



OPEN

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

Arom Choi^{1,3}, So Yeon Choi¹, Kyungsoo Chung^{2,3}, Hyun Soo Chung¹, Taeyoung Song⁴, Byunghun Choi⁴ & Ji Hoon Kim^{1,3}✉

This study aimed to develop a machine learning-based clinical decision support system for emergency departments based on the decision-making framework of physicians. We extracted 27 fixed and 93 observation features using data on vital signs, mental status, laboratory results, and electrocardiograms during emergency department stay. 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 were assessed. 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 unit admission, the range of AUROC curve change with the leading 6 was the highest according to different amounts of previous information (lagging). In this study, a human-centered approach to emulate the clinical decision-making process of emergency physicians has been adopted to enhance the use of the system. Machine learning-based clinical decision support systems customized according to clinical situations can help improve the quality of care.

Clinical decision support systems (CDSS) contribute to patient safety and improve clinical outcomes^{1,2}. Machine learning (ML) is being widely used in CDSS owing to its usefulness in diagnosis, prognosis, pattern recognition, and imaging classification with profound processing speed and the comprehensive nature of analytic methods³. The emergency care domain is particularly suitable for the challenge of adopting ML-based CDSS, because of the need for rapid clinical decision-making by physicians⁴. Accordingly, attempts at developing ML-based CDSS that enable efficient prediction for clinical practice have been reported in the setting of emergency departments (EDs)^{3,5–7}.

Globally, increased ED visits have led to resource saturation and crowding⁸, which affects both physicians and patients⁹. Workflows are delayed in frequently crowded EDs^{10–13}, 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 interventions⁴. However, previous studies have been limited to a narrow range of input data and were inappropriate for ED use since they dealt with broad and prolonged outcomes^{5,7}. Additionally, the validation of ML-based CDSS remains challenging in clinical practice¹⁷.

¹Department of Emergency Medicine, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea. ²Division 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. ³Institute for Innovation in Digital Healthcare, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea. ⁴LG Electronics, 128 Yeoui-daero, Yeongdeungpo-gu, Seoul 07336, Republic of Korea. ✉email: jichoon81@yuhs.ac

The trust of physicians in ML-based CDSS is essential for their application in clinical practice, which can be achieved through explainability for clinical relevance^{17,18}. Moreover, CDSS allow ED physicians as end-users to provide tailored control, specific to their clinical use. Nevertheless, the interpretability of ML-based CDSS is insufficient in their ED application. Limited studies have reflected the sequential processing of ED management³.

Therefore, we aimed to develop a practical ML-based CDSS for ED practice utilizing accessible clinical data according to the decision-making framework of physicians and to validate its clinical usefulness as a supportive tool in the ED.

Methods

Study design and setting. 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. By law, level 1 and 2 EDs must have 24 h/day staffing by board-certified emergency physicians¹⁹. During the study, an average of approximately 8000 adult patients visited the ED per month at the study site. Approximately 21% of the patients were admitted to the hospital from the ED, and an average of 2.5% patients were admitted to intensive care units (ICU) per month.

Upon arrival in this ED, vital signs and illness severity were assessed by a qualified triage nurse, and patients were shifted to appropriate treatment areas. The adult ED is divided into four different treatment areas: (1) 13 beds for patients who are extremely unstable and/or require intensive care and close monitoring, (2) 26 beds for patients who are medically stable but require continuous monitoring, (3) 20 beds for stable patients without monitoring, (4) and a fast track for simple evaluation or treatment. This study complied with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of Severance Hospital Human Research Protection Center, with a waiver for informed consent owing to its retrospective design and minimal harm to the patients (no. 4-2019-0555).

Selection of participants. Patients aged > 18 years who visited the ED were included. Patients with missing basic information, those who died on arrival, and those without any laboratory testing in the ED where clinical monitoring was deemed unnecessary or who decided to get discharged following simple and rapid treatment were excluded.

Data collection. Data were collected using a clinical research analysis portal system at the hospital's Digital Healthcare Department. Randomized identification numbers for research were used for each patient to anonymously extract clinical data. Data on the following variables were collected for all patients who visited the ED: age, sex, visit methods, traumatic or non-traumatic causes of visit, medical history such as hypertension, diabetes mellitus, tuberculosis, hepatitis, allergies and operation history, chief complaints and their duration, Korean Triage and Acuity Scale score, systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), body temperature (BT), respiratory rate (RR), and mental status. Data were extracted from the electronic medical records by both physicians and nurses upon arrival, and 27 fixed features were composed as fixed values for each patient during the ED stay. Changes in SBP, DBP, HR, BT, RR, and mental status, consent for "do not attempt resuscitation" order obtained during the ED stay, laboratory test results, and electrocardiogram results at the ED were extracted from the test result records and nursing records and used to derive 93 observation features. For the outcome variables, we extracted the dispositional order for decision to ICUs by an ED physician, interventions such as inotropes and vasopressors, intubation for airway maintenance or mechanical ventilation, in-hospital cardiac arrest (IHCA) requiring cardiopulmonary resuscitation, and the time of their occurrence. Consequently, 120 features (27 fixed features and 93 observation features) and four outcome variables were processed to build the prediction model (Table 1).

Data pre-processing. Fixed features, observation features, and outcome variables were combined to create a single data table for the analysis of each participant. The study period ranged from ED arrival as the starting point to the occurrence of outcomes for the four variables, each as an endpoint, such that it created several data

Features	Category	Number	Description
Fixed	Patient information	6	Sex, age, visit method(2), reason of visit(2)
	Past medical histories	7	Hypertension, diabetes mellitus, pulmonary tuberculosis, hepatitis viral carrier, allergies, medications, operations
	Chief complaints and duration	6	Chief complaints categorized into 15 groups; cardiovascular, pulmonary/respiratory, gastrointestinal organs, gastrointestinal bleeding, neurologic, genitourinary, obstetric/gynecological, musculoskeletal, ophthalmologic, ear/nose/throat, skin, orofacial, psychiatric, medical device related, other general issues. If multiple complaints exist in each patient, up to 3 categories (3) and its duration(3) were used
	Vital signs on arrival	8	Systolic blood pressure(2), diastolic blood pressure(2), respiratory rate, pulse rate, oxygen saturation, body temperature
Observation	Clinical observation	6	Systolic blood pressure, diastolic blood pressure, respiratory rate, pulse rate, oxygen saturation, body temperature
	Mental status monitoring	1	Change of mental status; alert, drowsy, stuporous, semi-comatose, comatose, confused and sedated
	Electrocardiogram diagnosis	23	Coronary(5), electrophysiologic(12), metabolic (5), unspecified(1)
	Electrocardiogram metrics	8	P axes, R axes, T axes, PR interval, QRS duration, QT, QTc, ventricular rate
	Laboratory test	55	Arterial blood gas analysis result(12), regular laboratory results(43)

