Submitted: 01 October, 2024 Accepted: 07 February, 2025 Published: 08 May, 2025 DOI:10.22514/sv.2025.062

Open Access

Signa Vitae

ORIGINAL RESEARCH

Federated learning for predicting critical intervention and poor clinical outcomes at emergency department triage stage

Sejin Heo[1,†], Geunho Choi[2,†], Su Min Kim[3], Hyung Joon Joo[3,4,5], Jong-Ho Kim[4,5], Soo-Yong Shin[2], Hee Yoon[1], Sung Yeon Hwang[1], Hansol Chang[1], Jae Yong Yu[6,7,8], Won Chul Cha[1,2], Se Uk Lee[1,*]

[1]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 06355 Seoul, Republic of Korea
[2]Department of Digital Health, Samsung Advanced Institute for Health Science & Technology (SAIHST), Sungkyunkwan University, 06355 Seoul, Republic of Korea
[3]Department of Medical Informatics, Korea University College of Medicine, 02708 Seoul, Republic of Korea
[4]Korea University Research Institute for Medical Bigdata Science, Korea University College of Medicine, 02708 Seoul, Republic of Korea
[5]Department of Cardiology, Cardiovascular Center, Korea University College of Medicine, 02841 Seoul, Republic of Korea
[6]Research Institute for Data Science and AI (Artificial Intelligence), Hallym University, 24252 Chuncheon-si, Republic of Korea
[7]Divison of Data Science, Hallym University, 24252 Chuncheon-si, Republic of Korea
[8]Department of Emergency Medicine, Hallym University, 24252 Chuncheon-si, Republic of Korea

***Correspondence**
seuk.lee@samsung.com
(Se Uk Lee)

[†] These authors contributed equally.

Abstract

Background: Early detection and timely intervention of patients at risk during the triage stage can significantly improve patient outcomes. This study aimed to predict requirements for critical respiratory or cardiovascular intervention and poor clinical outcome using federated learning (FL). **Methods**: Patients of two tertiary hospitals who visited the emergency department (ED) were included. Local models for each hospital and FL models to predict high flow nasal cannula or endotracheal intubation (model 1), central venous catheter insertion or vasopressor administration (model 2), and admission to intensive care unit or cardiac arrest during ED stay (model 3) were developed and internally validated with data from 2017 to 2020. These models were then externally validated with data from 2021. Available information such as underlying disease, recent blood test results, age, sex, and initial vital signs at triage stage were used as input variables. Performances of models were evaluated using area under the receiver operating characteristic (AUROC) with 95% confidence interval. **Results**: A total of 262,283 and 180,261 ED visits from Samsung Medical Center (hospital A) and Korea University ANAM Hospital (hospital B) respectively, were included. AUROC values of three local and three FL models in both hospitals all exceeded 0.85 in internal validation. For hospital B, local models showed better performance than the FL model, including model 2 (0.942 (0.938–0.946) *vs.* 0.890 (0.884–0.896)) and model 3 (0.910 (0.905–0.914) *vs.* 0.886 (0.881–0.891)). AUROC values of local and FL models also exceeded 0.85 in external validation. The FL model showed comparable performance except model 3 of hospital B. **Conclusions**: Federated learning models demonstrated comparable performance to local models in predicting critical interventions and poor clinical outcomes at triage.

Keywords

Critical intervention; Emergency department; Triage; Federated learning

1. Introduction

Rapid screening and timely interventions are important for managing critical patients in the emergency department (ED). Early detection of patients at risk during the triage stage can significantly improve patient outcomes [1–4]. With the advent of machine learning (ML), prediction models have demonstrated their ability to assist healthcare workers in triage [5, 6]. These models target various outcomes such as mortality and admissions to the intensive care unit (ICU), enabling earlier attention of physicians [7–9].

However, such models could not directly guide treatment decisions. Critical interventions should be performed properly, as these interventions can significantly impact patient prognosis [10–12]. Nevertheless, ED physicians often face challenges in making timely decisions due to factors such as crowding, lack of resources and/or low levels of experience [1, 3, 13]. Although previous studies have reported ML models targeting critical interventions, those models were derived from single center data with limited performances [14, 15].

Recent studies have revealed the potential of federated learning (FL) models for generalizability in settings of multinational heterogenous patient cohorts [16, 17]. FL has been applied to various cardiovascular conditions, including pulmonary thromboembolism (PTE) and atrial fibrillation (AFib). FL models have demonstrated robust performances in predicting PTE prognostic risks, highlighting their potential for seamless integration into clinical workflows. Additionally, FL methodologies have shown promise in AFib prediction by analyzing electrocardiogram (ECG) signals and other vital parameters, while addressing privacy concerns associated with patient data [18, 19].

To make one prediction model that works well in multiple hospitals, EDs would originally need to combine their patient data to create one centralized model. However, the data privacy policy could pose an issue [20–23]. FL allows distributed learning methods that keep data in its storage silo without the need to exchange data between locations for preserving data privacy [24]. However, studies applying FL in the field of emergency medicine are limited.

