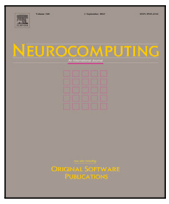
Neurocomputing 634 (2025) 129812

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Advancing DoA assessment through federated learning: A one-shot pseudo data approach

Thomas Schmierer *, Tianning Li, Di Wu, Yan Li

School of Mathematics, Physics and Computing, University of Southern Queensland, 487-535 West St, Darling Heights QLD 4350, Toowoomba, 4350, Australia

ARTICLE INFO

Communicated by M. Bianchini

Keywords:
EEG analysis
Federated learning
Depth of anaesthesia
Machine learning
Neural network

ABSTRACT

Accurately measuring the Depth of Anaesthesia (DoA) during surgical procedures is crucial for patient safety. A significant challenge in developing effective machine learning models for DoA assessment is the lack of data from single organisations and preserving data privacy between institutions. Federated learning offers a solution by enabling multiple parties to collaboratively train models without exchanging data. However, traditional federated learning algorithms perform poorly in data heterogeneous, non-identically distributed data distribution scenarios. To address these challenges, we propose a one-shot federated learning framework, DoAFedP-NN, which facilitates federated learning with heterogeneous model development. The framework is tested in a range of model and data heterogeneity environments. This method enables the training of a global DoA prediction model across different medical facilities without sharing local data.

The DoAFedP-NN model, utilising neural network design with entropy and spectral feature extraction, is compared to benchmark federated learning architectures, demonstrating its advantage in handling heterogeneous medical data. Experimental results show that DoAFedP-NN achieves robust DoA estimation when compared to the Bispectral (BIS) index, with high correlation coefficients of 0.8472 and 0.8542 across independent databases. The proposed model outperforms locally developed models, showing significant improvements when validated against external datasets from different medical facilities. This paper makes the key contributions: (1) introduces a one-shot pseudo-data method for federated learning; (2) demonstrates the effectiveness of this approach for EEG-based DoA using real-world databases; (3) showcases the model's ability to achieve high correlation with the BIS index while preserving patient privacy in a range of client distribution scenarios and under cross-validation.

1. Introduction

In contemporary surgical practice, the administration of anaesthesia is critical to ensure pain-free procedures through controlled, temporary loss of sensation and consciousness [1,2]. Overdosing on anaesthetics is associated with several postoperative complications including nausea, vomiting, hypotension, myocardial depression, and cognitive issues [3,4]. The evaluation of the depth of anaesthesia (DoA) is paramount to tailoring anaesthetic dosages to individual patient needs, thereby enhancing surgical safety and postoperative recovery [5]. Objective assessment techniques based on electroencephalography (EEG), such as the Bispectral Index (BIS), have gained prominence due to their ability to provide quantifiable DoA indices [6–9]. Despite the advancements, current EEG-based monitoring technologies face challenges including variability in response due to individual factors like age and health status. In addition, limitations inherent to commercial devices include sensitivity to external factors and data interpretation [10].

The development and application of quality of machine learning algorithms for EEG-based DoA analysis are constrained by the availability of diverse, high-quality datasets [11]. Regulatory and privacy concerns complicate data acquisition and sharing, posing a significant barrier to the development of generalised models. Federated learning (FL) has emerged as a promising solution by enabling collaborative model training across multiple institutions while preserving data privacy [12–14]. FL is a distributed computing method that solves data privacy issues by exchanging model parameters between clients and servers instead of sharing raw data. However, the data-heterogeneity of medical data, which is often non-independently and identically distributed (non-IID) across different institutions, poses substantial challenges for traditional FL models, which assume independent and identically distributed (IID) data [15]. These models also require multiple communication rounds, leading to high operational costs and increased security risks [16,17].

* Corresponding author.
E-mail addresses: thomas.schmierer@unisq.edu.au (T. Schmierer), tianning.li@unisq.edu.au (T. Li), di.wu@unisq.edu.au (D. Wu), yan.li@unisq.edu.au (Y. Li).

https://doi.org/10.1016/j.neucom.2025.129812
Received 23 June 2024; Received in revised form 11 January 2025; Accepted 21 February 2025
Available online 3 March 2025

To address these limitations, this paper introduces the DoAFedP-NN framework, a novel one-shot FL method that incorporates pseudo-data aggregation based on feature knowledge distillation. The framework presents two key innovations that distinguish it from traditional approaches. Firstly, the one-shot learning approach reduces communication overhead by requiring only a single communication round, thereby improving scalability and mitigating synchronisation issues typically faced in multi-round FL models [18]. Secondly, the pseudo-data aggregation method leverages the statistical distribution of local data to develop models in non-IID cases without the need for direct access to individual data, thus ensuring data privacy and enhancing model robustness in heterogeneous environments. This research demonstrates the potential of DoAFedP-NN to outperform traditional FL models, such as FedAvg, FedSGD, FedProx, FedNova and SCAFFOLD, in real-world, non-IID EEG-based DoA datasets [19]. By applying novel data aggregation methods and cross-validation, the model demonstrates superior predictive and estimation capabilities, delivering consistent and reliable results across diverse patient testing groups. Furthermore, the pseudo-data aggregation method is shown to have the facility to accommodate heterogeneous model development between clients and the global model.

This research focuses on EEG-based DoA assessment, emphasising the necessity of FL in addressing data isolation while maintaining patient data privacy and ensuring high analytical accuracy.

Our contributions are as follows:

1. We propose a one-shot pseudo-data FL aggregation method that can be applied effectively to medical applications to overcome the data isolation and model heterogeneity problems inherent to these applications. To the best of our knowledge, DoAFedP-NN is the first one-shot FL system to employ pseudo-data aggregation.
2. DoAFedP-NN was evaluated using real-world EEG data and compared to benchmark FL methods FedAvg, FedSGD, FedProx, FedNova and SCAFFOLD. The results demonstrate that DoAFedP-NN handles complex, non-homogeneous, non-IID medical data more effectively and achieves comparable outcomes to IID simulations.
3. A 2-client simulation demonstrates how DoAFedP-NN can be applied in real-world medical environments, overcoming data-sharing constraints posed by privacy regulations and enabling collaborative model development across institutions.
4. We employed 5-fold cross-validation and novel cross-database comparison to ensure the reliability of DoAFedP-NN's performance across diverse datasets with varying levels of data heterogeneity.

The rest of the paper will be organised as follows: Section 2 presents the related works in both EEG-based DoA and federated learning. Section 3 presents the research methodology. Section 4 presents the experimental results while Section 5 includes a more detailed discussion of these results.

2. Related work

This section reviews methodologies in federated learning for healthcare applications followed by techniques employed in EEG-based DoA assessment.

2.1. Federated learning in medical applications

Federated learning (FL) has emerged as a promising approach to address data isolation and privacy concerns in healthcare by enabling collaborative model training across multiple institutions without sharing raw data [18,20,21]. In healthcare, the adoption of FL necessitates an approach to address the challenges posed by data quality and data heterogeneity across clients and independent databases. There is a need in the space of EEG learning to overcome the problem of data isolation while maintaining the privacy preservation of trained models by limiting direct access to raw data [22]. Most healthcare and biomedical FL applications apply similarly generic aggregation methods for client models.

Google's popular FL algorithm, FedAvg, involves an iterative training process designed primarily for cross-device settings like mobile devices, where it manages ongoing user participation through multiple rounds. The model parameter updates are averaged to produce the global federated model in each learning iteration [23]. FedSGD (stochastic gradient descent) updates the global model using the averaged gradients from all clients after each local iteration. By employing gradient descent between iterations FedSGD methods can converge more quickly than other FL algorithms. Furthermore, these methods may reduce potential communication overhead and latency caused by infrequent model updates [24].

To overcome the challenges associated with non-IID data more recent FL methods introduce methods to avoid client drift and improve model convergence. Federated proximal (FedProx) learning is an extension of FedSGD designed to handle non-IID data by introducing a proximity term, proving particularly useful when data distributions vary across different hospitals or devices. The introduction of the proximal term acts to regulate the significance of local updates on the global model if they deviate too significantly [17]. SCAFFOLD has shown efficacy in addressing this issue by introducing control variates at both the server and client levels to correct client drift [25]. In contrast, FedNova, normalises client updates based on their effective gradient step size, ensuring that updates remain unbiased regardless of variations in local epochs, learning rates, or data heterogeneity [26]. Alternatively, federated transfer learning leverages a pre-trained model to assist clients with limited data, thereby improving model performance in settings where data availability is uneven among clients [17,27].

These iterative aggregation methods are effective in certain scenarios, however, they are limited in situations where parties are typically organisations or medical institutions rather than individual users. These challenges include the impracticality of requiring multi-round participation, vulnerability to inference attacks, and difficulties in maintaining a fair and trusted central server to coordinate the training process [18]. Furthermore, the inherent heterogeneity of medical data presents a significant challenge for traditional FL methods that rely on averaging techniques for model aggregation [19]. Data collected across different institutions are often non-independently and identically distributed (non-IID), meaning the data distributions vary significantly across clients [28]. This can lead to poor performance of global models trained using simple averaging techniques, as they may not adequately capture the nuances of each client's data. Several approaches have been developed to manage heterogeneous models in FL.

