

## Article

# Multicenter Analysis of Emergency Patient Severity through Local Model Evaluation Client Selection: Optimizing Client Selection Based on Local Model Evaluation

Yong-gyom Kim, SeMo Yang and KangYoon Lee \* 

Department of IT Convergence Engineering, Gachon University Seoul, Seongnam-Si 13120, Republic of Korea; gyom1204@gachon.ac.kr (Y.-g.K.); tpah20@gachon.ac.kr (S.Y.)

\* Correspondence: keylee@gachon.ac.kr

**Abstract:** In multi-institutional emergency room settings, the early identification of high-risk patients is crucial for effective severity management. This necessitates the development of advanced models capable of accurately predicting patient severity based on initial conditions. However, collecting and analyzing large-scale data for high-performance predictive models is challenging due to privacy and data security concerns in integrating data from multiple emergency rooms. To address this, our work applies federated learning (FL) techniques, maintaining privacy without centralizing data. Medical data, which are often non-independent and identically distributed (non-IID), pose challenges for existing FL, where random client selection can impact overall FL performance. Therefore, we introduce a new client selection mechanism based on local model evaluation (LMECS), enhancing performance and practicality. This approach shows that the proposed FL model can achieve comparable performance to centralized models and maintain data privacy. The execution time was reduced by up to 27% compared to the existing FL algorithm. In addition, compared to the average performance of local models without FL, our LMECS improved the AUC by 2% and achieved up to 23% performance improvement compared to the existing FL algorithm. This work presents the potential for effective patient severity management in multi-institutional emergency rooms using FL without data movement, offering an innovative approach that satisfies both medical data privacy and efficient utilization.

**Keywords:** federated learning; centralized learning; patient severity analysis; local model evaluation; client selection



**Citation:** Kim, Y.-g.; Yang, S.; Lee, K. Multicenter Analysis of Emergency Patient Severity through Local Model Evaluation Client Selection: Optimizing Client Selection Based on Local Model Evaluation. *Appl. Sci.* **2024**, *14*, 6876. <https://doi.org/10.3390/app14166876>

Academic Editors: Ioannis Doumanis and Daphne Economou

Received: 27 June 2024

Revised: 25 July 2024

Accepted: 30 July 2024

Published: 6 August 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

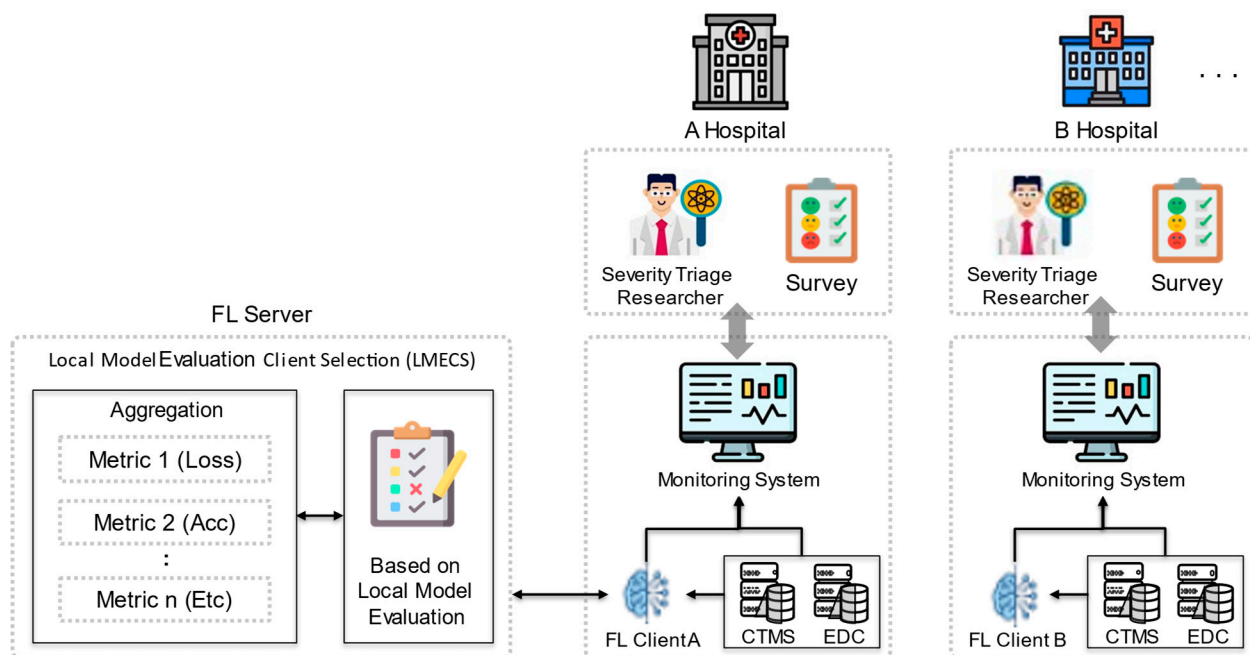
## 1. Introduction

Kondo Y et al. (2019) [1] state that a key issue in emergency rooms is classifying patient severity based on their initial state, enabling appropriate on-site treatment, and Hansen J et al. (2022) [2] and Gyu-Sung Ham et al. (2022) [3] analyze the early identification of high-risk patients to improve survival rates. According to Zhiqiang Liu et al. (2022) [4] and Yoojoong Kim et al. (2023) [5], this requires extensive data analysis, where the current method involves transferring data to a central data center for processing and analysis. However, centralized data analysis has drawbacks. Firstly, Sheller Micah J et al. (2020) [6] and Kaissis Georgios et al. (2021) [7] state that medical data are highly sensitive and centralizing them poses privacy risks. Secondly, Ahmadi N et al. (2022) [8] and Boggs et al. (2020) [9] propose that integrating data from multiple emergency rooms is challenging due to privacy regulations and inconsistencies in data formats among institutions. Thirdly, Dinh C. Nguyen (2022) [10] explores how delays and additional costs associated with data centralization can degrade model performance and utility.

According to Xu Jie et al. (2021) [11], Rieke Nicola et al. (2020) [12] and Lim Wei Yang Bryan et al. (2020) [13], federated learning (FL) is a solution to these issues. FL analyzes local data in situ at each institution and transmits only the learning results to a central

server, creating and sharing a global model while safeguarding data privacy and minimizing the risks and costs associated with data movement. Despite its advancements, X. Lian et al. (2017) [14] explain that FL faces challenges with computational efficiency and prediction performance. Existing FL often randomly involves clients in the learning process, potentially negatively impacting FL performance and introducing uncertainty. Furthermore, according to V. Smith et al. (2018) [15], disparate computational capacities among clients can increase computational load and extend training time. S. Ji et al. (2021) [16], W. Lin et al. (2022) [17] and Karimireddy et al. (2020) [18] state that not all clients contribute equally to FL performance; some may introduce noise or bias due to non-IID data quality and heterogeneity.

To address these issues, this work proposes a new client selection algorithm designed to improve the overall efficiency and effectiveness of FL, as shown in Figure 1. This algorithm performs client selection by evaluating the client's local model (LMECS). The global model is updated by aggregating the parameters of the selected local model. The proposed LMECS utilizes a metric-based evaluation system to determine the contribution of each client's local model to the improvement of the global model's performance. Clients possessing models that meet specific performance criteria such as loss, accuracy, and area under the ROC curve (AUC) are given priority in the aggregation process for creating the global model. This not only accelerates the performance improvement of the global model but also increases resilience against data discrepancies and anomalous client updates.



