Article

Federated Learning for Clinical Event Classification Using Vital Signs Data

Ruzaliev Rakhmiddin[1], KangYoon Lee[1]*

[1] Department of Computer Engineering, Gachon University, SongNamSi, 13120 South. Korea

* Correspondence: KangYoon Lee (e-mail: keylee@gachon.ac.kr)

Abstract: Effective healthcare relies on accurate and timely diagnosis; however, obtaining large amounts of training data while maintaining patient privacy remains challenging. This study introduces a novel approach utilizing federated learning (FL) and a cross-device multi-modal model for clinical event classification using vital signs data. Our architecture leverages FL to train machine learning models, including Random Forest, AdaBoost, and SGD ensemble model, on vital signs data from a diverse clientele at a Boston hospital (MIMIC-IV dataset). The FL structure preserves patient privacy by training directly on each client's device without transferring sensitive data. The study demonstrates the potential of FL in privacy-preserving clinical event classification, achieving an impressive accuracy of 98.9%. These findings underscore the significance of FL and cross-device ensemble technology in healthcare applications, enabling the analysis of large amounts of sensitive patient data while safeguarding privacy.

Keywords: Federated learning, clinical events, vital signs, classification, multimodal.

1. Introduction

Artificial intelligence (AI) techniques and technologies are used to improve various aspects of healthcare. This can include medical imaging, drug discovery, patient diagnosis, and treatment planning [1]. There is a growing body of research in this field, as AI can significantly improve the efficiency and accuracy of healthcare processes, ultimately leading to better patient outcomes. Some examples of related work include using AI to diagnose diseases such as cancer, using machine learning to analyze patient data and predict potential health issues, and using natural language processing to improve the efficiency of electronic medical records.

Big data [1] has recently become a buzzword in many industries, and healthcare is no exception. The healthcare sector generates vast amounts of data daily, including electronic health records, claims data, and clinical trial results [2,3]. Such data can be analyzed to identify patterns, trends, and associations that can help improve patient care, reduce costs, and advance medical research. The use of big data in healthcare is still in its initial stages, but it has already shown promise in several areas. For example, big data has been used to improve population health management by identifying patterns in patient health data that can help healthcare providers better understand the health needs of their patient population and develop strategies to improve population health. Big data has also been used to predict future patient needs and outcomes using predictive analytics and to develop clinical decision support systems that provide healthcare providers with real-time recommendations based on a patient's medical history and current condition. [4] Although there is a considerable improvement in the healthcare system, as mentioned above, privacy has been the main issue concerning big data, especially in the healthcare system. In addition, enhanced machine learning techniques and advanced pre-processing can be a positive approach to solving a problem using big data.

Machine learning, a branch of artificial intelligence, entails training computer algorithms to identify patterns within data and utilize those patterns to make informed decisions. In healthcare, machine learning is used to analyze substantial amounts of data from various sources, such as electronic health records, medical imaging, and wearable devices, to identify patterns and trends that can help improve patient care [5]. Predictive analytics: Machine learning algorithms can be used to analyze patient data to predict future health outcomes, such as the likelihood of developing a specific condi-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

tion or needing medical intervention. This can help healthcare providers make more informed decisions about patient care and allocate resources more efficiently in understanding the geographical inequalities of healthcare resources with Bayesian analysis [6], clinical data prediction using Random Forest classification [7], and disease pre-diction with XGBoost classification [8]. Clinical decision support: Machine learning can be used to develop clinical decision support systems, which provide healthcare providers with real-time recommendations based on a patient's medical history and current condition [9]. Diagnosis and treatment: Machine learning can analyze medical images, such as CT scans or X-rays, to assist in diagnosis and treatment planning. It can also analyze lab test results to identify potential health issues [10]. Personalized medicine: Machine learning can be used to develop personalized treatment plans for individual patients, considering their genetics, lifestyle, and medical history [11].

Federated learning (FL) [12] trains machine learning models on decentralized data. Instead of centralizing data in an individual location, federated learning allows data to remain on individual devices, such as smartphones or IoT devices. The model is trained across multiple devices by sending model updates to each device and receiving updated parameters. A global model is repeatable until it reaches a satisfactory level of performance. This allows for training on much larger datasets than possible with a centralized approach and helps protect users' privacy by keeping their data on their own cross devices such as electronic health records (EHRs), wearable devices (e.g., smartwatches and fitness trackers), and medical imaging devices. In the case of Federated Learning, cross-device functionality allows each of these devices to contribute to the learning process by training their own local models on the data they have, and then sharing the model parameters with a central server. The server then aggregates these parameters to update the global model, which is then sent back to each device. Figure 1 shows the general architecture of using Federated learning in the healthcare system with components and connection with FL.

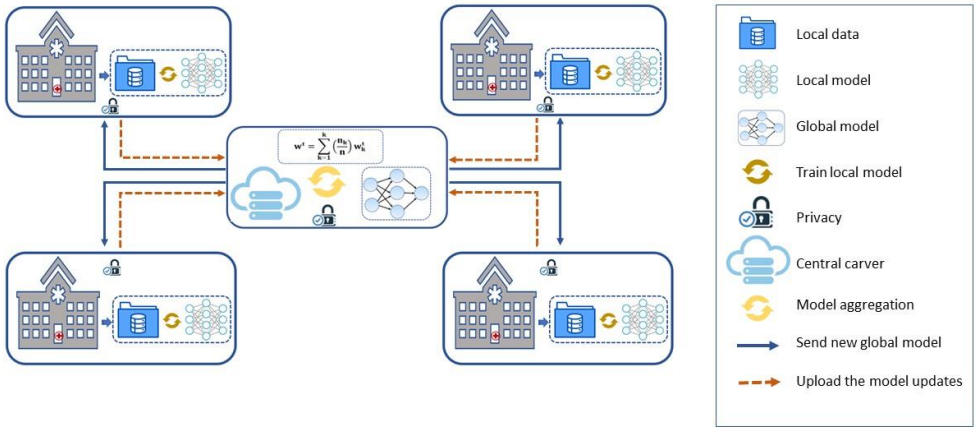

Figure 1. The general concept of Federated learning in the healthcare system.

Federated learning has the potential to be particularly useful in the healthcare industry, where data privacy and security are of paramount importance. With it, sensitive patient data can be kept on individual devices and hospital servers rather than centralized in a single location [13,14, 15]. This can help to protect patient privacy and comply with regulations such as HIPAA [16]. In addition, federated learning can train more accurate models by allowing for data aggregation from a more significant number of patients. This can be especially beneficial in rare disease research [17], where a centralized dataset may not have enough examples to train a reliable model. Federated learning can also enable the training of models on a more diverse patient population, which can lead to more generalizable and, therefore, more valuable models [18].