A recently proposed heterogeneous FL (HFL) framework, FedTKD, employs adaptive knowledge distillation to enhance knowledge transfer between the server and clients. This involves transferring knowledge from a larger, pre-trained "teacher" model to a smaller "student" model. This helps bridge the gap between diverse models, facilitating knowledge sharing even when their structures differ [29]. This method aims to improve the accuracy of client models without compromising data privacy. This method specifically uses a selectivity knowledge fusion method to ensure high-quality global logit computation and an adaptive knowledge distillation method to improve knowledge transfer between the server and clients [18]. In other works, local clients contribute pseudo-data, derived from their original datasets, for model training on a federated server. This approach demonstrates the effectiveness of pseudo-data in improving communication rounds within FL contexts [30,31]. Alternatively, the CFSL model is specifically designed to address the challenges of non-IID data in FL by leveraging a personalised FL approach. The model is designed to adapt to the heterogeneity in data by leveraging hypernetworks to generate neural networks tailored to the unique data characteristics of each client [20]. These HFL models pose challenges related to computational complexity

and communication efficiency, particularly when implemented on non-IID data. Training and aggregating HFL models can be computationally expensive, particularly when dealing with many clients and complex models [18]. This complexity can hinder the scalability of HFL models because these models may involve frequent communication between clients and the server to exchange model parameters or knowledge representations.

One-shot FL presents a promising alternative in constrained data-sharing environments [32]. This approach involves a single round of communication where parties upload their local models to a centralised platform once, which then aggregates these models to create a final global model. This paradigm has demonstrated efficacy when dealing with image datasets in medical applications [33]. One-shot FL not only circumvents the issues of multi-round dependency and security concerns but also aligns with the practicalities of organisational participation, where entities are not continuously engaged but may still benefit from shared models. One-shot FL, therefore, offers a robust solution by reducing reliance on repeated party availability and enhancing the security and feasibility of FL in sensitive and critical sectors like healthcare [32].

2.2. EEG-based DoA assessment

Signal decomposition plays a pivotal role in the proficient analysis of electroencephalography (EEG) signals. The utilisation of the Fast Fourier Transform (FFT) technique stands as a reliable method adopted for such analyses [6]. Alternative decomposition methods employing discrete wavelet transform and power spectral density successfully demonstrated the capacity to capture real-time transitions from consciousness to unconsciousness during anaesthesia induction [34].

A range of feature extractions have shown effectiveness in discerning patterns in EEG signals associated with DoA levels. The second-order difference plot (SODP) is an effective graphical method for representing data variability in EEG signal analysis for the DoA. This method plots successive rates of variability against each other and has shown consistently strong results in classifying epileptic signals and in DoA EEG analysis [35,36]. SODP's ability to highlight non-linear dynamics makes it useful for uncovering patterns and anomalies in EEG data. Entropy-based metrics quantify EEG signal unpredictability or randomness. Permutation and spectral entropy are notably associated with the BIS index; observing correlations greater than 0.7 with the BIS in existing literature [36,37]. Additional entropy measures employed in recent work include wave entropy, hierarchical dispersion entropy (HDE), sample entropy, Hurst entropy, singular value decomposition entropy, and fuzzy entropy [6].

Studies on EEG-based DoA analysis have achieved variable success with linear regression modelling. The limiting factors often identified in these studies are the feature extraction method and sample size examined. This method is shown to have low computational intensity if appropriate feature extraction is employed [6]. Deep learning models, such as artificial neural networks (ANN), can extract complex features from EEG signals and learn hierarchical representations, which can be instrumental in accurately determining the DoA [6]. This group of methods may require substantial data and computational resources and may be prone to overfitting [38]. Feed-forward neural networks (FFNNs) and multilayer perceptron (MLP) demonstrate high effectiveness in estimating the DoA from raw EEG signals as observed with a correlation of 0.94 with the BIS [39]. In a 56-patient, single-channel EEG study, long short-term memory (LSTM) modelling was observed to achieve a correlation of 0.70 and an area under the curve (AUC) of 0.93 based on the BIS [40]. These methods effectively learn and remember over long sequences making use of the temporal features present in the EEG signal.

3. Methodology

3.1. Federated learning methodology

The methodology underpinning this study's FL approach, termed DoAFedP-NN, is a process specifically tailored for the DoA index design using EEG data. This method employs a one-shot FL architecture with heterogeneous model development to account for the challenges placed on medical institution data management. The FL aggregation method is based on a feature knowledge distillation approach and is divided into six steps, as shown in Fig. 1, where the key processes are described as follows. The flow of this algorithm is described in Algorithm 1.

Algorithm 1: DoAFedP-NN Algorithm, k denotes local client, $(X_{\text{train}}, y_{\text{train}})$ represents local training data and (S_x, S_y) the pseudo data

Input: Local data $(X_{\text{train}}, y_{\text{train}})$
Output: Trained global model θ_{global}

Local Training: ▷ Local Client
for each client k **do**
  Initialise NN for $(X_{\text{train}}, y_{\text{train}})$
  $f(\theta_k) \leftarrow \text{TrainModel}(X_{\text{train}}, y_{\text{train}}, E_{\max}, P_{\text{goal}})$
  Extract mean μ_k and standard deviation σ_k from X_{train}: $\phi_k = (\mu_k, \sigma_k)$
end for
Federated Aggregation: ▷ Central Server
for each client k **do**
  Fetch θ_k, ϕ_k from local client
  Generate N pseudo data points, S_x^k, for each client $S_x^k = \{x_1, x_2, \ldots, x_N\} \sim \mathcal{N}(\mu_k, \sigma_k)$
  Generate N pseudo labels, S_x^k, for each client, k, based on ϕ_k at central server $S_y^k = \{y_1, y_2, \ldots, y_N\} = f(S_x^k; \theta_k)$
end for
Combine pseudo data, $S = \bigcup_{k=1}^{K}(S_x, S_y)$
Global Model Training:
Train global model θ_{global} using combined pseudo data S
Model Update: ▷ Local Client
for each client k **do**
  Update local model θ_k using θ_{global}: $\theta_k \leftarrow \theta_{\text{global}}$
end for

(1) Local model development: Initially, each participating client, k, undertakes the extraction of the key features, denoted as $X_k = (x_1, x_2, \ldots, x_5)$, from single-channel EEG recordings based on the methods discussed in Section 3.2. The local model, $f(\theta_k)$, is trained on the extracted features using an MLP FFNN to predict the anaesthesia depth (as measured by BIS) from these features. The model parameters, θ_k, are updated using the scaled conjugate gradient algorithm, (*trainscg*), and the training process continues until either the performance goal, P_{goal}, or the maximum number of epochs, $E_{\max}$, is reached:

$$f(\theta_k) \leftarrow \text{TrainModel}(X_{\text{train}}, y_{\text{train}}, E_{\max}, P_{\text{goal}})$$

(2) Upload client model information: Following the local training phase, each client transmits its model's parameters, θ_k, to a Central Server (CS). In addition, the client sends the summary statistics, $\phi_k = (\mu_k, \sigma_k)$, where μ_k is the univariate mean of each feature and σ_k is the standard deviation of each feature for each client k. By having clients share their model parameters only once and transmit minimal aggregate statistics, this approach results in a substantial reduction in communication size compared to traditional FL models.

(3) Pseudo data generation: The CS synthesises the pseudo-feature data set based on the statistical information received from each client,

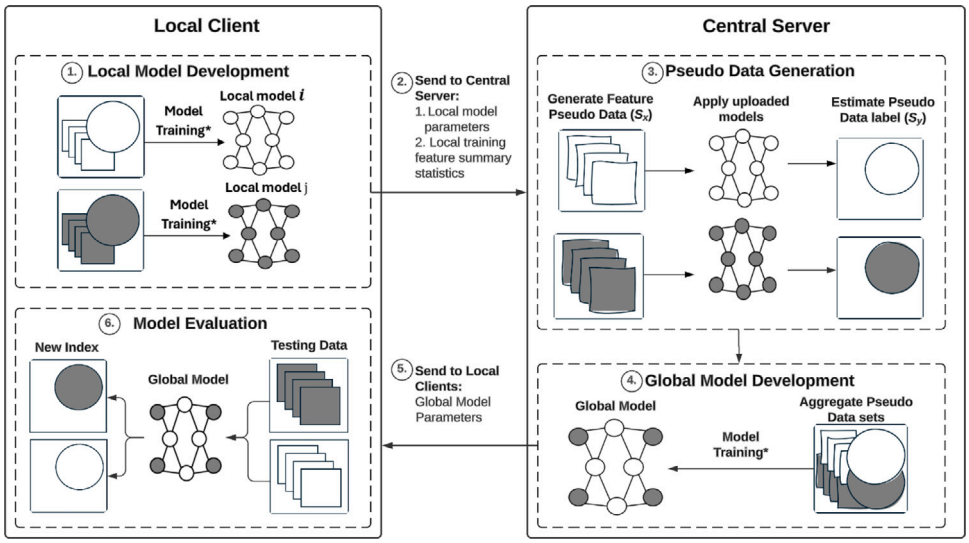


Fig. 1. Pseudo data federated learning (DoAFedP-NN) architecture. * Model training as specified in Fig. 2.

S_x^k, representative of each client's original feature distribution. The pseudo-feature data points are generated from a normal distribution $\mathcal{N}(\mu_k, \sigma_k)$ using the means, μ_k, and standard deviations, σ_k, contained in ϕ_k to capture the local feature distributions without compromising privacy. For each client k, pseudo data representing the BIS label $S_y^k = y_1, y_2, \ldots, y_N$ are generated by applying the respective client model $f(\theta_k)$ to the pseudo-feature data, such that $S_y^k = f(S_x^k, \theta_k)$.

(4) Global model development: The CS aggregates the pseudo datasets from all clients k, expressed as $S = \bigcup_{k=1}^{K}(S_x, S_y)$, to form a combined dataset. Using this aggregated pseudo data, the CS trains the global model, θ_{global}, over the combined dataset S.

(5) Send global model to clients: The parameters of the global model θ_{global} are transmitted back to each client for model evaluation. The global model is a product of a one-shot FL process, avoiding the need for iterative communication typical in traditional FL frameworks.