**Figure 1.** Federated learning scenario with LMECS that applies to real-world environments.

By proposing this LMECS, this work aims to not only comply with the stringent privacy requirements for medical data but also present an advanced FL model that shows superior performance metrics compared to existing FL algorithms. Through this work, the goal is to extend to the rapid and accurate identification of high-risk patients in emergency rooms without data sharing, potentially improving patient outcomes and operational efficiency within the healthcare system. This work contributes in the following ways:

- **FL-based Severity Analysis of Patients:** By utilizing FL technology, this work overcomes the limitations of traditional centralized data collection and analysis, enabling accurate analysis of the severity of patients in multi-institutional emergency rooms while maintaining data privacy.
- **Flexible and Efficient Client Selection:** A new algorithm is proposed to evaluate clients' local models during the FL process and select the optimal clients based on custom

metrics. This accelerates performance improvement speed, reduces communication costs, and enhances the overall predictive performance of the model.

- **Validation Using Real Multi-Institutional Emergency Room Data:** The utility and validity of the proposed client selection algorithm are demonstrated through performance comparisons of centralized learning, standalone learning, and FL models using real multi-institutional emergency room data from Korea. The experimental validation strengthens the validity of the performance comparison and increases the practical applicability of the research findings.

The paper is organized as follows: Section 2 covers the related work pertinent to this work. In Section 3, the data, models, and FL methods employed in the research are introduced. Section 4 presents the environment in which the research experiments were conducted. The results of the research are discussed in Section 5. Finally, Section 6 summarizes the research and proposes directions for future research.

## 2. Related Works

### 2.1. Patient Severity Analysis Based on Medical Data

To analyze patient severity, hospitals utilize clinical data and various machine learning algorithms. Studies are being conducted in areas such as predicting cardiovascular events in patients [19], analyzing the severity of COVID-positive patients [20,21], and assessing the condition of emergency patients [22,23] using data and predictive models. Sánchez-Salmerón et al. [24] explain that artificial intelligence models applied in these studies include logistic regression, random forests, XGBoost, and deep learning, among other types. According to Álvaro Valencia-Parra et al. [25], the outcomes of such models are determined by the quality of the data used. Haitao Song et al. [26] state that clinical data from hospitals typically involve a small number of patients and exhibit a non-IID nature. Li Y et al. [27] explain that when data are imbalanced, models may not learn effectively, leading to decreased performance in predictions. To address this issue, it is necessary to increase the amount of trainable data and minimize bias to enhance predictions. For this reason, recent research starts with collecting as much data as possible [28]. However, in the field of medical research, there is opposition to centralizing data storage, and concerns about data sharing and privacy protection remain [29,30].

In this study, we used the NEDIS dataset. NEDIS is a system designed to support emergency medical policy decisions related to the occurrence, transportation, treatment, and discharge of emergency patients. Data from the emergency rooms of 172 hospitals, collected from 2017 to 2022, were used. The dataset includes vital information about emergency patients, with personally identifiable information and codes removed. The data are labeled using the Korean Triage and Acuity Scale (KTAS) to classify the severity levels of patients, and a total of 2,051,908 data entries were used.

### 2.2. Federated Learning in Healthcare

FL is structured in a client–server architecture, with clients often set up in hospital environments, applied in multiple institutions in the healthcare sector. Research has been conducted using basic algorithms like FedAvg, fundamental to FL, to predict the risk status of patients in hospitals with non-IID data [31,32]. Additionally, studies have been performed using FL for personalized healthcare services [10], like individual health management and the analysis of mortality rates within hospitals, based on electronic health records (EHRs) [33,34].

FL consists of two main components: local updates by distributed clients and global updates, where a central server aggregates the parameters from each client. The basic local update in FL is carried out as follows.

$$C = \{C^1, C^2, C^3, \dots, C^i\} \quad (1)$$

$$\mathcal{W}_{t+1}^i = \mathcal{W}_t - \alpha \nabla \mathcal{L}(C^i, \mathcal{W}_t) \quad (2)$$

$C$  represents the local dataset held by each client.  $W$  denotes the parameters of the learning model, while  $t$  and  $i$  represent the round and client, respectively.  $\alpha$  is the learning rate, and  $L$  represents the loss function. The basic global model update in FL is carried out as follows:

$$\mathcal{W}_{t+1} = \sum_i^k P_i \mathcal{W}_{t+1}^i \quad (3)$$

$P_i$  represents the importance or weight of a client's local model in the global update. In foundational FL approaches like FedSGD and FedAvg [35],  $P_i$  can be  $k_i/k$ . Ultimately, the goal of FL is to optimize the objective function.

$$\begin{aligned} & \min_{\mathcal{W}} \mathcal{F}(\mathcal{W}), \text{ where } \mathcal{F}(\mathcal{W}) \\ & = E[\mathcal{L}(\mathcal{W}, C)] + \lambda * \mathcal{G}(\mathcal{W}) \end{aligned} \quad (4)$$

$E[L(W, D)]$  denotes the expected value of the update from the local client's model and dataset, while  $\lambda$  and  $G(W)$  represent the hyperparameters controlling the intensity of regularization and the regularization term, respectively. A notable example is FedProx [36]. In the case of FedSGD and FedAvg, there is no separate regularization term, but various algorithms add, while  $\lambda$  and  $G(W)$  to aim for improved model performance. These FL algorithms have no criteria for client selection and usually select clients at random.

### 2.3. Client Selection in Federated Learning

In typical FL, clients participating in the training of the global model are randomly selected. However, since each client's data are non-IID, simply selecting clients randomly does not guarantee the improved performance of the global model [37]. To address this issue, it is necessary to select clients in a way that enhances the performance of the global model. Research has been conducted focusing on two primary types of client selection: resource-centric and performance-centric. Resource-centric client selection has been studied in environments where client resources are limited, such as IoT settings [38,39]. In environments where data statistics are more critical than device performance, such as with medical data, research has mainly focused on selecting clients based on the performance of their local models. Studies related to performance-centric client selection include those selecting clients based on the local model's loss value to achieve rapid convergence and better performance [40], as well as research selecting clients based on the accuracy of their local models [41]. The findings of these studies are empirical, based on experimental results [42,43]. Therefore, this work also aims to derive empirical results through various experiments.

## 3. Method

To predict the severity of patients in multi-institutional emergency rooms based on FL and to verify the utility of FL, data from various hospital emergency rooms in Korea were utilized. Additionally, a severity analysis model layer demonstrating superior performance was employed. This work also showcases a differentiated client selection algorithm from existing FL algorithms, demonstrating its superior performance compared to the conventional algorithms.

### 3.1. Dataset

The dataset used in this work is the Korean national emergency department information system (NEDIS). NEDIS supports emergency medical policy decision-making by tracking the emergency medical system process from the occurrence of an emergency patient to their transfer, treatment, and discharge [44]. The data used in this work consist of real-time collected patient visit information from emergency rooms of 172 hospitals in South Korea, spanning from 2017 to 2021. The variables provided in the dataset are described in Table 1. To differentiate FL clients (hospitals), emergency institution numbers were used. These institution numbers are anonymized values of the institution names. For predicting the severity of a patient's condition, basic patient information, symptom onset

time, initial vital signs, response, main symptoms, and KTAS (Korean Triage and Acuity Scale) severity indicators were utilized. The target variable for severity analysis was based on the emergency treatment outcomes. Patients admitted to the ICU or who had died were defined as requiring critical care, while all other cases were defined as general patients. These definitions were used to establish the target for patient severity analysis. Information regarding patient transfers was excluded since the final emergency treatment outcomes for these patients could not be ascertained.