Table 1. Summary of features.

labels with different timelines following multiple outcomes for each patient. Additionally, to analyze the data over time, we resampled the data labels into units with intervals of 1 h. We pre-processed the resampled 1 h-units without any additional observation features using the carry-forward method to fill the unit with the closest values in chronological order. This approach is grounded in the assumption that until a physician detects a change in a patient's condition, the previous condition remains unchanged, and decisions are made based on previous data until the next random observation. Physicians make comprehensive decisions regarding the frequency of vital sign measurements, or the provision of immediate emergency intervention based on the patient's physical examination and previously measured vital signs.

For the outliers in vital sign data, we set the upper and lower limits of each variable considering the physiological range. Subsequently, the dataset was split into the training and test sets at 2:3 and 1:3 ratios, respectively, ensuring sufficient data for training and testing purposes. Owing to the imbalanced characteristics of the dataset, we performed randomized under-sampling in the training set to compensate for a considerably small proportion of outcomes. A dataset for each outcome was separately under-sampled using specific ratios. To verify the performance of the model in an environment similar to that of the real world, we did not apply under-sampling to the test set (Fig. 1).

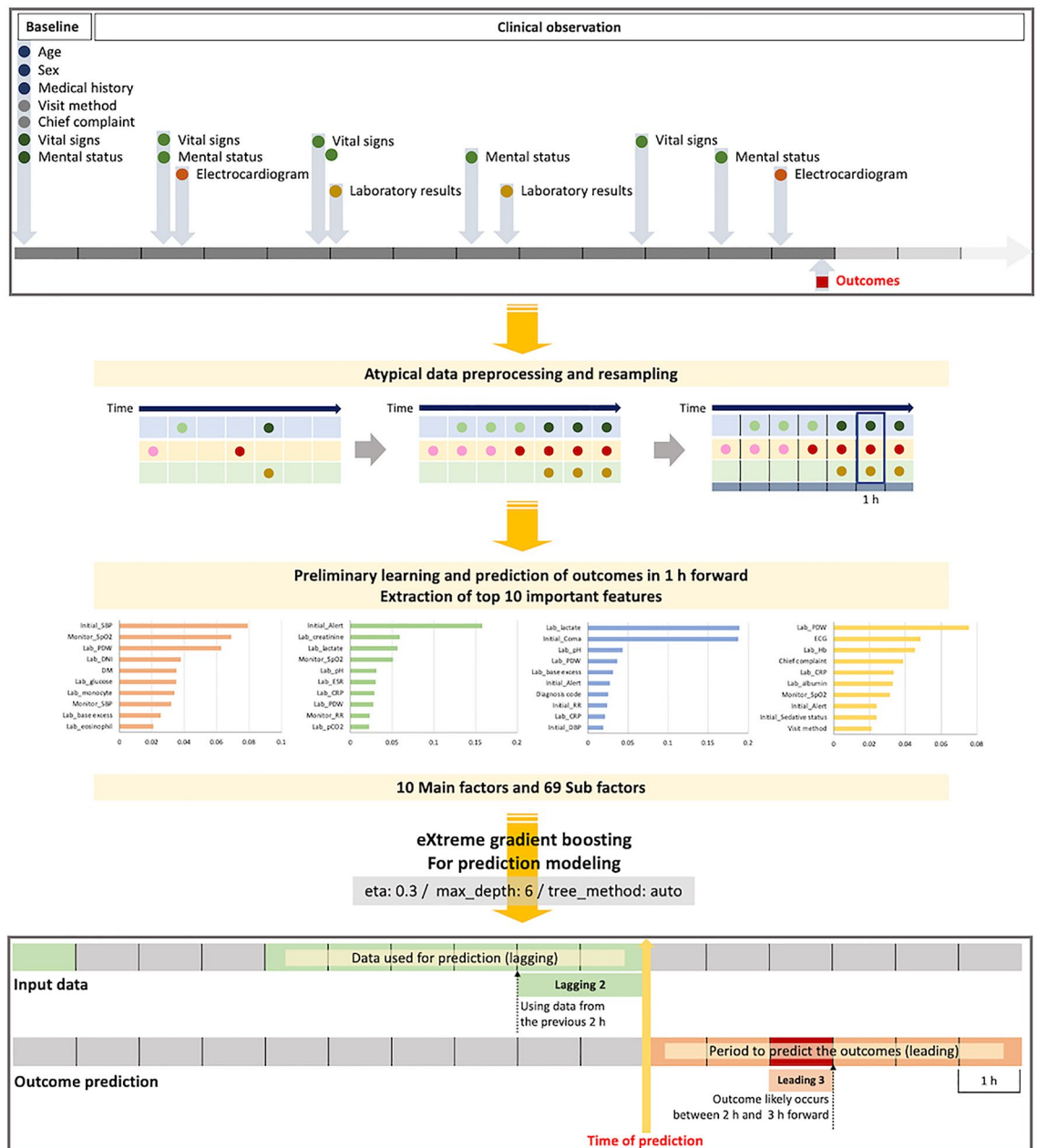


Figure 1. Data pre-processing and model development process.

Model development. eXtreme gradient boosting (XGBoosting) was used to develop the model. XGBoosting is an algorithm with a combination of good prediction performance and fast processing time, compared with other ML algorithms. The term "gradient boosting" refers to the improvement of a single weak model by combining several weak models to collectively produce a powerful model that minimizes errors by setting goals for the following models. This algorithm is used in a wide range of applications, including regression, classification, ranking, and user-defined prediction, as it minimizes bias and underfitting²⁰. First, we derived 10 important predictors that were highly associated with the outcome upon its occurrence. Next, we developed a baseline 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.