(6) Model evaluation: The global model θ_{global} is evaluated by comparing the DoAFedP-NN index against the BIS index of the testing group. The evaluation includes 5-fold cross-validation, where performance metrics such as correlation and root mean square error (RMSE) between the predicted BIS values and the actual values are computed.

3.2. Model building and feature extraction

The performance of the DoAFedP-NN algorithm in modelling DoA is evaluated by comparing performance to the BIS index, given its established role as a benchmark in clinical practice [6]. However, while BIS provides a useful reference point, it has limitations that can be addressed by integrating a broader range of patient data through FL. By aligning the DoAFedP-NN index with BIS while leveraging diverse patient datasets, this approach aims to develop a more robust and accurate model for DoA assessment. To capture and represent the relationship between the EEG signal and the BIS index a process of feature extraction and model building is undertaken as outlined in Fig. 2. This is because the non-stationary nature of EEG signals necessitates advanced feature extraction techniques to extract meaningful information [6]. A multilayer perceptron (MLP) neural network is utilised with a network architecture comprising three hidden layers, each containing ten neurons. The scaled conjugate gradient (SCG) algorithm, known for its efficiency, low memory requirements, and speed with large datasets, is used for training [39]. Identical model structure and feature extraction methods are used in the global model and client models in the DoAFedP-NN algorithm. Application of heterogeneous models between clients and central server is discussed in Section 5.3.

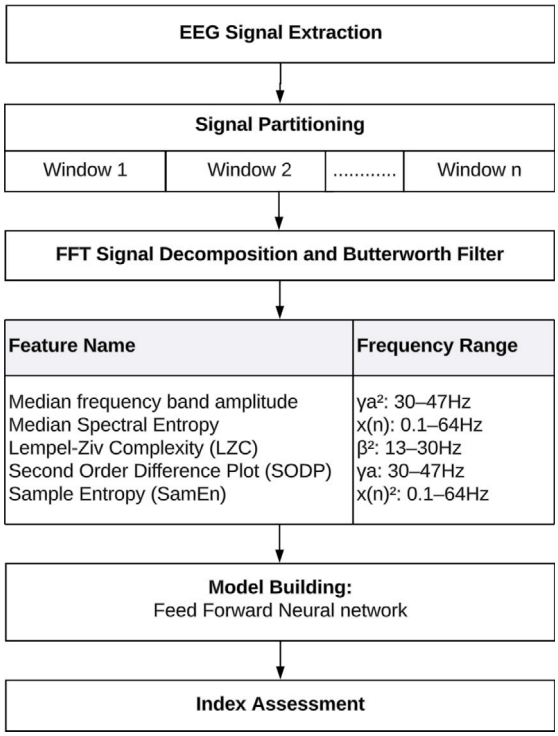


Fig. 2. Model building and feature extraction flowchart.

Before feature extraction, the signal is denoised with a high and low pass Butterworth filter to remove line interference from the signal. The signal is then decomposed into 56-second windows with 55 s of overlap with the previous segment. These windows are selected to allow for a suitable quantity of historical signal to produce a stable, yet responsive signal [36]. Each of these windows is then segmented further into 15 frequency bands and an additional corresponding 15 power functions using the fast Fourier transform method (FFT). Four key feature extraction methods are applied to these decomposed signal windows to identify features that effectively represent the BIS index. These features, second order difference plot (SODP), Lempel–Ziv complexity (LZC), sample entropy (SampEn), and spectral entropy, are shown in Fig. 2. In addition, the median frequency band amplitude for the 30–47 Hz

frequency band is also selected. The independence of these methods enables the capture of distinct information, which has been recognised in existing literature as a strong indicator of depth of anaesthesia, allowing for a comprehensive assimilation of the data [6].

The SODP is calculated separately for each second within the input window, following the established methods in EEG-based DoA research [35,36,41,42]. The SODP for the signal, $x(n)$, is found by plotting $X(n) = x(n+1) - x(n)$ against $Y(n) = x(n+2) - x(n+1)$, where:

$$SODP = |\log(3\pi\sqrt{(SX^2 + SY^2 + D)(SX^2 + SY^2 - D)})| \quad (1)$$

where SX and SY are defined as:

$$SX = \sqrt{\sum_{N=0}^{n-1} \frac{X(n)^2}{N}}, \quad SY = \sqrt{\sum_{N=0}^{n-1} \frac{Y(n)^2}{N}}. \quad (2)$$

SXY is defined as:

$$SXY = \frac{1}{N} \sum (X(n) * Y(N)) \quad (3)$$

The distance, D, is hence calculated as:

$$D = \sqrt{(SX^2 + SY^2) - 4(SX^2 SY^2 - SXY^2)} \quad (4)$$

Spectral entropy and SampEn are used to quantify a time series's complexity or regularity. SampEn is calculated by measuring the probability that sequences of length m match, within a tolerance r, and how that probability changes when the sequences are extended by one point. For a signal of length N, $B_m(r)$ is the average probability that two sequences of length m match within r, and $A_m(r)$ is the average probability that two sequences of length $m+1$ match within the same tolerance. SampEn is defined as

$$\text{SampEn}(m, r, N) = -\ln\left(\frac{A_m(r)}{B_m(r)}\right) \quad (5)$$

The lower SampEn value indicates greater regularity and predictability, while a higher value signifies more complexity [43]. The parameters m and r are typically set based on the application, with $m = 1$ and $r = 0.2 \times \sigma$ (where σ is the standard deviation of the signal) being common choices for physiological signals [44].

The spectral entropy takes the signal's normalised power spectrum distribution in the frequency domain as a probability distribution and calculates its Shannon entropy [37]. For a given signal $x(n)$, $X(n)$ is the discrete Fourier transform. The probability distribution, $P(n)$, is:

$$P(n) = \frac{X(n)}{\sum_i X(i)} \quad (6)$$

Where $\sum_{m=1}^{M} P(n) = 1$. Hence, the spectral entropy at time t, can be expressed in terms of the time-variant probability distribution $P(t, n)$ at time t:

$$SE_n(t) = -\sum_{n=1}^{N} P(t, n) \log_2 P(t, n) \quad (7)$$

The median values of the spectral entropy over the 56-second window is used as the extracted feature in alignment with [36,45].

The LZC is a non-parametric measure that quantifies the complexity of data sequences by identifying new pattern generation rates. A pattern is defined as a sub-sequence that has not been seen before in the sequence. The LZC of a sequence is then defined as the total number of distinct patterns in the sequence. LZC has been used successfully in DoA applications with EEG in the past to quantify the complexity or irregularity of the EEG signal [46].

The features selected are pivotal in capturing the diverse characteristics of brain activity during anaesthesia. The specific features selected for modelling with the frequency bands they are extracted through are detailed in Fig. 2. This selection includes both frequency and time-domain characteristics as well as linear and non-linear properties [6].

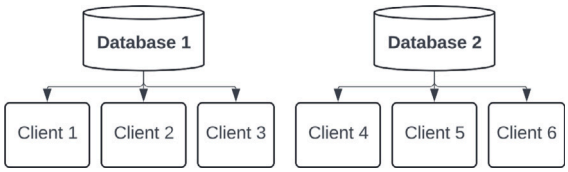


Fig. 3. Client Allocation for non-IID experiment.

4. Results

The following section explores the efficacy of the DoAFedP-NN in performing FL across non-IID distributed clients. This model employs homogeneous model building between clients and the central server as described in Section 3. The model is evaluated against the industry benchmark DoA index, the BIS index, by assessing regression metrics including mean squared error (MSE) and correlation coefficient. These metrics are compared to state-of-the-art FL benchmark methods.

4.1. Experiment environment

This FL simulation utilised two real-world EEG and BIS databases. The VitalDB dataset is an open-access resource from Seoul National University Hospital, Republic of Korea [47], comprised of 6388 cases. The VitalDB dataset offers high-resolution intraoperative signals and clinical information from surgical patients, with anaesthesia durations ranging from 90 to 245 min. It includes a balanced sample of male and female patients aged 48 to 68 years, weighing between 53 and 69 kg. The dataset provides EEG signals recorded at 500 Hz and an extensive collection of additional clinical parameters, such as drug dosages and administration times. Similarly, the UniSQ dataset contains EEG data from adult patients aged 22 to 83 years, weighing between 60 and 130 kg. EEG signals were recorded using a two-channel setup at 128 Hz, with bispectral index (BIS) values documented at one-second intervals during surgery. Anaesthetist notes detailing drug dosages and administration times were available for these patients, providing additional context for the EEG data. For our analysis, we selected 20 patients from each database. This selection ensured patients had similar surgery durations and that each patient exhibited a comprehensive representation of all anaesthesia stages. In addition, high signal quality index (SQI) throughout the procedure was a critical factor in patient inclusion to ensure minimal periods of high noise or disruption, thereby ensuring the reliability and comparability of the analysis. The FL simulation was implemented in MATLAB. The experiments were conducted on a system equipped with an Intel i5-10500 CPU and 32 GB of memory, running MATLAB R2021a version 9 with the Deep Learning toolbox.

4.2. Client allocation

Patients from UniSQ and VitalDB databases are distributed among 6 clients (Clients 1 to 3 for UniSQ and Clients 4 to 6 for VitalDB) as shown in Fig. 3. Within each client patients were randomly divided into training and testing sets, with at least two patients in each testing set and the remaining data designated for training. This allocation was cycled through 5-fold cross-validation to ensure comprehensive model evaluation. The client allocation described here is used in Sections 4.3 and 4.4

The non-IID nature of the data is central to this analysis. As shown in Fig. 4, the BIS distributions vary distinctly across the 6 clients, a finding further supported by Table 1, which presents client-specific BIS statistics. A Kolmogorov–Smirnov test confirmed the non-identical nature of these distributions, reinforcing the non-IID assumption.