Thus, this study aimed to predict requirements for critical respiratory or cardiovascular intervention and poor clinical outcome defined as either direct admission to an intensive care unit (ICU), cardiac arrest or ED death during the triage stage [9, 25]. In addition, the feasibility of applying FL to two EDs with distinct heterogenous patient groups was investigated.

2. Methods

2.1 Study design, settings and population

This was a retrospective bicenter study performed in two EDs of tertiary university hospitals in Seoul, Korea. One hospital is the Samsung Medical Center which has one of the biggest cancer centers in South Korea. It is located in the southern part of Seoul. It is designated as a local emergency medical center. The hospital has 2000 beds. Its annual visit volume of ED is about 70,000. The other hospital is Korea University ANAM Hospital. It is located in the northern part of Seoul. It is designated as a regional emergency medical center and a representative ED of that district. The hospital has a total of 1000 beds. Its annual visit volume of ED is about 40,000.

Patients who visited ED from January 2017 to December 2021 were included. Patients who were younger than 20 years old, who visited with trauma, who visited for purposes other than medical treatment, who left without being seen, and who were already in cardiac arrest or dead on arrival at visit were excluded. This study was approved by the Institutional Review Boards of Korea University ANAM Hospital (No. 2021AN0545) and Samsung Medical Center (No. 2022-08-175), both of which waived informed consent based on institutional guidelines. In accordance with deliberation results of data review boards, neither of the two hospitals exported the original data.

2.2 Input variables

Data were extracted from clinical data warehouse of each hospital. Clinical data of ED visits, along with data from the preceding year, were collected and organized into two categories.

Underlying medical conditions were defined as having a medical history within the past year involving cardiovascular disease, pulmonary disease, neurology disease, or cancers based on Korean classification of diseases 8th revision (KCD-8). KCD-8 is based on the international classification of disease 10th revision. Most recent blood laboratory test results of aspartate aminotransferase, total bilirubin, C-reactive protein, creatinine, total white blood cell count, hemoglobin, platelet count, lactic acid, n-terminal pro b-type natriuretic peptide, sodium and potassium values were collected.

Age, sex, chief complaint, mode of ED visit, Korean triage and acuity scale (KTAS) level, initial vital signs of systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse rate (PR), respiratory rate (RR), saturation of partial pressure oxygen (SpO_2), oxygen supply status while ED visit and alertness of patients on (Alert, Verbal, Pain, Unresponsive) AVPU scale were collected at triage stage.

2.3 Study outcomes

We derived three models with different outcomes. The study outcome of model 1 was events of critical respiratory support defined as applying high flow nasal cannula (HFNC) or endotracheal intubation during ED stay. Model 2 predicted events of critical circulatory support. Circulatory support was defined as needs for central venous catheter insertion for any reasons or needs for a vasopressor including dopamine, epinephrine, phenylephrine and vasopressin [26]. Model 3 targeted poor clinical outcomes defined as admission to ICU, cardiac arrest or death during ED stay.

2.4 Local model development and federated learning

Data from both hospitals were divided into two sets. The data set from 2017 to 2020 was used to develop local models and FL models (derivation set). The data set from 2021 was used only for external validation (external validation set) of local and FL models as described above. Data from 2017 to 2020 were split into training set, validation set and test set at a ratio of 7:1:2.

Local models were developed using a shallow neural network with the Tensorflow 2.13.0 package. The batch sized for a neural network was 128. There was only one hidden layer consisting of 256 rectified linear units. RMSprop was the optimizer for local models. The learning rate was 0.001. The loss function was binary crossentropy. FL models were developed using Federated Averaging (FedAvg), a widely used aggregation algorithm in FL.

$$f(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \; where \; F_k(w) = \frac{1}{n_k} \sum_{i \in p_k} f_i(w)$$

This equation explains the aggregation algorithm of FedAvg. It is based on a FedAvg paper [27]. K is defined as the total number of hospitals. The data index set for hospital k is P_k, with $n_k = |P_k|$ representing the number of data samples held by hospital k. In our case, K was 2 because the study was conducted in two institutions. The n_k represents the number of learning data. It was 148,664 and 108,949, respectively. This formula applies weights proportional to the amount of data each hospital possesses. However, when hospitals are uniformly random, $E_{P_k}[F_k(w)] = f(w)$.

Communication rounds of FL were set to 10 rounds, with 10 local epochs per communication round. The FL model was developed with 10 communication rounds and a total of 100 local rounds. A user-interface based website accessible only to learning participants was built with NodeJS 16.17.0. FedAvg operation and communication were implemented in Flask 2.0.3. Learning participants accessed the aggregation server implemented in the form of a website, sent parameters, and received a FL model.

