Heliyon 9 (2023) e19210

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Clinical support system for triage based on federated learning for the Korea triage and acuity scale

Hansol Chang[a,b,1], Jae Yong Yu[c,1], Geun Hyeong Lee[b,d], Sejin Heo[a,b], Se Uk Lee[a], Sung Yeon Hwang[a], Hee Yoon[a], Won Chul Cha[a,b,e], Tae Gun Shin[a], Min Seob Sim[a], Ik Joon Jo[a], Taerim Kim[a,b,*]

[a] Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 115 Irwon-ro Gangnam-gu, Seoul, 06355, South Korea
[b] Department of Digital Health, Samsung Advanced Institute for Health Science & Technology (SAIHST), Sungkyunkwan University, 115 Irwon-ro Gangnam-gu, Seoul, 06355, South Korea
[c] Department of Biomedical System Informatics, Yonsei University College of Medicine, Seoul, South Korea
[d] Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon 16419, South Korea
[e] Digital Innovation Center, Samsung Medical Center, Seoul, Korea. 81 Irwon-ro Gangnam-gu, Seoul 06351, South Korea

A R T I C L E I N F O

Keywords:
Triage
Emergency department
Emergency medical service
Clinical decision-making
Machine learning

A B S T R A C T

Background and aims: This study developed a clinical support system based on federated learning to predict the need for a revised Korea Triage Acuity Scale (KTAS) to facilitate triage.
Methods: This was a retrospective study that used data from 11,952,887 patients in the Korean National Emergency Department Information System (NEDIS) from 2016 to 2018 for model development. Separate cohorts were created based on the emergency medical center level in the NEDIS: regional emergency medical center (REMC), local emergency medical center (LEMC), and local emergency medical institution (LEMI). External and temporal validation used data from emergency department (ED) of the study site from 2019 to 2021. Patient features obtained during the triage process and the initial KTAS scores were used to develop the prediction model. Federated learning was used to rectify the disparity in data quality between EDs. The patient's demographic information, vital signs in triage, mental status, arrival information, and initial KTAS were included in the input feature.
Results: 3,626,154 patients' visits were included in the regional emergency medical center cohort; 8,278,081 patients' visits were included in the local emergency medical center cohort; and 48,652 patients' visits were included in the local emergency medical institution cohort. The study site cohort, which is used for external and temporal validation, included 135,780 patients visits. Among the patients in the REMC and study site cohorts, KTAS level 3 patients accounted for the highest proportion at 42.4% and 45.1%, respectively, whereas in the LEMC and LEMI cohorts, KTAS level 4 patients accounted for the highest proportion. The area under the receiver operating characteristic curve for the prediction model was 0.786, 0.750, and 0.770 in the external and temporal validation. Patients with revised KTAS scores had a higher admission rate and ED mortality rate than those with unaltered KTAS scores.

* Corresponding author. Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, (06355) 115 Irwon-ro Gangnam-gu, Seoul, South Korea.
E-mail address: taerim.j.kim@gmail.com (T. Kim).
[1] Hansol Chang and Jae Yong Yu contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2023.e19210
Received 8 May 2023; Received in revised form 11 August 2023; Accepted 16 August 2023
Available online 17 August 2023

Conclusions: This novel system might accurately predict the likelihood of KTAS acuity revision and support clinician-based triage.

1. Introduction

Triage is a strategy for rapidly classifying patients upon their arrival in an emergency department (ED) to identify the urgency of further assessment or care [1,2]. Promptly classifying and assigning patients to the appropriate resources in the proper location is critical to the overall operation of an ED [3–5]. Furthermore, because the triage system helps classify patients with potentially deteriorated status, timely and accurate triaging of patients is crucial to patient outcomes [6–8].

While triage skills can be acquired through training and education, accuracy can vary based on the nurse's or physician's experience or emergency department (ED) crowding [2,9–11]. Such systems are often challenging in actual practice, when both time and information are limited and patients have various medical conditions, so triage staff often rely on their intuition and clinical experience with triage [12]. Several studies reported reasons for re-evaluation of triage [13–15]. Over-triage exhausts ED resources prematurely, and under-triage of patients can cause delays in proper treatment, prolonged length of stay (LOS), and compromised patient safety [5,16–19].

Several studies have applied deep learning (DL) or other artificial intelligence-based models on triage acuity scales for better prediction [12,20–23]. These studies on DL-based triage and acuity scores predicted in-hospital mortality and hospitalization during triage [20,22,24–28]. However, there have been a few attempts to predict miss-triaged patients at triage. Furthermore, data quality of local emergency centers is often compromising due to each institution's manpower, size and condition of medical center [29]. Difference of data quality among each medical center's levels causes difficulty in conducting conventional validation.

Recently, federated learning (FL) has been studied in the medical field to protect patient data, satisfy hospital security policies, make it possible to apply machine learning to real multi-institutional training data, and overcome differences in data quality among institutions [30–34]. The FL process requires a server–client structure [33]. The client (each institution) sends a weight of training results to an aggregation server without raw data exposure. The aggregation server collects the weights from each institution and then performs federated averaging to average the weight and update the new global model [34]. After that, each institution downloads a new global model from the aggregation server, performs validation, and repeats the local training process.

For this study, we developed a federated learning–based clinical support system to assist clinicians in using a five-level triage acuity scale, the Korea Triage Acuity Scale (KTAS), by predicting the probability of revised KTAS acuity in advance and classifying presumed severe patients more accurately to provide them with timely care.

2. Methods

2.1. Study setting and population

We conducted a retrospective study using the National Emergency Department Information System (NEDIS) database for model derivation. NEDIS collects data from patients who visit 151 EDs in Korea in real time. For this study, patient information was included from January 1, 2016, to December 31, 2018. We additionally collected data from the ED of a 1,989-bed tertiary referral hospital from 2019 to 2021 for temporal validation. We excluded patients whose visits did not involve treatment (patient registration for medical certification or prescription); who left without being seen; or were younger than 20 years from both populations. Patients with missing triage information, including KTAS score, mental status, discharge date, or mode of arrival, were also excluded.