**Table 1.** Data characteristics information.

Characteristics	Description	Details
Emergency Room (ER) ID	Classification number for each ER	Composed of 172 hospitals in total
Basic Information	Patient's basic demographic information	Gender, Age
Time of Symptom Onset	Time from onset of symptoms to arrival at the ER	Distributed from 1 h to more than 30 days
Early Biosignals	Initial vital signs information measured in the ER	Blood Pressure (Systolic, Diastolic), Heart rate, Respiration rate, Body Temperature, Oxygen Saturation rate
Response Upon Visit	Patient's initial reaction seen in the emergency room	Alert, Reacting to voice, Reacting to pain, Unresponsive
Main Symptoms	Main symptoms of patients who came to the ER	Pain, Fever, Dizziness, etc.
KTAS Severe Index	Korean emergency patient classification information	Resuscitation, Urgent, Emergency, Semi-emergency, non-emergency
Outcomes	Patient final status outcome in the ER	Return Home, Transfer, Hospitalization (General, ICU), Death

### 3.2. Model

To predict the severity of emergency room patients, a model based on deep neural networks (DNNs) was created. The model's structure references a feedforward neural network structure that has shown excellent performance in severity analysis problems [45]. The model includes a total of five hidden layers, with each layer comprising 89 neurons. This configuration provides sufficient complexity to learn various patterns necessary for predicting severity. Each hidden layer undergoes a linear transformation followed by batch normalization to standardize the layer's output, and a ReLU activation function is applied. Additionally, a dropout rate of 0.5 is set for each layer to contribute to the prevention of overfitting. In the output layer, a sigmoid activation function is applied to adjust the result value to between 0 and 1, thereby predicting the patient's severity.

### 3.3. Proposed Federated Learning Algorithm

Existing FL algorithms aggregate local models from multiple clients using simple averaging, random sampling, or by adding regularization terms, and then update the global model. These approaches have several inherent issues in terms of efficiency and practicality [46]. Simple averaging, which aggregates under the assumption that all clients are equally important, can be inefficient for model optimization as it does not consider the quality, quantity, and distribution of data. Additionally, random sampling is reliant on chance and cannot eliminate the possibility that some inefficient local models might adversely affect the global model. The method of adding regularization terms can also

increase the complexity of the model, and finding the appropriate regularization term can be challenging. To solve these issues, this paper proposes a client selection algorithm based on local model evaluation. The proposed algorithm flexibly selects clients according to various evaluation metrics that determine the model's performance. This dynamic selection mechanism evaluates and compares each client's local model using metrics such as loss, accuracy, AUC, or other custom metrics. To express the LMECS algorithm mathematically, it is necessary first to define the process of evaluating each client's local model and the criteria for their selection. The LMECS algorithm is as follows. The method for evaluating a client's local model is as follows:  $S_i$  represents the score of client  $i$ 's local model.  $E$  evaluates how well the local model performs on the dataset  $C_i$  using the global model parameters  $W_t$ .

$$S_i = E(C^i, W_t) \quad (5)$$

A set of clients  $C_s$  that meet a user-defined criterion  $\theta$  is defined. A client  $i$  is selected when its score  $S_i$  is equal to or exceeds the threshold  $\theta$ .

$$C_s = \{ i \mid S_i \geq \theta \} \quad (6)$$

The local model parameters of the selected clients are aggregated to update the global model. Each client updates its model based on its local data and sends these updated model parameters  $W_{t+1}^i$  to the central server. The FL server aggregates these updates, applying weights divided by the number of selected clients  $|C_s|$ . Consequently, this formula updates the global model using the parameters of the selected local models, creating a new model  $W_{t+1}$  to be used in the next round of training.

$$W_{t+1} = \frac{1}{|C_s|} \sum_{i \in C_s} W_{t+1}^i \quad (7)$$

The pseudocode of the proposed algorithm is as follows, as illustrated in Algorithm 1. During the selection process,  $S_i$  represents the performance score of each client's local model, while  $S_T$  is the threshold value used to select clients based on this score. A client is only selected if its score exceeds this threshold, and it becomes part of the set  $C_s$ . Only the local updates from clients in this set are used to improve the global model.

The proposed algorithm evaluates clients' local models based on metrics, selecting only those that meet a user-defined criterion for inclusion in the aggregation process to update the global model. Updates from clients that could degrade the performance of the global model due to lower quality are excluded. This approach ensures that only updates from clients contributing positively to the model's performance and efficiency are considered.

---

#### Algorithm 1 Local Model Evaluation based Client Selection

---

- 1: **Input:** Set of all clients  $C$ , datasets  $D_i$  of each client  $i$
  - 2: **Output:** Aggregate global model parameters from selected clients  $W$
  - 3: Initialize global model parameters  $W$
  - 4: for each round  $t$  do
  - 5: if  $t == 0$  then
  - 6:   Select all subset of clients  $C_s$
  - 7: else
  - 8:   Select clients  $C_s$  from previously selected clients
  - 9: end if
  - 10: for each client  $i \in C_s$  in parallel do
  - 11: if client  $i$  is selected then
  - 12:   Local training on client  $i$
  - 13:   for each local epoch  $e$  to  $E$  do
  - 14:     for each batch  $b$  in local dataset  $D_i$  do
  - 15:       Update local model  $W_i$
-

**Algorithm 1** *Cont.*


---

```

16:   end for
17:   end for
18:   end if
19:   Evaluate local model and calculate score  $S_i = E(C_i, W_t)$ 
20:   Communicate the score  $S_i$  to the server
21:   Send updated model  $W_{t+1}^i$  to server
22:   end for
23:   Server-side client selection based on score  $S_i$ 
24:   Selection threshold is  $\theta$ 
25:   for each client  $i \in C_s$  do
26:     if  $S_i \geq \theta$  then
27:       Add  $i$  to  $C_s$ 
28:     end if
29:   end for
30:   Aggregate updates from selected clients to update global model  $W_{t+1} =$ 
       $Aggregate(C_s, W_{t+1}^i)$ 
31:   end for
32:   Function Aggregate:
33:   Input: Set of selected clients  $C_s$ , updated model parameters  $W_{t+1}$ 
34:   Output: Aggregate global model parameters  $W_{t+1}$ 
35:    $W_{t+1} = \frac{1}{|C_s|} \sum_{i \in C} W_{t+1}^i$ 
36:   End Function

