Decentralized Federated Learning for Epileptic Seizures Detection in Low-Power Wearable Systems

Saleh Baghersalimi, Tomas Teijeiro, Amir Aminifar *Senior IEEE Member*,
David Atienza *IEEE Fellow*

***Abstract*—In healthcare, data privacy of patients regulations prohibits data from being moved outside the hospital, preventing international medical datasets from being centralized for AI training. Federated learning (FL) is a data privacy-focused method that trains a global model by aggregating local models from hospitals. Existing FL techniques adopt a central server-based network topology, where the server assembles the local models trained in each hospital to create a global model. However, the server could be a point of failure, and models trained in FL usually have worse performance than those trained in the centralized learning manner when the patient's data are not independent and identically distributed (Non-IID) in the hospitals. This paper presents a decentralized FL framework, including training with adaptive ensemble learning and a deployment phase using knowledge distillation. The adaptive ensemble learning step in the training phase leads to the acquisition of a specific model for each hospital that is the optimal combination of local models and models from other available hospitals. This step solves the non-IID challenges in each hospital. The deployment phase adjusts the model's complexity to meet the resource constraints of wearable systems. We evaluated the performance of our approach on edge computing platforms using EPILEPSIAE and TUSZ databases, which are public epilepsy datasets.**

***Index Terms*—Federated Learning, Wearable systems, Deep learning, Electrocardiogram, Epilepsy, Knowledge distillation, Multi-biosignal processing, Seizure detection.**

I. Introduction

Epilepsy is a widespread neurological condition that affects approximately 65 million individuals in all age groups globally [1]. This disorder presents itself in various forms, ranging in severity from mild to severe. There are not only different types of seizures associated with epilepsy, but the consequences and effects of this condition also vary widely. As such, individuals with epilepsy and their families often face diverse challenges and experiences depending on the specifics of the condition.

Manuscript received 29 December 2022; revised 9 August 2023; accepted 25 September 2023. This work has been partially supported by the PEDESITE Swiss NSF Sinergia project (GA No. CRSII5-193813 / 1), the RESoRT project (GA No. REG-19-019) from the Botnar Foundation, and the WASP Program funded by the Knut and Alice Wallenberg Foundation, the grant RYC2021-032853-I funded by MCIN/AEI/ 10.13039/501100011033 and by European Union NextGenerationEU/PRTR.

Saleh Baghersalimi and David Atienza are with the Institute of Electrical Engineering, Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland, (email: saleh.baghersalimi@epfl.ch; david.atienza@epfl.ch).

Tomas Teijeiro is with the Basque Center for Applied Mathematics (BCAM), Bilbao, Spain (email: tteijeiro@bcamath.org)

Amir Aminifar is with the Department of Electrical and Information Technology at Lund University, Sweden, (email: amir.aminifar@eit.lth.se)

Among the numerous challenges faced by those living with epilepsy—such as access to quality healthcare, information and coordination of services, and societal stigma—a potentially fatal consequence known as SUDEP (Sudden Unexpected Death in Epilepsy) looms [2]. SUDEP, which generally occurs during or after a seizure, accounts for unanticipated deaths within the epilepsy community [3]. Although rare, prevention of deaths caused by SUDEP is possible through emergency alerts to caregivers and family [4]. Implementing real-time seizure detection through continuous EEG or ECG monitoring represents a promising avenue to mitigate the impact of seizures and enhance the quality of life.

The current frontier in machine learning for seizure detection is deep neural networks (DNNs) [5, 6, 7, 8]. DNNs can automatically discern high-level features from biomedical signals, distinguishing seizures without preliminary feature engineering. To construct an efficient epileptic seizure detection system, a large-scale dataset is required. However, transmitting raw patient data to a central server could contravene privacy laws such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

Federated Learning (FL) has emerged as a robust method for handling extensive distributed data without the need to transfer raw data to a centralized location, which could raise major privacy concerns [9, 10, 11]. Traditional FL involves a centralized approach, with a central server aggregating trained model parameters from various devices without transmitting raw patient data. This centralization, however, introduces risks, including server failures or delays due to server overload or an increasing number of devices [12].