This study was approved by the Institutional Review Board (IRB) of the Samsung Medical Center (IRB No. 2021-03-198). The need for informed consent was waived by the institutional Review Board (IRB) of the Samsung Medical Center Samsung Medical Center because of the retrospective, observational, and anonymous nature of the study. All methods were performed in accordance with the relevant guidelines and regulations. It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

2.2. Korean Triage Acuity Scale

The KTAS is a nationwide triage tool in Korea created based on the Canadian Triage Acuity Scale and the five-level triage system [8,35]. KTAS is a 5-level triage and acuity scale with the assessor selecting from a chief complaint list and then selecting the primary modifiers or special modifiers to assign an appropriate score. It uses 155 main complaints for adults and 165 complaints for children. After complaints are selected, common and general characteristics of all complaints are the primary consideration. Secondary considerations are the specific characteristics of each complaint [8,15]. The KTAS includes both severity and urgency [8,36]. Priority of care is determined on the basis of KTAS classification results. The KTAS is used by nurses, emergency medical technicians, or doctors who have been certified by The Korean Society of Emergency Medicine [8,36].

2.3. Comparison of ED clinical outcomes between rKTAS and uKTAS

Patients whose KTAS scores were revised were categorized into the revised KTAS acuity (rKTAS) group, whereas patients with unchanged KTAS acuity were categorized into the unchanged KTAS acuity (uKTAS) group. According to a previous study, revision of KTAS acuity is performed for patients who are thought to be under- or over-triaged, which can occur for several reasons [13,14]. Secondary outcomes included LOS, discharge to home, ED death, ward admission, and transfer after ED, and these outcomes were compared between the KTAS rKTAS and uKTAS groups [37]. LOS is expressed as median with interquartile range (IQR), and ED results are expressed as counts and percentages. We compared the secondary outcomes between patients in the rKTAS and uKTAS group based on initial KTAS. We sought to identify differences between rKTAS and uKTAS clinical outcomes to demonstrate the importance of predicting the probability of KTAS acuity revision.

2.4. Prediction outcomes

This research attempted to predict whether original KTAS scores would be revised prior to ED discharge [38]. Initial KTAS and final KTAS are both documented in the NEDIS. As a result of many stages, some acuity may be reassessed and if indicated, revised. For instance, nurses can re-evaluate the KTAS scale immediately after the first triage scale decision if they thought that initial KTAS was under- or overestimated, and some KTAS adjustments occur during the ED stage due to patient circumstances [13,14]. Therefore, we sought to predict the possibility of KTAS acuity revision to support initial triage and minimize under- or overestimation of initial triage stage.

2.5. Data source

Patient data (sex, age, vital signs, ED visits, and derived features) were used for the modeling. All the elements could be collected during the initial triage time, which is measured initially and is mandatory when patients visit an ED. Demographic information (age and sex), vital signs (temperature, heart rate, systolic blood pressure, respiratory rate, oxygen saturation), ED visit information (disease onset to ED visit time, initial and final KTAS scores) were extracted from the clinical data warehouse of the study site and NEDIS.

2.6. Cohort development based on the level system used for Korean emergency departments

Separate cohorts for model development were made according to the level of each emergency medical center in the NEDIS database. In Korea, emergency medical centers are divided into three levels: regional emergency medical centers (REMCs), local emergency medical centers (LEMCs), and local emergency medical institutions (LEMIs) [39]. Those levels reflect the size and number of back-up medical personnel and available departments, so patient characteristics also differ by level. Therefore, cohort were made for emergency medical centers at each level for additional data quality analysis, population comparison, and model validation. After conducting conventional validation, federated learning was added to account for the differences in data quality among medical center levels.

2.7. Data preparation

We built a prediction model to quantify the probability of the primary outcome. All data processing and statistical analyses were conducted using R software, version 3.6.1, and Python, version 3.6.8. We used the following information from patients who visited an ED from 2016 to 2018 as input variables for the model: age, sex, systolic blood pressure, diastolic blood pressure, pulse rate, oxygen saturation, body temperature, initial KTAS score, route of arrival, and method of transportation.

Age was dichotomized as patients older and younger than 65. The alert, response to verbal output, response to pain, and unresponsive (AVPU) scale was used to quantify mental condition. The AVPU scale was utilized instead of the GCS since it saves time and was demonstrated to be effective in multiple prior studies [40–42]. In addition, the NEDIS only collects AVPU data. The reason for the ED visit was categorized as disease or trauma. Vital signs were categorized as normal or abnormal. The normal ranges for vital signs were systolic blood pressure, 100–150 mmHg; diastolic blood pressure, 60–90 mmHg; pulse rate, 50–100 bpm; respiratory rate, 12–20 breaths/min; body temperature, 36–37.5 °C; SpO2, 95%–100% [43–45]. The visit time was categorized in three groups based on shift changes for the nurses: 07:00 to 14:59, 15:00 to 22:59, and 23:00 to 06:59. Visit date was categorized as weekdays or weekend. The method of transportation to the ED was categorized as the 119 (number of emergency services in Korea) group, private ambulance group, and other group, which includes walk-in patients. The route of arrival indicates whether the patient visited the ED directly (without another hospital visit) or transferred from another hospital.

2.8. Machine learning

We used four modeling methods. First, a multivariate logistic regression analysis estimated the likelihood of clinical outcomes after adjusting for other potential factors. Next, machine-learning methods known to be good for classification (random forest, XGBoost, and DL with the Python packages "Skitlearn" and "Tensorflow") were used with the following hyper-parameters: number of total trees in the random forest and the number of layers and number of hidden units which were validated and selected using the validation set.

We used the FL-based machine-learning approach to predict whether patients were under- or over-classified during triage in the ED using NEDIS data. The FL approach can be used to build an optimal model by sharing only the weights for each institution without needing to share raw data. However, FL does require the standardization of patient data for the model. Fortunately, NEDIS already uses a standardized format for its whole nationwide ED database. Therefore, additional standardization was unnecessary.

In order to perform FL, a client that performs learning in an actual institution and a server that collects learning results of the client are required. The client is responsible for performing learning on its own by each institution and can transmit the learning result weight to the server. The server collects weight values sent from each institution and calculates the federated average. This averages the total weight values and, through this, creates a new global model with advantages over previous models. After that, the client downloads this global model and proceeds with re-learning. This is called a round, and it is FL to improve the global model by repeating the round.

The weight values exchanged between the server and the client are numerical values generated as a result of learning and are generally meaningless. This has the advantage of protecting personal information because the original data is not shared.