This study conducted clinical event classification using vital signs data with federated learning. The Flower FL algorithm was selected for the implementation of keeping the privacy of the dataset. Several machine learning techniques, such as Random Forest classification, XGBoost, and multi-model ensemble Random Forest with SGD, Random Forest with XGBoost were used to get the optimal result. A custom fine-tuned method in federated learning was used to acquire the best results.

 doi:10.20944/preprints202306.0806.v1

Section 2 presents the related works, Section 3 the materials and methods, Section 4 the experimental results, and the last section the conclusion.

2. Related work

Clinical event classification using vital signs [19] data is critical in healthcare as it allows for early detection and management of various medical conditions. Researchers globally have extensively explored computational techniques, including machine learning and predictive modeling, to develop accurate and reliable methods for such predictions. Effective classification can identify health risks or critical events earlier, allowing for timely intervention and potentially preventing severe outcomes. Also, automated classification systems can quickly analyze a high volume of patient data, assisting healthcare providers in making more accurate and faster diagnoses with using Machine Learning models.

Machine learning is a popular approach in this field, as it allows for the analysis of vast amounts of historical and current data from various sources in healthcare to make predictions [1,20]. Medical machine learning contributes significantly to reducing the investment spent on health care and renewing the relationship between doctor and patient by reducing investment in this field [21]. A wireless radar, for example, collects vital signs data using radar technology and categorizes healthy and infected people using five machine-learning models [22]. In 2019 years, Juan-Jose Beunza et al. [23], to predict clinical events, compared several supervised classification machine learning algorithms for internal validity and accuracy. The Framingham open database used new methods in the data preparation process and got women an accuracy value of 0.81 while men a value of 0.78. However, their performance in the degree of accuracy is not considered sufficient and is often hindered by the lack of large, diverse, and labeled data. Yuanyuan et al. [24] introduced the system for using a convolutional neural network (CNN) with enhanced deep learning techniques to predict heart disease on an Internet of Medical Things (IoMT) platform. The "enhanced deep learning" aspect likely refers to using advanced techniques such as transfer learning or ensemble methods to improve the performance of the CNN. The IoMT platform uses medical devices connected to the Internet to collect and transmit data for analysis.

Jie Xu et al. [12] wrote that the survey aims to examine the use of federated learning in the biomedical field. It will provide an overview of various solutions for dealing with federated learning's statistical system and privacy challenges. Another example is highlighting these technologies' potential applications and impacts in healthcare is that of Thanveer Shaik et al. [25], who proposed a decentralized privacy-protected system for monitoring in-patient activity in hospitals using sensors and AI models to classify twelve routine activities with the FedStack system. FedStack is a proposed system for using stacked federated learning for personalized activity monitoring. Federated learning is a technique for training machine learning models on decentralized data, where data is distributed across multiple devices or locations. Stacked federated learning refers to a specific technique where multiple federated models are trained and combined to form a final model. This paper suggests using this approach for activity monitoring, which involves collecting data from sensors or other devices worn by individuals to track their physical activity and utilizing the trained models to personalize the monitoring and analysis of such data. Similarly, Ittai Dayan et al. [26] worked on predicting the future oxygen requirements for symptomatic COVID-19 patients using vital signs, laboratory data, and chest X-rays with the FL model. Also, the research proposed using federated learning for predicting clinical outcomes in patients with COVID-19. Federated learning is a technique for training machine learning models on decentralized data, in which information is distributed across multiple devices or locations. In this case, the authors suggest this approach to train models on data from different hospitals or clinics and improve the accuracy of predictions for patients with COVID-19. They also claim that this approach can help make predictions in real-time, improving the models' performance by sharing knowledge across different institutions.

The proposed cross-device ensemble method offers advantages over existing methods by combining and building upon the related approaches mentioned above. Firstly, it provides privacy pro-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

tection by training models on decentralized data, whereas FL safeguards sensitive patient information as data never leave individual devices or institutions. Secondly, this method ensures robustness by enabling data integration from various sources, leading to more accurate and robust models. These advantages make this approach a promising solution for healthcare applications that require enormous amounts of sensitive patient data while ensuring privacy and robustness.

3. Materials and Methods

The overall concept of the architecture covers the dataset description, pre-processing of the dataset, machine learning part, and, eventually, federated learning.

3.1 Dataset description

This research uses the Medical Information Mart for Intensive Care (MIMIC-IV) [27,28] dataset. This dataset contains de-identified electronic health record (EHR) data from patients admitted to intensive care units (ICUs) at the Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts, USA, between 2008 and 2019. With data on more than 300,000 hospital admissions, the MIMIC-IV dataset was chosen since it is one of the world's most extensive publicly available ICU datasets, making it an invaluable resource for researchers studying critical care medicine, health outcomes, and medical informatics. The MIMIC-IV dataset is used to gather information on patient demographics, diagnoses, medications, laboratory results, vital signs, and more, providing a highly detailed view of patients' medical histories. Despite being de-identified to protect patient privacy, the dataset retains a high degree of clinical detail, making it useful for various research applications.

Table 1. Description of Vital Signs and Typical Normal Ranges.

Term	Description	Normal Range
SpO2	The oxygen saturation level in the patient's blood.	95%–100%
BPM	The heart rate, measured in beats per minute.	60–100 beats per min
RR	The number of breaths the patient takes per minute, providing insights into respiratory function.	12–18 breaths per min
SBP	The highest pressure exerted on the arterial walls during the cardiac cycle.	90–120 mmHg
DBP	The lowest pressure exerted on the arterial walls when the heart is at rest between beats.	60–90 mmHg
MBP	The average pressure within the patient's arteries over a complete cardiac cycle.	60–110 mmHg

The MIMIC-IV dataset was a vital resource for this research into critical care medicine, health outcomes, and medical informatics. Its vast size, clinical detail, and open availability make it an ideal dataset for various research applications. It should be noted that access to the MIMIC-IV dataset is restricted and requires approval from the PhysioNet Data Use Agreement (DUA). This study selected six vital signs (SpO2, BPM, RR, SBP, DBP, and MBP) from the MIMIC-IV dataset for analysis. Table 1 provides a concise overview of vital signs commonly used in healthcare along with their descriptions and typical normal ranges. These vital signs include SpO2 (oxygen saturation level), BPM (heart

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

rate), RR (respiratory rate), SBP (systolic blood pressure), DBP (diastolic blood pressure), and MBP (mean blood pressure). The table serves as a reference for healthcare professionals to assess and monitor patients' physiological parameters within the expected normal range. Table 2 illustrates the initial version for data pre-processing which includes vital sign measurements extracted from the main dataset. The table displays the distribution of these vital signs over the duration of a patient's stay in the intensive care unit (ICU). It provides valuable insights into the variations and trends of these physiological parameters during the patient's ICU stay.