```

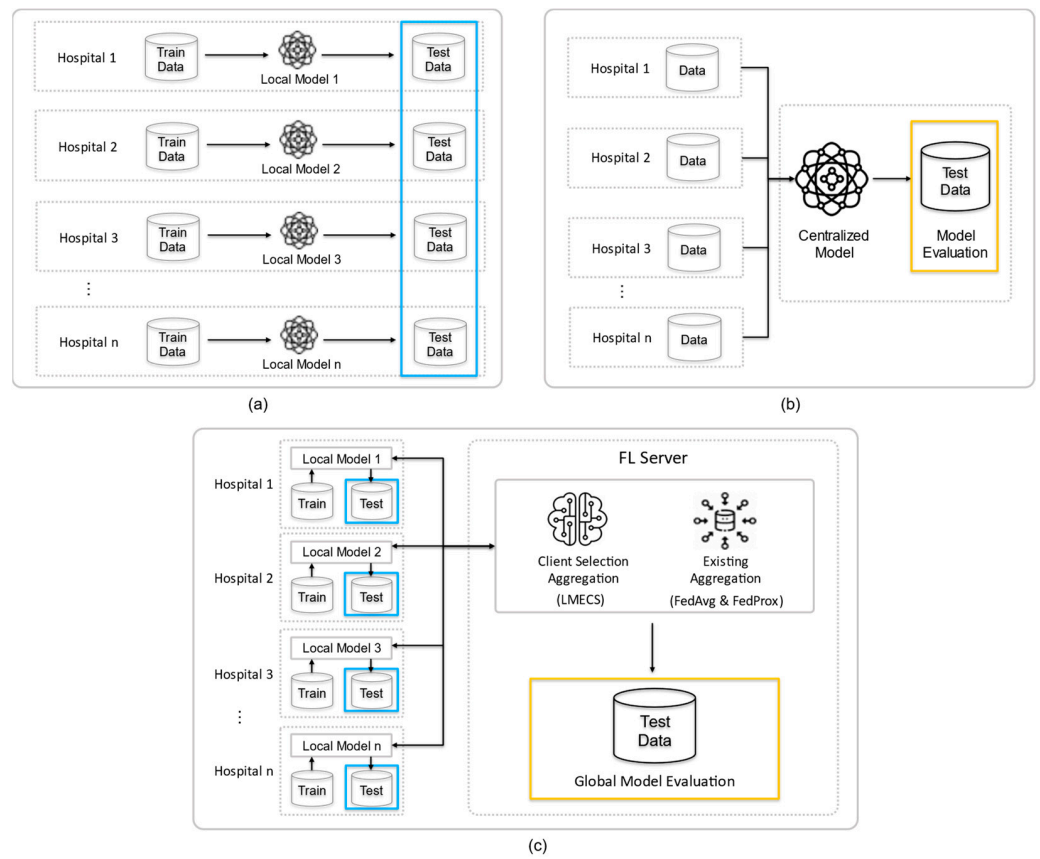
---

**4. Experimental Design**

To validate the severity analysis scenario for patients in multi-institutional emergency rooms, we established three experimental environments, as shown in Figure 2: standalone learning (SL), centralized learning (CL), and FL. All experiments utilized the previously mentioned NEDIS data. To ensure comparability, each experiment under the three conditions used the same configuration of the dataset for each hospital. This approach allowed us to evaluate the efficiency of each model in a controlled manner, providing a clear comparison of how these different learning environments perform in the context of emergency room patient severity analysis. This controlled approach was crucial in determining the most effective method for analyzing patient severity in a multi-institutional emergency setting.

**4.1. Environment Setup**

We performed FL simulations based on Ray [47] using the Flower framework [48] in a server environment equipped with an Intel(R) Core (TM) i9-10980XE CPU (3.00 GHz) (Intel Corporation, Santa Clara, CA, USA) and two NVIDIA GeForce RTX 3090 GPUs (NVIDIA Corporation, Santa Clara, CA, USA). Additionally, we were able to monitor and observe the results of this experiment by utilizing the FedOps platform (version 1.1). [49–51], which can manage the FL lifecycle. The data processing and modeling were carried out using Pytorch (version 2.2.0). To facilitate unbiased and accurate comparative analysis across the three experimental environments, we defined a rigorous experimental design. Data from 50 randomly selected hospitals were used, and to ensure reliability and statistical significance, 10 trials were conducted for each set. Despite the random selection of hospitals, each experimental environment was structured with the same dataset to ensure comparability. The hyperparameter information for all experiments is presented in Table 2. The data configuration for each experiment is detailed in Table 3, and a non-IID data distribution was observed in all experimental environments. This thorough experimental setup ensured that the comparative analysis between standalone learning, centralized learning, and FL environments was robust, allowing for a comprehensive understanding of each approach in the context of emergency room patient severity analysis.



**Figure 2.** Experimental design of this study: (a) is standalone learning (b) is centralized learning (CL) and (c) is federated learning (FL). The blue frame is the test data for testing the client’s model, and the yellow frame is the data for testing the global model.

**Table 2.** Hyperparameter configuration used in the experiment.

Hyperparameter	Notation	Value
Number of CPUs per Client	$c_{cpu}$	4
Number of GPUs per Client	$c_{gpu}$	0
Server GPU Resources	$s_{gpu}$	1
Communication Round/Iteration	$r$	100
Number of Epochs per Round	$e$	2
Batch Size	$b$	1024
Criterion for Loss	$L$	BinaryCross Entropy
Optimization	$opt$	Adam
Learning Rate	$\eta$	0.001

**Table 3.** Non-IID data configuration for each experiment.

Experiment Number	Client Data (Train)			Evaluation Data (Test)		
	Non-Critical Care	Critical Care	Total Size	Non-Critical Care	Critical Care	Total Size
1	93.2%	6.8%	3,866,342	93.09%	6.91%	773,268
2	93.35%	6.65%	3,301,971	93.03%	6.97%	660,394
3	93.71%	6.29%	3,492,681	92.94%	7.06%	698,536
4	93.62%	6.38%	2,476,352	92.91%	7.09%	495,270
5	93.8%	6.2%	3,574,151	92.83%	7.17%	714,830
6	92.55%	7.45%	2,418,920	93.4%	6.6%	483,784
7	94.1%	5.9%	3,595,783	92.74%	7.26%	719,156
8	91.93%	8.07%	3,547,194	93.59%	6.41%	709,438
9	93.4%	6.6%	2,665,113	93.05%	6.95%	544,022
10	92.75%	7.25%	3,763,022	93.26%	6.74%	752,604



#### 4.2. Standalone Learning

In the standalone environment, each client trains its local model  $C$  using its local dataset  $D_i$ . This scenario reflects a situation where clients operate independently without sharing data. Each client's dataset is split into a training set and a test set in an 8:2 ratio, with the test set used to evaluate the performance of each client's local model (Figure 2a).

#### 4.3. Centralized Learning

In the centralized environment, all participating clients' data  $AD_i = \{D_1, D_2, \dots, D_i\}$  are aggregated to train a centralized model  $AC$ . The training set in this model comprises 10 parts, and the test set has two parts. The training set includes data from each hospital, and the test set is randomly composed of data not included in the training set. The performance of the centralized model is evaluated using this test set (Figure 2b).

#### 4.4. Federated Learning

In the federated environment, each client trains and evaluates its local model without data sharing and shares the results to create and assess a global model. The local models are updated based on the updated global model, and the clients then retrain and re-evaluate these models. The performance comparison of FL is performed through the existing algorithms FedAvg and FedProx, which randomly select clients, and the proposed LMECS algorithm. To compare the performance of FL and CL, a test set for CL and a global model evaluation dataset for FL were constructed. Both datasets comprise the same data. Additionally, to compare the performance of FL and SL, the ratio of training to test sets for each client in FL was structured the same as in SL, and a performance evaluation of each local model was carried out using the test set (Figure 2c).

### 5. Results

In our work, we utilized a comprehensive set of performance metrics—accuracy, sensitivity, specificity, PPV, NPV, and AUC—with a 95% confidence interval (95% CI), to assess model effectiveness. The 95% CI is particularly critical in clinical settings, ensuring a robust evaluation of the model's diagnostic capabilities and allowing for accurate comparison between models. These metrics were selected due to their relevance in clinical settings, where a holistic view of model performance is essential. Accuracy measures the proportion of true results among the total number of cases examined. Sensitivity and specificity provide insight into the model's ability to correctly identify positive and negative cases, respectively. PPV and NPV offer a perspective on the model's predictive power, while AUC provides a cumulative measure of performance across all classification thresholds. These metrics collectively offer a robust evaluation framework for the predictive models' diagnostic capabilities. In comparing CL and FL, we included execution time. Execution time comparison provides valuable insights into the trade-offs between model performance and the computational cost associated with the training of each model. This method also allows us to fine-tune the algorithm based on the performance (loss, accuracy, AUC) of the metrics, and ensures that the improvements seen in the LMECS over existing FL algorithms are statistically significant and not due to random chance.