2.5 Performance evaluation and statistical analysis

The performance of the local model was evaluated only with data from the institution that derived the local model. The performance of the FL model was evaluated with data from both hospitals. The model performance was evaluated with area under the receiver operating characteristic (AUROC). The 95% confidence interval was obtained by bootstrapping 1000 times. In addition, accuracy, precision, recall, F1-score and area under precision recall curve (AUPRC) were analyzed.

All continuous and categorical variables were reported as medians (interquartile range (IQR)) after normality test and numbers (percentages), respectively. To test differences in characteristics between two hospitals, square test or Fisher's exact test, whichever appropriate, was used for comparing categorical variables. A two-sided p-value < 0.05 was considered statistically significant. All statistical analyses were performed using Python Software (version 3.9.7, Python Software Foundation, OR, USA, https://www.python.org/).

3. Results

3.1 Patients' demographics in the derivation set

A total of 212,378 and 155,642 who visited ED of the two hospitals from 2017 to 2020 were included in the derivation set (Fig. 1). Distribution of initial mental status, mode of arrival, underlying medical conditions showed significant differences between the two hospitals ($p < 0.01$). Among four underlying disease categories, cancer accounted for the most in hospital A (34.4%) versus hospital B (8.3%). In addition, the proportion of patients who had underlying disease was significantly different between the two hospitals (60.7% in hospital A *vs.* 36.0% in hospital B, Table 1).

Incidences of study outcomes also differed between the two hospitals. The incidence of those needing critical respiratory support was 1.8% in hospital A and 5.3% in hospital B. Incidence of those needing critical circulatory support was 3.5% in hospital A and 8.3% in hospital B. Incidence of those with poor clinical outcomes was 2.8% in hospital A and 13.3% in hospital B. Differences of all outcomes between the two hospitals were significant (p-values < 0.01 for all outcomes).

3.2 Performances of local models and FL models with the derivation test set

Three local prediction models and three FL prediction models were derived and their performances were evaluated with the internal test set. AUROC values of local and FL models all exceeded 0.85 (Table 2). The performances of the following local models and FL models were not significantly different: three models of hospital A and model 1 of hospital B. However, FL model showed significantly lower performances than local models (model 2 and model 3 of hospital B, which were derived from outcomes with highest incidences). Model 2, a local model, had an AUROC of 0.942 (95% confidence interval (CI): 0.938–0.946), higher than the AUROC of 0.890

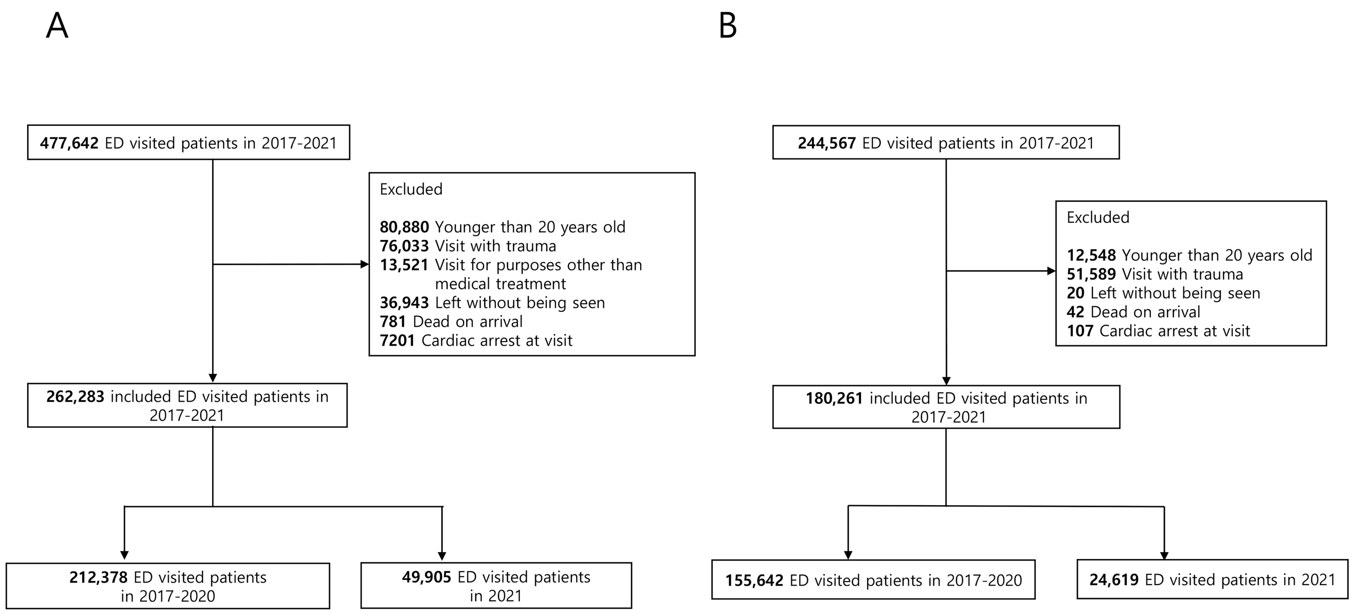

FIGURE 1. Flow diagram for patient selection. (A) Hospital A: Samsung Medical Center; (B) Hospital B: Korea University ANAM Hospital. ED: Emergency department.