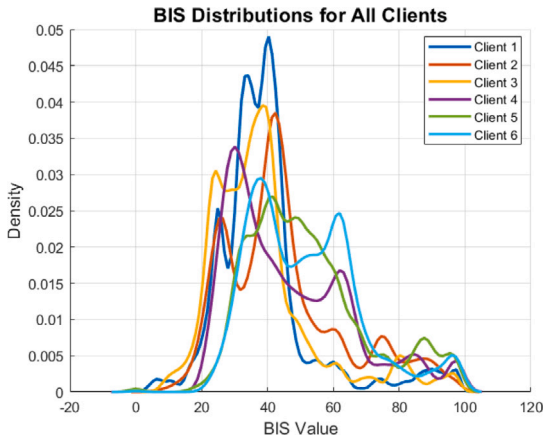


Fig. 4. BIS distribution across clients.

Table 1
Client statistics: target mean, target variance, data length and number of patients per client.

Client	Mean	Var	Length	Patient count
Client 1	39.09	234.2	19 774	6
Client 2	44.84	332.3	20 586	7
Client 3	38.13	265.2	17 155	7
Client 4	46.59	353.5	28 457	6
Client 5	51.81	318.2	25 357	7
Client 6	51.49	279.5	28 026	7

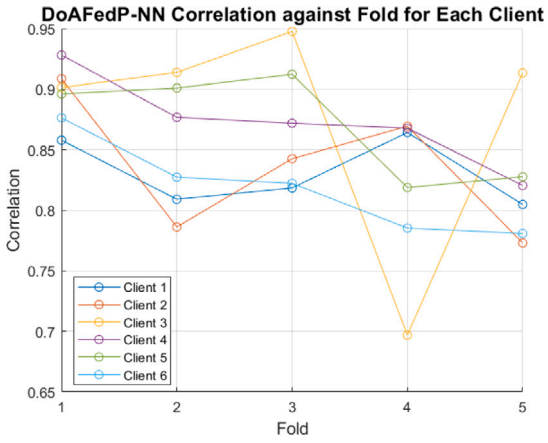


Fig. 5. DoAFedP-NN cross-validation correlation results.

4.3. Experimental results — DoAFedP-NN

The performance of the DoAFedP-NN algorithm in modelling DoA is evaluated using the BIS as the benchmark. As BIS is the standard for DoA analysis, closer alignment with BIS values indicates a more accurate representation of DoA by the model. The correlation and RMSE between the DoAFedP-NN index and the BIS index serves as a measure of the model's effectiveness in capturing patients' DoA.

The experimental results for the DoAFedP-NN model demonstrate its effectiveness when applied to EEG data from UniSQ and VitalDB clients. These results evaluate the effectiveness of the proposed model to capture the true state of DoA as measured by the BIS index. All results were obtained using a 5-fold cross-validation procedure to ensure robustness. The average cross-validation results for each client are presented in Table 2 and detailed in Figs. 5 and 6. The local training iterations to achieve convergence for each local model varied between client and database. The number of iterations required ranged from 37 iterations for Client 3 to 133 iterations for Client 4. This variation illustrates the differences in data characteristics and complexities across clients. To illustrate the quality of the DoAFedP-NN model in representing the DoA, Figs. 7 and 8 show the DoAFedP-NN alongside the BIS index for a single patient for each database. These figures highlight the close relationship between these indexes over the course of a single anaesthetic procedure.

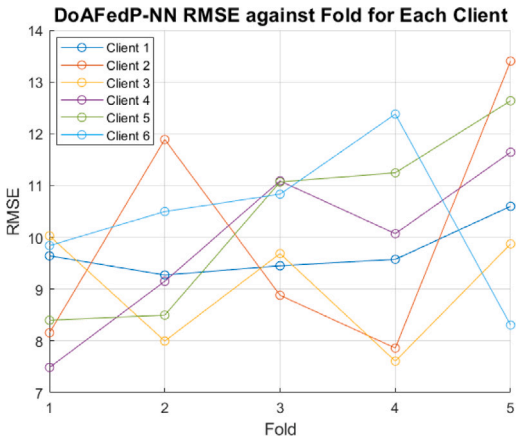


Fig. 6. DoAFedP-NN cross-validation RMSE results.

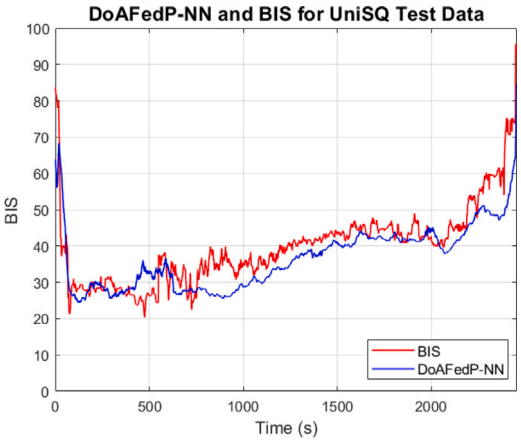


Fig. 7. DoAFedP-NN model prediction against actual BIS for a single test case in the UniSQ database.

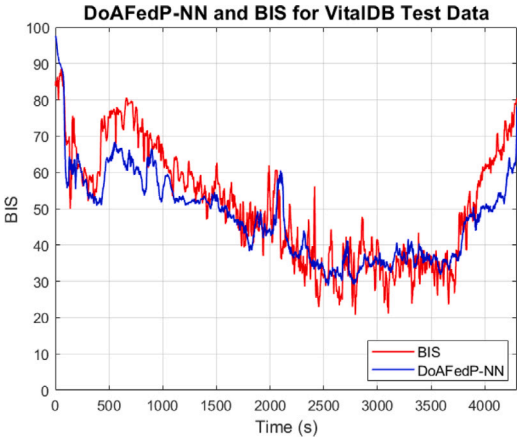


Fig. 8. DoAFedP-NN model prediction against actual BIS for a single test case in the VitalDB database.

Table 2
Comparison of Local and Global, DoAFedP-NN, model performance for UniSQ and VitalDB clients. Information presented is based on the average of 5 fold cross-validation result for each client.

Client	Local corr	Local RMSE	DoAFedP-NN Corr	DoAFedP-NN RMSE
Client 1	0.7518	11.7010	0.8309	9.7082
Client 2	0.8030	10.8044	0.8360	10.0385
Client 3	0.7956	10.1749	0.8746	9.0388
UniSQ Overall	**0.7835**	**10.8934**	**0.8472**	**9.5952**
Client 4	0.8476	11.2848	0.8730	9.8876
Client 5	0.8534	11.4683	0.8711	10.3693
Client 6	0.7958	10.5973	0.8184	10.3707
VitalDB Overall	**0.8322**	**11.1168**	**0.8542**	**10.2092**

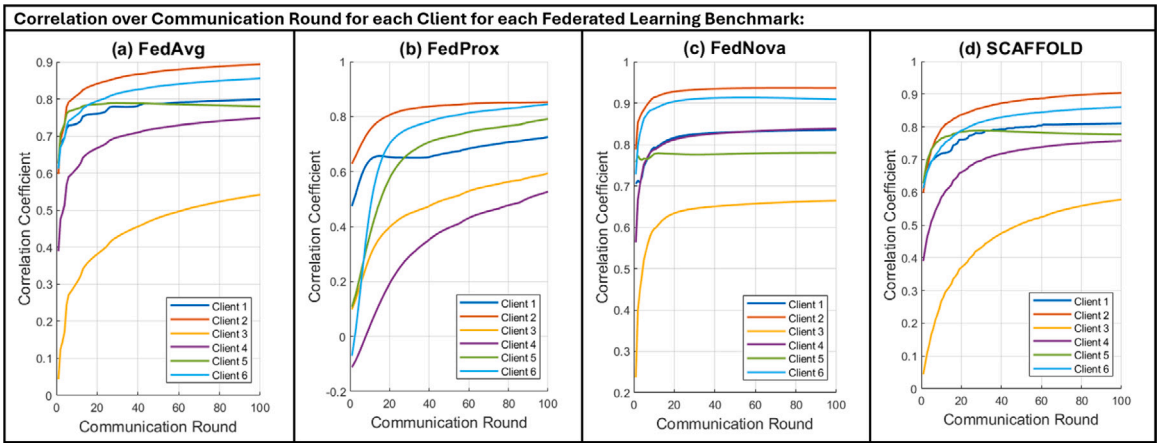


Fig. 9. Accuracy against communication round for leading federated learning benchmarks (a) FedAvg, (b) FedProx, (c) FedNova, and (d) SCAFFOLD for each client.

The average performance of the DoAFedP-NN model across all folds shows a correlation of 0.8475 and an RMSE of 10.1551 with the BIS index, indicating its strong predictive capability. This result reflects the ability of the FL approach to integrate data from multiple clients and build a comprehensive model that generalises well across diverse data distributions.

For the UniSQ clients, the DoAFedP-NN model consistently outperformed the locally trained models. Specifically, the global model achieved an average correlation of 0.8472 with an RMSE of 9.5952, compared to the local models, which had an average correlation of 0.7835 and an RMSE of 10.8934. This improvement for the DoAFedP-NN relative to the local models was mirrored in the VitalDB clients. The average correlation for the DoAFedP-NN was 0.8542 with an RMSE of 10.2092, compared to the local models, which had a correlation of 0.8322 and an RMSE of 11.1168. These results underscore the performance of the proposed FL model, which benefits from the collective data of all clients and mitigates the limitations of training on individual, non-IID datasets. Furthermore, the DoAFedP-NN model is required to transmit an additional amount of additional information because of the statistical information sent to the central server. In these experiments, each client transmitted a total of only 2.51 KB of information, consisting of 2.27 KB (90.65%) for model parameters and 240 bytes (9.35%) for aggregate statistical information.