For example, in the case of inotropic use with leading 2 and lagging 3, the model predicted the use of inotropic medications between 1 and 2 h from the prediction time point using information collected in the previous 3 h (Fig. 1). The hyperparameters in each algorithm were tuned based on tenfold cross-validation during model development.

Outcome measures. The outcome measures included IHCA, inotropic use, intubation, and ICU admission. The selection of the four indicators for clinical deterioration to be predicted in the ED clinical setting was informed by previous studies, which identified these outcomes as the most critical^{21–24}. IHCA was defined as the development of witnessed or unwitnessed cardiac arrest with all causes within the ED. Inotropic use was defined as inotrope and vasopressor administration, such as noradrenaline, adrenaline, dopamine, dobutamine, or vasopressin, to overcome shock after adequate fluid administration. Intubation was defined as airway maintenance that protected the airway with positive pressure ventilation or conversion to a conventional ventilator from a home ventilator. ICU admission was defined as the physician's order of admission to the ICU, regardless of the patient's diagnosis.

Analysis. Continuous data were expressed as mean values with standard deviations. Categorical data were expressed as frequencies and percentages. All tests were two-sided, with a statistical significance of $P < 0.05$. To assess the model's performance, we used specificity, sensitivity, precision, F1 scores to assess the class imbalance, area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC). All statistical analyses were conducted with R Statistical Package (version 3.4.3) (www.R-project.org). Furthermore, we used XGBoosting (version 1.0.2) and Python (version 3.6.9, www.python.org) programming environment for experiments and modeling^{20,25}.

Results

Characteristics of study subjects. Of the 490,549 patients, 21,342 were excluded because of missing basic information. A total of 469,207 patients were enrolled. Of these, 165,862 patients without laboratory tests were excluded. Eventually, we used the data of 303,345 patients with 4,787,121 input data, which were resampled into 24,148,958 resampled 1 h-units. This input dataset was split into a training set and a test set in 2:3 and 1:3 ratios, respectively, for each outcome (Fig. 2). Figure 2 depicts the data processing flow and the number of data handled. Table 2 summarizes the baseline characteristics of the study population.

Main results. Figure 3 depicts the 10 important predictors that were derived for the baseline predictive model for each outcome in the order of the feature importance score. The SBP upon arrival for inotropic use, alert mental status for intubation, lactate levels in IHCA, and platelet distribution width in ICU admission displayed the highest feature importance score.

The AUROC of the model with lagging 6 and leading 0 displayed the highest value among all models derived with lagging from 0 to 6 and leading 0 to 6 to predict all four outcomes (Fig. 4). Specifically, the AUROC increased upon acquiring more information from the past (lagging), except for predicting IHCA. The change in AUROC of IHCA was the smallest for all four outcomes with increased lagging. In cases with inotropic use, intubation, and ICU admission, the range of AUROC change with the leading 6 was the highest according to different amounts of previous information (lagging). Supplementary Table S1 summarizes the test characteristics, such as the specificity, recall, F1 score, precision, AUROC, and AUPRC, according to each leading and lagging for the four outcomes. Moreover, to enhance interpretability, we utilized explainable artificial intelligence with Shapley values for each model and provided the results in Supplementary Fig. S1, which illustrates how the model arrived at its predictions²⁶. We subsequently performed external validation of the model using a separate dataset collected over a period of 2 years, and the results were shown in Supplementary Table S2 and Supplementary Fig. S2. Consistent with the findings of our internal validation, the external validation showed the same pattern, confirming our model's comparable predictive power and generalizability.

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% (Fig. 5).

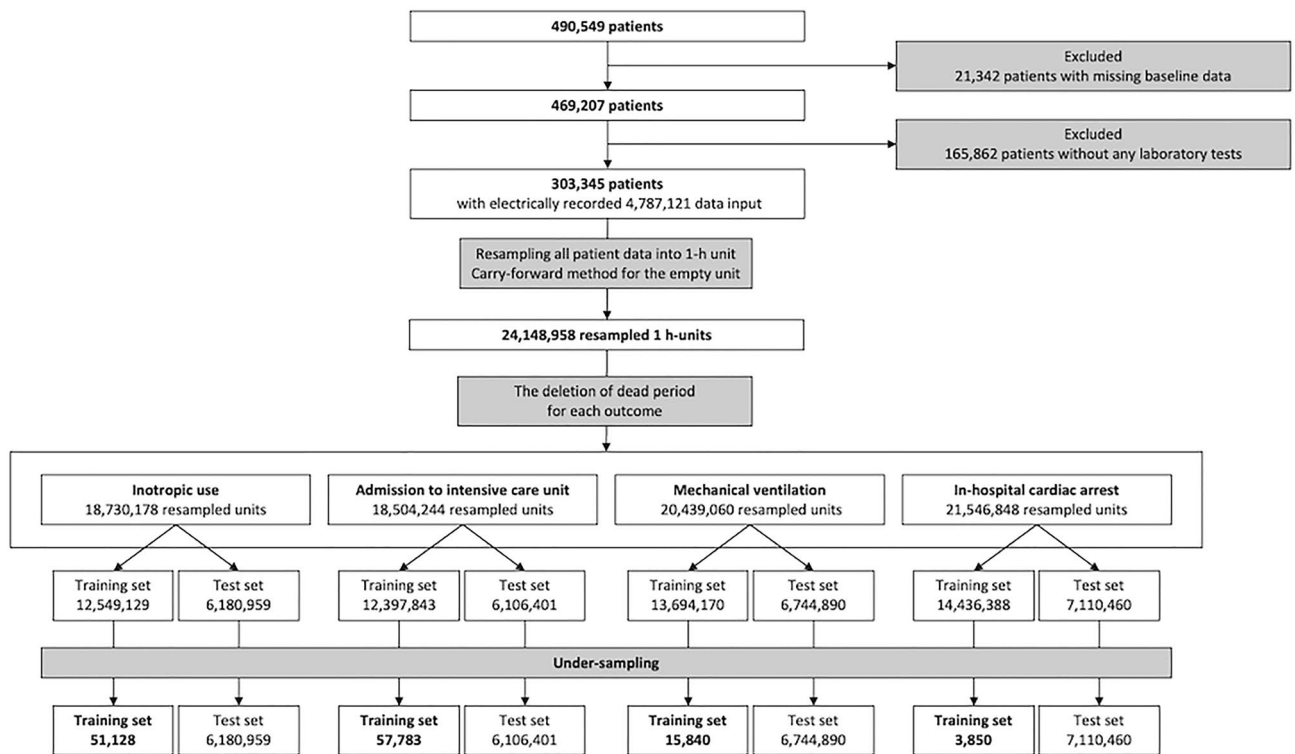


Figure 2. Flow chart for the study process.