Table 2. Head of the initial version of the MIMIC IV dataset for federated learning process.

Index	Subject_id	Charttime	Storetime	Valuenum	Valueuom
0	10003700	2165-04-24 05:28:00	2165-04-24 05:37:00	152.0	mmHg (SBP)
1	10003700	2165-04-24 05:28:00	2165-04-24 05:37:00	97.0	mmHg (DBP)
2	10003700	2165-04-24 05:28:00	2165-04-24 05:37:00	110.0	mmHg (MBP)
3	10003700	2165-04-24 05:30:00	2165-04-24 05:37:00	65.0	bpm
4	10003700	2165-04-24 05:30:00	2165-04-24 05:37:00	14.0	insp/min
5	10003700	2165-04-24 05:31:00	2165-04-24 05:37:00	100.0	%
6	10003700	2165-04-24 05:37:00	2165-04-24 05:37:00	120.0	bpm
7	10003700	2165-04-24 05:37:00	2165-04-24 05:37:00	50.0	bpm
8	10003700	2165-04-24 05:37:00	2165-04-24 05:37:00	160.0	mmHg (SBP)
9	10003700	2165-04-24 05:37:00	2165-04-24 05:37:00	90.0	mmHg (DBP)

3.2 Data pre-processing

Data pre-processing is a crucial step in machine learning as it helps prepare the data for analysis and modeling—some of the critical reasons for data pre-processing. For example, data cleaning helps to identify and remove any errors, inconsistencies, or missing values in the data. This helps to ensure that the data are accurate and reliable for analysis and modeling. The first step of pre-processing is to remove or fill missing values and noise in the dataset. MIMIC-IV dataset contains a lot of missing values. These missing values can be filled with measures like mean, median, or mode, or using model-based imputation methods. The next step is feature selection that identify features are relevant to the prediction. For example, for predicting a clinical event, features might include SBP or BMP, etc. Unnecessary features might include patient ID, which is not predictive. To bring all the features to a similar level, normalization is an essential process that includes the data that is often normalized or standardized. This prevents features with larger scales from dominating the model. The z-score (1) is a common method of normalization, and it's calculated using the following formula.

$$z = (x - \mu) / \sigma \quad (1)$$

In the given equation, x represents a data point, μ denotes the mean of the dataset, and σ represents the standard deviation of the dataset. In data analysis, the dataset is divided into a training set and a testing set. The training set is used to train the model, while the testing set evaluates its performance. A common split ratio is 70% for training and 30% for testing. This ensures effective learning and unbiased evaluation of the model's generalization.

In the initial version of the dataset, there were no clinic event targets, whereas PEACE-Home [29] proposed a system for monitoring patients in a home-based setting using vital signs such as heart rate, blood pressure, and respiratory rate. The system used probabilistic estimation to identify abnormal clinical events, such as deterioration in a patient's condition, by analyzing correlations among vital signs and separating clinic events as target data while clustering and using a relied on expert system. Data labeling is a process of assigning labels or tags to data to be used for training or evaluating machine learning models. In the context of PEACE-Home, data labeling was likely to involve identifying and tagging instances of abnormal clinical events within the vital signs data collected from patients in a home-based setting. This can be done through manual annotation by healthcare

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

professionals or algorithms to identify and label events of interest automatically. If a vital sign is out of its expected range for a prolonged period of time, cannot be treated promptly, and persists, that is a clinical event in patient care. The expected ranges are often tailored to each patient, based on their specific health condition and history, although there are general medical guidelines that outline the typical boundaries of various vital signs. For instance, consider a patient with a history of hypertension. The patient's normal blood pressure may consistently register above the typically accepted "normal" range (SBP 80-120 and DBP 60-90 mmHg). A clinical event occurs if their blood pressure spikes to a dangerously high level, above their usual expected maximum or in bradycardia, this event refers to a slower-than-normal heart rate, defined as a heartbeat of 60 beats per minute (bpm) or less. For example, if a patient's heart rate drops to 55 bpm and stays there for a significant period without intervention, it would qualify as a Bradycardia clinical event. The study specifically examined simultaneous changes in four vital signs from generalized normal values and developed techniques to predict these changes in advance.

The labeled data was generated from the initial version of dataset as normal and abnormal clinical events using threshold values. The model can then monitor patients in a home-based setting and identify potential health problems early on. Table 3 shows the labeled clinic event data from the MIMIC IV dataset using the PEACE-Home method.

Table 3. Characteristics and threshold values for each clinical event, indicating the presence or absence of specific abnormalities in vital signs.

Labels	Reason	Threshold values	0	1	2	3	4
Hypertension	High BP	(SBP≥120 and DBP≥80) or MBP≥105	X	X	X	○	X
Hypotension	Low BP	(SBP≤90 and DBP≤60) or MBP≤70	X	○	○	X	○
Tachycardia	High HR	HR ≥ 100	X	○	X	○	○
Bradycardia	Low HR	HR ≤ 60	X	X	○	X	X
Tachypena	High RR	RR ≥ 17	X	○	○	○	X
Bradypena	Low RR	RR ≤ 12	X	X	X	X	○
Hypoxia	Low SPO_2	SPO_2 ≤ 93%	X	○	○	○	○
Acronym			NNNN	THTH	BHTH	TTTH	THBH
Number of samples			145085	45186	31132	27915	12840

Table 4 categorizes unique clinical events in a patient's health status, based on simultaneous occurrences of specific vital signs deviations. Each clinical event is associated with an acronym and a distinct label. The acronyms THTH, BHTH, TTTH, and THBH denote combinations of abnormal vital signs including heart rate, blood pressure, breathing rate, and oxygen saturation. The NNNN category represents a normal state, where all vital signs are within the expected range. This classification is intended to facilitate the rapid and accurate identification of a patient's health condition, supporting timely and effective medical intervention.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

Table 4. Classification of clinical events based on concurrent abnormalities or normalcy in patient vital signs.