4.4. Benchmark analysis

In this section, we compare the DoAFedP-NN model with several widely recognised FL aggregation methods: FedAvg, FedSGD, FedProx, FedNova and SCAFFOLD. These benchmarks were chosen due to their established efficacy in handling distributed data. In particular, FedNova and SCAFFOLD have been shown in recent work to have high-efficacy FL scenarios that exhibit non-IID data structures and in medical applications [25,26]. The performance is evaluated based on the correlation with the BIS index across communication rounds. The benchmarks were conducted under identical conditions to ensure comparability. Patients from UniSQ and VitalDB databases are allocated to each client in the same way as in the DoAFedP-NN results as described in Section 4.2 and Table 1. Similarly, training patients are allocated within each client ensuring that at least two patients are reserved for testing in each cross-validation round. Initially, a global model was calibrated using 100 data points randomly selected from each client to form a public dataset. This model was then distributed to the clients as the starting condition. For each benchmark method, a single local training epoch was used, with 100 communication rounds between the central server and each client in alignment with existing literature [18]. For each training epoch, all training patients are utilised in each client. The FedProx, FedSGD, FedNova and SCAFFOLD methods use a learning rate of 0.1, with FedProx also incorporating a proximal term of 0.1 to mitigate local model drift. SCAFFOLD mitigates client drift through active gradient correction, FedNova enhances fairness and stability by addressing imbalances in client updates. In contrast, the DoAFedP-NN method utilised the one-shot method with 2000 pseudo data points during the central server update, as detailed in Section 3.

The results from the 5-fold cross-validation show that DoAFedP-NN consistently outperforms benchmark and state-of-the-art FL methods. Compared to the highest benchmark, FedNova, DoAFedP-NN achieved a 7.8% improvement in correlation for the UniSQ database (clients 1 to 3) and a 9.2% improvement for the VitalDB database (clients 4 to 6) when compared to BIS. These results underscore the ability of DoAFedP-NN to model accurate estimations of DoA based on distributed, non-IID data. The progression of accuracy across these communication rounds for the leading benchmark methods is illustrated in Fig. 9, showing the correlation of the FL model with each client's testing data after each round. Table 3 indicates each model's percentage improvement relative to the base FL benchmark, FedAvg. It can be seen that all methods, except FedSGD outperform this benchmark method with the DoAFedP-NN outperforming this base method 10.3% and 16.1% for UniSQ and VitalDB clients respectively. Furthermore, the

Table 3
Comparison of Model Performance Across UniSQ and VitalDB Data Centres. Results show the average correlation (Corr) and RMSE across all clients following 5-fold cross-validation. Percentage improvement calculated relative to FedAvg.

Model	UniSQ		VitalDB	
	Corr	RMSE	Corr	RMSE
FedAvg	0.768 (Base)	11.78 (Base)	0.736 (Base)	12.29 (Base)
FedSGD	0.659 (−14.2%)	14.03 (+19.0%)	0.720 (−2.17%)	13.65 (+11.1%)
FedProx	0.773 (+0.65%)	13.52 (+14.8%)	0.730 (−0.82%)	14.90 (+21.2%)
SCAFFOLD	0.761 (−0.91%)	11.27 (−4.33%)	0.727 (−1.22%)	12.10 (−1.55%)
FedNova	0.786 (+2.34%)	9.939 (−15.6%)	0.782 (+6.25%)	11.435 (−7.00%)
DoAFedP-NN	**0.847** (**+10.3%**)	**9.60** (**−18.5%**)	**0.854** (**+16.1%**)	**10.21** (**−16.9%**)

one-shot approach of the DoAFedP-NN model drastically reduces the communication overhead and computational cost for FL in non-IID scenarios. Compared to each of the communication rounds for the traditional methods, the DoAFedP-NN model's single communication round requires an additional 10% size to communicate model parameters. This cost is offset by the reduced number of communication rounds required in the one-shot FL approach. Consequently, the DoAFedP-NN method achieved a communication reduction of approximately 98.9% per client across the entire training process. Despite this significantly lower communication burden, DoAFedP-NN outperforms these traditional methods in both correlation and RMSE, making it a more efficient and effective solution for handling non-IID data in FL scenarios.

5. Discussion

5.1. IID data simulation with Dirichlet data partitioning

A series of simulations with varying levels of heterogeneity was conducted using real EEG signals to approximate the closest possible IID scenario for EEG analysis. IID simulations of EEG signals are inherently constrained by their time-dependent nature, the presence of short- and long-term correlations, and temporal dependencies introduced by overlapping windows. To address these challenges, datasets from the UniSQ and VitalDB databases were pooled, shuffled, and redistributed across six clients using a Dirichlet distribution (Fig. 10) [18]. This method ensured the controlled allocation of data to each client and facilitates the simulation of varying levels of data heterogeneity.

To manage the continuous nature of the target variable (BIS), the pooled data were first divided into 10 bins based on the corresponding ranges of BIS index values. The Dirichlet distribution was then used to allocate data proportionally between clients within each bin. The degree of heterogeneity was governed by the Dirichlet parameter (α), where smaller values of α resulted in skewed, heterogeneous distributions favouring a few clients, and larger values of α produced more balanced, homogeneous distributions approaching IID distributions [18].

This approach effectively simulated varying levels of data heterogeneity, enabling the evaluation of FL models under both non-IID and IID conditions. For this study, α values of 0.1, 0.5, 1, 5, and 10 were selected to represent a comprehensive range of heterogeneity levels. In addition, a full IID simulation was conducted using a uniform allocation without the Dirichlet distribution. The data distribution of the simulations for each client is shown in Fig. 11. Furthermore, the authentic case (discussed in Section 4) was compared with these simulated heterogeneity distributions. All results were derived from the average of 5-fold cross-validation for each α level.

The results demonstrate that the DoAFedP-NN model performs exceptionally well across the tested range of α values, underscoring its

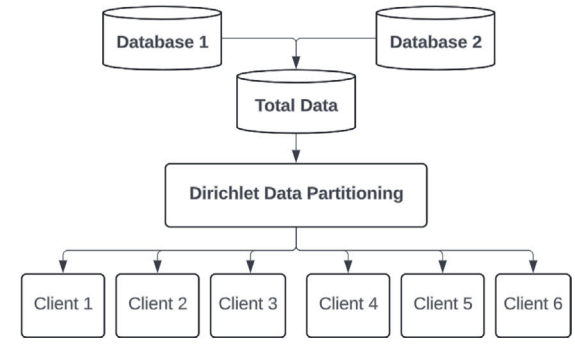


Fig. 10. Client allocation with Dirichlet Data Partitioning for heterogeneous data distribution simulations.

Table 4
Comparative performance of database cross-validation, local, and global model evaluations for the neural network model with the BIS index.

			Train			
			UniSQ	VitalDB	DoAFedP-NN	Global model
Test	UniSQ	Corr	**0.8546**	0.8331	**0.8657**	**0.8776**
		RMSE	7.1473	8.1431	7.1304	6.8921
	VitalDB	Corr	0.7884	**0.8395**	**0.8469**	**0.8459**
		RMSE	11.2275	8.9384	8.956	9.0406

capability to facilitate FL under varying degrees of data heterogeneity (Fig. 12). At high levels of heterogeneity ($\alpha = 0.25$), a correlation of 0.813 was observed between the DoAFedP-NN index and the BIS index. Conversely, at high levels of homogeneity ($\alpha = 10$), a correlation of 0.8539 was recorded. Despite the non-IID nature of the original allocation in the authentic case, the DoAFedP-NN achieved performance comparable to that of the homogeneous data distribution, with a correlation reduction of less than 0.4%. The robustness of this method is further supported by the minor difference observed in the IID simulation, which had an average correlation of 0.8588 with the BIS observed in all clients.

The highest-performing FL benchmark, FedNova, was evaluated under these heterogeneity scenarios and shown in Fig. 12. Unlike the DoAFedP, FedNova's modelling efficacy is substantially reduced as data heterogeneity increases. The DoAFedP-NN can be seen to outperform traditional and state-of-the-art FL methods in both IID and non-IID conditions.

5.2. Application of DoAFedP-NN in overcoming data isolation

In developing robust machine learning models for medical applications, collaboration across data stores is essential to ensure generalisation to diverse patient populations [48]. To assess the efficacy of the DoAFedP-NN model, we performed a two-client simulation, using distributed data across the UniSQ and VitalDB databases. This scenario mirrors real-world medical data isolation issues and highlights how DoAFedP-NN can facilitate database cohesion without violating patient privacy. The DoAFedP-NN model was implemented with this client allocation according to Section 3. Further validation was provided by cross-database testing, where each database's locally trained model was applied to the testing data of the other database.

In this case, localised models showed better performance on their native data but struggled when applied to external datasets, reflecting the lack of generalisation that arises from isolated training. As shown in Table 4, the average correlation for locally trained models was 0.848, while cross-trained models exhibited a lower correlation of 0.811.