We used a simple artificial neural network with two parts. The first block contained a fully connected layer and a dropout layer and used the Rectified Linear Unit activation function with 5 repetitions. The second block was the output layer and used the softmax activation function. The number of units for each layer was 64, 128, 256, 128, 64 and 3, with a 0.3 dropout rate. We used the stochastic gradient descent for the optimizer with a learning rate of 0.01. We also considered 5 epochs and a batch size of 64 for local learning; we performed 20 rounds through 3 individual clients for the FL.

2.9. *Model performance measure and validation*

To evaluate model discrimination, we plotted the receiver operating characteristic (ROC) and precision recall (PR) curves and calculated the area under the ROC curve (AUROC), which is informative when the outcome is unbalanced because it depends on the outcome prevalence. The model calibration plot was assessed by comparing the predicted and empirical probabilities.

We performed validation using the NEDIS population at each ED level in the NEDIS database. Additionally, we performed temporal validation with information from the study site about patients who visited the ED from 2019 to 2021.

2.10. *Statistical analysis*

Descriptive statistics are provided for the demographic features and characteristics of ED visits. Categorical variables are expressed as counts and percentages of the total data available within the database. We compared initial KTAS score at triage with KTAS score at discharge to identify patterns in outcomes. We performed univariable and multivariable analyses with logistic regression method to evaluate the risk factors for revision of KTAS scores to further explain differences in the association between features and outcomes. $P < 0.01$ was considered statistically significant for all statistical tests. All data processing and statistical analyses were conducted using R, version 3.6.1.

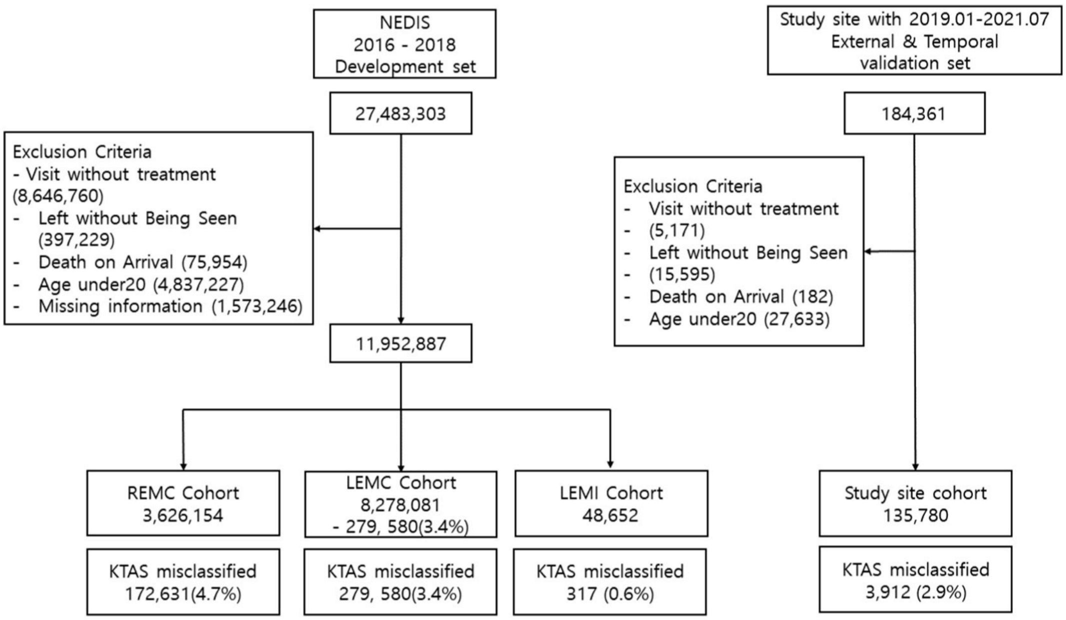

Fig. 1. Flowchart of the study participant selection process and the deviation in each cohort set. NEDIS: National Emergency Department Information System, REMC: Regional Emergency Medical Center, LEMC: Local Emergency Medical Center, LEMI: Local Emergency Medical Institution, KTAS: Korean Triage Acuity Scale.

3. Results

3.1. Population

The initial NEDIS data included 27,483,303 ED visits from January 2016 to December 2018. Data for patients who visited without treatment (8,646,760), left without being seen (397,229), dead on arrival (75,954), were younger than 20 years (4,837,227) and having missing information (1,573,246) were excluded from the study population. The final NEDIS data population (11,952,887) was divided into three cohorts: REMC (3,626,154), LEMC (8,278,081) and LEMI (48,652). These three cohorts were used as the prediction model development set. The study site data included 184,361 ED visits from January 2019 to July 2021. Data for patients who visited without treatment (5,171), left without being seen (15,595), were dead on arrival (182), or were younger than 20 years (27,633) were excluded from the study population. Therefore, 135,780 ED visits were included in the external and temporal validation data set (Fig. 1).

3.2. Demographic characteristics

Overall, we used data for 11,952,887 patients from NEIDS and 135,780 patients from the study site (Fig. 1). Among the NEDIS patients, 3,626,156 patients were in the REMC cohort, 8,278,081 patients were in the LEMC cohort, and 48,652 patients were in the LEMI cohort. The distribution of each NEDIS cohort is shown in Table 1. Among the patients in the REMC and study site cohorts, KTAS level 3 patients accounted for the highest proportion at 42.4% and 45.1%, respectively, whereas in the LEMC and LEMI cohorts, KTAS level 4 patients accounted for the highest proportion. The population characteristics of each cohort differed significantly ($p < 0.01$).

Supplementary Table 1 shows distribution of KTAS change among rKTAS group.

Table 1
Demographic characteristics of the study population.