Delving deeper into FL algorithms, Federated Averaging (FedAvg) stands out as one of the notable strategies. In FedAvg, devices locally update a global model before the updated parameters are transmitted to the central server [9, 13, 14, 15]. However, FedAvg has its own set of challenges. It struggles with non-independent and not identically distributed (non-IID) data distribution commonly found in different devices, which has been shown to hinder the FL process's efficiency [16, 17, 18]. Centralized FL is further complicated by varying computational resources across devices, leading to unequal local training times.

In this paper, we address the crucial concern of failure points in centralized Federated Learning (FL) systems, particularly in the context of healthcare settings like hospitals, and propose a novel decentralized FL framework that eliminates the need for a central server. This decentralized approach is illustrated in Figure 1, focusing on the scenario involving various hospitals, each dealing with patients displaying diverse seizure types, ages, or genders.

The core of our work lies in the following main components:

- **Real-time Decentralized FL Framework:** We develop a serverless, real-time decentralized Federated Learning (FL) framework for seizure detection in an interconnected network of hospitals. This innovative design facilitates a cooperative and synergistic learning environment among hospitals, where they can train complex teacher Deep Neural Networks (DNNs) using both EEG and ECG signals, and lighter student DNNs using only ECG signals. This approach ensures collaborative learning without sharing sensitive patients' data.
- **Adaptive Ensembling for Non-Identical Data Distribution in Decentralized FL:** In our study, we conducted a comprehensive examination of non-identical patients' data distributions across individual hospitals, highlighting the significant challenges that the non-Independently and Identically Distributed (non-IID) nature presents to learning accuracy. We identify that the performance of the decentralized FL system is deeply influenced by the underlying distribution of patients' data. To address these challenges, our framework introduces an adaptive ensembling phase. This phase allows for the learning of a specific teacher DNN for each hospital, combining the strength of local models with models from other hospitals. This approach not only mitigates the challenges posed by non-IID data distribution but also substantiates the improvements in seizure detection. We demonstrate the improvement in seizure detection on the EPILEPSIAE and TUSZ public datasets.
- **Wearable Device Implementation Study:** Our research extends to evaluating the feasibility of deploying the seizure detection model on wearable IoT devices with limited resources. We create a knowledge distillation-based [19] (teacher-student) approach to design high-precision, low-power devices capable of long-term patient monitoring after hospital discharge, relying solely on ECG signal input.

By weaving together these interconnected components, we provide a comprehensive, decentralized solution to the centralized FL's shortcomings. The practical applications of our work have profound implications in medical scenarios, emphasizing collaboration among healthcare facilities while ensuring patient privacy, dealing with inherent data disparities, and adapting to emerging technological platforms. Our model, grounded in real-world needs and validated through rigorous experimentation, serves as a blueprint for future advances in decentralized FL in healthcare.

The rest of this article is organized as follows. Section II describes the problem of interest and discusses related work. Section III presents the training of our proposed formulation for a decentralized FL framework for epileptic seizure detection. Section IV describes seizure detection model deployment to low-power wearable IoT systems for the patient monitoring stage. In Section V, we discuss the experimental setup. Section VI analyzes our decentralized FL's computational and energy consumption characteristics and studies the benefit of the proposed ensembling stage. Finally, in Section VII, we summarize the main conclusions of this work.

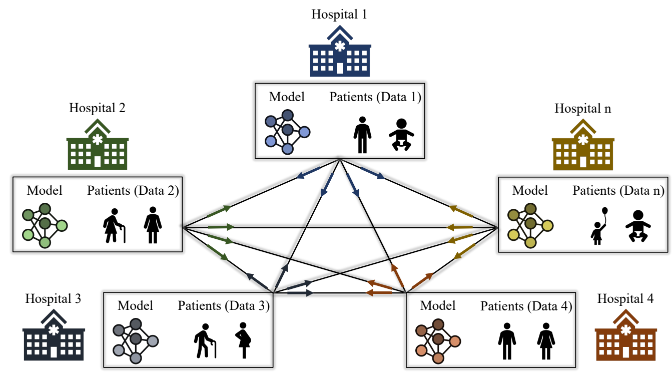


Fig. 1: Structure of our proposed decentralized FL. Each hospital is responsible for different patients with different types of seizures, gender, and age. The hospitals collaborate with each other by exchanging model updates.