Clinical events	Acronym	label
Simultaneous Tachycardia, Hypotension, Tachypnea, and Hypoxia	THTH	1
Simultaneous Bradycardia, Hypotension, Tachypnea, and Hypoxia	BHTH	2
Simultaneous Tachycardia, Hypertension, Tachypnea, and Hypoxia	TTTH	3
Simultaneous Tachycardia, Hypotension, Bradypnea, and Hypoxia	THBH	4
All six bio-signals are within the normal range	NNNN	0

3.3. MACHINE LEARNING PART

Machine learning can be applied to clinical event classification tasks in several ways. One common approach is to use supervised machine learning algorithms, such as decision trees, random forests, or support vector machines, to predict the class of a given clinical event based on a set of features or attributes. The algorithm is trained on a labeled dataset of past clinical events and their corresponding classes and then used to make predictions on new, unseen data. In a clinical event classification task, the features used as inputs to the machine learning algorithm could include demographic information, vital signs, laboratory test results, medications, and other relevant information. The target variable or output of the algorithm would be the class of the clinical event, such as sepsis, pneumonia, or a heart attack. Overall, using machine learning in the clinical event classification task can enhance the accuracy and efficiency of healthcare delivery by enabling the rapid and reliable identification of patients with specific conditions. This study implemented several ML methods to compare and get the best result on clinical event classification tasks, such as Random Forest Classifier, XGBoost classifier, AdaBoost classifier, Stochastic Gradient Decent, and Bayesian Ridge classifier.

3.3.1. RANDOM FOREST CLASSIFIER

Random Forest Classifier is a machine learning algorithm used in healthcare to predict outcomes, classify patients, and identify disease risk factors [30, 31]. It is a method of ensemble learning that combines multiple decision trees to enhance the accuracy and robustness of the model. In healthcare, Random Forest Classifier is often used in medical image analysis to detect and diagnose diseases like cancer, Alzheimer's, and cardiovascular disease. It can also predict readmission rates, length of hospital stays, and mortality rates. The algorithm works by randomly selecting subsets of the features and building a decision tree based on the selected features. The process is repeated multiple times to create a forest of decision trees. Each decision tree in the forest predicts the outcome during prediction, and the majority vote determines the final prediction. Random Forest Classifier is known for its ability to handle high-dimensional data, missing values, and noisy data. It is also less prone to overfitting compared to other machine learning algorithms. Overall, Random Forest Classifier is a powerful tool in healthcare for improving diagnosis and treatment outcomes.

3.3.2. LOGISTIC REGRESSION

Logistic regression is a statistical technique used to analyse datasets where independent variables determine outcomes. In healthcare, it is commonly employed to predict the probability of an event based on patient characteristics such as age, gender, medical history, and laboratory results. For example, logistic regression is employed to predict the probability of patients developing a particular disease using their demographic and clinical information [31]. It also aids in assessing the effectiveness of treatments or interventions by examining the relationship between the treatment and the outcome. A key advantage of logistic regression in healthcare is its simplicity and interpretability, allowing for informed decision-making and predictions. However, it assumes a linear relationship between the independent variables and the outcome, which may not always hold true in complex healthcare scenarios. Additionally, logistic regression's performance can be influenced by the quality and completeness of the training data. Biased or inaccurate predictions can arise from missing or

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

incomplete data. Hence, it is crucial to thoroughly evaluate the data's quality and completeness when utilizing logistic regression in healthcare settings.

3.3.3. ADABOOST CLASSIFIER

Adaptive Boosting, or AdaBoost, is a boosting algorithm that can be used for both binary and multi-class classification problems. AdaBoost is also an ensemble learning method that combines multiple weak classifiers to create a strong and robust classifier [32]. The idea behind AdaBoost is to adjust the weights of the samples in the training data at each iteration to give more emphasis to the samples that are misclassified by the current ensemble of classifiers. In AdaBoost, a weak classifier is first trained on the data and used to make predictions. The samples misclassified by the weak classifier are given a higher weight, and a new weak classifier is trained on the reweighted data. This process is repeated multiple times, and the predictions of each weak classifier are combined to form the final prediction. AdaBoost is a simple and effective algorithm in various applications, including image and speech recognition, bioinformatics, and medical diagnosis.

3.3.4. STOCHASTIC GRADIENT DESCENT

Stochastic Gradient Descent (SGD) [36] is a widely utilized optimization algorithm for training various machine learning models, including classifiers. In SGD-based classifiers, the model learns to make predictions by updating its weights iteratively to minimize a loss function that measures the difference between the predicted outputs and the actual outputs. In the case of SGD classifiers, the weights are updated based on the gradient of the loss function, computed on a small subset of the training data called a mini batch. This contrasts with batch gradient descent, where the gradient is computed on the entire training set. Using mini batches makes SGD computationally efficient and allows the model to converge faster. SGD is particularly useful when dealing with large datasets, where computing the gradient on the entire dataset can be very expensive. Additionally, SGD is a flexible algorithm that can be used with several loss functions and regularization methods, making it suitable for a wide range of classification tasks. To use SGD for classification, one needs to define the loss function, the regularization method, and other hyperparameters, such as the learning rate and the size of the mini batches. In practice, a common approach is cross-entropy loss and L2 regularization, although other choices are possible depending on the task and the data.

3.3.5. GAUSSIAN CLASSIFIER

The Gaussian classifier, or the Gaussian naive Bayes classifier [37], is a probabilistic classification model used in ML. It is based on Bayes' theorem and assumes that the features of a dataset are independent and normally distributed. The model calculates the probability of a data point belonging to each class and assigns it to the class with the highest probability. The model is extensively employed in tasks such as text classification, spam filtering, and image recognition. It is a simple yet effective model and can manage large datasets with high-dimensional feature spaces. However, its assumption of feature independence may not hold in some datasets, which can lead to decreased accuracy.

3.4. FEDERATED LEARNING

Multiple parties can train a shared model using federated learning without sharing raw data. Instead, the raw data remains on the participants' devices, and only the model parameters are communicated and aggregated to form the final model. Each participant has a local model trained on its own data in a federated learning structure. The local models are then used to make predictions on new data, and the weights of the loss function concerning the model parameters are calculated. These weights are then communicated to a central server, which aggregates the weights and updates the global model parameters. The updated model parameters are then sent back to the participants, and the process is repeated until the model has converged.

Federated learning trains models on data distributed across many parties or devices, like hospitals in healthcare or individual devices. It allows for shared model training, such as in clinical event

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

classification, without compromising privacy, as raw data never leaves the local device. This cross-institution and cross-device learning approach ensures data security and privacy, making it valuable in a data-driven world.