Variable	REMC	LEMC	LEMI	Study site	*P-value*
	(N = 3,626,154)	(N = 8,278,081)	(N = 48,652)	(N = 135,780)	
Age, over 65	1,110,082 (30.6%)	2,236,952 (27.0%)	12,507 (25.7%)	48,519 (35.7%)	<0.001
Sex, male	1,823,351 (50.3%)	4,051,740 (48.9%)	23,432 (48.2%)	66,560 (49.0%)	<0.001
Type of chief complaint					<0.001
- Disease	2,718,053 (75.0%)	5,982,939 (72.3%)	34,757 (71.4%)	117,185 (86.3%)	
- Trauma	908,101 (25.0%)	2,295,142 (27.7%)	13,895 (28.6%)	18,595 (13.7%)	
AVPU scale (Mental status)					<0.001
- A, Alert	3,457,769 (95.4%)	8,021,294 (96.9%)	47,982 (98.6%)	132,902 (97.9%)	
- V, Verbal	89,414 (2.5%)	136,917 (1.7%)	304 (0.6%)	1,293 (1.0%)	
- P, Pain	62,631 (1.7%)	92,789 (1.1%)	249 (0.5%)	1,056 (0.8%)	
- U, Unresponsive	16,340 (0.5%)	27,081 (0.3%)	117 (0.2%)	529 (0.4%)	
Initial KTAS					<0.001
−1	52,020 (1.4%)	53,123 (0.6%)	128 (0.3%)	609 (0.4%)	
−2	356,180 (9.8%)	463,794 (5.6%)	1,347 (2.8%)	8,036 (5.9%)	
−3	1,535,829 (42.4%)	2,862,422 (34.6%)	7,447 (15.3%)	61,237 (45.1%)	
−4	1,372,001 (37.8%)	3,876,870 (46.8%)	27,743 (57.0%)	59,481 (43.8%)	
−5	310,124 (8.6%)	1,021,872 (12.3%)	11,987 (24.6%)	6,417 (4.7%)	
Final KTAS					<0.001
−1	56,291 (1.6%)	59,074 (0.7%)	132 (0.3%)	634 (0.5%)	
−2	369,474 (10.2%)	493,956 (6.0%)	1,395 (2.9%)	8,649 (6.4%)	
−3	1,595,266 (44.0%)	3,002,662 (36.3%)	7,533 (15.5%)	62,731 (46.2%)	
−4	1,302,486 (35.9%)	3,725,031 (45.0%)	27,672 (56.9%)	57,712 (42.5%)	
−5	302,637 (8.3%)	997,358 (12.0%)	11,920 (24.5%)	6,054 (4.5%)	
Method of Transportation					<0.001
−119 ambulance	766,448 (21.1%)	1,708,162 (20.6%)	10,612 (21.8%)	16,648 (12.3%)	
-Private ambulance	286,746 (7.9%)	350,271 (4.2%)	1,404 (2.9%)	8,868 (6.5%)	
-Walk in and other	2,572,960 (71.0%)	6,219,648 (75.1%)	36,636 (75.3%)	110,264 (81.2%)	
Route of arrival					<0.001
-Direct	2,921,850 (80.6%)	7,382,579 (89.2%)	46,485 (95.5%)	107,369 (79.1%)	
-Transfer	607,284 (16.7%)	741,460 (9.0%)	1,924 (4.0%)	19,200 (14.1%)	
-Other	97,020 (2.7%)	154,042 (1.9%)	243 (0.5%)	9,211 (6.8%)	
Weekend	1,193,247 (32.9%)	2,923,958 (35.3%)	18,946 (38.9%)	37,896 (27.9%)	<0.001
KTAS Changed	172,631 (4.7%)	279,580 (3.4%)	317 (0.6%)	3,912 (2.9%)	<0.001

Disease type: whether visit by disease or trauma, KTAS: Korean Triage Acute Scale, AVPU scale: mental status scaling, A refers Alert, V refers Response to verbal output, P refers Response to pain, U refers Unresponsive. Method of Transportation: Transportation that patient take for Emergency Department visit, Route of arrival: whether visit Emergency Department directly without other hospital visit, or transfer from other hospital.
***P*-values were calculated using *t*-test for continuous variables and chi-square test for categorical variables.

3.3. Comparison of ED clinical outcomes between rKTAS and uKTAS

Table 2 shows outcome differences between the rKTAS group and the uKTAS group. Among patients initially triaged as KTAS 4 or KTAS 5, the rKTAS group had a higher proportion of admissions than the uKTAS group. Among patients initially triaged as KTAS 3, the rKTAS group had a higher proportion of ED deaths than the uKTAS group. Patients in the rKTAS group had a longer LOS than the uKTAS group, especially among initial KTAS 4 and 5 patients. In contrast, the proportion of patients discharged home was lower in the rKTAS group among those with an initial KTAS of 3, 4, or 5. Among patients initially triaged as KTAS 2, the rKTAS group also had a higher ED mortality rate (Table 2, $p < 0.001$).

3.4. Prediction model

Table 3 summarizes the prediction results using the AUROC with 95% confidence interval by input and machine learning type. The AUROC values for the machine learning model were 0.777, 0.775, and 0.774 in the internal validation with the REMC, LEMC, and LEMI datasets, respectively, and 0.786, 0.750, and 0.770 in the external temporal validation with the study site data from 2019, 2020, and 2021, respectively. The area under the recall precision curve (AUPRC) values of the model for data from the REMC, LEMC, and LEMI datasets were 0.363, 0.366, and 0.308, respectively.

Remarkably, the AUROC values for the external validation of a prediction model built with the LEMI database were only 0.659, 0.609, and 0.627 in the study site data from 2019, 2020, and 2021, respectively, without using FL. (Supplementary Table 2). However, they increased to 0.786, 0.750, and 0.770, respectively, after FL was used.

3.5. Risk factor analysis

Table 4 shows the results of our analysis of risk factors associated with KTAS acuity revision in univariable and multivariable analyses. All factors, including patient demographic information and vital signs, were associated with acuity revision. All vital signs, especially an abnormal pulse rate, were associated with KTAS acuity revision (odds ratio [OR]: 1.37; confidence interval [CI]: 1.36–1.39). An initial classification of KTAS 4 (OR 4.10, CI: 4.06–4.14) or 5 (OR: 3.09, CI: 3.04–3.14) was also associated with revision of KTAS. The AVPU scale, especially the V (response to verbal output) scale, was associated with KTAS acuity revision (OR: 1.43, CI: 1.39–1.46). Arrival by a 119 ambulance (OR: 1.75; CI: 1.73–1.77) was also related to KTAS acuity revision. Subgroup analysis was also performed by level of the emergency center (Supplementary Table 3. Supplementary Table 4. Supplementary Table 5.).

Table 2
Outcomes between patients whose KTAS was revised and patients whose KTAS remained unchanged.