TABLE 1. Baseline characteristics of patients in derivation set.

Characteristic	Samsung Medical Center (n = 212,378)	Korea University ANAM Hospital (n = 155,642)
Age, median (IQR)	59.0 (26.0)	60.0 (33.0)
Female, n (%)	107,615 (50.6)	79,889 (51.3)
Korean triage and acuity scale, n (%)*		
1	1149 (0.5)	6554 (4.2)
2	13,488 (6.3)	29,350 (18.9)
3	99,698 (46.9)	96,106 (61.7)
4	83,113 (39.1)	17,884 (11.5)
5	14,930 (7.0)	5748 (3.7)
Initial vital signs at visit, median (IQR)		
Systolic blood pressure (mmHg)	127 (33)	130 (34)
Diastolic blood pressure (mmHg)	77 (20)	80 (20)
Pulse rate (/min)	88 (26)	90 (26)
Respiratory rate (/min)	18 (4)	20 (0)
Body Temperature (°C)	36.9 (0.8)	36.7 (0.9)
Saturation of partial pressure oxygen (%)	98 (2)	98 (3)
Initial mental status, n (%)*		
Awake	207,327 (97.6)	134,734 (86.5)
Verbal	2429 (1.1)	14,217 (9.1)
Pain	1902 (0.8)	5932 (3.8)
Unresponsive	720 (0.3)	759 (0.4)
Mode of ED visit, n (%)*		
Emergency medical service	25,946 (12.2)	49,394 (31.7)
Hospital or private ambulance	16,238 (7.6)	13,499 (8.6)
Walk-in	170,194 (80.1)	92,749 (59.5)
Underlying medical conditions, n (%)*		
Cardiovascular disease	44,386 (20.8)	27,973 (17.9)
Cancer	73,068 (34.4)	13,057 (8.3)
Neurology disease	726 (0.3)	9756 (6.2)
Pulmonary disease	10,695 (5.0)	5173 (3.3)
Outcome incidence, n (%)		
Critical respiratory support*	3960 (1.8)	8352 (5.3)
Critical circulatory support*	7515 (3.5)	12,967 (8.3)
Poor clinical outcome*	5974 (2.8)	20,831 (13.3)

**Distributions of categorical variables are significantly different ($p < 0.01$) between the two hospitals.*
Critical respiratory support includes high flow nasal cannula and endotracheal intubation.
Critical circulatory support includes central venous catheter insertion and vasopressor administration.
Poor clinical outcome includes admission to intensive care unit and cardiac arrest or death during emergency department stay.
IQR: Interquartile range; ED: emergency department.

TABLE 2. Performances of local models and federated learning models in derivation test set and external validation set using area under receiver operating characteristic.

		Derivation test set (2017∼2020)		External validation set (2021)	
		Samsung Medical Center (n = 42,689)	Korea University ANAM Hospital (n = 31,285)	Samsung Medical Center (n = 49,905)	Korea University ANAM Hospital (n = 24,619)
Model 1					
	Incidence	1.8%	5.3%	2.2%	6.5%
	Local model	0.906 (0.895, 0.917)	0.922 (0.916, 0.927)	0.903 (0.894, 0.913)	0.930 (0.923, 0.936)
	FL model	0.911 (0.901, 0.922)	0.920 (0.914, 0.925)	0.915 (0.907, 0.923)	0.941 (0.937, 0.946)
Model 2					
	Incidence	3.5%	8.3%	4.1%	11.2%
	Local model	0.881 (0.873, 0.889)	0.942 (0.938, 0.946)	0.881 (0.873, 0.888)	0.921 (0.917, 0.926)
	FL model	0.870 (0.860, 0.880)	0.890 (0.884, 0.896)	0.870 (0.862, 0.879)	0.923 (0.918, 0.928)
Model 3					
	Incidence	2.8%	13.3%	2.5%	10.5%
	Local model	0.868 (0.857, 0.878)	0.910 (0.905, 0.914)	0.855 (0.844, 0.866)	0.910 (0.905, 0.915)
	FL model	0.870 (0.860, 0.879)	0.886 (0.881, 0.891)	0.864 (0.853, 0.874)	0.896 (0.891, 0.902)

Model 1 is a prediction model for critical respiratory support which includes high flow nasal cannula and endotracheal intubation.
Model 2 is a prediction model for critical circulatory support which includes central venous catheter insertion and vasopressor administration.
Model 3 is a prediction model for poor clinical outcome which includes admission to intensive care unit and cardiac arrest or death during emergency department stay.
FL: federated learning.