II. Related Work

A. Seizure detection using EEG signal

The benchmark for non-invasive seizure detection is EEG monitoring [20], which has been used for decades in highly specialized and costly hospital environments. Several techniques, such as wavelet transform [21, 22], entropies [23], Hilbert marginal spectrum [24], Hilbert Huang Transform [25], fusion features [26], and tunable Q-factor [27] are typically used to extract information from the EEG signal. Deep learning has recently received considerable attention for specific applications, such as epileptic seizure detection and prediction. For instance, [28] proposed a method for the epileptic focus localization problem by merging an autoencoder and a K-means algorithm. In [29], they presented a pseudo-prospective seizure prediction method from an EEG signal. The authors in [30] performed epileptic seizure detection using EEG signals with an adaptive implementation of CNNs. In [31], the authors presented a framework using DNN to capture brain abnormalities based on multi-channel scalp EEG signals. In [32], the authors proposed a hybrid bilinear DNN using surface EEG for epilepsy classification diagnosis. However, these studies demand patients to wear a cap to acquire EEG signals, creating social stigma and discomfort [4]. Furthermore, EEG recordings are highly susceptible to artifacts created by patients' movements.

B. Seizure detection using ECG signal

In addition to abnormal brain activities during epileptic seizures, other biosignals can get affected. In [33, 34, 35, 36, 37], they show that epileptic seizures are associated with increased heart rate and cardiovascular alterations. In [38], the authors used ECG and PPG signals and reached a sensitivity of 70% and a corresponding false-alarm rate of 2.11 per hour. In [39], they proposed a technique that combines time and frequency-domain features of heart-rate variability. It was limited to the specific case of newborns and reached a sensitivity