In contrast, the DoAFedP-NN model achieved an average correlation of 0.857 across both databases, outperforming even the globally trained model, which was constructed by combining all data from both

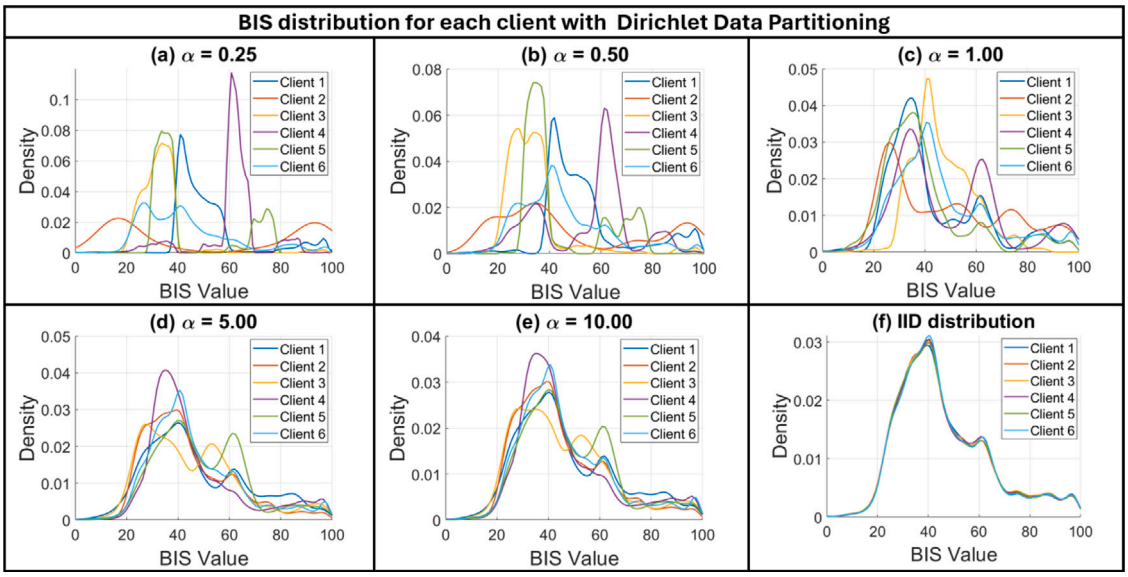


Fig. 11. BIS distribution for each client: (a) $\alpha = 0.25$, (b) $\alpha = 0.50$, (c) $\alpha = 1.00$, (d) $\alpha = 5.00$, (e) $\alpha = 10.00$, (f) IID distribution.

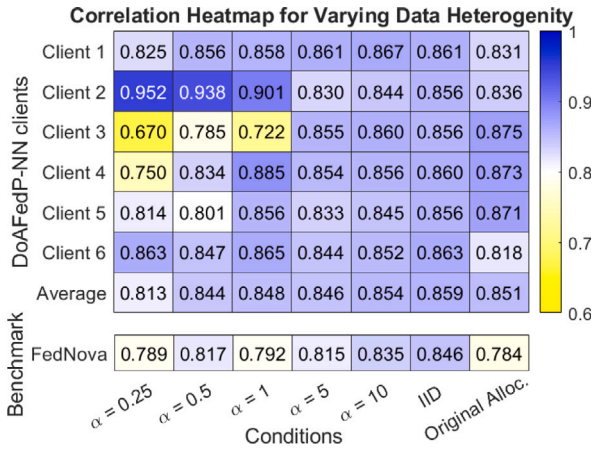


Fig. 12. Correlation of DoAFedP-NN and FedNova with the BIS index for varying levels of data heterogeneity. Includes comparison to the original allocation outlined in Section 4.2.

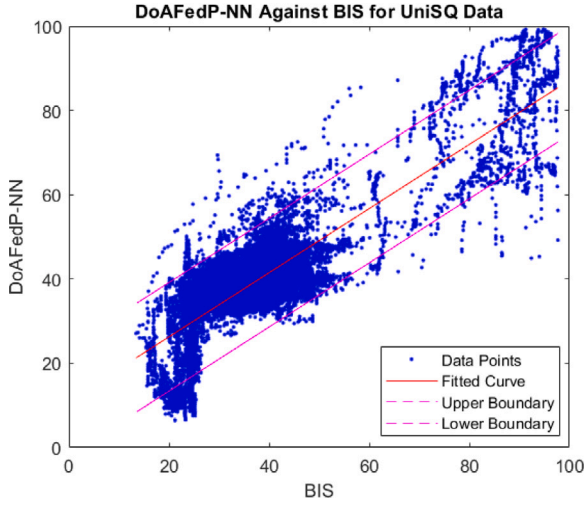


Fig. 13. Scatterplots with 95% confidence interval for estimate and actual DoA values for UniSQ database.

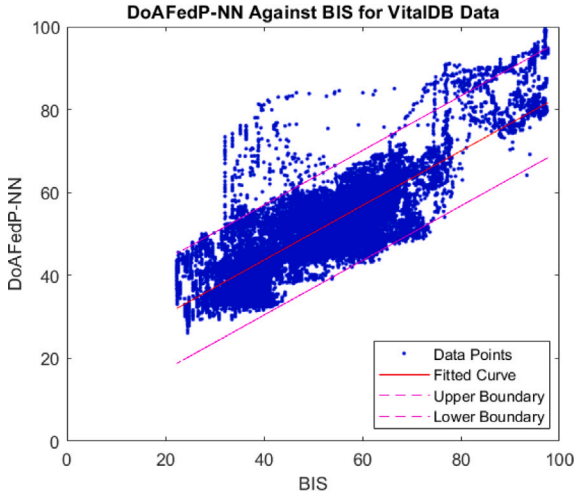


Fig. 14. Scatterplots with 95% confidence interval for estimate and actual DoA values for VitalDB database.

databases. The FL pseudo-aggregation method used in DoAFedP-NN effectively normalises and stratifies training data, reducing bias and ensuring that the model generalises well to unseen data (Table 4). The DoAFedP-NN index against the BIS index for each database is shown in Figs. 13 and 14, with 95% confidence intervals.

These findings highlight the robustness of the DoAFedP-NN algorithm, especially in real-world scenarios where data isolation between institutions is a significant obstacle. By providing more accurate and generalised predictions compared to locally or cross-trained models, DoAFedP-NN proves to be an effective solution for distributed medical data.

5.3. Ablation experiments

An ablation study was conducted on both databases to assess the relative contributions of model features and parameter configurations to the performance of the DoAFedP-NN model. The results presented in Table 5 are the average correlation and RMSE across all clients, obtained via 5-fold cross-validation. In each case, the global model employed a 3-layer MPL FFNN with 10 neurons each, as described in

Table 5
Performance metrics for ablation experiments of the DoAFedP-NN model. Results are average across 6 clients and 5-fold cross-validation.

Variation	Description	Correlation	RMSE
Without LZC feature	Baseline model without LZC	0.8374	10.0480
Without SODP feature	Baseline model without SODP	0.7702	11.9450
NN model variation	All clients use 2 layer NN	0.8400	10.2239
Linear regression model	LR modelling used in all clients	0.8166	10.4312
Support vector machine model	SVM modelling used in all clients	0.7968	10.9689
Heterogeneous model development	LR for UniSQ and NN for VDB clients	0.8283	10.4415
Heterogeneous model development	LR for VDB and NN for UniSQ clients	0.8410	9.7914
Baseline model	**DoAFedP-NN model**	**0.8475**	**10.1551**

Section 3. The analysis demonstrated that feature reduction negatively impacted model performance. Specifically, removing LZC and SODP led to a decline in model quality, with the correlation dropping to 0.8374 and 0.7702, respectively.

In addition, variations in model-building approaches were shown to affect performance. Reducing the number of hidden layers in the NN model design resulted in a small reduction correlation and increase in the RMSE observed. The experiments investigated the efficacy of fitting linear regression and SVM models which observed correlations of 0.8166 and 0.7968 respectively.

In this case, the NN model consistently outperforms all other modelling experiments. However, practical scenarios may benefit from a heterogeneous modelling approach. The capacity of this FL architecture with pseudo data generation to facilitate heterogeneous model development was illustrated by combining linear regression and NN modelling across different clients. In particular, applying linear regression to the VitalDB dataset and NN modelling to the UniSQ dataset resulted in only a marginal reduction in model quality—less than one per cent—compared to the uniform modelling approach of the standard DoAFedP-NN implementation.

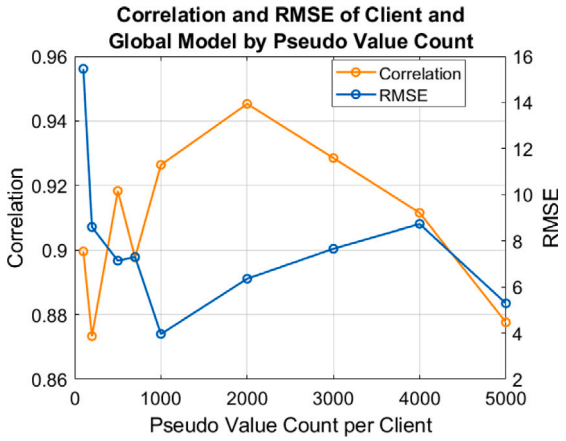


Fig. 15. Association of client and global models with variation in the number of pseudo data points generated.

5.4. Data optimisation

The quantity of pseudo data generated by client information is crucial for effective FL, as it captures the nuances of local models in developing a global model. Fig. 15 illustrate that increasing pseudo data points enhances the correlation and reduces the RMSE between local and global models up to a certain point. This occurs at 2000 pseudo data points per client. Beyond this threshold, additional pseudo data yields minimal improvement, indicating diminishing returns for computational expense. Identifying the optimal volume of pseudo data is essential for computational efficiency, achieving desired accuracy levels without incurring excessive computational costs. Balancing resource optimisation with model accuracy is key for the effective deployment of the DoAFedP-NN model, enhancing the practical viability of FL applications in diverse real-world settings.

5.5. Privacy protection measures

This FL approach preserves patient privacy by only sharing model parameters and aggregated statistics outside the local client [21]. This ensures that raw patient data remains on local clients. Sharing only the mean and standard deviation of feature distributions does not threaten privacy because these summary statistics are aggregated over multiple patients and do not reveal individual data. This minimal information is too general to reconstruct personal details, ensuring that patient confidentiality is maintained while still enhancing the collaborative model [49]. Patient privacy is strengthened through a one-shot FL approach, which restricts client communication to a single round. This approach significantly reduces the encryption overhead, improving both efficiency and security. Additionally, this method minimises opportunities for data attacks or manipulation during data exchanges, further enhancing the overall security of the FL process [32,50]. Further supporting privacy, real-world applications would incorporate advanced encryption methods, such as transport layer security (TLS), to secure data transmissions [51]. Additionally, techniques like differential privacy can be employed to further protect individual data points during model updates.