3.4.1. Hyperparameters in Federated Learning

Hyperparameters in federated learning are crucial settings that determine the behavior and performance of the learning algorithm. They are chosen prior to training and play a significant role in achieving optimal model performance. Common hyperparameters in federated learning include the learning rate, the number of communication rounds, the aggregation method, regularization parameters, batch size, and the number of local training epochs. The learning rate controls the step size during optimization, while the number of communication rounds determines the total number of iterations. The aggregation method specifies how updates from multiple clients are combined, and regularization parameters help prevent overfitting. The batch size determines the number of data samples used for local updates, and the number of local training epochs defines the iterations on each client's data. Choosing appropriate hyperparameter values is crucial for achieving efficient convergence, accuracy, and robustness in federated learning models.

3.4.2. Flower Framework

Flower [38] is a federated learning method that aims to improve the performance and fairness of federated learning models. It stands for fairness, accuracy, and privacy in federated learning and is based on differential privacy. In the Flower federated learning method, the participants first locally train their models on their own data and then send their model parameters to the central server. The central server then computes a global model by aggregating the model parameters while adding noise to the aggregated weights to ensure differential privacy, as shown in figure 2.

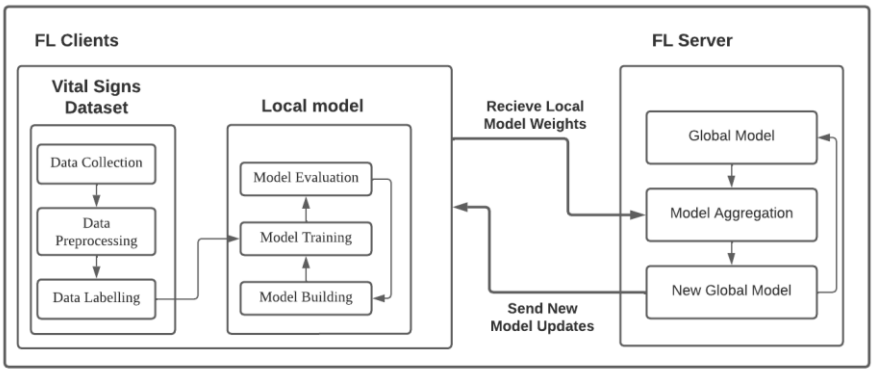

Figure 2. Configuration diagram of FL operation that manages the FL lifecycle.

Flower is a flexible, friendly, and fast framework designed for Federated Learning (FL). The primary components of the Flower FL architecture include the server, the clients, and the gRPC communication layer which interconnects them.

In this setup, the server plays a central role in coordinating the entire FL process. Its responsibilities encompass orchestrating the federated learning process, facilitating communication with participating clients, receiving model updates from clients, and aggregating these updates using a designated strategy, such as Federated Averaging (FedAvg). Once the updates have been aggregated, the server sends the newly updated global model back to the clients, keeping them synchronized and allowing the learning process to continue iteratively.

3.4.3. Federated optimization algorithm FedAvg

FedAvg [39] is an algorithm commonly used in Federated Learning (FL). It's used to aggregate the model updates sent by different clients to the server in an FL setup. The server initializes a global

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

model and sends it to the selected clients for training. Each participating client receives a copy of the global model and trains it on their local data for several epochs, producing a local model. After local training, each client sends the weights of their local model back to the server. The server then aggregates these local model weights into a new global model. The aggregation is usually a weighted average, where the weights could be proportional to the number of samples each client has. This is the key step that constitutes Federated Averaging (1). It ensures that the new global model is representative of the data from all clients. The aggregated model becomes the new global model and is sent back to the clients for the next round of training. The last stage is iteration, that process is repeated over multiple rounds until the global model's performance reaches a satisfactory level or other stopping criteria are met.

$$\boldsymbol{w_{t+1}} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \boldsymbol{w_{t+1}^k} \quad (1)$$

$\boldsymbol{w_t}$ – model weights on communication rounds #*t*, $\boldsymbol{w_{t+1}^k}$ - model weights on communication rounds #*t* on client *k*, μ learning rate,

Algorithm 1. FederatedeAveraging. The K clients are indexed by *k*; B is local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:

Initialize $\boldsymbol{w_0}$

for each round *t* = 1,2,… **do**

$m \leftarrow max(C \cdot K, 1)$

$\boldsymbol{S_t} \leftarrow$ (random set of *m* clients)

for each round $k \in \boldsymbol{S_t}$ **in parallel do**

$\boldsymbol{w_{t+1}^k} \leftarrow$ ClientUpdate $(k, \boldsymbol{w_0})$

$\boldsymbol{m_t} \leftarrow \sum_{k \in S_t} \boldsymbol{n_k}$

$\boldsymbol{w_{t+1}} \leftarrow \sum_{k \in S_t} \frac{n_k}{n} \boldsymbol{w_{t+1}^k}$ // *Erratum*[4]

ClientUpdate(*k, w*): // *Run on client k*

B← (split $\boldsymbol{P_k}$ into batches of size *B*)

for each local epoch *i* from 1 to *E* **do**

for batch $b \in B$ **do**

$w \leftarrow w - \eta \Delta \ell(w; b)$

return *w* to server

C is the fraction of clients (devices or servers with local data samples) that are randomly selected to participate in the computation during each round of training. E is the number of local epochs, or the number of times each client passes through its entire local dataset in each round. B is the size of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

the local minbatch that the client uses for its updates. If B = ∞, the entire local dataset is treated as a single batch. B = ∞ (used in experiments) implies full local dataset is treated as the minibatch as given pseudo-code is given in Algorithm 1.

The main idea behind the Flower method is to ensure that the model parameters are updated evenly across all participants, regardless of the size and quality of their data. This is achieved by weighing the participants' contributions to the global model based on their data quality and the model performance on their local data. The Flower has several advantages compared to traditional federated learning methods as it ensures fairness in the model training by weighing the participants' contributions based on their data quality and the performance of the local model on their data. This helps to prevent the dominance of participants with larger and more diverse data, which can result in a suboptimal global model. Also, Flower incorporates differential privacy by adding noise to the aggregated gradients or weights before sending them to the central server. This helps to ensure the privacy of participants' data even if the central server is compromised.

One of its unique advantages is improved accuracy. By weighing the participants' contributions based on their data quality and the performance of the local model, Flower can improve the accuracy of the global model because the model parameters contributing the most to the global model are updated more frequently, resulting in a more accurate model. The Flower federated learning method provides a privacy-preserving and fair solution for training shared models on distributed data. It is beneficial in clinical settings where data are collected and stored in different hospitals or clinics.