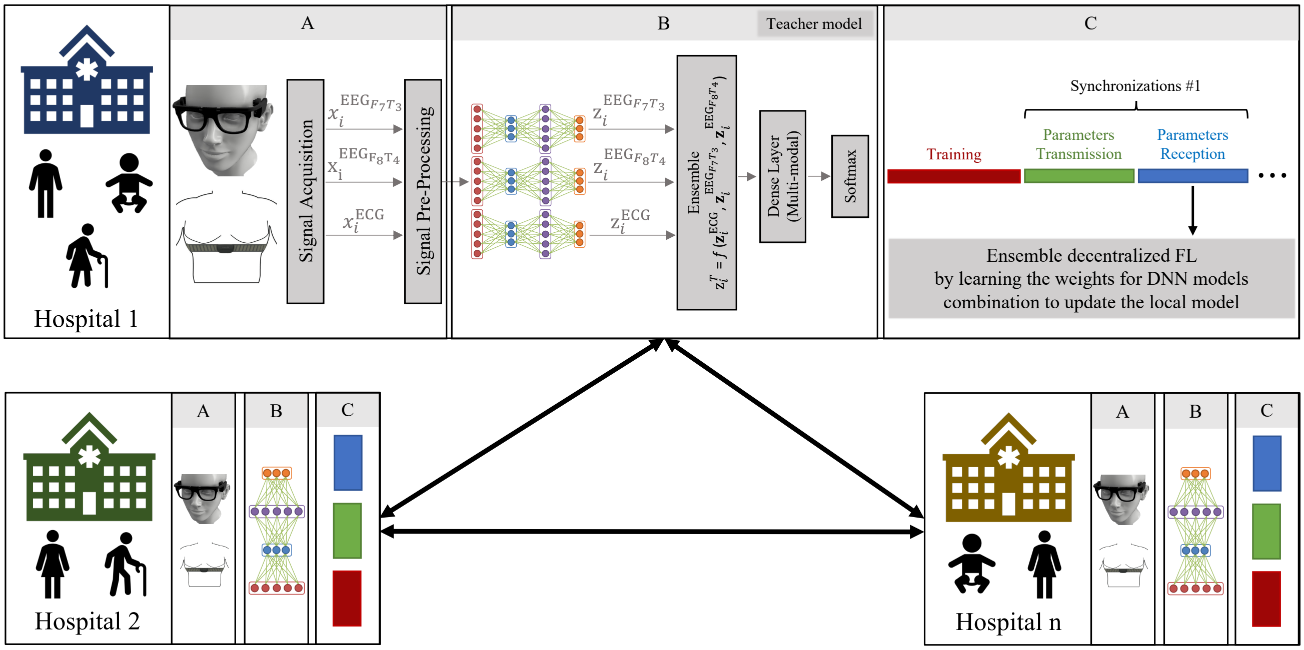


Fig. 2: Training phase: The overall flow of the proposed decentralized FL system with n hospitals. Each hospital trains an identical teacher model for epileptic seizure detection on its data locally and independently. The teacher model is composed of three DNNs to extract valuable information from ECG, $\text{EEG}_{F_7T_3}$ and $\text{EEG}_{F_8T_4}$ signals that are acquired from a chest strap and two electrode pairs F_7T_3 and F_8T_4. Each hospital transmits the current weights of the teacher model to other hospitals during the training process and after a specific number of iterations. Then, each hospital learns the appropriate combination of local and global weights received from other hospitals.

of 60% and specificity of 60%. In [40, 41], they proposed a multi-parametric machine-learning approach to detect epileptic seizures by examining the cardiac and respiratory responses to seizures in the ECG signal. However, these studies did not attempt to find a personalized solution for each patient to achieve satisfactory and comparable detection accuracy to seizure detection by EEG signals.

C. Seizure detection using multi-biosignal combination

Associating other biosignals with EEG can better detect different types of seizures. In [42], the authors introduced a combination of electrodermal activity and accelerometer signals. In [43], the authors have used EEG, Electromyography (EMG), and ECG signals and have shown an improvement in sensitivity compared to using each sensor separately. In [44], the authors performed seizure detection using an SVM model on multi-channel EEG and single-channel ECG individually and then fused them into one final decision. In [32], the authors employed CNN in a combination of EEG, ECG, and respiration. However, it is impossible to execute these techniques on IoT wearable devices due to computation and memory constraints.

D. Model compression using knowledge distillation

In 2015, Hinton et al. [45] introduced the concept of knowledge distillation in neural networks. This involves training a single model (the student) using the continuous outputs of a more complex model or ensemble (the teacher), rather than relying solely on the discrete labels of the dataset. Works such as [46, 47] have applied knowledge distillation to image datasets, demonstrating its potential as a robust regularization technique. These studies showed that a student DNN trained using knowledge distillation often generalizes better than one trained directly with dataset labels. Yet, these referenced systems did not merge data from multiple types or sources.

It's pertinent to differentiate between "Multi-modal Data" and "Multi-biosignal Combination". Multi-modal data merges distinct modalities like text, images, and sound, whereas multi-biosignal combination integrates various biological signals like EEG and ECG. In our study, we employ a Multi-biosignal Combination approach. Specifically, the teacher DNN is trained by integrating both EEG and ECG signals, enhancing its depth and precision, and distinguishing our work from purely multi-modal systems.

E. Patient privacy preservation using federated learning

In the early days of deep learning, the procedure for model training required distinct steps. It began with the assembly of significant samples to form a comprehensive and balanced dataset. This was followed by the creation of a singular, centralized model that analyzed these samples and produced a result in line with the ground truth. Post-analysis, adjustments could be introduced, such as the procurement of extra samples, dataset refinement, and hyperparameter tuning. However, the accumulation of data for a specific task can be quite costly, and raises privacy concerns, along with related legal issues, particularly in the healthcare field.

Centralized Federated Learning (FL) offers a training process for a global model by aggregating multiple local models. In a centralized FL network, a server mediates model parameters between clients. The server consolidates the local model parameters through an averaging algorithm, resulting in global model parameters, as outlined by FedAvg [9].

Researchers have implemented various approaches within centralized FL. For instance, [48] personalized FedAvg for specific ECG features for epileptic seizure detection. Similarly, [49] designed an unsupervised gradient aggregation method to tackle drift and convergence variability. Additionally, the work presented in [50] introduced a privacy-preserving federated learning framework, Fed-ESD, which employs fog nodes as local aggregators and a spatiotemporal transformer network to enable sharing of location-based EEG signals for comparable IoMT applications. However, servers can be vulnerable and face high bandwidth and energy costs [51, 52], or diminish learning performance [12]. While centralized methods have shown promising results, there are inherent challenges that give rise to the consideration of alternative architectures, such as Decentralized Federated Learning (DFL).