#### 5.1. Comparison of CL and FL

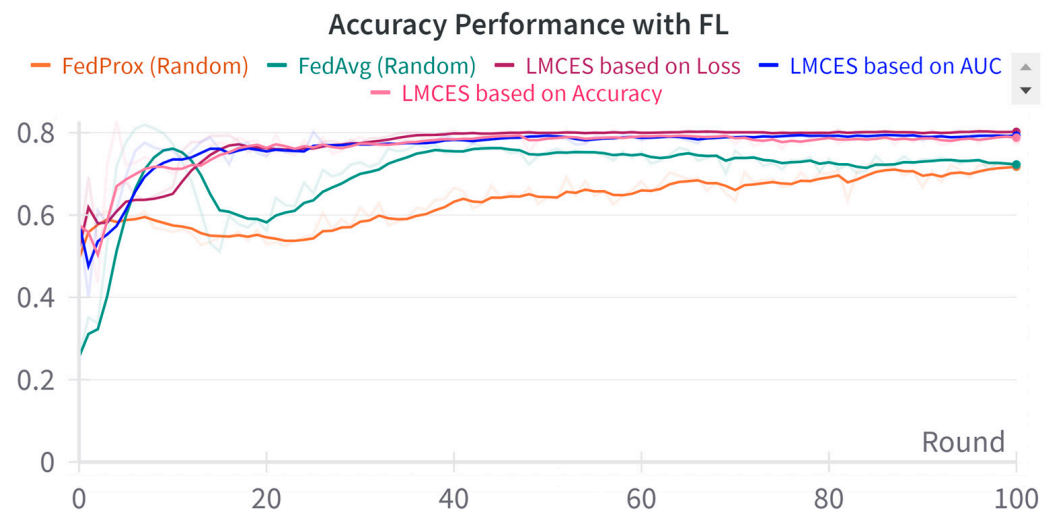
A performance comparison between CL and FL was conducted using the evaluation data presented in Table 3. As seen in Figure 2, in the CL environment, the test set was used, while in FL, a dataset was employed specifically for evaluating the global model. The performance comparison between CL and FL is shown in Table 4. The accuracy-based LMECS showed slightly higher accuracy ( $81.56 \pm 1.63$ ) than CL ( $81.16 \pm 0.29$ ) and outperformed other existing FL algorithms (Figure 3). In sensitivity, which is an indicator of how well the intensive care unit was identified, accuracy-based LMECS ( $83.81 \pm 0.33$ ) showed the most similar performance to CL ( $84.02 \pm 0.16$ ) and was superior to the existing FL algorithm (Figure 4). Specificity was highest in FedAvg ( $83.23 \pm 2.58$ ) compared to CL,

LMECS, and FedProx. However, the lower sensitivity of FedAvg ( $79.13 \pm 1.73$ ) suggests that the model did not adequately capture the nuances of critical care and tended to train in the direction of typically non-critical care. FedAvg’s selection of clients was random, leading to a model trained with clients having less critical care information (severe non-IID data). The PPV and NPV metrics indicated that LMECS was almost similar to CL and superior to other existing FL algorithms. The low performance of PPV and high performance of NPV can be attributed to the relatively low proportion of critical care cases in the total dataset, as shown in Table 3. In terms of AUC, which considers both the sensitivity and specificity for critical care, accuracy-based LMECS ( $81.33 \pm 0.21$ ) was competitive with CL ( $81.92 \pm 0.14$ ) and significantly outperformed FedAvg and FedProx (Figure 5). In terms of execution time, FL showed a 29.7% reduction ( $3001.1 \pm 85.2$ ) compared to CL ( $4268.8 \pm 151$ ), demonstrating that a distributed client environment is more efficient than analyzing large amounts of data at once. This assessment was conducted in a simulation environment, and results may vary in a real-world distributed client setting. However, similar execution times can be expected with stable network conditions. Furthermore, the performance of the proposed FL approach showed a 27.12% decrease in execution time compared to existing FL algorithms (maximum  $4117.7 \pm 107$  s).

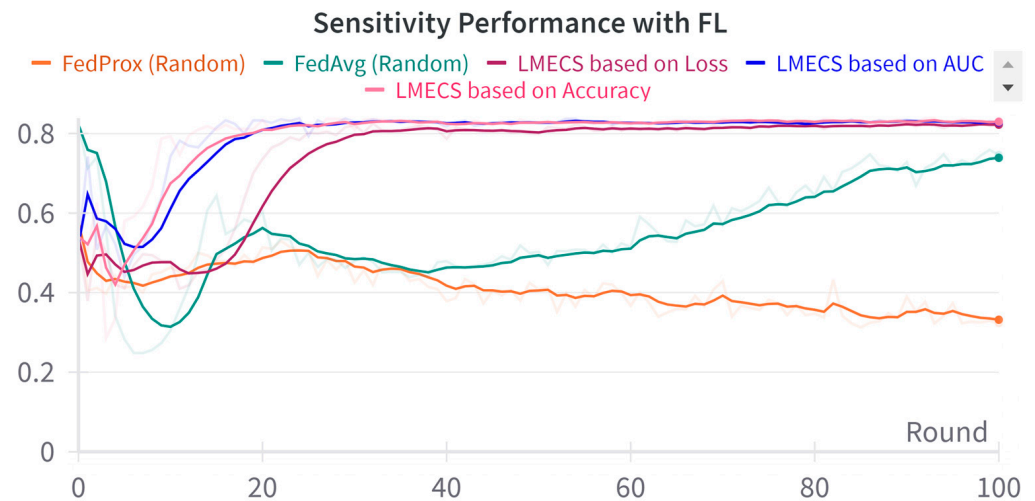
**Table 4.** Comparison of global model performance between centralized learning and federated learning.

Metrics (95% CI)	Centralized Learning	Federated Learning			
		LMECS (Loss)	LMECS (Accuracy)	LMECS (AUC)	FedAvg
Accuracy (%)	$81.16 \pm 0.29$	$80.82 \pm 0.24$	<b><math>81.56 \pm 1.63</math></b>	$81.09 \pm 1.25$	$80.32 \pm 1.92$
Sensitivity (%)	<b><math>84.02 \pm 0.16</math></b>	$83.03 \pm 0.34$	$83.81 \pm 0.33$	$83.78 \pm 2.22$	$79.13 \pm 1.73$
Specificity (%)	$81.15 \pm 0.35$	$80.97 \pm 0.37$	$82.45 \pm 2.76$	$81.57 \pm 2.28$	<b><math>83.23 \pm 2.58</math></b>
PPV (%)	<b><math>24.30 \pm 0.53</math></b>	$23.78 \pm 0.53$	$23.58 \pm 0.45$	$23.49 \pm 0.48$	$22.11 \pm 0.46$
NPV (%)	<b><math>98.53 \pm 0.05</math></b>	$98.44 \pm 0.12$	$98.45 \pm 0.05$	$98.45 \pm 0.04$	$96.25 \pm 0.22$
AUC (%)	<b><math>81.92 \pm 0.14</math></b>	$81.33 \pm 0.21$	$81.18 \pm 0.26$	$81.28 \pm 0.21$	$75.96 \pm 1.34$
Execution Time (s)	$4268.8 \pm 151$	<b><math>3001.1 \pm 85.2</math></b>	$3291 \pm 83.2$	$3278.4 \pm 69.4$	$3207.6 \pm 101$