Several hyperparameters in federated learning can impact the performance and convergence of the model. For example, the learning rate determines the size of the step taken toward the negative gradient during model parameter updates. Overshooting the optimal solution can occur with a high learning rate, while slow convergence can occur with a low learning rate. The number of communication rounds determines how often the model parameters are updated and aggregated between the participants and the central server. More communication rounds can result in better convergence as the local batch size also determines the number of examples each participant uses to calculate the gradients or weights for its local model. They follow regularization, which adds a penalty term to the loss function to prevent overfitting. This can help improve the model's generalization performance, especially when dealing with insignificant amounts of data. The distribution of data across the participants can impact the performance and convergence of the model. A skewed distribution, where one participant has significantly more data than others, can result in suboptimal convergence. The last parameter of federated learning is that the heterogeneity of the data across the participants can impact the convergence and generalization performance of the model. This includes differences in the data's distribution, quality, and label balance.

4. Experimental Results.

This study used the Gachon University Laboratory as the environment for applying performance metrics in machine learning. Federated learning for clinical event classification tasks featured the following environment: 24 GB 3090 RTX GPU, 64 GB RAM, core-i9 4.5Ghz, Python, Cuda. The choice of model parameters can also impact the machine learning model's performance. For example, the number of trees in a random forest model or the regularization parameter in a logistic regression model can affect the model's performance. Also, the choice of evaluation metrics is an integral part of the environment. Different metrics may be more appropriate for different types of problems and data. There are several ways to compare machine learning models, such as employing performance metrics. One of the most common ways to compare machine learning models is to evaluate their performance using relevant metrics, including accuracy, precision, recall, and F1-score. These metrics quantitatively assess the model's ability to solve a specific problem. Model accuracy refers to how often the model makes correct predictions. Precision (3) is the proportion of the model's accurate positive predictions among all positive predictions. Recall (4) (Sensitivity) is the proportion of accurate positive predictions the model makes among all positive cases. Eventually, F1-Score (5) is the harmonic mean of precision and recall. Overall, it is essential to consider combining these factors when comparing machine learning models to determine which model best suits a specific problem.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\mathbf{TP} + \mathbf{FP}} \quad (\mathbf{2})$$

$$Recall = \frac{TP}{\mathrm{TP} + \mathrm{FP}} \quad (3)$$

$$F1 = 2 \times \frac{Precision * Recall}{\mathrm{Precison} + \mathrm{Recall}} \quad (4)$$

Table 5 illustrates the machine learning performance on the MIMIC IV dataset on clinic event classification using Flower federated learning techniques. This study investigated the performance of various machine learning models, including Random Forest, Logistic Regression, Support Vector Machines (SVM), AdaBoost, and Gaussian Naïve Bayes, in a federated learning setting. The models were tested with different numbers of clients (3, 5, and 10) and communication rounds (5, 10, and 15). The goal was to assess the impact of these factors on the overall performance and determine the most effective combination for clinical event classification.

Table 5. Evaluating the performance of machine learning models in federated learning with varying rounds and clients.

ML model	Number of rounds	Number of clients	Train acc (F1)	Test acc (F1)
Random Forest	5	3	97.7	94.3
	10	5	98.9	98.9
	15	10	97.2	90.3
Logistic Regression	5	3	93.2	90.1
	10	5	94	92.3
	15	10	92.3	89.3
SGD	5	3	70.1	65.3
	10	5	75.4	60.3
	15	10	70.3	68.3
AdaBoost	5	3	97.6	90.3
	10	5	97.4	92.3
	15	10	90.1	87.2
Gaussian	5	3	80.1	80.3
	10	5	89.7	78.3
	15	10	82.3	74.3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

The results of this study indicate that the highest accuracy across all ML models was achieved when using 10 communication rounds and 5 clients in the FL environment. Figure 3 illustrates the experimental result of classification as this optimal combination was observed for Random Forest, Logistic Classifier, Support Vector Machines (SVM), Ada-Boost, and Gaussian Naïve Bayes in the context of clinical event classification.

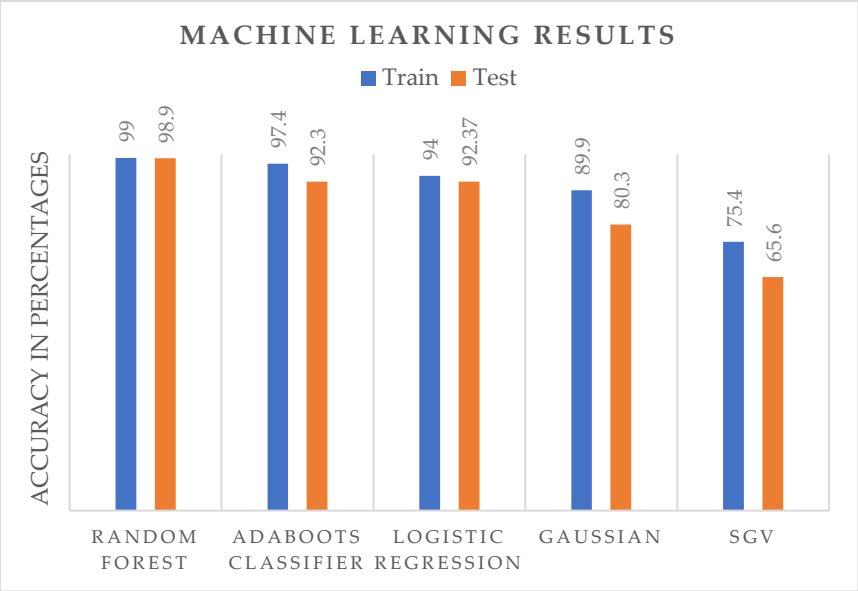

Figure 3. Optimal performance achieved with 10 rounds and 5 clients for various machine learning models.

The results of this study demonstrate a significant improvement in classification accuracy compared to other research approaches in the field of clinical event classification, as shown in Table 6. The method used, incorporating FL, achieved an impressive 98.9% accuracy, outperforming all other methods investigated. This finding highlights the effectiveness and potential of FL in enhancing the performance of ML models for clinical event classification. The superior performance of the FL-based method can be attributed to its ability to leverage distributed datasets, maintain data privacy, and facilitate collaborative learning among multiple clients. This approach allows for the development of robust models that can generalize better and adapt to diverse data sources, leading to improved classification accuracy.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2023 doi:10.20944/preprints202306.0806.v1

Table 6. Superior performance of federated learning-based method in clinical event classification.