Initial KTAS	Variable	rKTAS patients	uKTAS patients	*P*-value **
KTAS5	LOS, min	153 [77–305]	57 [21–118]	<0.001
	ED disposition, n(%)			<0.001
	Home	30945 (0.61)	1206055 (0.93)	
	ER Death	47 (0)	46 (0)	
	Admission	18468 (0.36)	81714 (0.06)	
	Transfer	1383 (0.03)	5325 (0)	
KTAS4	LOS, min	213 [124–396]	104 [55–174]	<0.001
	ED disposition, n(%)			<0.001
	Home	116809 (0.4)	4323641 (0.87)	
	ER death	418 (0)	251 (0)	
	Admission	159039 (0.55)	615053 (0.12)	
KTAS3	Transfer	12872 (0.04)	48531 (0.01)	
	LOS, min	179 [106–339]	174 [109–302]	<0.001
	ED disposition, n(%)			<0.001
	Home	52119 (0.59)	2631637 (0.61)	
	ER death	1003 (0.01)	3904 (0)	
	Admission	32166 (0.36)	1567979 (0.36)	
	Transfer	3592 (0.04)	113298 (0.03)	
KTAS2	LOS, min	216 [123–387]	215 [122–398]	<0.001
	ED disposition, n(%)			<0.001
	Home	11211 (0.52)	287697 (0.36)	
	ER death	1001 (0.05)	7179 (0.01)	
	Admission	8354 (0.39)	459327 (0.57)	
	Transfer	1033 (0.05)	45519 (0.06)	
KTAS1	[a]LOS, min	255 [139–481]	233 [124–474]	<0.001
	ED disposition, n(%)			<0.001
	Home	804 (0.39)	8626 (0.08)	
	ER death	23 (0.01)	7832 (0.08)	
	Admission	1084 (0.52)	76473 (0.74)	
	Transfer	157 (0.08)	10272 (0.1)	

[a] LOS: length of stay, IQR and median was calculated in []; P-values were calculated with t-tests for continuous variables (LOS) and chi-square tests for categorical variables (Others).

Table 3
AUROC in predicting changed KTAS by using the federated learning.

Type	Cohort	Time	AUROC (95% CI)	AUPRC (95% CI)
Internal Validation	NEDIS (REMC)	2016–2018	0.777 (0.776, 0.777)	0.363 (0.363, 0.363)
Internal Validation	NEDIS (LEMC)	2016–2018	0.775 (0.775, 0.776)	0.366 (0.366, 0.366)
Internal Validation	NEDIS (LEMI)	2016–2018	0.774 (0.767, 0.781)	0.308 (0.308, 0.308)
External & Temporal validation	Study site	2019	0.786 (0.783, 0.789)	0.336 (0.336, 0.336)
External & Temporal validation	Study site	2020	0.750 (0.746, 0.755)	0.335 (0.335, 0.335)
External & Temporal validation	Study site	2021	0.770 (0.766, 0.773)	0.335 (0.334, 0.335)

AUROC: Area Under the Receive Operating Curve, AUPRC: Area Under the Recall Precision Curve, CI: Confidence Interval, KTAS: Korea Triage Acuity Scale, NEDIS: National Emergency Department Information System, REMC: Regional Emergency Medical Center, LEMC: Local Emergency Medical Center, LEMI: Local Emergency Medical Institution.

Table 4
Factors associated with KTAS acuity revision.

Variables (reference)	Univariate Analysis			Multivariate Analysis		
	OR	95% CI	*P*-value	OR	95% CI	*P*-value
Age (below 65)			<0.001			<0.001
65 or more	1.47	1.46–1.48		1.32	1.31–1.33	
Sex (female)			<0.001			<0.001
Male	1.11	1.10–1.12		1.15	1.14–1.15	
Initial KTAS (3)			<0.001			<0.001
1	0.97	0.93–1.02		0.36	0.34–0.38	
2	1.31	1.29–1.33		0.74	0.72–0.75	
4	2.82	2.79–2.84		4.10	4.06–4.14	
5	1.91	1.89–1.93		3.09	3.04–3.14	
AVPU scale (A, alert)			<0.001			<0.001
V (Response to verbal output)	1.49	1.46–1.52		1.43	1.39–1.46	
P (Response to pain)	1.25	1.22–1.28		1.41	1.37–1.45	
U (Unresponsive)	0.76	0.72–0.81		1.69	1.57–1.82	
Trauma	1.04	1.04–1.05		0.77	0.76–0.78	
visit time (07:00–14:59)			<0.001			<0.001
15:00–22:59	0.87	0.86–0.88		0.93	0.92–0.93	
23:00–06:59	0.87	0.86–0.87		0.90	0.89–0.91	
Visit date (week)			<0.001			<0.001
Weekend	0.87	0.87–0.88		0.91	0.90–0.92	
Route of arrival (direct)			<0.001			<0.001
Transfer	1.28	1.27–1.29		1.46	1.44–1.48	
Other	1.21	1.18–1.23		1.31	1.28–1.34	
Method of transportation (Walk in and other)			<0.001			<0.001
119 ambulance	1.57	1.56–1.58		1.75	1.73–1.77	
Other ambulance	1.37	1.36–1.39		1.17	1.15–1.20	
SBP (100–150 mmHg)			<0.001			<0.001
Out of range	1.15	1.14–1.16		1.18	1.17–1.20	
DBP (60–90 mmHg)			<0.001			<0.001
Out of range	1.11	1.10–1.11		1.13	1.12–1.14	
PR (50–120 bpm)			<0.001			<0.001
Out of range	1.29	1.28–1.30		1.37	1.36–1.39	
RR (12–20/min)			<0.001			<0.001
Out of range	1.20	1.19–1.21		1.16	1.15–1.17	
TEMP (36–37.5 °C)			<0.001			<0.001
Out of range	1.16	1.15–1.17		1.16	1.14–1.17	
SPo2 (95–100%)			<0.001			<0.001
Out of range	1.22	1.21–1.24		1.33	1.31–1.35	

4. Discussion

This study achieved an AUROC of more than 0.75, regardless of the level of ED center, by using FL. Therefore, our model overcomes the data disparities between different levels of emergency medical centers. Especially in LEMIs, in which hospital management systems are less likely to be standardized or supported than in REMCs and LEMCs due to a lack of resources and personnel, this model shows improved outcomes after using FL (Supplementary Table 2, Table 3).

This study has many advantages due to its data sources. Validation was conducted using both national-level data and data from separate cohorts divided by ED level to address data quality issues and population imbalances between cohorts. Our results demonstrate our model's potential usability regardless of an ED's characteristics or center level. Temporal validation was also conducted at the study site, and the AUROC was more than 0.75. Even though the COVID pandemic has affected Korea since December 2019 and caused many changes in patient characteristics and ED operations, model performance was still higher than 0.75 [46].