6. Conclusion

This study introduced the DoAFedP-NN framework, which leverages FL to develop robust EEG-based depth of anaesthesia (DoA) models while addressing data isolation issues common in medical applications. By employing a novel pseudo-data aggregation method, the framework enables one-shot FL, preserving data privacy while maintaining strong performance across diverse datasets. The results, based on a 5-fold cross-validation with a 6-client model, demonstrated that DoAFedP-NN consistently outperforms benchmark and state-of-the-art FL methods such as FedProx, FedNova and SCAFFOLD. The model achieved average correlations with the BIS index of 0.8472 for UniSQ and 0.8542 for VitalDB clients, significantly outperforming the traditional benchmarks. This performance underscores the framework's capacity to handle heterogeneous data and model situations, ensuring generalisation across clients. The ability of DoAFedP-NN to function effectively in non-IID cases marks it as a practical and scalable solution for real-world medical applications, where direct data sharing is often not feasible due to privacy regulations. By drastically reducing communication efficiency by approximately 98.9% compared to traditional FL, the DoAFedP-NN method effectively maintains privacy and model performance.

The proposed FL approach overcomes data-isolation issues in machine learning development for medical applications to facilitate the development of more robust analytical models. Future work should explore more advanced pseudo-data generation methods, test the framework on larger and more diverse datasets, and further investigate the impact of heterogeneous data and model architectures across clients. Further research to demonstrate the privacy and security benefits of the one-shot FL model in medical data analysis may be needed. The

findings of this study advocate for a paradigm shift in medical data analysis towards more collaborative and privacy-preserving methods to improve model quality and, in turn, patient outcomes.

The code used in this study has been released at https://github.com/ThomasSchmierer/DoAFedP.

CRediT authorship contribution statement

Thomas Schmierer: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Tianning Li:** Writing – review & editing, Software, Conceptualization. **Di Wu:** Writing – review & editing, Supervision, Conceptualization. **Yan Li:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Data availability

Data will be made available on request.

References

[1] M.T.V. Chan, T.L. Hedrick, T.D. Egan, P.S. García, S. Koch, P.L. Purdon, M.A. Ramsay, T.E. Miller, M.D. McEvoy, T.J. Gan, the Perioperative Quality Initiative, Workgroup, American society for enhanced recovery and perioperative quality initiative joint consensus statement on the role of neuromonitoring in perioperative outcomes: Electroencephalography, Anesth. Analg. 130 (5) (2020) 1278–1291, http://dx.doi.org/10.1213/ANE.0000000000004502.

[2] W.Z. Chew, W.Y. Teoh, N. Sivanesan, P.S. Loh, I. Shariffuddin, L.K. Ti, K.T. Ng, Bispectral index (BIS) monitoring and postoperative delirium in elderly patients undergoing surgery: A systematic review and meta-analysis with trial sequential analysis, J. Cardiothorac. Vasc. Anesth. 36 (12) (2022) 4449–4459, http://dx.doi.org/10.1053/j.jvca.2022.07.004.

[3] M. Berger, S.S. Eleswarpu, M. Cooter Wright, A.M. Ray, S.A. Wingfield, M.T. Heflin, S. Bengali, A.D. Udani, Developing a real-time electroencephalogram-guided anesthesia-management curriculum for educating residents: A single-center randomized controlled trial, Anesth. Analg. 134 (1) (2022) 159–170, http://dx.doi.org/10.1213/ane.0000000000005677.

[4] A. Shander, G.P. Lobel, D.M. Mathews, Brain monitoring and the depth of anesthesia: Another goldilocks dilemma, Anesth. Analg. 126 (2) (2018) 705–709, http://dx.doi.org/10.1213/ANE.0000000000002383.

[5] M.S. Avidan, G.A. Mashour, Prevention of intraoperative awareness with explicit recall: making sense of the evidence, Anesthesiology 118, http://dx.doi.org/10.1097/ALN.0b013e31826904a6.

[6] T. Schmierer, T. Li, Y. Li, Harnessing machine learning for EEG signal analysis: Innovations in depth of anaesthesia assessment, Artif. Intell. Med. 151 (2024) 102869, http://dx.doi.org/10.1016/j.artmed.2024.102869.

[7] T. Nguyen-Ky, P. Wen, Y. Li, Consciousness and depth of anesthesia assessment based on Bayesian analysis of EEG signals, IEEE Trans. Biomed. Eng. 60 (6) (2013) 1488–1498, http://dx.doi.org/10.1109/TBME.2012.2236649.

[8] T. Nguyen-Ky, P. Wen, Y. Li, R. Gray, An improved de-trended moving average method for accurately monitoring the depth of anaesthesia, IEEE Trans. Biomed. Eng. 57 (10) (2010).

[9] T. Nguyen-Ky, P. Wen, Y. Li, Theoretical basis for identification of different anesthetic states based on routinely recorded EEG during operation, Comput. Biol. Med. 39 (1) (2009) 40–45.

[10] Z. Hajat, N. Ahmad, J. Andrzejowski, The role and limitations of EEG-based depth of anaesthesia monitoring in theatres and intensive care, Anaesthesia 72 (2017) 38–47, http://dx.doi.org/10.1111/anae.13739.

[11] J. Xu, B.S. Glicksberg, C. Su, P. Walker, J. Bian, F. Wang, Federated learning for healthcare informatics, J. Heal. Informatics Res. 5 (2021) 1–19, http://dx.doi.org/10.1007/s41666-020-00082-4.

[12] X. Xu, H. Peng, M.Z.A. Bhuiyan, Z. Hao, L. Liu, L. Sun, L. He, Privacy-preserving federated depression detection from multisource mobile health data, IEEE Trans. Ind. Informatics 18 (7) (2022) 4788–4797, http://dx.doi.org/10.1109/TII.2021.3113708.

[13] N. Rieke, J. Hancox, W. Li, F. Milletari, H.R. Roth, S. Albarqouni, S. Bakas, M.N. Galtier, B.A. Landman, K. Maier-Hein, et al., The future of digital health with federated learning, NPJ Digit. Med. 3 (1) (2020) 119, http://dx.doi.org/10.1038/s41746-020-00323-1.

[14] A. Padthe, R. Ashtagi, S. Mohite, P. Gaikwad, R. Bidwe, H. Naveen, Harnessing federated learning for efficient analysis of large-scale healthcare image datasets in IoT-enabled healthcare systems, Int. J. Intell. Syst. Appl. Eng. 12 (2024) 253–263.

[15] C.S. Sindhu, N.P. Hegde, A framework to handle data heterogeneity contextual to medical big data, in: 2015 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC, 2015, pp. 1–7, http://dx.doi.org/10.1109/ICCIC.2015.7435779.

[16] L. Li, Y. Fan, M. Tse, K.-Y. Lin, A review of applications in federated learning, Comput. Ind. Eng. 149 (2020) 106854, http://dx.doi.org/10.1016/j.cie.2020.106854.

[17] M. Ghader, B. Farahani, Z. Rezvani, M. Shahsavari, M. Fazlali, Exploiting federated learning for EEG-based brain-computer interface system, in: 2023 IEEE International Conference on Omni-Layer Intelligent Systems, COINS 2023, 2023, http://dx.doi.org/10.1109/COINS57856.2023.10189325.

[18] L. Chen, W. Zhang, C. Dong, D. Zhao, X. Zeng, S. Qiao, Y. Zhu, C.W. Tan, FedTKD: A trustworthy heterogeneous federated learning based on adaptive knowledge distillation, Entropy 26 (1) (2024) http://dx.doi.org/10.3390/e26010096.

[19] D.C. Nguyen, Q.-V. Pham, P.N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. Dobre, W.-J. Hwang, Federated learning for smart healthcare: A survey, ACM Comput. Surv. (Csur) 55 (3) (2022) 1–37, http://dx.doi.org/10.1145/3501296.

[20] W. Zhang, Z. Bao, Y. Liu, L. Xu, Q. Lu, H. Ning, X. Wang, S. Yang, F.-Y. Wang, Z. Li, Cfsl: A credible federated self-learning framework, IEEE Internet Things J. 10 (24) (2023) 21349–21362, http://dx.doi.org/10.1109/JIOT.2023.3286398.

[21] Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, B. He, A survey on federated learning systems: Vision, hype and reality for data privacy and protection, IEEE Trans. Knowl. Data Eng. 35 (4) (2023) 3347–3366, http://dx.doi.org/10.1109/TKDE.2021.3124599.

[22] C.A. Fidas, D. Lyras, A review of EEG-based user authentication: Trends and future research directions, IEEE Access 11 (2023) 22917–22934, http://dx.doi.org/10.1109/ACCESS.2023.3253026.

[23] W. Ding, M. Abdel-Basset, H. Hawash, S. Abdel-Razek, C. Liu, Fed-ESD: Federated learning for efficient epileptic seizure detection in the fog-assisted internet of medical things, Inform. Sci. 630 (2023) 403–419, http://dx.doi.org/10.1016/j.ins.2023.02.052.

[24] S. Bn, S. Abdullah, Privacy Sensitive Speech Analysis Using Federated Learning to Assess Depression, vol. 2022-, IEEE International Conference on Acoustics, Speech and Signal Processing, 2022, pp. 6272–6276, http://dx.doi.org/10.1109/ICASSP43922.2022.9746827.

[25] T.E.T. Djaidja, B. Brik, A. Boualouache, S.M. Senouci, Y. Ghamri-Doudane, Federated learning for 5G and beyond, a blessing and a curse-an experimental study on intrusion detection systems, Comput. Secur. 139 (2024) 103707.