(95% CI: 0.884–0.896) for the FL model. Similarly, model 3, a local model, showed a higher AUROC of 0.910 (95% CI: 0.905–0.914) than the FL model, which had an AUROC of 0.886 (95% CI: 0.881–0.891).

3.3 External validation with a time-split validation set

Time-split external validation sets were established with 49,905 and 24,619 patients at hospital A and hospital B, respectively. Incidences of study outcomes between derivation and external validation sets were significantly different in all six models (all $p < 0.01$, Fig. 2).

Performances of three local prediction models and three FL prediction models were evaluated with external validation sets. AUROC values of all local and FL models also exceeded 0.85. FL models showed performances comparable to those of local models except model 3 of hospital B (Table 2). Detailed performances of FL models with cutoff using Youden index are shown in Table 3. The recall was above 0.79 for all models. AUPRC values of models ranged from 0.26 to 0.36 in hospital A and 0.50 to 0.62 in hospital B.

4. Discussion

This study investigated the feasibility of using FL for predicting critical interventions in EDs based on real-world data from two hospitals. Performances of FL models were comparable to those of local models across three predictive models. Considering significant data heterogeneity in EDs, which varies by region and time, results of this study highlight the applicability of FL in the field of emergency medicine.

Developing and utilizing models specified to each ED may yield the best performance. One study has compared centralized, FL, and local ED models for predicting admission and demonstrated that the local model has the highest performance, while the FL model has the lowest performance [28]. However, many EDs lack resources to independently develop their own local models using their own data. Furthermore, a multicenter approach is recommended to improve model generalizability and performance [29, 30]. Nonetheless, stringent data regulations make aggregating data across centers difficult. FL aligns with emerging hospital data privacy trends and regulations by enabling training without data sharing between sites.

We selected our targeted outcomes based on common practices during ED stays and categorized them by affected organs. Moreover, we focused on interventions that, depending on their application, could significantly influence clinical outcomes regardless of disease type. These interventions are heavily reliant on clinical judgment [30–33]. Predictive models, particularly outcomes of Models 1 and 2, can address situations requiring substantial clinical expertise for accurate and timely decision-making [32, 34, 35]. Additionally, proficiency levels can vary depending on the frequency of case encounters and characteristics of the ED where they work. In the triage stage, if the need for respiratory or circulatory support is evident, or if a physician determines that such intervention is necessary, prompt action should follow. However, the