Discussion

We developed a CDSS that helped ED physicians predict four critical deterioration events that should be detected pre-emptively and confirmed favorable prediction performances for each outcome. The prediction model developed in this study was designed to imitate the decision-making framework of an ED physician, which consists of three consecutive steps: observation, reasoning, and action²⁷. Generally, physicians collect clinical findings from patients examined in the ED and perform reasoning based on their empirical expertise. Clinical findings during arrival occasionally change along with the random addition of novel findings during stays in the ED. Thus, physicians should perform additional reasoning to detect critical events after updating the clinical findings. Following this framework, we trained the model to recognize and reason with recent inputs by updating every hour with all clinical findings that were sequentially added or changed during the ED stay. Through this training, the overall AUROC of our models for the three critical events, except cardiac arrest, were increased when the leading was shorter (predicting the near future), and the lagging was longer (prediction based on more sequential information). In addition, in predicting events after 6 h, the longer the lagging was (the more sequential the information), the more the AUROC had improved. In other words, sequential information is more important for improving the accuracy of predicting events in the distant future. Previous studies that predicted critical events used several fixed variables, such as vital signs, as the predictors^{17,28–31}. Contrarily, our prediction model was created by utilizing all clinical findings that occurred during ED practice and included laboratory tests and electrocardiograms, which have rarely been included as input variables in previous studies. Moreover, our study recognized updates in all clinical variables as recent observations. In practice, serially assessed clinical findings often contain missing values for various reasons and should be handled properly and efficiently³. Li et al. included patients with complete records to develop a machine-learning model to predict early mortality in the ED using electronic health records, potentially limiting the generalizability of their findings³². In contrast to this study, we assumed that ED physicians re-evaluated patients every hour and predicted critical events based on clinical findings in those reassessments. Therefore, the missing values observed in the 1 h-resampled unit were filled with the most recent values based on the physician's framework for decision-making. Thus, our model for ML-based CDSS was designed to imitate the sequential decision-making framework of ED physicians. Using all observations for referring to inferences in the real world, we attempted to induce ED physicians to use it in their practice. Moreover, we believe that our novel attempts can be extended to various clinical settings or outcome predictions with the same decision-making structure.

Models predicting clinical deterioration target two incompatible goals: the early detection of outcomes and fewer false-positive alerts to prevent alarm fatigue³³. In the ED, it could be more fatal for a critical outcome to occur in patients without a CDSS-predicted deterioration than a false CDSS alarm. The failure to predict critical outcomes during ED practice is associated with unexpected death, and false alarms cause wastage of medical resources^{34,35}. Unlike wards with relatively difficult access to human and material resources for critical patients, ED have easier access to these resources since they are staffed with resident medical personnel and equipped with resources for immediate resuscitation. Consequently, the negative impact of alarm fatigue is relatively less

	Values	Missing data (%)
Age (mean \pm SD)	52.21 \pm 20.00	0
Sex (male), n (%)	142,208 (46.9)	0
Visit method (EMS), n (%)	76,497 (25.2)	0
Past medical history, n (%)		
Hypertension	28.54	1.35
Diabetes mellitus	15.68	1.35
Pulmonary tuberculosis	2.54	1.36
Hepatitis	3.2	1.37
Allergies	2.23	1.37
Operative history	23.96	1.48
Initial vital signs at arrival (mean \pm SD)		
Systolic blood pressure (mmHg)	132.60 \pm 26.60	1.45
Diastolic blood pressure (mmHg)	77.49 \pm 14.72	1.45
Pulse rate (/min)	87.81 \pm 18.66	1.44
Respiratory rate (/min)	17.03 \pm 2.67	1.44
Oxygen saturation (%)	97.60 \pm 7.42	2.62
Body temperature ($^{\circ}$ C)	36.78 \pm 1.65	1.28
Mental status at arrival, n (%)		
Alert	291,868 (96.2)	0
Drowsy	5038 (1.7)	
Stupor	2361 (0.8)	
Semicoma	338 (0.1)	
Coma	468 (0.2)	
Sedated	3272 (1.1)	
Outcome, n (%)		
Intubation	2196 (0.7)	0
Inotropic use	6939 (2.3)	
ICU admission	7914 (2.6)	
IHCA	543 (0.0)	

Table 2. Baseline characteristics of the patients. *SD* standard deviation, *EMS* emergency medical services, *ICU* intensive care unit, *IHCA* in-hospital cardiac arrest.

than that for other units. Therefore, we demonstrated a change in the number of false alarms after adjusting the sensitivity for predicting significant events from 95 to 100%. These data can be used as a reference by ED physicians in selecting an acceptable rate of false alarms that increases in proportion to accurate predictions of catastrophic events. Additionally, physicians supported by the ML-based CDSS can apply the optimal threshold for each lagging and leading based on practicality and the clinical environment. Therefore, the model was developed as a practical tool that could be customized per clinical unit.