This model differs from other DL-based triage models in several ways because it uses initial KTAS. The variables required for this prediction model can be obtained during a patient's initial encounter with an ED. This model is distinct from other DL-based triage methods in that it utilizes initial KTAS. The data necessary for this prediction model can be collected during a patient's initial visit to an ED. This model does not attempt to forecast outcomes or serve as a replacement for the original KTAS system. As this is a clinical decision support system for the original KTAS revision prediction, the original KTAS is required. Therefore, adjustment of KTAS is not the outcome of this model.

This model is more focused on KTAS and short-term outcomes predicting acuity revision in the ED than on long-term patient outcome prediction; this model differs from previous ones in that it is not attempting to compete with a conventional triage system, but rather improve clinical decision making by detecting KTAS acuity revision, and is thus more focused on supporting original triage than previous models [20]. This model uses the traditional KTAS result itself as a feature for prediction. Therefore, this model needs an initial triage result to predict revision probability. It cannot substitute for traditional triage at all.

The KTAS is scaled by humans. Also, no scoring system can achieve 100% accuracy. This model tries to prevent under or over-triage of KTAS in advance by predicting the probability of revision of the KTAS result during the ED stay. By comparing rKTAS and uKTAS, study try to show the reason why prediction of revision of KTAS is needed (Table 2). As we showed in the study, the rKTAS group shows a higher ED death and ICU admission rate. By demonstrating clinical outcomes such as length of stay, ED death, ICU admission, and admission, this study explores the significance of predicting KTAS acuity revision. Th Also, Authors try to show that under- or over-triage, which is partially shown as revision of KTAS in the conclusion of retrospective data, might likely result in adverse outcomes. This might be because of the late recognition of a patient in a potentially critical state. Therefore, by predicting revision of KTAS, which is predicting under or over triage, critical patients might be discovered earlier and decreases adverse outcome by preventing under- or over-triage. By doing this, this model seeks to trigger reassessment earlier for patients whose condition might deteriorate by predicting the probability of KTAS acuity revision.

This study might help clinicians identify patients likely to require KTAS acuity revision, thereby supporting timely triage operations, particularly in under-resourced EDs. Furthermore, by predicting KTAS acuity revision, our model seeks to predict changes in patient status indirectly in advance. Despite the high performance of scaling tools, there will be cases where revision is necessary for various reasons [38]. In our study, patients who initially had KTAS levels of 3 or 4 but whose levels were later changed had higher ED mortality than patients whose KTAS scores remained unchanged. In the initial KTAS 5 group, the rKTAS group had a higher admission rate. Because those levels, KTAS 3, 4 or 5are considered to have normal or lower acuity, which is not immediately noticeable to clinicians, patients at those levels are under- or overestimated, which can result in loss of opportunities for prompt, appropriate management.

In addition, rKTAS patients at all KTAS levels had a longer LOS compared to uKTAS patients. Despite the fact that there is only a one- or 2-min difference (patients with an initial KTAS of 4), minimal time difference should also be considered due to the large sample size of this study [47]. Additionally, other groups show more marked differences. Therefore, under- or over-triage may be associated with prolonged LOS.

Our study shows risk factors for KTAS acuity revision and attempts to explain which factors are associated with KTAS acuity revision. DL-based models are difficult to explain because the algorithm process is hidden,[30,31] but our univariable and multivariable logistic regression analyses substantiate our algorithm's results. An initial KTAS score of 4 was highly associated with misclassification. Additionally, a patient's vital signs, especially the pulse rate and V and P scales in the AVPU, were associated with KTAS misclassification.

The method of transportation was another risk factor for KTAS acuity revision. That might relate to how patients came by the ambulance are shown in the first presentation, which could affect the judgment of the triaging medical staff. That provides further evidence of the subjective characteristics of KTAS and shows why this study is needed: our model can support the KTAS system characteristics that rely on subjective judgment.

4.1. Limitations

This was a retrospective study, so there might have been selection bias. Additionally, we were unable to validate prospectively. We performed temporal validation with data of different period as well as external validation to reduce that bias. Future research should include additional prospective validation to allow for the use of this model in actual practice. This study also excluded injured pediatric patients. Many systems triage adult and pediatric patients separately [7,48,49]. Therefore, pediatric patients require a separate study [7]. Finally, because we used a nationwide database to develop our model, the data quality might have varied among medical centers, depending on their size and staffing levels. To address that possibility, we used the federated learning method to minimize the data quality gaps among REMC, LEMC, and LEMI data.

5. Conclusions

We developed a federated learning–based triage assistant system for EDs. Our novel system accurately predicted KTAS level changes and was useful for detecting patients whose KTAS acuity would be revised. This study's findings could help clinicians with triage decisions in situations in which the medical staff's judgment is unclear, such as in crowded or understaffed EDs. As this is a study about model development, further research will be needed for real-world implementation [50–52].

Author contribution statement

Hansol Chang; Jae Yong Yu: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Geun Hyeong Lee: Analyzed and interpreted the data.

Sejin Heo; Se Uk Lee; Sung Yeon Hwang; Hee Yoon; Won Chul Cha; Tae Gun Shin; Min Seob Sim; Ik Joon Jo: Contributed reagents, materials, analysis tools or data.

Taerim Kim: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Data availability statement

Data are available in the study site clinical data warehouse. The datasets generated and analyzed during the current study are not publicly available due dataset includes although is de-identified, part of patient information, but are available from the corresponding author on reasonable request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments and funding

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI21C136702.

Abbreviations

(ED)	Emergency department
(LOS)	Length of stay
(DL)	Deep learning
(FL)	Federated learning
(KTAS)	Korea Triage Acuity Scale
(NEDIS)	National Emergency Department Information System
(IRB)	Institutional Review Board
(IQR)	Interquartile range
(REMC)	Regional emergency medical center
(LEMC)	Local emergency medical center
(LEMI)	Local emergency medical institution
(AVPU)	Alert, response to Verbal output, response to Pain, Unresponsive
uKTAS	unchanged KTAS acuity
rKTAS	revised KTAS acuity
(ROC)	Receiver operating characteristic
(PR)	Precision recall
(AUROC)	Area under the receiver operating characteristic curve

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e19210.

References

[1] C.M. Fernandes, et al., Five-level triage: a report from the ACEP/ENA five-level triage task force, J. Emerg. Nurs. 31 (1) (2005) 39–50, https://doi.org/10.1016/j.jen.2004.11.002, quiz 118.