	Research in [29]	**Research in [30]**	**Research in [31]**	**Our model**
Number of fixtures	6	1	2	6
Vital signs	HR, BP, RR, SPO	BP	HR, BP	HR, BP, RR, SPO
Clinical event	Any	Any	Any	Any
Number of normal samples	1300	30	571	145085
Number of abnormal samples	130	30	116	117073
Accuracy	95.5 average	94%	ROC max 0.86	98.9
Federated learning	No	No	No	Yes

5. Conclusions

The classification of clinical events using vital signs data from cross-devices multi-modal model is crucial in healthcare, as it allows for the early detection and management of various medical conditions. This study employed FL to classify clinical events using vital signs data, utilizing datasets from multiple clients of X hospital, and employing cross-device ensemble ML classification models, such as Random Forest, AdaBoost, and SGD. Flower FL offered several advantages for clinic event classification, including privacy-preserving capabilities, enabling collaboration between multiple parties to train ML models, and safeguarding the privacy of each party's data. This happens because each party is only required to share encrypted model updates with other parties rather than sharing raw data. Furthermore, Flower FL is designed to scale to many participants, making it particularly suitable for clinic event classification problems where multiple hospitals or clinics may have data to contribute. By combining data and insights from multiple parties, Flower FL can help improve the ML model's performance for clinic event classification because the model can leverage the combined data and insights from various sources. By aggregating model updates from multiple parties, Flower FL can help make ML models for clinic event classification more robust and less susceptible to overfitting to a single party's data. Overall, this study demonstrated that using Flower FL for clinic event classification with ML classification can significantly improve the performance and robustness of the model while preserving the privacy of each party's data. This makes it an essential tool for addressing complex and sensitive healthcare problems.

This study achieved a high accuracy rate of 98.9% for clinic event classification on the MIMIC IV dataset. A client management system is planned to proactively address errors during training, such as data quality or communication issues. This system will ensure that each client's data is adequately incorporated into the ML model for clinical event classification using vital signs data, further improving accuracy and robustness. This study also aims to explore advanced techniques, such as deep learning and ensemble methods, to enhance model performance. These future developments will make this study's approach even more valuable for healthcare providers and researchers.

Funding: This work was supported in part by the Commercialization's 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 by the National Research Foundation of Korea (NRF) Grant funded by MSIT under Grant NRF-2022R1F1A1069069.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bote-Curiel, L.; Muñoz-Romero, S.; Gerrero-Curieses, A.; Rojo-Álvarez, J.L. Deep Learning and Big Data in Healthcare: A Double Review for Critical Beginners. Appl. Sci. 2019, 9, 2331. DOI: 10.3390/app9112331.
2. Xia, Q.; Sifah, E.B.; Smahi, A.; Amofa, S.; Zhang, X. BBDS: Blockchain-Based Data Sharing for Electronic Medical Records in Cloud Environments. Inf. 2017, 8, 44. DOI: 10.3390/info8020044.
3. Gallagher, D.; O'Halloran, P.; De Barra, M.; Davy, A.; Silke, B.; Ward, M.; McNicholas, B. Implementation and Continuous Monitoring of an Electronic Health Record Embedded Readmissions Clinical Decision Support Tool. J. Pers. Med. 2020, 10, E103. DOI: 10.3390/jpm10030103.
4. Albahri, O.S.; Zaidan, A.A.; Zaidan, B.B.; Hashim, M.; Albahri, A.S.; Alsalem, M.A. Real-Time Remote Health-Monitoring Systems in a Medical Centre: A Review of the Provision of Healthcare Services-Based Body Sensor Information, Open Challenges and Methodological Aspects. J. Med. Syst. 2018, 42, 1-47. DOI: 10.1007/s10916-018-1006-6.
5. Siddique, S.; Chow, J.C.L. Machine Learning in Healthcare Communication. Encycl. 2021, 1, 220-239. DOI: 10.3390/encyclopedia1010021.
6. Song, C.; Zeng, X.; Nie, H.; Huang, S.; Hu, M.; Huang, F.; Liu, X. Spatial and Temporal Impacts of Socioeconomic and Environmental Factors on Healthcare Resources: A County-Level Bayesian Local Spatiotemporal Regression Modeling Study of Hospital Beds in Southwest China. Int. J. Environ. Res. Public Health 2020, 17, 5890. DOI: 10.3390/ijerph17165890.
7. Wang, F.; Wang, Y.; Ji, X.; Wang, Z. Effective Macrosomia Prediction Using Random Forest Algorithm. Int. J. Environ. Res. Public Health 2022, 19, 3245. DOI: 10.3390/ijerph19063245.
8. Abdullah, T.A.A.; Zahid, M.S.M.; Ali, W. A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions. Symmetry 2021, 13, 2439. DOI: 10.3390/sym13122439.
9. Mazo, C.; Kearns, C.; Mooney, C.; Gallagher, W.M. Clinical Decision Support Systems in Breast Cancer: A Systematic Review. Cancer 2020, 12, 369. DOI: 10.3390/cancers12020369.
10. Sallam, M.; Almaghaslah, D.; Alsaddik, A.; Alam, S.; Almaghaslah, E.; Al-Mendalawi, M.D. Assessing Healthcare Workers' Knowledge and Their Confidence in the Diagnosis and Management of Human Monkeypox: A Cross-Sectional Study in a Middle Eastern Country. Healthcare 2022, 10, 1722. DOI: 10.3390/healthcare10091722.
11. Guk, K.; Han, G.; Lim, J.; Jeong, K.; Kang, T.; Lim, E.; Jung, J. Evolution of Wearable Devices with Real-Time Disease Monitoring for Personalized Health Care. Nanomaterials 2019, 9, 813. DOI: 10.3390/nano9060813
12. Li, L.; Fan, Y.; Tse, M.; Lin, K.Y. A review of applications in federated learning. Comput. Ind. Eng. 2020, 149, 106854. DOI: 10.1016/j.cie.2020.106854
13. Xu, J.; Glicksberg, B.S.; Su, C.; Walker, P.; Bian, J.; Wang, F. Federated Learning for Healthcare Informatics. J. Healthcare Inform. Res. 2021, 5, 1-9.
14. Brisimi, T.S.; Chen, R.; Mela, T.; Olshevsky, A.; Paschalidis, I.C.; Shi, W. Federated learning of predictive models from federated electronic health records. Int. J. Med. Inform. 2018, 112, 59-67.
15. Antunes, R.S., André da Costa, C., Küderle, A., Yari, I.A., Eskofier, B. Federated learning for healthcare: Systematic review and architecture proposal. ACM Trans. Intell. Syst. Technol. (TIST) 2022, 13(4), 1-23.
16. Choudhury, O.; Gkoulalas-Divanis, A.; Salonidis, T.; Sylla, I.; Park, Y.; Hsu, G.; Das, A. Anonymizing data for privacy-preserving federated learning. arXiv preprint arXiv:2002.09096. 2020, Feb 21.
17. Pati, S.; Baid, U.; Edwards, B.; Sheller, M.; Wang, S.H.; Reina, G.A.; Foley, P.; Gruzdev, A.; Karkada, D.; Davatzikos, C.; Sako, C. Federated learning enables big data for rare cancer boundary detection. Nat. Commun. 2022, 13, 7346. DOI: 10.1038/s41467-022-33407-5.