Previous studies on the prediction of clinical deterioration in the ED have generally set outcomes with longer time windows, such as in-hospital mortality within 30 days³⁶. In contrast, our study set prediction times of up to 6 h forward, that were appropriate to the ED. The reason for developing predictive models for the timely identification of patients at a high risk of clinical deterioration is to prioritize the point-of-care, effectively allocate resources, and prevent adverse outcomes¹⁷. However, previous ML-based predictive models have not been adequately tailored for context-specific patterns of care³⁷. The primary priority of ED physicians is to preemptively detect critical events that require immediate interventions and to make resource-allocation decisions during a patient's stay in the ED^{5,38}. Accordingly, we intended to develop a predictive model that could support physicians during point-of-care clinical decisions by presenting the outcomes within a relatively shorter time window. Since customization to the study setting is important for the feasibility of the predictive model, we set the time range based on the average ED length of stay in the study setting. Moreover, we determined three critical events requiring rapid resuscitation and ICU admission associated with a disposition decision were the most practical outcomes for developing an ML-based CDSS that supported ED physicians' practice.

For an application to the medical field, researchers should prepare ML-based CDSS for clinically actionable explanations. The physicians did not trust the predictions when the logic behind those was unclear^{17,39–42}. We attempted to increase the explainability to physicians by presenting features that were highly influential in predicting the occurrence of the four critical events in our study and found that the features of the influence of each event were different. Additionally, we used an imbalanced dataset with a low event occurrence frequency. Therefore, we performed an under-sampling method and presented the results with AUROC and AUPRC as the performance metrics. AUPRC is a relative indicator because the baseline comprises the fraction of positives in the population⁴³. The baseline AUPRC ranged from 0.000006 to 0.000424 according to each outcome, thereby

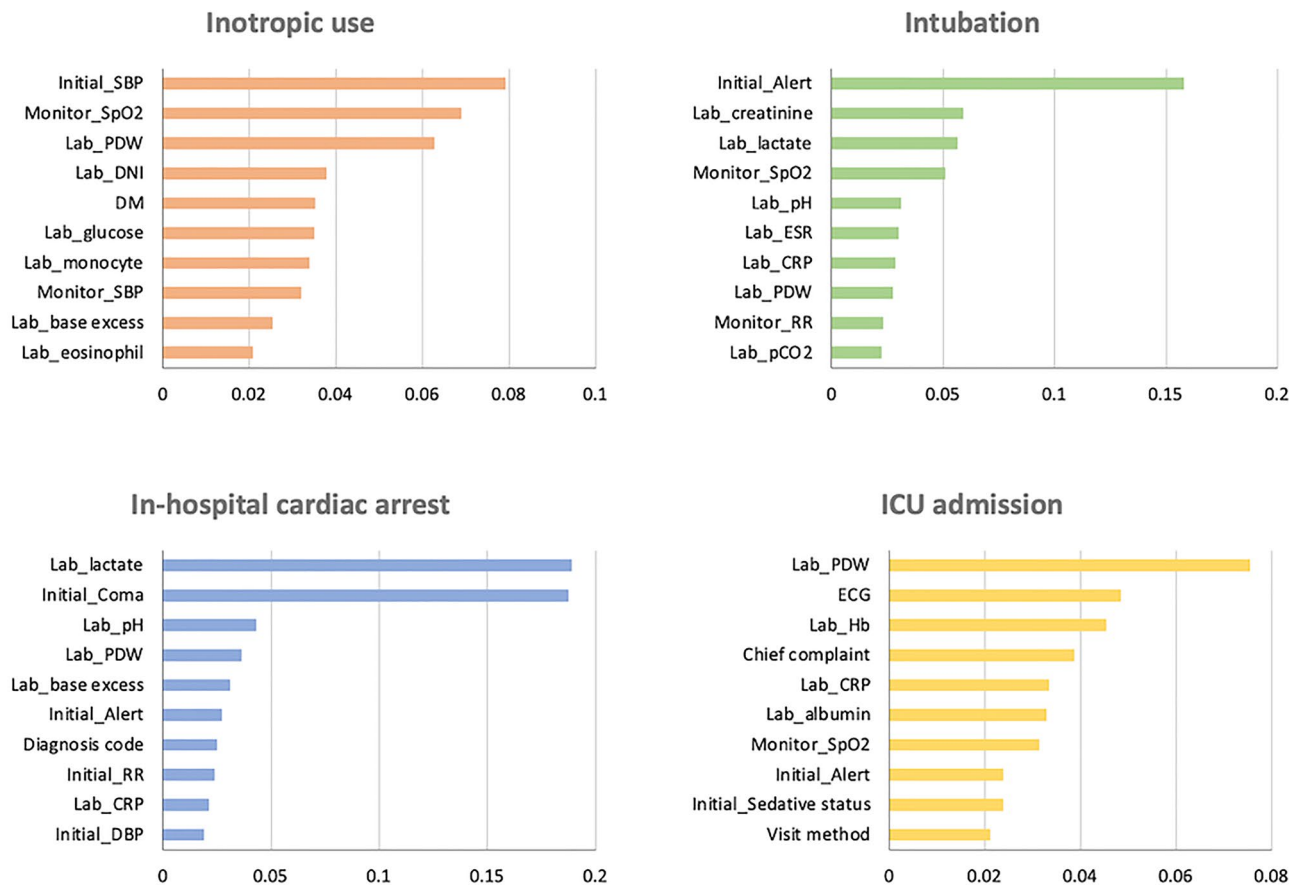


Figure 3. Top 10 important features in predicting each outcome. *ICU* intensive care unit, *SBP* systolic blood pressure, *SpO₂* oxygen saturation, *PDW* platelet distribution width, *DNI* delta neutrophil index, *DM* diabetes mellitus, *ESR* erythrocyte sedimentation rate, *CRP* C-reactive protein, *RR* respiratory rate, *DBP* diastolic blood pressure, *ICU* intensive care unit, *ECG* electrocardiogram, *Hb* hemoglobin.

representing an extremely small proportion of outcome occurrences in this study. However, the AUPRC in the present model showed superior values to the baseline AUPRC, with a maximum of 0.113, 0.234, 0.214, and 0.032 for cardiac arrest, inotropic use, intubation, and ICU admission, respectively.