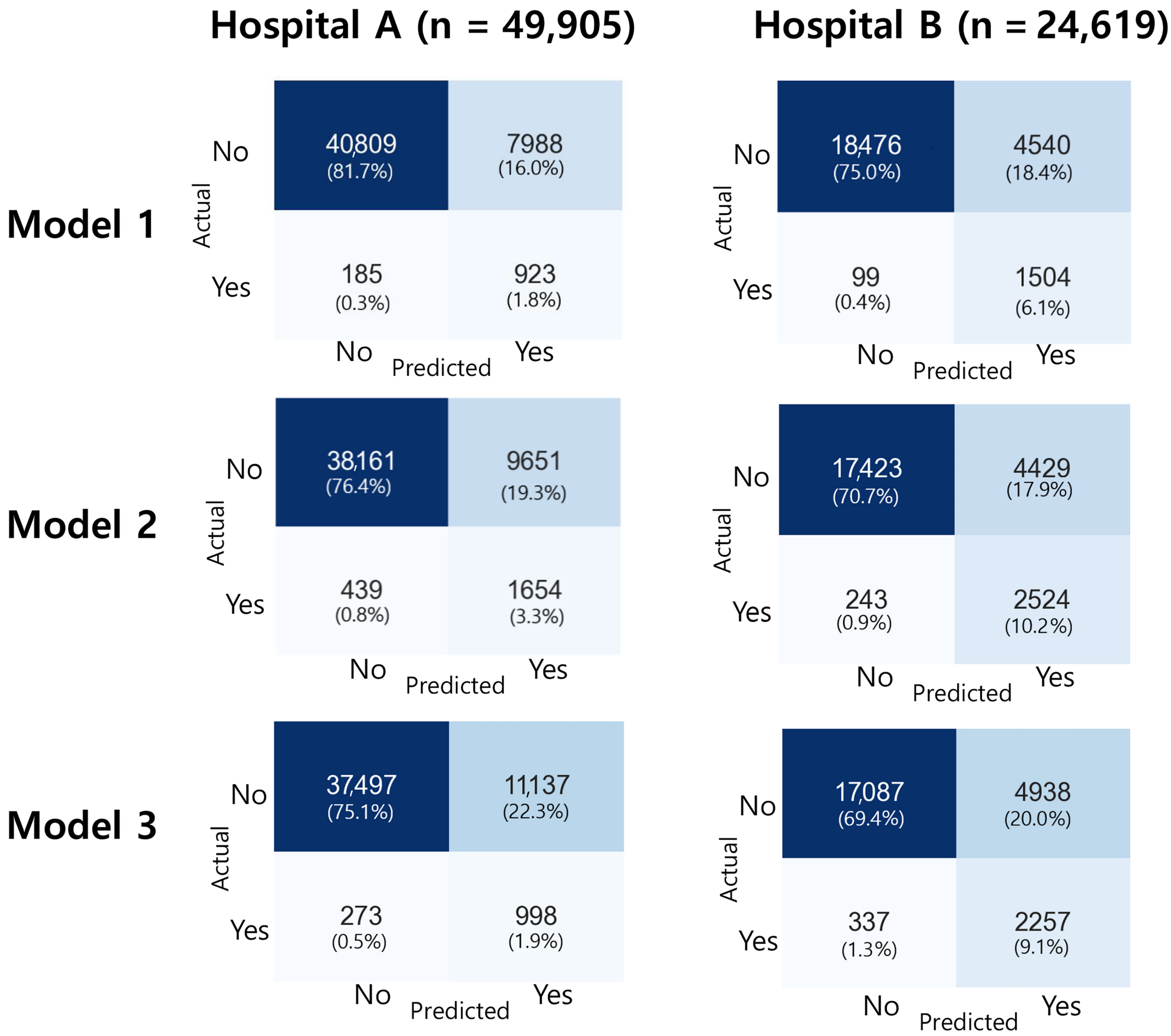

FIGURE 2. Confusion matrix of a federated learning model using an external validation set. Hospital A: Samsung Medical Center; Hospital B: Korea University ANAM Hospital.

TABLE 3. Detailed performance of federated learning model in an external validation set.

	Accuracy	Precision	Recall	F1-score	AUPRC	AUROC
Model 1						
Samsung Medical Center	0.84	0.10	0.83	0.18	0.32	0.92
Korea University ANAM Hospital	0.81	0.24	0.93	0.39	0.54	0.94
Model 2						
Samsung Medical Center	0.80	0.15	0.79	0.25	0.36	0.87
Korea University ANAM Hospital	0.81	0.36	0.91	0.51	0.62	0.92
Model 3						
Samsung Medical Center	0.77	0.08	0.79	0.15	0.26	0.86
Korea University ANAM Hospital	0.78	0.31	0.87	0.46	0.50	0.89

AUROC: Area under the receiver operating characteristic; AUPRC: Area under precision recall curve.
Model 1 is a prediction model for critical respiratory support which includes high flow nasal cannula and endotracheal intubation.
Model 2 is a prediction model for critical circulatory support which includes central venous catheter insertion and vasopressor administration.
Model 3 is a prediction model for poor clinical outcome which includes admission to intensive care unit and cardiac arrest or death during emergency department stay.

purpose of this study was to assist physicians who might lack sufficient clinical experience or face challenges in clinical decision-making by providing predictive values to support their decisions.

Therefore, these models are thought to be beneficial for supporting less experienced physicians' decision-making. For skilled physicians, these models have the potential to reduce the time and cognitive load involved in making decisions. This efficiency allows for quicker preparation and intervention, thereby enabling attention to be directed towards other patients [36, 37]. Moreover, some EDs have limited resources to provide critical interventions to critically ill patients. In such cases, our models can also aid decision-making for triage and transfer of patients during ED crowding situations.

An additional strength of our study was the use of longitudinal data including laboratory test results from 1 year prior and underlying disease history, which were considered during triage but not recorded in triage data. Incorporating past medical history enables more in-depth models beyond what could be achieved with triage data alone. Patients with chronic diseases often visit EDs of their follow-up care hospitals [38]. Our models incorporated past medical history data. They might be generalized to other hospitals with substantial populations of follow-up patients from their outpatient clinics.

This study has some limitations. We used time-split data from the same hospitals for conducting external validation. Although incidence and characteristics differed over time, validating our models using data from unseen EDs would provide stronger evidence in future studies. Moreover, large-scale multicenter validation is needed to demonstrate real-world utility of our models. We selected shallow neural net models with only one hidden layer. Exploring various backbone model architecture would be valuable. Our model revealed that the FL approach underperformed other models when it was applied to data from specific institutions. This limitation could potentially be addressed by incorporating transfer learning into the FL model. Furthermore, several FL algorithms have been proposed to address challenges of non-independent and identically distributed (non-IID) data in FL. While we applied the commonly used FedAvg algorithm, future studies could explore alternative algorithms such as federated proximal (FedProx), stochastic controlled averaging for federated learning (SCAFFOLD) and federated normalized averaging (FedNova), which have been specifically designed to address non-IID issues [39].

5. Conclusions

FL models for predicting critical interventions and poor clinical outcomes at triage stage presented performances comparable to those of local models. Further studies with multiple sites and advanced algorithms will be needed to improve the reliability and generalizability of our models.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this study are available on reasonable request from the corresponding author.