[2] M. Christ, F. Grossmann, D. Winter, R. Bingisser, E. Platz, Modern triage in the emergency department, Dtsch Arztebl Int 107 (50) (2010) 892–898, https://doi.org/10.3238/arztebl.2010.0892.

[3] O. Miro, M. Sanchez, J. Milla, Hospital mortality and staff workload, Lancet 356 (9238) (2000) 1356–1357, https://doi.org/10.1016/S0140-6736(05)74269-0.

[4] O. Miro, et al., Decreased health care quality associated with emergency department overcrowding, Eur. J. Emerg. Med. 6 (2) (1999) 105–107, https://doi.org/10.1097/00063110-199906000-00003.

[5] K.K. Yang, S.S.W. Lam, J.M.W. Low, M.E.H. Ong, Managing emergency department crowding through improved triaging and resource allocation, Oper. Res. Heal. Care 10 (2016) 13–22, https://doi.org/10.1016/j.orhc.2016.05.001.
[6] J.Y. Kim, et al., Reliability of Korean triage and acuity scale-based triage system as a severity index in emergency patients, J. Kor. Soc. Emerg. Med. 28 (6) (2017) 552–556.
[7] T. Lim, J.J.S Park, Pediatric Korean triage and acuity scale, Ped. Emerg. Med. J. 2 (2) (2015) 53–58, https://doi.org/10.1080/15563650.2021.1939881.
[8] J. Park, T. Lim, Korean triage and acuity scale (KTAS), J. Kor. Soc. Emerg. Med. 28 (6) (2017) 547–551.
[9] M.J. Bullard, et al., Revisions to the Canadian emergency department triage and acuity scale (CTAS) guidelines, CJEM 19 (S2) (2016) S18–S27, https://doi.org/10.1017/cem.2017.365, 2017.
[10] I. Kim, et al., Use of the National Early Warning Score for predicting in-hospital mortality in older adults admitted to the emergency department, Clin. Exp. Emerg. Med. 7 (1) (2020) 61–66, https://doi.org/10.15441/ceem.19.036.
[11] S.B. Cetin, O. Eray, F. Cebeci, M. Coskun, M. Gozkaya, Factors affecting the accuracy of nurse triage in tertiary care emergency departments, Turk. J. Emerg. Med. 20 (4) (2020) 163–167, https://doi.org/10.4103/2452-2473.297462.
[12] S. Levin, 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 (5) (2018) 565–574 e562, https://doi.org/10.1016/j.annemergmed.2017.08.005.
[13] E.S. Lee, H. Oh, Re-evaluation characteristics of the Korean Triage and Acuity Scale (KTAS): the relationship between overcrowding and KTAS re-evaluation, J. Kor. Soc. Emerg. Med. 32 (2) (2021) 179–188.
[14] J.S. Hinson, et al., Accuracy of emergency department triage using the Emergency Severity Index and independent predictors of under-triage and over-triage in Brazil: a retrospective cohort analysis, Int. J. Emerg. Med. 11 (1) (2018) 3, https://doi.org/10.1186/s12245-017-0161-8.
[15] S.H. Moon, J.L. Shim, K.S. Park, C.S. Park, Triage accuracy and causes of mistriage using the Korean Triage and Acuity Scale, PLoS One 14 (9) (2019), e0216972, https://doi.org/10.1371/journal.pone.0216972.
[16] A.J. Singer, H.C. Thode Jr., P. Viccellio, J.M. Pines, The association between length of emergency department boarding and mortality, Acad. Emerg. Med. 18 (12) (2011) 1324–1329, https://doi.org/10.1111/j.1553-2712.2011.01236.x.
[17] K.E. Kocher, W.J. Meurer, J.S. Desmond, B.K. Nallamothu, Effect of testing and treatment on emergency department length of stay using a national database, Acad. Emerg. Med. 19 (5) (2012) 525–534, https://doi.org/10.1111/j.1553-2712.2012.01353.x.
[18] J.S. Kim, et al., Prolonged length of stay in the emergency department and increased risk of in-hospital cardiac arrest: a nationwide population-based study in South Korea, 2016-2017, J. Clin. Med. 9 (7) (2020) 2284, https://doi.org/10.3390/jcm9072284.
[19] P. Tanabe, R. Gimbel, P.R. Yarnold, D.N. Kyriacou, J.G. Adams, Reliability and validity of scores on the emergency severity index version 3, Acad. Emerg. Med. 11 (1) (2004) 59–65, https://doi.org/10.1197/j.aem.2003.06.013.
[20] J.M. Kwon, Y. Lee, Y. Lee, S. Lee, H. Park, J. Park, Validation of deep-learning-based triage and acuity score using a large national dataset, PLoS One 13 (10) (2018), e0205836, https://doi.org/10.1371/journal.pone.0205836.
[21] S.W. Choi, T. Ko, K.J. Hong, K.H. Kim, Machine learning-based prediction of Korean triage and acuity scale level in emergency department patients, Heal. Inform. Res. 25 (4) (2019) 305–312, https://doi.org/10.4258/hir.2019.25.4.305.
[22] Y. Raita, T. Goto, M.K. Faridi, D.F.M. Brown, C.A. Camargo Jr., K. Hasegawa, Emergency department triage prediction of clinical outcomes using machine learning models, Crit. Care 23 (1) (2019) 64, https://doi.org/10.1186/s13054-019-2351-7.
[23] C.-Y. Kang, J.H. Yoon, Current challenges in adopting machine learning to critical care and emergency medicine, Clin. Exp. Emerg. Med. 10 (2) (2023) 132–137, https://doi.org/10.15441/ceem.23.041.
[24] M. Fernandes, S.M. Vieira, F. Leite, C. Palos, S. Finkelstein, J.M.C. Sousa, Clinical decision support systems for triage in the emergency department using intelligent systems: a Review, Artif. Intell. Med. 102 (2020), 101762, https://doi.org/10.1016/j.artmed.2019.101762.
[25] W.S. Hong, A.D. Haimovich, R.A. Taylor, Predicting hospital admission at emergency department triage using machine learning, PLoS One 13 (7) (2018), e0201016, https://doi.org/10.1371/journal.pone.0201016.
[26] T. Goto, C.A. Camargo Jr., M.K. Faridi, R.J. Freishtat, K. Hasegawa, Machine learning-based prediction of clinical outcomes for children during emergency department triage, JAMA Netw. Open 2 (1) (2019), e186937, https://doi.org/10.1001/jamanetworkopen.2018.6937.
[27] H.G. Van Spall, et al., Prediction of emergent heart failure death by semi-quantitative triage risk stratification, PLoS One 6 (8) (2011), e23065, https://doi.org/10.1371/journal.pone.0023065.
[28] H. Chang, W.C. Cha, Artificial intelligence decision points in an emergency department, Clin. Exp. Emerg. Med. 9 (3) (2022) 165–168.
[29] Korean Law Information Center. https://www.law.go.kr/LSW/eng/engMain.do?eventGubun=060124. (Accessed 25 February 2023).
[30] G.A. Kaissis, M.R. Makowski, D. Rückert, R.F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nat. Mach. Intell. 2 (6) (2020) 305–311, https://doi.org/10.1038/s42256-020-0186-1.
[31] N. Rieke, et al., The future of digital health with federated learning, NPJ Dig. Med. 3 (1) (2020) 119, https://doi.org/10.1038/s41746-020-00323-1.
[32] M.J. Sheller, G.A. Reina, B. Edwards, J. Martin, S. Bakas, Multi-institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 92–104.
[33] G.H. Lee, S.Y. Shin, Federated learning on clinical benchmark data: performance assessment, J. Med. Internet Res. 22 (10) (2020), e20891, https://doi.org/10.2196/20891.
[34] A. Vaid, et al., Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: machine learning approach, JMIR Med. Inform. 9 (1) (2021), e24207, https://doi.org/10.2196/24207.
[35] J.H. Ryu, et al., Changes in relative importance of the 5-level triage system, Korean triage and acuity scale, for the disposition of emergency patients induced by forced reduction in its level number: a multi-center registry-based retrospective cohort study, J. Kor. Med. Sci. 34 (14) (2019) e114, https://doi.org/10.3346/jkms.2019.34.e114.
[36] The Korea Society of Emergency Medicine, Korean Triage and Acuity Scale Homepage. http://www.ktas.org/. (Accessed 25 February 2023).
[37] B.A. Lentz, et al., Validity of ED: addressing heterogeneous definitions of over-triage and under-triage, Am. J. Emerg. Med. 35 (7) (2017) 1023–1025, https://doi.org/10.1016/j.ajem.2017.02.012.
[38] J.B. Park, J. Lee, Y.J. Kim, J.H. Lee, T.H. Lim, Reliability of Korean triage and acuity scale: interrater agreement between two experienced nurses by real-time triage and analysis of influencing factors to disagreement of triage levels, J. Kor. Med. Sci. 34 (28) (2019) e189, https://doi.org/10.3346/jkms.2019.34.e189.
[39] S. Ham, Y.G. Min, M.K. Chae, H.H. Kim, Epidemiology and regional differences of acute poisonings of eight cities in Gyeonggi-do province in Korea using data from the National Emergency Department Information System of Korea, Clin. Exp. Emerg. Med. 7 (1) (2020) 43–51, https://doi.org/10.15441/ceem.19.014.
[40] S.R. Janagama, J.A. Newberry, M.A. Kohn, G.V.R. Rao, M.C. Strehlow, S.V. Mahadevan, Is AVPU comparable to GCS in critical prehospital decisions?–A cross-sectional study, Am. J. Emerg. Med. 59 (2022) 106–110.
[41] A.G. Nuttall, K.M. Paton, A.M. Kemp, To what extent are GCS and AVPU equivalent to each other when assessing the level of consciousness of children with head injury? A cross-sectional study of UK hospital admissions, BMJ Open 8 (11) (2018), e023216, https://doi.org/10.1136/bmjopen-2018-023216.
[42] C.A. Kelly, A. Upex, D.N. Bateman, Comparison of consciousness level assessment in the poisoned patient using the alert/verbal/painful/unresponsive scale and the Glasgow Coma Scale, Ann. Emerg. Med. 44 (2) (2004) 108–113, https://doi.org/10.1016/j.annemergmed.2004.03.028.
[43] C. Barfod, et al., Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency department - a prospective cohort study, Scand. J. Trauma Resuscitation Emerg. Med. 20 (2012) 28, https://doi.org/10.1186/1757-7241-20-28.
[44] K.E. Philip, E. Pack, V. Cambiano, H. Rollmann, S. Weil, J. O'Beirne, The accuracy of respiratory rate assessment by doctors in a London teaching hospital: a cross-sectional study, J. Clin. Monit. Comput. 29 (4) (2015) 455–460, https://doi.org/10.1007/s10877-014-9621-3.
[45] C.A. Dinarello, R. Porat, Fever, J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, in: J. Loscalzo (Ed.), Harrison's Principles of Internal Medicine, twentieth ed., McGraw-Hill Education, New York, NY, 2018.
[46] S.I. Lee, S.B. Kang, S.Y. Lee, D.S. Choi, The effect of regional distribution of isolation rooms in emergency departments on ambulance travel time during the COVID-19 pandemic, Clin. Exp. Emerg. Med. 10 (2) (2023) 191–199, https://doi.org/10.15441/ceem.22.355.