The present study had several limitations. We selected the XGBoosting algorithm considering the study design using the tabular dataset because of its ease of use and overall good processing performance⁴⁴. However, the process could be applied using other algorithms to develop predictive models that supposedly perform better. Second, the present study had a retrospective design, which presents a possibility of potential bias. Particularly, the medical staff manually updated the clinical findings of our dataset and did not record the findings that they were unaware of. This limitation can be compensated for by applying a real-time patient information acquisition system in clinical practice to obtain the dataset. Therefore, additional studies that prospectively evaluate the model using real-time datasets should be continued to prove the feasibility of our predictive model in daily ED practice. Third, the management of patients in a clinical setting may affect the study outcomes; however, our predictive model did not include this feature as an input. This necessitates expanding the modeling to include the clinical efforts of physicians in the future.

In summary, we developed a practical ML-based CDSS for ED practice by utilizing accessible clinical data according to the decision-making framework of physicians. CDSS should be customized according to clinical situations, and ML algorithms can help improve their performance.

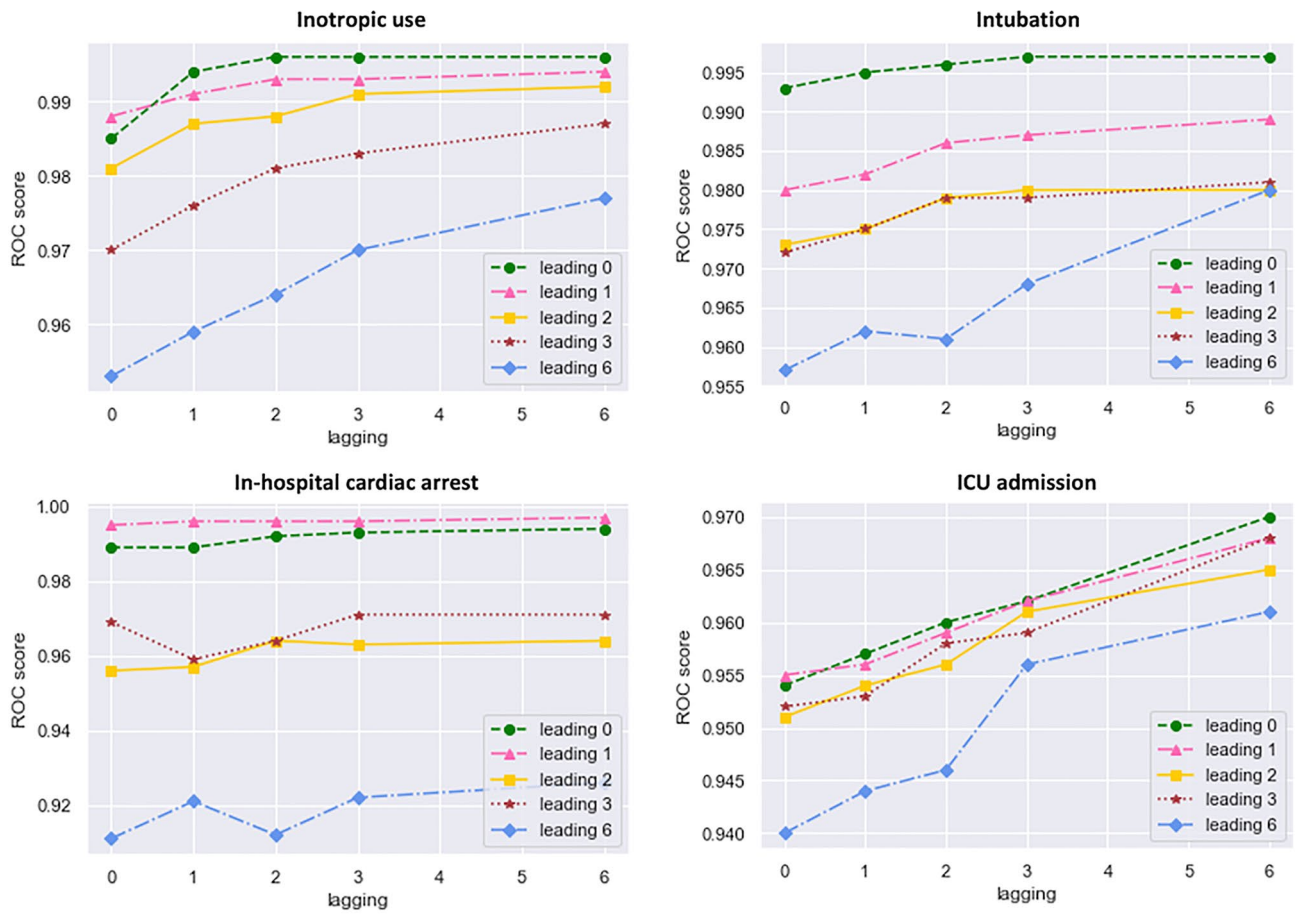


Figure 4. Area under receiver operating curve value for each outcome. ROC receiver operating curve, ICU intensive care unit.

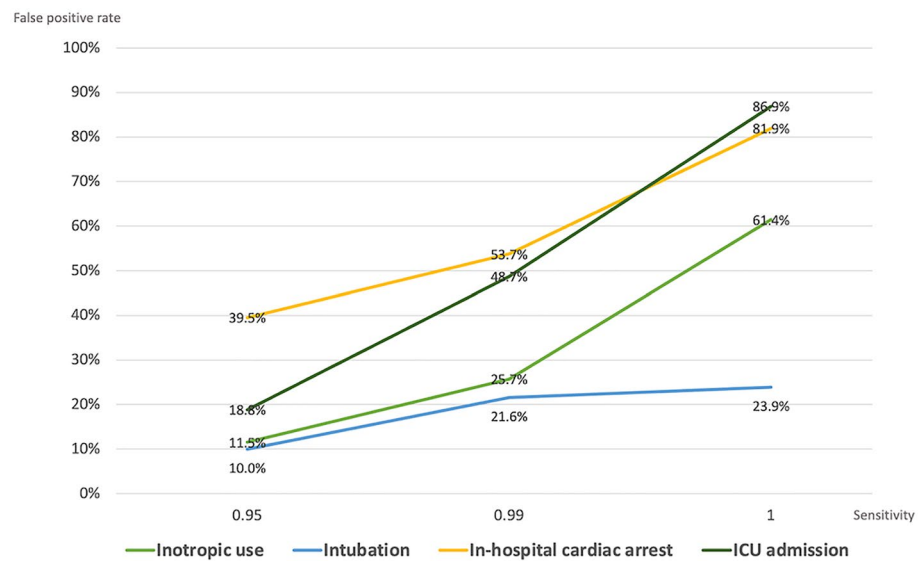


Figure 5. Fraction of false-positive cases in the model with leading 6 and lagging 6 for each outcome. ICU intensive care unit.