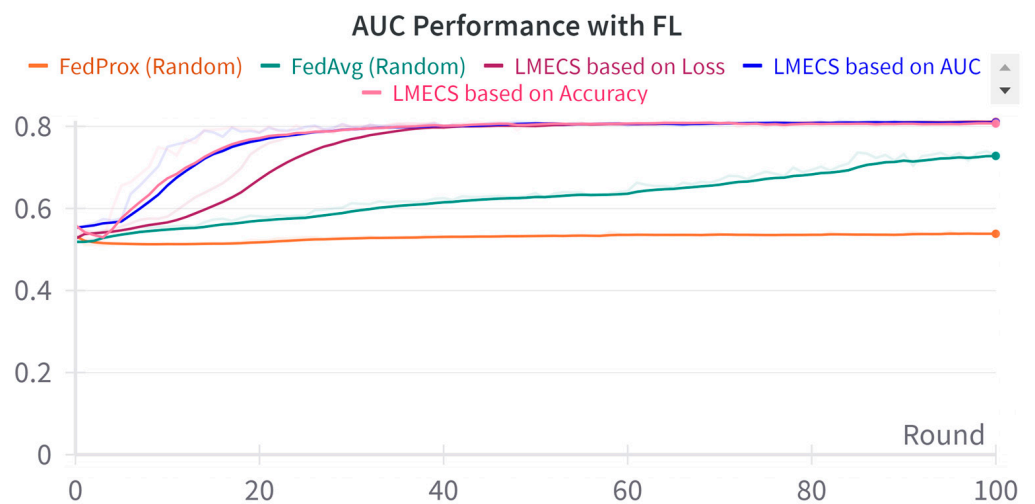
Bold denotes the highest performance in each Metrics.



**Figure 3.** Accuracy performance comparison between LMECS and existing FL algorithms.



**Figure 4.** Sensitivity performance comparison between LMECS and existing FL algorithms.



**Figure 5.** AUC performance comparison between LMECS and existing FL algorithms.

### 5.2. Comparison of SL and FL

SL and FL were compared in terms of the performance of local models using the test data from the client data in Table 3, as shown in Figure 2. Since there were 50 client local models, their performance was averaged for evaluation. The performance comparison between SL and FL is shown in Table 5. The accuracy-based LMECS outperformed both SL and other FL algorithms in overall performance. In terms of accuracy, our algorithm ( $81.42 \pm 2.33$ ) showed a 3% improvement over SL ( $78.17 \pm 0.42$ ) and up to a 9% improvement over existing FL algorithms ( $62.81 \pm 14.8$ ). Sensitivity, an indicator of critical care prediction, improved by 3% ( $83.44 \pm 0.63$ ) compared to SL ( $80.70 \pm 1.55$ ) and also outperformed the existing algorithms. Specificity, PPV, and NPV also showed better performance than both SL and the existing FL models. In terms of AUC performance, the AUC-based LMECS was the highest, but not significantly different from other LMECS algorithms. This LMECS showed a 2% improvement over SL ( $79.68 \pm 0.74$ ) and 6% and 23% improvements over FedAvg ( $75.72 \pm 1.69$ ) and FedProx ( $75.72 \pm 1.69$ ), respectively. These experimental comparisons demonstrate that applying various metrics within the LMECS algorithm can provide adaptability and excellent performance suitable for user-specific tasks. Our experimental results indicate that LMECS can achieve performance levels similar to centralized learning, highlighting its potential to reach the capabilities of a centralized approach while maintaining the benefits of FL. Additionally, when compared to standalone

learning, the use of LMECS-based FL positively impacted the performance of client local models, showcasing an enhancement in functionality over existing FL algorithms.

**Table 5.** Comparison of local model average performance between standalone learning and federated learning.

Metrics (95% CI)	Standalone Learning	Federated Learning			
		LMECS (Loss)	LMECS (Accuracy)	LMECS (AUC)	FedAvg
Accuracy (%)	78.17 ± 0.42	80.60 ± 0.64	<b>81.42 ± 2.33</b>	81.35 ± 1.62	78.11 ± 1.12
Sensitivity (%)	80.70 ± 1.55	82.74 ± 0.58	<b>83.44 ± 0.63</b>	83.02 ± 0.69	77.56 ± 1.93
Specificity (%)	77.66 ± 0.50	80.57 ± 0.66	<b>81.91 ± 3.71</b>	81.72 ± 2.74	80.29 ± 1.47
PPV (%)	21.11 ± 1.43	23.32 ± 0.36	<b>23.62 ± 0.12</b>	23.21 ± 0.31	21.82 ± 0.72
NPV (%)	95.13 ± 1.52	98.12 ± 0.56	<b>98.72 ± 0.38</b>	98.24 ± 0.16	96.12 ± 0.88
AUC (%)	79.68 ± 0.74	81.38 ± 0.41	81.43 ± 0.56	<b>81.45 ± 0.39</b>	75.72 ± 1.69

Bold denotes the highest performance in each Metrics.

## 6. Conclusions

We underscore the transformative potential of the Local Model Evaluation Client Selection (LMECS) algorithm within FL for patient severity analysis in multi-institutional emergency room settings. This work primarily focused on addressing the challenges posed by existing FL and centralized data analysis methods, particularly in terms of data privacy, computational efficiency, and the diverse nature of medical data. The LMECS algorithm, an innovative contribution to FL, has demonstrated its ability to adapt to various user-specific tasks, offering not just compatibility with the stringency of medical data privacy but also a substantial improvement in performance metrics compared to existing FL algorithms. Significantly, the experimental results revealed that LMECS could match, and in some instances surpass, the performance levels of centralized learning (CL). This is a critical breakthrough, suggesting that it is possible to achieve the efficiency and accuracy of centralized models while retaining the inherent benefits of FL, such as data privacy and reduced data movement. Moreover, our findings highlighted a significant reduction in execution time with the implementation of the LMECS algorithm. This reduction was observed when compared both to CL and existing FL algorithms. Such efficiency is especially crucial in clinical settings, where prompt decision-making based on accurate and timely data analysis can have a substantial impact on patient outcomes. This aspect of LMECS underlines its suitability for real-world healthcare applications, where both speed and accuracy are paramount. Another crucial aspect of our research findings is the enhanced performance of the proposed LMECS algorithm compared to standalone learning (SL), where FL is not applied. The experimental results indicate that LMECS significantly improved the overall performance of local models. This improvement was not limited to specific metrics but was observed across a range of performance indicators, demonstrating the algorithm's ability to enhance model accuracy and reliability in diverse medical data environments. In terms of metric-based performance evaluation, LMECS's adaptability was evident. The ability to tailor the algorithm based on specific performance metrics like loss, accuracy, and AUC enables a more nuanced approach to model evaluation. In the future, applying LMECS to multiple classification and prediction models involving data from a variety of hospitals will add significant weight to the work. This real-world applicability is important to state the effectiveness of the algorithm beyond theoretical or simulated environments. In summary, this work not only contributes to the field of medical data analysis and patient care but also opens new avenues for further research. The advancements in FL brought about by LMECS can be explored in other domains where privacy concerns and data heterogeneity are predominant. The future of FL, particularly in healthcare, looks promising, with LMECS paving the way for more efficient, private, and accurate data analysis methods.

**Author Contributions:** Design and writing, Y.-g.K. and S.Y.; conceptualization Y.-g.K., S.Y. and K.L.; software, Y.-g.K. and S.Y.; writing—original draft preparation, K.L.; supervision and contribution, K.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the Commercialization Promotion Agency for R&D Outcome (COMPA) Grant funded by the Korean Government (MSIT) under Grant 2022-Future research service development support-1-SB4-1, and in part 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: HI22C1569). (Corresponding author: KangYoon Lee).