18. Sannara, E.K.; Portet, F.; Lalanda, P.; German, V.E. A federated learning aggregation algorithm for pervasive computing: Evaluation and comparison. In Proceedings of the 2021 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2021, 1-10.
19. Awad, F.H.; Hamad, M.M.; Alzubaidi, L. Robust Classification and Detection of Big Medical Data Using Advanced Parallel K-Means Clustering, YOLOv4, and Logistic Regression. Life 2023, 13, 691. https://doi.org/10.3390/life13030691
20. Kumar, S., and Singh, M. "Big data analytics for healthcare industry: impact, applications, and tools." Big Data Min. Analyt. 2018, vol. 2, no. 1, pp. 48-57. DOI: 10.26599/BDMA.2018.9020031.
21. Dolley, S. "Big Data Solution to Harnessing Unstructured Data in Healthcare." IBM Report, 2015.
22. Han, T.T.; Pham, H.Y.; Nguyen, D.S.; Iwata, Y.; Do, T.T.; Ishibashi, K.; Sun, G. Machine learning based classification model for screening of infected patients using vital signs. Informatics in Medicine Unlocked 2021, 24, 100592. DOI: 10.1016/j.imu.2021.100592.
23. Beunza Nuin, J.J.; Puertas Sanz, E.; García Ovejero, E.; Villalba, G.; Condés Moreno, E.; Koleva, G.; Hurtado, C.; Landecho, M. Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). J. Biomed. Inform. 2019, 97, 103257. DOI: 10.1016/j.jbi.2019.103257.
24. Pan, Y., Fu, M., Cheng, B., Tao, X., and Guo, J. "Enhanced deep learning assisted convolutional neural network for heart disease prediction on the Internet of medical things platform." IEEE Access, 2020, vol. 8, pp. 189503-189512. DOI: 10.1109/ACCESS.2020.3026214.
25. Shaik, T.; Tao, X.; Higgins, N.; Gururajan, R.; Li, Y.; Zhou, X.; Acharya, U.R. FedStack: Personalized activity monitoring using stacked federated learning. Knowl. -Based Syst. 2022, 257, 109929. DOI: 10.1016/j.knosys.2022.109929.
26. Dayan, I.; Roth, H.R.; Zhong, A.; Harouni, A.; Gentili, A.; Abidin, A.Z.; Liu, A.; Costa, A.B.; Wood, B.J.; Tsai, C.S.; Wang, C.H. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat. Med. 2021, 27, 1735-1743.
27. Budrionis, A.; Miara, M.; Miara, P.; Wilk, S.; Bellika, J.G. Benchmarking PySyft federated learning framework on MIMIC-III dataset. IEEE Access 2021, 9, 116869-116878. DOI: 10.1109/ACCESS.2021.3105929.
28. Johnson, A.E.W.; Bulgarelli, L.; Shen, L.; Gayles, A.; Shammout, A.; Horng, S.; Pollard, T.J.; Moody, B.; Gow, B.; Lehman, L.W.; Celi, L.A. MIMIC-IV, a freely accessible electronic health record dataset. Sci. Data 2023, 10, 1. DOI: 10.1038/s41597-022-01899-x.
29. Forkan, A. R. M. and Khalil, I. "PEACE-Home: probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring." Pervasive Mob. Comp., 2017, vol. 38, pp. 296-311. DOI: 10.1016/j.pmcj.2016.12.009.
30. Hauschild, A.C.; Lemanczyk, M.; Matschinske, J.; Frisch, T.; Zolotareva, O.; Holzinger, A.; Baumbach, J.; Heider, D. Federated Random Forests can improve local performance of predictive models for various healthcare applications. Bioinformatics 2022, 38, 2278-2286.DOI: 10.1093/bioinformatics/btac065.
31. Lu, H.; Uddin, S.; Hajati, F.; Moni, M.A.; Khushi, M. A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. Appl. Intell. 2022, 52, 2411-2422. DOI: 10.1007/s10489-021-02533-w.
32. Tabosa de Oliveira, T.; da Silva Neto, S.R.; Teixeira, I.V.; Aguiar de Oliveira, S.B.; de Almeida Rodrigues, M.G.; Sampaio, V.S.; Endo, P.T. A Comparative Study of Machine Learning Techniques for Multi-Class Classification of Arboviral Diseases. Front. Trop. Dis., 2022, vol. 2, no. 71. DOI: 10.3389/fitd.2021.769968.
33. Clifton, L.; Clifton, D.A.; Watkinson, P.J.; Tarassenko, L. Identification of Patient Deterioration in Vital-Sign Data Using One-Class Support Vector Machines. In Proceedings of the 2011 Federated Conference on Computer Science and Information Systems (FedCSIS), 18 September 2011; pp. 125-131
34. Rocha, T.; Paredes, S.; Carvalho, P.; Henriques, J.; Harris, M. Wavelet Based Time Series Forecast with Application to Acute Hypotensive Episodes Prediction. In Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, 31 August 2010; pp. 2403-2406.
35. Cao, H.; Eshelman, L.; Chbat, N.; Nielsen, L.; Gross, B.; Saeed, M. Predicting ICU Hemodynamic Instability Using Continuous Multiparameter Trends. In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 20 August 2008; pp. 3803-3806.

 doi:10.20944/preprints202306.0806.v1

36. Netrapalli, P. Stochastic gradient descent and its variants in machine learning. J. Indian Inst. Sci. 2019, 99, 201-213.
37. Bi, Z.J.; Han, Y.Q.; Huang, C.Q.; Wang, M. Gaussian naive Bayesian data classification model based on clustering algorithm. In Proceedings of the 2019 International Conference on Modeling, Analysis, Simulation Technologies and Applications (MASTA 2019); Atlantis Press: 2019 Jul; pp. 396-400.
38. Beutel, D.J.; Topal, T.; Mathur, A.; Qiu, X.; Fernandez-Marques, J.; et al. Flower: A Friendly Federated Learning Framework. 2022, HAL Id: hal-03601230.
39. Li, X.; Huang, K.; Yang, W.; Wang, S.; Zhang, Z. On the Convergence of FedAvg on Non-IID Data. arXiv preprint, 2019. Available online: https://arxiv.org/abs/1907.02189 (accessed on 4 July 2019).