Data availability

Data are available from the corresponding author upon reasonable request.

Received: 22 January 2023; Accepted: 21 May 2023

Published online: 26 May 2023

References

- Osheroff, J. A. *et al.* A roadmap for national action on clinical decision support. *J. Am. Med. Inform. Assoc.* **14**, 141–145. <https://doi.org/10.1197/jamia.M2334> (2007).
- Sutton, R. T. *et al.* An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digit. Med.* **3**, 17. <https://doi.org/10.1038/s41746-020-0221-y> (2020).
- Hong, S., Lee, S., Lee, J., Cha, W. C. & Kim, K. Prediction of cardiac arrest in the emergency department based on machine learning and sequential characteristics: Model development and retrospective clinical validation study. *JMIR Med. Inform.* **8**, e15932. <https://doi.org/10.2196/15932> (2020).
- Vogel, S. *et al.* Development of a clinical decision support system for smart algorithms in emergency medicine. *Stud. Health Technol. Inform.* **289**, 224–227. <https://doi.org/10.3233/SHTI210900> (2022).
- Fernandes, M. *et al.* Clinical decision support systems for triage in the emergency department using intelligent systems: A review. *Artif. Intell. Med.* **102**, 101762. <https://doi.org/10.1016/j.artmed.2019.101762> (2020).
- Kwon, J. M. *et al.* Validation of deep-learning-based triage and acuity score using a large national dataset. *PLoS ONE* **13**, e0205836. <https://doi.org/10.1371/journal.pone.0205836> (2018).
- Mendo, I. R., Marques, G., de la Torre Diez, I., Lopez-Coronado, M. & Martin-Rodriguez, F. Machine learning in medical emergencies: A systematic review and analysis. *J. Med. Syst.* **45**, 88. <https://doi.org/10.1007/s10916-021-01762-3> (2021).
- Levin, S. *et al.* Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. *Ann. Emerg. Med.* **71**, 565–574 e562. <https://doi.org/10.1016/j.annemergmed.2017.08.005> (2018).
- Eitel, D. R., Rudkin, S. E., Malvey, M. A., Killeen, J. P. & Pines, J. M. Improving service quality by understanding emergency department flow: A White Paper and position statement prepared for the American Academy of Emergency Medicine. *J. Emerg. Med.* **38**, 70–79. <https://doi.org/10.1016/j.jemermed.2008.03.038> (2010).
- Hwang, U. *et al.* Emergency department patient volume and troponin laboratory turnaround time. *Acad. Emerg. Med.* **17**, 501–507. <https://doi.org/10.1111/j.1553-2712.2010.00738.x> (2010).
- Kulstad, E. B., Sikka, R., Sweis, R. T., Kelley, K. M. & Rzechula, K. H. ED overcrowding is associated with an increased frequency of medication errors. *Am. J. Emerg. Med.* **28**, 304–309. <https://doi.org/10.1016/j.ajem.2008.12.014> (2010).
- Mills, A. M., Shofer, F. S., Chen, E. H., Hollander, J. E. & Pines, J. M. The association between emergency department crowding and analgesia administration in acute abdominal pain patients. *Acad. Emerg. Med.* **16**, 603–608. <https://doi.org/10.1111/j.1553-2712.2009.00441.x> (2009).
- Jo, S. *et al.* ED crowding is associated with inpatient mortality among critically ill patients admitted via the ED: Post hoc analysis from a retrospective study. *Am. J. Emerg. Med.* **33**, 1725–1731. <https://doi.org/10.1016/j.ajem.2015.08.004> (2015).
- Elliott, D. J. *et al.* An interdepartmental care model to expedite admission from the emergency department to the medical ICU. *Jt. Commun. J. Qual. Patient Saf.* **41**, 542–549. [https://doi.org/10.1016/s1553-7250\(15\)41071-2](https://doi.org/10.1016/s1553-7250(15)41071-2) (2015).
- Chalfin, D. B. *et al.* Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Crit. Care Med.* **35**, 1477–1483. <https://doi.org/10.1097/01.CCM.0000266585.74905.5A> (2007).
- Tsai, M. T. *et al.* The influence of emergency department crowding on the efficiency of care for acute stroke patients. *Int. J. Qual. Health Care* **28**, 774–778. <https://doi.org/10.1093/intqhc/mzw109> (2016).
- Muralitharan, S. *et al.* Machine learning-based early warning systems for clinical deterioration: Systematic scoping review. *J. Med. Internet Res.* **23**, e25187. <https://doi.org/10.2196/25187> (2021).
- Gao, H. *et al.* Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med.* **33**, 667–679. <https://doi.org/10.1007/s00134-007-0532-3> (2007).
- Ahn, K. O. *et al.* Association between deprivation status at community level and outcomes from out-of-hospital cardiac arrest: A nationwide observational study. *Resuscitation* **82**, 270–276. <https://doi.org/10.1016/j.resuscitation.2010.10.023> (2011).
- Chen, T. Q., Guestrin, C. XGBoost: A scalable tree boosting system. In *KDD'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* 785–94 (2016). <https://doi.org/10.1145/2939672.2939785>
- Goulden, R. *et al.* qSOFA, SIRS and NEWS for predicting in-hospital mortality and ICU admission in emergency admissions treated as sepsis. *Emerg. Med. J.* **35**, 345–349. <https://doi.org/10.1136/emmermed-2017-207120> (2018).
- Gearhart, A. M., Furmanek, S., English, C., Ramirez, J. & Cavallazzi, R. Predicting the need for ICU admission in community-acquired pneumonia. *Respir. Med.* **155**, 61–65. <https://doi.org/10.1016/j.rmed.2019.07.007> (2019).
- Sun, J. T. *et al.* External validation of a triage tool for predicting cardiac arrest in the emergency department. *Sci. Rep.* **12**, 8779. <https://doi.org/10.1038/s41598-022-12781-6> (2022).
- Nakwan, N. & Prateepchaiboon, T. Predicting the requiring intubation and invasive mechanical ventilation among asthmatic exacerbation-related hospitalizations. *J. Asthma* **59**, 507–513. <https://doi.org/10.1080/02770903.2020.1853768> (2022).
- Van Rossum, G. & Drake, F. L. *The Python Language Reference Manual*. (Network Theory Limited, 2011).
- Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67. <https://doi.org/10.1038/s42256-019-0138-9> (2020).
- Wilk, S. *et al.* A task-based support architecture for developing point-of-care clinical decision support systems for the emergency department. *Methods Inf. Med.* **52**, 18–32. <https://doi.org/10.3414/ME11-01-0099> (2013).
- Wu, T. T., Lin, X. Q., Mu, Y., Li, H. & Guo, Y. S. Machine learning for early prediction of in-hospital cardiac arrest in patients with acute coronary syndromes. *Clin. Cardiol.* **44**, 349–356. <https://doi.org/10.1002/clc.23541> (2021).
- Barnes, S., Saria, S. & Levin, S. An evolutionary computation approach for optimizing multilevel data to predict patient outcomes. *J. Healthc. Eng.* **2018**, 7174803. <https://doi.org/10.1155/2018/7174803> (2018).
- Henning, D. J. *et al.* Derivation and validation of predictive factors for clinical deterioration after admission in emergency department patients presenting with abnormal vital signs without shock. *West J. Emerg. Med.* **16**, 1059–1066. <https://doi.org/10.5811/westjem.2015.9.27348> (2015).
- Raita, Y. *et al.* Emergency department triage prediction of clinical outcomes using machine learning models. *Crit. Care* **23**, 64. <https://doi.org/10.1186/s13054-019-2351-7> (2019).
- Li, C. *et al.* Machine learning based early mortality prediction in the emergency department. *Int. J. Med. Inform.* **155**, 104570. <https://doi.org/10.1016/j.ijmedinf.2021.104570> (2021).
- Guillame-Bert, M. *et al.* Learning temporal rules to forecast instability in continuously monitored patients. *J. Am. Med. Inform. Assoc.* **24**, 47–53. <https://doi.org/10.1093/jamia/ocw048> (2017).
- Leenen, J. P. L. *et al.* Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: Systematic review. *J. Med. Internet Res.* **22**, e18636. <https://doi.org/10.2196/18636> (2020).

