lagging). We subsequently performed external validation of the model using a separate dataset collected over a period of 2 years. Consistent with the findings of our internal validation, the external validation showed the same pattern, confirming the comparable predictive power and generalizability of our model. The systolic blood pressure (SBP) upon arrival for inotropic use, alert mental status for intubation, lactate levels in in-hospital cardiac arrest (IHCA), and platelet distribution width in intensive care units (ICU) admission displayed the highest feature importance score. The number and percentage of false-positive cases with the sensitivity fixed at 95%, 99%, and 100% were analyzed to identify the accuracy with which the model predicted the deterioration of patients while allowing a certain level of false alarm. To predict inotropic use, 11.5% of false-positive cases with 95% sensitivity increased to 61.4% of cases with 100% sensitivity. Likewise, the false-positive rate increased from 10% to 23.9%, 39.5% to 81.9%, and 18.8% to 86.9% for the prediction of mechanical ventilation, in-hospital arrest, and ICU admission, respectively, upon changing the sensitivity of the predictive model from 95% to 100%.

Conclusion.
We developed a practical ML-based CDSS for ED practice by utilizing available information according to the decision-making framework of physicians. CDSS should be customized according to clinical situations, and ML algorithms can help improve their performance.