**Institutional Review Board Statement:** We conducted a study involving human subjects and a retrospective dataset collected from the emergency room of a South Korean hospital. This study was confirmed to be exempt from review by the Gachon University Bioethics Review Committee in Korea, under the approval number 1044396-202303-HR-044-01. The results of this study are based on the National Emergency Department Information System (N2023-27-2-06-21).

**Informed Consent Statement:** This retrospective study conducted between 2017 and 2022 in South Korea utilized data from the National Emergency Department Information System. The requirement for informed consent was waived by the IRB due to the retrospective nature of the study. Patient information was anonymized prior to analysis.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from national emergency medical center in South Korea and are available <https://dw.nemc.or.kr/nemcMonitoring/data/user/DataUseInfoMain.do> with the permission of national emergency medical center in South Korea.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Kondo, Y.; Fukuda, T.; Uchimido, R.; Kashiura, M.; Kato, S.; Sekiguchi, H.; Zamami, Y.; Hifumi, T.; Hayashida, K. Advanced life support vs basic life support for patients with trauma in prehospital settings: A systematic review and meta-analysis. *Front. Med.* **2021**, *8*, 660367. [[CrossRef](#)] [[PubMed](#)]
2. Hansen, J.; Rasmussen, L.S.; Steinmetz, J. Prehospital triage of trauma patients before and after implementation of a regional triage guideline. *Injury* **2022**, *53*, 54–60. [[CrossRef](#)] [[PubMed](#)]
3. Ham, G.-S.; Kang, M.; Joo, S.-C. A Study on Finding Emergency Conditions for Automatic Authentication Applying Big Data Processing and AI Mechanism on Medical Information Platform. *KSII Trans. Internet Inf. Syst.* **2022**, *16*, 2772–2786.
4. Liu, Z.; Shi, X.; Jin, H. Data-Efficient Performance Modeling for Configurable Big Data Frameworks by Reducing Information Overlap Between Training Examples. *Big Data Res.* **2022**, *30*, 100358. [[CrossRef](#)]
5. Kim, Y.; Kim, J.-H.; Kim, Y.-M.; Song, S.; Joo, H.J. Predicting medical specialty from text based on a domain-specific pre-trained BERT. *Int. J. Med. Inform.* **2023**, *170*, 104956. [[CrossRef](#)] [[PubMed](#)]
6. Sheller, M.J.; Edwards, B.; Reina, G.A.; Martin, J.; Pati, S.; Kotrotsou, A.; Milchenko, M.; Xu, W.; Marcus, D.; Colen, R.R.; et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Sci. Rep.* **2020**, *10*, 12598. [[CrossRef](#)] [[PubMed](#)]
7. Georgios, K.; Alexander, Z.; Passerat-Palmbach, J.; Ryffel, T.; Usynin, D.; Trask, A.; Lima, I.; Mancuso, J.; Jungmann, F.; Steinborn, M.-M. End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nat. Mach. Intell.* **2021**, *3*, 473–484.
8. Ahmadi, N.; Peng, Y.; Wolfien, M.; Zoch, M.; Sedlmayr, M. OMOP CDM Can Facilitate Data-Driven Studies for Cancer Prediction: A Systematic Review. *Int. J. Mol. Sci.* **2022**, *23*, 11834. [[CrossRef](#)] [[PubMed](#)]
9. Boggs, K.M.; Teferi, M.M.; Espinola, J.A.; Sullivan, A.F.; Hasegawa, K.; Zachrisson, K.S.; Samuels-Kalow, M.E.; Camargo, C.A., Jr. Consolidating Emergency Department-specific Data to Enable Linkage with Large Administrative Datasets. *West. J. Emerg. Med.* **2020**, *21*, 141–145. [[CrossRef](#)]
10. Nguyen, D.C.; Pham, Q.-V.; Pathirana, P.N.; Ding, M.; Seneviratne, A.; Lin, Z.; Dobre, O.; Hwang, W.-J. Federated Learning for Smart Healthcare: A Survey. *ACM Comput. Surv.* **2022**, *55*, 60. [[CrossRef](#)]
11. Xu, J.; Glicksberg, B.S.; Su, C.; Walker, P.; Bian, J.; Wang, F. Federated learning for healthcare informatics. *J. Healthc. Inform. Res.* **2021**, *5*, 1–19. [[CrossRef](#)] [[PubMed](#)]
12. Rieke, N.; Hancox, J.; Li, W.; Milletari, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.N.; Landman, B.A.; Maier-Hein, K. The future of digital health with federated learning. *npj Digit. Med.* **2020**, *3*, 119. [[CrossRef](#)] [[PubMed](#)]
13. Lim, W.Y.B.; Luong, N.C.; Hoang, D.T.; Jiao, Y.; Liang, Y.-C.; Yang, Q.; Niyato, D.; Miao, C. Federated learning in mobile edge networks: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2031–2063. [[CrossRef](#)]
14. Lian, X.; Zhang, C.; Zhang, H.; Hsieh, C.-J.; Zhang, W.; Liu, J. Can decentralized algorithms outperform centralized algorithms? A case study for decentralized parallel stochastic gradient descent. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.