35. Green, M. *et al.* Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. *Resuscitation* **123**, 86–91. <https://doi.org/10.1016/j.resuscitation.2017.10.028> (2018).
36. Guan, G., Lee, C. M. Y., Begg, S., Crombie, A. & Mnatzaganian, G. The use of early warning system scores in prehospital and emergency department settings to predict clinical deterioration: A systematic review and meta-analysis. *PLoS ONE* **17**, e0265559. <https://doi.org/10.1371/journal.pone.0265559> (2022).
37. Petitgand, C., Motulsky, A., Denis, J. L. & Regis, C. Investigating the barriers to physician adoption of an artificial intelligence-based decision support system in emergency care: An interpretative qualitative study. *Stud. Health Technol. Inform.* **270**, 1001–1005. <https://doi.org/10.3233/SHTI200312> (2020).
38. Hoot, N. R. & Aronsky, D. Systematic review of emergency department crowding: Causes, effects, and solutions. *Ann. Emerg. Med.* **52**, 126–136. <https://doi.org/10.1016/j.annemergmed.2008.03.014> (2008).
39. Sandhu, S. *et al.* Integrating a machine learning system into clinical workflows: Qualitative study. *J. Med. Internet Res.* **22**, e22421. <https://doi.org/10.2196/22421> (2020).
40. Stroh, L., Hehakaya, C., Ranschaert, E. R., Boon, W. P. C. & Moors, E. H. M. Implementation of artificial intelligence (AI) applications in radiology: Hindering and facilitating factors. *Eur. Radiol.* **30**, 5525–5532. <https://doi.org/10.1007/s00330-020-06946-y> (2020).
41. Romero-Brufau, S. *et al.* A lesson in implementation: A pre-post study of providers' experience with artificial intelligence-based clinical decision support. *Int. J. Med. Inform.* **137**, 104072. <https://doi.org/10.1016/j.ijmedinf.2019.104072> (2020).
42. Schwartz, J. M. *et al.* Factors influencing clinician trust in predictive clinical decision support systems for in-hospital deterioration: Qualitative descriptive study. *JMIR Hum. Factors* **9**, e33960. <https://doi.org/10.2196/33960> (2022).
43. Saito, T. & Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **10**, e0118432. <https://doi.org/10.1371/journal.pone.0118432> (2015).
44. Shwartz-Ziv, R. & Armon, A. Tabular data: Deep learning is not all you need. *Inf. Fusion* **81**, 84–90. <https://doi.org/10.1016/j.inffus.2021.11.011> (2022).

Acknowledgements

The authors thank Sang Wook Park, B.S., Eunmin Lim, M.S., and Jingu Choi, M.S., from C&M Standard Lab., LG electronics for their technical support.

Author contributions

A.C. and J.H.K. conceived and designed the study, and J.H.K. and K.C. obtained research funding. A.C. and J.H.K. supervised the conduct of the study and data collection. A.C. and S.Y.C. managed the data, including quality control. A.C. and J.H.K. was in charge of initial data mining and cleansing, H.S.C. supervised the statistical analysis section, and T.S. and B.C. were in charge of machine-learning analysis. A.C. and J.H.K. drafted the manuscript, and all authors contributed substantially to its revision. J.H.K. assumes the overall responsibility for the study and the manuscript.

Funding

This research was funded by the Korea Medical Device Development Fund by the Korean government (the MSIT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety (Project number: KMDF_PR_20200901_0089, 9991006762)), by the Medical Data-Driven Hospital Support Project through the Korea Health Information Service (KHIS), funded by the Ministry of Health & Welfare, Republic of Korea and also supported by the Severance Hospital Research fund for Clinical excellence (Grant numbers C-2022-0010).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-35617-3>.

Correspondence and requests for materials should be addressed to J.H.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023