[47] S. Yuzeng, L.L. Hui, Improving the wait time to triage at the emergency department, BMJ Open Qual. 9 (1) (2020), e000708, https://doi.org/10.1136/bmjoq-2019-000708.
[48] M. van Veen, et al., Manchester triage system in paediatric emergency care: prospective observational study, BMJ 337 (2008) a1501, https://doi.org/10.1136/bmj.a1501.
[49] A.R. Allen, M.J. Spittal, C. Nicolas, E. Oakley, G.L. Freed, Accuracy and interrater reliability of paediatric emergency department triage, Emerg. Med. Australasia (EMA) 27 (5) (2015) 447–452, https://doi.org/10.1111/1742-6723.12455.
[50] H. Chang, J.Y. Yu, S. Yoon, T. Kim, W.C. Cha, Machine learning-based suggestion for critical interventions in the management of potentially severe conditioned patients in emergency department triage, Sci. Rep. 12 (1) (2022), 10537, https://doi.org/10.1038/s41598-022-14422-4.
[51] F. Magrabi, et al., Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications, Yearb Med. Inform. 28 (1) (2019) 128–134, https://doi.org/10.1055/s-0039-1677903.
[52] J.H. Yoon, M.R. Pinsky, G. Clermont, Artificial intelligence in critical care medicine, Crit. Care 26 (1) (2022) 75, https://doi.org/10.1186/s13054-022-03915-3.