AUTHOR CONTRIBUTIONS

SH—Conceptualization, Study Design, Drafting of the Manuscript, Data Interpretation. GC—Data Curation, Statistical Analysis, Algorithm Development, Drafting of the Manuscript. SMK—Data Curation, Statistical Analysis, Visualization. HJJ—Data Curation, Methodology. JHK—Data Curation, Validation. SYS—Data Curation, Critical Review. HY—Data Curation, Resources. SYH—Data Curation, Investigation. HC—Data Curation, Investigation. JYY—Critical Review, Manuscript Review & Editing. WCC—Data Curation, Investigation. SUL—Conceptualization, Study Design, Supervision, Data Interpretation, Final Approval.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Institutional Review Boards of Korea University ANAM Hospital (No. 2021AN0545) and Samsung Medical Center (No. 2022-08-175), both of which waived informed consent based on institutional guidelines.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research was supported by a grant (grant number: HI21C1367010023) 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.

CONFLICT OF INTEREST

The authors declare no conflict of interest. Won Chul Cha is serving as one of the Editorial Board members of this journal. We declare that Won Chul Cha had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to YK.

REFERENCES

[1] Gaieski DF, Agarwal AK, Mikkelsen ME, Drumheller B, Cham Sante S, Shofer FS, *et al*. The impact of ED crowding on early interventions and mortality in patients with severe sepsis. The American Journal of Emergency Medicine. 2017; 35: 953–960.

[2] Gruen RL, Jurkovich GJ, McIntyre LK, Foy HM, Maier RV. Patterns of errors contributing to trauma mortality: lessons learned from 2,594 deaths. Annals of Surgery. 2006; 244: 371–380.

[3] Hasegawa K, Sullivan AF, Tsugawa Y, Turner SJ, Massaro S, Clark S, *et al*. Comparison of US emergency department acute asthma care quality: 1997–2001 and 2011–2012. The Journal of Allergy and Clinical Immunology. 2015; 135: 73–80.

[4] Rathore SS, Curtis JP, Chen J, Wang Y, Nallamothu BK, Epstein AJ, *et al*. Association of door-to-balloon time and mortality in patients admitted to hospital with ST elevation myocardial infarction: national cohort study. The BMJ. 2009; 338: b1807.

[5] Goto T, Camargo CA III, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. JAMA Network Open. 2019; 2: e186937.

[6] Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA III, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Critical Care. 2019; 23: 64.

[7] Khera R, Haimovich J, Hurley NC, McNamara R, Spertus JA, Desai N, *et al*. Use of machine learning models to predict death after acute myocardial infarction. JAMA Cardiology. 2021; 6: 633–641.

[8] Sánchez-Salmerón R, Gómez-Urquiza JL, Albendín-García L, Correa-Rodríguez M, Martos-Cabrera MB, Velando-Soriano A, *et al*. Machine learning methods applied to triage in emergency services: a systematic review. International Emergency Nursing. 2022; 60: 101109.

[9] Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, *et al*. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. Annals of Emergency Medicine. 2018; 71: 565–574.e2.

[10] Kamunga BLG, Bearnot CJ, Martin KD, Uwamahoro DL, Cattermole GN. Epidemiology and outcomes of critically ill patients in the emergency department of a tertiary teaching hospital in Rwanda. International Journal of Emergency Medicine. 2024; 17: 170.

[11] Mohr NM, Wessman BT, Bassin B, Elie-Turenne MC, Ellender T, Emlet LL, *et al*. Boarding of critically ill patients in the emergency department. Critical Care Medicine. 2020; 48: 1180–1187.

[12] Pang PS, Levy P, Shah SJ. Treatment of acute heart failure in the emergency department. Expert Review of Cardiovascular Therapy. 2013; 11: 1195–1209.

[13] Hwang U, Weber EJ, Richardson LD, Sweet V, Todd K, Abraham G, *et al*. A research agenda to assure equity during periods of emergency department crowding. Academic Emergency Medicine. 2011; 18: 1318–1323.

[14] Liu Y, Gao J, Liu J, Walline JH, Liu X, Zhang T, *et al*. Development and validation of a practical machine-learning triage algorithm for the detection of patients in need of critical care in the emergency department. Scientific Reports. 2021; 11: 24044.

[15] Chang H, Yu JY, Yoon S, Kim T, Cha WC. Machine learning-based suggestion for critical interventions in the management of potentially severe conditioned patients in emergency department triage. Scientific Reports. 2022; 12: 10537.

[16] Dou Q, So TY, Jiang M, Liu Q, Vardhanabhuti V, Kaissis G, *et al*. Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study. NPJ Digital Medicine. 2021; 4: 60.

[17] Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, *et al*. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific Reports. 2020; 10: 12598.

[18] Alreshidi FS, Alsaffar M, Chengoden R, Alshammari NK. Fed-CL—an atrial fibrillation prediction system using ECG signals employing federated learning mechanism. Scientific Reports. 2024; 14: 21038.