15. Smith, V.; Forte, S.; Chenxin, M.; Taka, M.; Jordan, M.I.; Jaggi, M. Cocoa: A general framework for communication-efficient distributed optimization. *J. Mach. Learn. Res.* **2018**, *18*, 1–49.
16. Ji, S.; Jiang, W.; Walid, A.; Li, X. Dynamic Sampling and Selective Masking for Communication-Efficient Federated Learning. *IEEE Intell. Syst.* **2021**, *37*, 27–34. [[CrossRef](#)]
17. Lin, W.; Xu, Y.; Liu, B.; Li, D.; Huang, T.; Shi, F. Contribution-based Federated Learning client selection. *Int. J. Intell. Syst.* **2022**, *37*, 7235–7260. [[CrossRef](#)]
18. Karimireddy, S.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; Suresh, A. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. In Proceedings of the 37th International Conference on Machine Learning 2020, Virtual, 13–18 July 2020; pp. 5132–5143.
19. Ngufor, C.; Caraballo, P.J.; O’byrne, T.J.; Chen, D.; Shah, N.D.; Pruinelli, L.; Steinbach, M.; Simon, G. Development and Validation of a Risk Stratification Model Using Disease Severity Hierarchy for Mortality or Major Cardiovascular Event. *JAMA Netw. Open* **2020**, *3*, e208270. [[CrossRef](#)]
20. Patel, D.; Kher, V.; Desai, B.; Lei, X.; Cen, S.; Nanda, N.; Gholamrezanezhad, A.; Duddalwar, V.; Varghese, B.; Oberai, A.A. Machine learning based predictors for COVID-19 disease severity. *Sci. Rep.* **2021**, *11*, 4673. [[CrossRef](#)]
21. De Souza, F.S.H.; Hojo-Souza, N.S.; Santos, E.B.D.; Da Silva, C.M.; Guidoni, D.L. Predicting the Disease Outcome in COVID-19 Positive Patients Through Machine Learning: A Retrospective Cohort Study With Brazilian Data. *Front. Artif. Intell. Sec. Med. Public Health* **2021**, *4*, 579931. [[CrossRef](#)]
22. Lee, S.; Kang, W.; Kim, D.; Seo, S.; Kim, J.; Jeong, S.; Yon, D.; Lee, J. An Artificial Intelligence Model for Predicting Trauma Mortality Among Emergency Department Patients in South Korea: Retrospective Cohort Study. *J. Med. Internet Res.* **2023**, *25*, e49283. [[CrossRef](#)]
23. Hwang, S.; Lee, B. Machine learning-based prediction of critical illness in children visiting the emergency department. *PLoS ONE* **2022**, *17*, e0264184. [[CrossRef](#)] [[PubMed](#)]
24. Sánchez-Salmerón, R.; Gómez-Urquiza, J.L.; Albendín-García, L.; Correa-Rodríguez, M.; Martos-Cabrera, M.B.; Velando-Soriano, A.; Suleiman-Martos, N. Machine learning methods applied to triage in emergency services: A systematic review. In *International Emergency Nursing*; Elsevier: Amsterdam, The Netherlands, 2022.
25. Valencia-Parra, A.; Parody, L.; Varela-Vaca, Á.J.; Caballero, I.; Gómez-López, M.T. DMN4DQ: When data quality meets DMN. *Decis. Support Syst.* **2021**, *141*, 113450. [[CrossRef](#)]
26. Song, H.; Leng, H.; Hou, Z.; Gao, R.; Chen, C.; Meng, C.; Sun, J.; Li, C.; Ma, B. Grouped-sampling technique to deal with unbalance in Raman spectral data modeling. *Photodiagnosis Photodyn. Ther.* **2022**, *40*, 103059. [[CrossRef](#)] [[PubMed](#)]
27. Li, Y.; Zhao, Y.; Xixin Niu Zhou, W.; Tian, J. The Efficiency Evaluation of Municipal-Level Traditional Chinese Medicine Hospitals Using Data Envelopment Analysis after the Implementation of Hierarchical Medical Treatment Policy in Gansu Province, China. *INQUIRY J. Health Care Organ. Provis. Financ.* **2022**, *59*. [[CrossRef](#)] [[PubMed](#)]
28. Naemi, A.; Mansourvar, M.; Schmidt, T.; Wiil, U.K. Prediction of Patients Severity at Emergency Department Using NARX and Ensemble Learning. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 2793–2799.
29. Caine, K.; Hanania, R. Patients want granular privacy control over health information in electronic medical records. *J. Am. Med. Inform. Assoc.* **2013**, *20*, 7–15. [[CrossRef](#)] [[PubMed](#)]
30. Yang, J.-J.; Li, J.-Q.; Niu, Y. A hybrid solution for privacy preserving medical data sharing in the cloud environment. *Fut. Gener. Comput. Syst.* **2015**, *43*, 74–86. [[CrossRef](#)]
31. Holub, P.; Kohlmayer, F.; Prasser, F.; Mayrhofer, M.T.; Schlünder, I.; Martin, G.M.; Casati, S.; Koumakis, L.; Wutte, A.; Kozera, Ł.; et al. Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health. *Biopreserv. Biobank.* **2018**, *16*, 97–105. [[CrossRef](#)] [[PubMed](#)]
32. Huang, L.; Shea, A.L.; Qian, H.; Masurkar, A.; Deng, H.; Liu, D. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. *J. Biomed. Inform.* **2019**, *99*, 103291. [[CrossRef](#)] [[PubMed](#)]
33. Deng, T.; Hamdan, H.; Yaakob, R.; Kasmiran, K.A. Personalized Federated Learning for In-Hospital Mortality Prediction of Multi-Center ICU. *IEEE Access* **2023**, *11*, 11652–11663. [[CrossRef](#)]
34. Wu, Q.; Chen, X.; Zhou, Z.; Zhang, J. FedHome: Cloud-edge based personalized federated learning for in-home health monitoring. *IEEE Trans. Mob. Comput.* **2020**, *21*, 2818–2832. [[CrossRef](#)]
35. McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S.; Arcas, B.A.Y. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL, USA, 20–22 April 2017.
36. Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated optimization in heterogeneous networks. In Proceedings of the Third Conference on Machine Learning and Systems (MLSys), Austin, TX, USA, 2–4 March 2020.
37. Huang, T.; Lin, W.; Shen, L.; Li, K.; Zomaya, A.Y. Stochastic Client Selection for Federated Learning with Volatile Clients. *IEEE Internet Things J.* **2022**, *9*, 20055–20070. [[CrossRef](#)]
38. Nishio, T.; Yonetani, R. Client selection for federated learning with heterogeneous resources in mobile edge. In Proceedings of the ICC 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–7.
39. Zeng, Q.; Du, Y.; Huang, K.; Leung, K.K. Energy-efficient radio resource allocation for federated edge learning. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.

40. Cho, Y.J.; Wang, J.; Joshi, G. Client selection in federated learning: Convergence analysis and power-of-choice selection strategies. *arXiv* **2020**, arXiv:2010.01243.
41. Putra, M.A.P.; Putri, A.R.; Zainudin, A.; Kim, D.-S.; Lee, J.-M. ACS: Accuracy-based client selection mechanism for federated industrial IoT. *Internet Things* **2023**, *21*, 100657. [[CrossRef](#)]
42. Goetz, J.; Malik, K.; Bui, D.; Moon, S.; Liu, H.; Kumar, A. Active federated learning. *arXiv* **2019**, arXiv:1909.12641.
43. Laguel, Y.; Pillutla, K.; Malick, J.; Harchaoui, Z. Device heterogeneity in federated learning: A superquantile approach. *arXiv* **2020**, arXiv:2002.11223.
44. Central Emergency Medical Center. Korean National Emergency Department Information System. 2013. Available online: <https://dw.nemc.or.kr/nemcMonitoring/mainmgr/Main.do> (accessed on 17 July 2023).
45. Kang, D.-Y.; Cho, K.-J.; Kwon, O.; Kwon, J.-M.; Jeon, K.-H.; Park, H.; Lee, Y.; Park, J. Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. *Scand. J. Trauma Resusc. Emerg. Med.* **2020**, *28*, 17. [[CrossRef](#)]
46. Hsieh, K.; Phanishayee, A.; Mutlu, O.; Gibbons, P. The non-IID data quagmire of decentralized machine learning. *arXiv* **2019**, arXiv:1910.00189.
47. Moritz, P.; Nishihara, R.; Wang, S.; Tumanov, A.; Liaw, R.; Liang, E.; Elibol, M.; Yang, Z.; Paul, W.; Jordan, M.I.; et al. Ray: A Distributed Framework for Emerging AI Applications. In Proceedings of the 13th USENIX Symposium on Operating Systems Design and Implementation, Carlsbad, CA, USA, 8–10 October 2018.
48. Beutel, D.J.; Topal, T.; Mathur, A.; Qiu, X.; Fernandez-Marques, J.; Gao, Y.; Sani, L.; Li, K.H.; Parcollet, T.; de Gusmão, P.P.B.; et al. Flower: A Friendly Federated Learning Framework. *arXiv* **2022**, arXiv:2007.14390.
49. Moon, J.; Yang, S.; Lee, K. FedOps: A Platform of Federated Learning Operations with Heterogeneity Management. *IEEE Access* **2024**, *12*, 4301–4314. [[CrossRef](#)]
50. Yang, S.; Moon, J.; Kim, J.; Lee, K. FLScalize: Federated Learning Lifecycle Management Platform. *IEEE Access* **2023**, *11*, 47212–47222. [[CrossRef](#)]
51. Cognitive Computing Lab in Gachon University. Federated Learning Lifecycle Operations Management Platform. 2023. Available online: <https://github.com/gachon-CCLab/FedOps> (accessed on 17 July 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.