[26] U. Milasheuski, L. Barbieri, B.C. Tedeschini, M. Nicoli, S. Savazzi, On the impact of data heterogeneity in federated learning environments with application to healthcare networks, in: 2024 IEEE Conference on Artificial Intelligence, CAI, 2024, pp. 1017–1023, http://dx.doi.org/10.1109/CAI59869.2024.00185.

[27] C. Ju, D. Gao, R. Mane, B. Tan, Y. Liu, C. Guan, Federated Transfer Learning for EEG Signal Classification, IEEE, 2020, pp. 3040–3045, http://dx.doi.org/10.1109/EMBC44109.2020.9175344.

[28] R. Hu, Y. Guo, H. Li, Q. Pei, Y. Gong, Personalized federated learning with differential privacy, IEEE Internet Things J. 7 (10) (2020) 9530–9539, http://dx.doi.org/10.1109/JIOT.2020.2991416.

[29] D. Jiang, C. Shan, Z. Zhang, Federated learning algorithm based on knowledge distillation, in: 2020 International Conference on Artificial Intelligence and Computer Engineering, ICAICE, IEEE, 2020, pp. 163–167, http://dx.doi.org/10.1109/ICAICE51518.2020.00038.

[30] Y. Zhang, Z. Li, X. Han, S. Ding, J. Li, J. Wang, S. Ying, J. Shi, Pseudo-data based self-supervised federated learning for classification of histopathological images, IEEE Trans. Med. Imaging (2023) http://dx.doi.org/10.1109/TMI.2023.3323540.

[31] M.M. Rahman, S. Purushotham, FedPseudo: Privacy-preserving pseudo value-based deep learning models for federated survival analysis, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 1999–2009, http://dx.doi.org/10.1145/3580305.3599348.

[32] S. Su, B. Li, X. Xue, One-shot federated learning without server-side training, Neural Netw. 164 (2023) 203–215, http://dx.doi.org/10.1016/j.neunet.2023.04.035.

[33] M. Kang, P. Chikontwe, S. Kim, K.H. Jin, E. Adeli, K.M. Pohl, S.H. Park, One-shot federated learning on medical data using knowledge distillation with image synthesis and client model adaptation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 521–531.
[34] T. Nguyen-Ky, P. Wen, Y. Li, R. Gray, Measuring and reflecting depth of anesthesia using wavelet and power spectral density, IEEE Trans. Inf. Technol. Biomed. 15 (4) (2011) 630–639, http://dx.doi.org/10.1109/TITB.2011.2155081.
[35] R.B. Pachori, S. Patidar, Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions, Comput. Methods Programs Biomed. 113 (2) (2014) 494–502, http://dx.doi.org/10.1016/j.cmpb.2013.11.014.
[36] T. Schmierer, T. Li, Y. Li, A novel empirical wavelet SODP and spectral entropy based index for assessing the depth of anaesthesia, Heal. Inf. Sci. Syst. 10 (1) (2022) 10, http://dx.doi.org/10.1007/s13755-022-00178-8.
[37] J.S. Ra, T. Li, Y. Li, A novel spectral entropy-based index for assessing the depth of anaesthesia, Brain Inf. 8 (1) (2021) 10, http://dx.doi.org/10.1186/s40708-021-00130-8.
[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[39] H.C. Lee, H.G. Ryu, E.J. Chung, C.W. Jung, Prediction of bispectral index during target-controlled infusion of propofol and remifentanil, Anesthesiology 128 (3) (2018) 492–501, http://dx.doi.org/10.1097/ALN.0000000000001892.
[40] Y.F. Chen, S.Z. Fan, M.F. Abbod, J.S. Shieh, M. Zhang, Nonlinear analysis of electroencephalogram variability as a measure of the depth of anesthesia, IEEE Trans. Instrum. Meas. 71 (2022) http://dx.doi.org/10.1109/TIM.2022.3167793.
[41] R.B. Pachori, S. Patidar, Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions, Comput. Methods Programs Biomed. 113 (2) (2014) 494–502, http://dx.doi.org/10.1016/j.cmpb.2013.11.014.
[42] T. Li, P. Wen, Depth of anaesthesia assessment using interval second-order difference plot and permutation entropy techniques, IET Signal Process. 11 (2) (2017) 221–227, http://dx.doi.org/10.1049/iet-spr.2015.0114.
[43] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiology- Hear. Circ. Physiol. 278 (6) (2000) H2039–H2049, http://dx.doi.org/10.1152/ajpheart.2000.278.6.H2039.
[44] X. Jie, R. Cao, L. Li, Emotion recognition based on the sample entropy of EEG, Bio- Med. Mater. Eng. 24 (1) (2014) 1185–1192, http://dx.doi.org/10.3233/BME-130919.
[45] Y. Pan, J. Chen, X. Li, Spectral entropy: A complementary index for rolling element bearing performance degradation assessment, Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci. 223 (5) (2009) 1223–1231, http://dx.doi.org/10.1243/09544062JMES1224.
[46] A. Shalbaf, M. Saffar, J.W. Sleigh, R. Shalbaf, Monitoring the depth of anesthesia using a new adaptive neurofuzzy system, IEEE J. Biomed. Heal. Inf. 22 (3) (2018) 671–677, http://dx.doi.org/10.1109/jbhi.2017.2709841.
[47] H.-C. Lee, Y. Park, S.B. Yoon, S.M. Yang, D. Park, C.-W. Jung, VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients, Sci. Data 9 (1) (2022) 279, http://dx.doi.org/10.1038/s41597-022-01411-5.
[48] W.G. Van Panhuis, P. Paul, C. Emerson, J. Grefenstette, R. Wilder, A.J. Herbst, D. Heymann, D.S. Burke, A systematic review of barriers to data sharing in public health, BMC Public Health 14 (2014) 1–9, http://dx.doi.org/10.1186/1471-2458-14-1144.
[49] P. Kairouz, H.B. McMahan, B. Avent, A. Bellet, M. Bennis, A.N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al., Advances and open problems in federated learning, Found. Trends Mach. Learn. 14 (1–2) (2021) 1–210, http://dx.doi.org/10.1561/2200000083.
[50] N. Wang, Y. Deng, W. Feng, S. Fan, J. Yin, S.-K. Ng, One-shot sequential federated learning for non-IID data by enhancing local model diversity, in: Proceedings of the 32nd ACM International Conference on Multimedia, 2024, pp. 5201–5210, http://dx.doi.org/10.1145/3664647.3681054.
[51] O. Zeidler, J. Sturm, D. Fraunholz, W. Kellerer, Performance evaluation of transport layer security in the 5G core control plane, in: Proceedings of the 17th ACM Conference on Security and Privacy in Wireless and Mobile Networks, WiSec '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 78–88, http://dx.doi.org/10.1145/3643833.3656140.

Thomas Schmierer holds a Bachelor of Science and a Bachelor of Commerce from Monash University (2013), a Graduate Diploma in Education from Queensland University of Technology (2014), and a Master of Science in Applied Data Science from the University of Southern Queensland (2021). Currently a Ph.D. candidate at the University of Southern Queensland, Thomas' research focuses on machine learning algorithms, signal processing, and electroencephalography (EEG) for medical monitoring applications, with a specific interest in depth of anaesthesia.

Tianning Li received the B.ISM. degree from Nanjing University, Nanjing, China, in 2009, the M.AccFin. degree from The University of Adelaide, Adelaide, SA, Australia, in 2012, and the Ph.D. degree from the University of Southern Queensland (UniSQ), Toowoomba, QLD, Australia, in 2015. He is currently a Computing Lecturer and the First Year Experience Lead with the School of Mathematics, Physics and Computing, UniSQ. His research interests include machine learning algorithms, signal processing, and electroencephalography (EEG) research.

Di Wu received his B.Sc. degree in Information Technology (Honours) as a first class honours student and M.Sc. degree in Information Technology (Professional) from Deakin University, Australia in 2013 and 2012, respectively. He received Ph.D. degree from the School of Computer Science & Australian Artificial Intelligence Institute(AAII), University Technology Sydney, Sydney, Australia in 2020. He is a Senior Lecturer at the School of Mathematics, Physics, and Computing, the University of Southern Queensland (UniSQ). Prior to that, he was a Lecturer at the same School. Prior to that, he was a Researcher Fellow at the Australian Institute for Machine Learning (AIML) and School of Computer Science, University of Adelaide, Adelaide, Australia. Previous to this, he was an Associate Research Fellow, Artificial Intelligence at Deakin Blockchain Innovation Lab, School of Information Technology, Deakin University, Melbourne, Australia, and worked as a Postdoc Fellow at the School of Computer Science, University of Technology Sydney (UTS), Sydney, Australia. He has more than 10 years of experience in research & development and academia. He has substantial industry experience in large project management, software development, and large system maintenance experience while working on various projects at China Telecom (Global 500), Shanghai. His research area focuses on applying federated learning, AI security and privacy, and trustworthy AI. He has published papers in high-quality refereed books, conferences, and journals, including top-tier venues such as ICLR, IJCAI, TKDE etc. He also serves as an associate editor in NLPJ and a reviewer for many high-quality academic conferences and journals, such as ICLR, SIGKDD, ACM MM, CoRL, TMC, TNNLS, TETCI, PR, etc.

Yan Li is a Professor in Computer Science with the School of Mathematics, Physics and Computing at the University of Southern Queensland, Australia. Her research interests include artificial intelligence, big data analytics, signal and image processing, biomedical engineering, and computer networking technologies and security.

Prof Yan Li has published 258 high quality publications, supervised dozens of PhD students to completion, and obtained significant research grants from Australia government, industry and through international collaborations.