[19] Zhou J, Wang X, Li Y, Yang Y, Shi J. Federated-learning-based prognosis assessment model for acute pulmonary thromboembolism. BMC Medical Informatics and Decision Making. 2024; 24: 141.

[20] Jordan S, Fontaine C, Hendricks-Sturrup R. Selecting privacy-enhancing technologies for managing health data use. Frontiers in Public Health. 2022; 10: 814163.

[21] Malin B, Goodman K; Section Editors for the IMIA Yearbook Special Section. Between access and privacy: challenges in sharing health data. Yearbook of Medical Informatics. 2018; 27: 55–59.

[22] Yazijy S, Schölly R, Kellmeyer P. Towards a toolbox for privacy-preserving computation on health data. Studies in Health Technology and Informatics. 2022; 290: 234–237.

[23] Zhang P, Kamel Boulos MN. Privacy-by-design environments for large-scale health research and federated learning from data. International Journal of Environmental Research and Public Health. 2022; 19: 11876.

[24] Rodríguez-Barroso N, Stipcich G, Jiménez-López D, Ruiz-Millán JA, Martínez-Cámara E, González-Seco G, *et al*. Federated learning and differential privacy: software tools analysis, the Sherpa.ai FL framework and methodological guidelines for preserving data privacy. Information Fusion. 2020; 64: 270–292.

[25] Dugas AF, Kirsch TD, Toerper M, Korley F, Yenokyan G, France D, *et al*. An electronic emergency triage system to improve patient distribution by critical outcomes. The Journal of Emergency Medicine. 2016; 50: 910–918.

[26] Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. Critical care. 2019; 23: 374.

[27] McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. Proceedings of Machine Learning Research. 2017; 54: 1273–1282.

[28] Van der werf AJ. Federated learning in the emergency department [Bachelor's thesis]. Tilburg University. 2022.

[29] Sarma KV, Harmon S, Sanford T, Roth HR, Xu Z, Tetreault J, *et al*. Federated learning improves site performance in multicenter deep learning without data sharing. Journal of the American Medical Informatics Association. 2021; 28: 1259–1264.

[30] Cho KJ, Kim JS, Lee DH, Lee SM, Song MJ, Lim SY, *et al*. Prospective, multicenter validation of the deep learning-based cardiac arrest risk management system for predicting in-hospital cardiac arrest or unplanned intensive care unit transfer in patients admitted to general wards. Critical Care. 2023; 27: 346.

[31] Scheeren TWL, Bakker J, De Backer D, Annane D, Asfar P, Boerma EC, *et al*. Current use of vasopressors in septic shock. Annals of Intensive Care. 2019; 9: 20.

[32] Sanfilippo F, La Via L, Carpinteri G, Astuto M. Timing of intubation, beds in intensive care and inter-hospital transfer: rings of a complex chain during pandemic conditions. Critical Care. 2022; 26: 44.

[33] Gopinath B, Sachdeva S, Kumar A, Kumar G. Advancing emergency airway management by reducing intubation time at a high-volume academic emergency department. BMJ Open Quality. 2021; 10: e001448.

[34] Guerci P, Belveyre T, Mongardon N, Novy E. When to start vasopressin in septic shock: the strategy we propose. Critical Care. 2022; 26: 125.

[35] Evans HL, Zonies DH, Warner KJ, Bulger EM, Sharar SR, Maier RV, *et al*. Timing of intubation and ventilator-associated pneumonia following injury. Archives of Surgery. 2010; 145: 1041–1046.

[36] Oredsson S, Jonsson H, Rognes J, Lind L, Göransson KE, Ehrenberg A, *et al*. A systematic review of triage-related interventions to improve patient flow in emergency departments. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine. 2011; 19: 43.

[37] Fekonja Z, Kmetec S, Fekonja U, Mlinar Reljić N, Pajnkihar M, Strnad M. Factors contributing to patient safety during triage process in the emergency department: a systematic review. Journal of Clinical Nursing. 2023; 32: 5461–5477.

[38] Chiu YM, Vanasse A, Courteau J, Chouinard MC, Dubois MF, Dubuc N, *et al*. Persistent frequent emergency department users with chronic conditions: a population-based cohort study. PLOS ONE. 2020; 15: e0229022.

[39] Wang J, Liu Q, Liang H, Joshi G, Poor HV. A novel framework for the analysis and design of heterogeneous federated learning. IEEE Transactions on Signal Processing. 2021; 69: 5234–5249.

How to cite this article: Sejin Heo, Geunho Choi, Su Min Kim, Hyung Joon Joo, Jong-Ho Kim, Soo-Yong Shin, *et al*. Federated learning for predicting critical intervention and poor clinical outcomes at emergency department triage stage. Signa Vitae. 2025; 21(5): 21-28. doi: 10.22514/sv.2025.062.