DFL provides a novel approach to training global models by facilitating direct interaction between clients, eliminating the need for a central server. This architecture enhances privacy protection by minimizing the exposure of data and model information to a central entity. By allowing clients to interact directly, DFL also improves efficiency and robustness, reducing overall network communication and computation costs, and providing resilience against single-point failures.

Despite these advantages, DFL introduces unique challenges, such as coordination across a decentralized network, maintaining consistency, and ensuring data integrity and trust. Innovative solutions like merging local client models using transfer learning have been suggested by [53], while [54] proposed assigning weights to each model from other clients, creating a weighted model combination.

DFL has also found applications in various fields, including healthcare and edge computing, and its exploration remains relatively uncharted, showing promising avenues for innovations, particularly in environments where centralized coordination is infeasible or undesirable.

Both centralized and decentralized Federated Learning (FL) strategies tackle the essential problem of model learning in the face of varied or non-Independently and Identically Distributed (non-IID) client data distributions. While many studies, including [55, 56, 57, 58, 59, 60], have harnessed DNN training using FL across a multitude of applications, the intricate challenge of model learning with inconsistent or non-IID client data distributions still stands as a comparatively uncharted area in both centralized and decentralized environments.

In our method, each client explores the layers from other client models, identifying the optimal personalized layer combination for their own model.

III. Decentralized Federated Learning Framework

We divide our framework into two phases: decentralized federated learning framework and seizure detection DNN deployment to low-power wearable IoT systems. This section describes the first phase, decentralized FL framework shown in Figure 2, consisting of three parts:

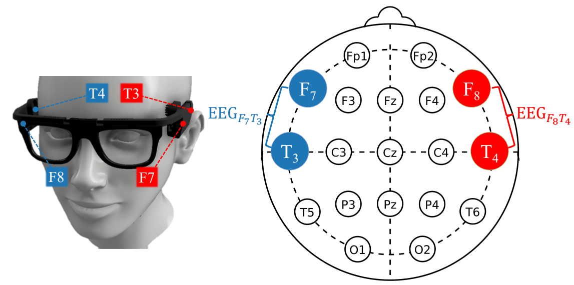


Fig. 3: Electrode locations of F_7T_3 and F_8T_4 for EEG monitoring using the e-Glass wearable system for epileptic seizure detection [62].

(A) Signal acquisition and pre-processing.
(B) Epileptic seizure detection DNN.
(C) Proposed decentralized federated learning by weighting the contributions of each hospital to create a more accurate DNN.

A. *Signal Acquisition and Pre-Processing*

In our proposed decentralized federated learning system, each hospital includes a teacher and a student network. The teacher network is formed using EEG and ECG data, whereas the student network requires only ECG signal [61]. The EEG signals for the teacher network are gathered from two channels (F_7T_3 and F_8T_4) via four electrodes, as depicted in Figure 3. These electrodes were specifically selected to capture EEG signals via e-Glass, a wearable system designed for real-time scenarios that relies on these four EEG electrodes [62]. e-Glass is intended to be an unobtrusive system that helps patients evade the social stigma tied to wearing traditional EEG head caps. In section VI-B, we demonstrated the feasibility of epileptic seizure detection in real-time using either the teacher or student network with edge AI systems.

The decision to utilize the F_7T_3 and F_8T_4 EEG channels for our teacher DNN-based epileptic seizure detection system was influenced by multiple factors, which accommodate the seizure type variability among different patients:

1) Seizure Detection: F_7T_3 and F_8T_4 channels are located in the frontotemporal regions of the brain, known for their involvement in seizure activity. They're useful in detecting various types of seizures.
2) Signal Quality: These channels offer cleaner signals with less noise and interference from muscle activity or eye movements, enhancing detection accuracy.
3) Computational Efficiency: Limiting to two EEG channels reduces computational load, making real-time seizure detection feasible, even on wearable devices with limited power.
4) Social Acceptance: Using F_7T_3 and F_8T_4 channels in wearables like e-Glass offers a discreet alternative to traditional EEG caps, reducing associated social stigma.
5) Generalizability: Despite seizure type variability, the choice of these channels and our DNN design ensure reliable performance across a broad patient range.

The teacher DNN requires considerable data for the training process. The data on actual epileptic seizures for a particular

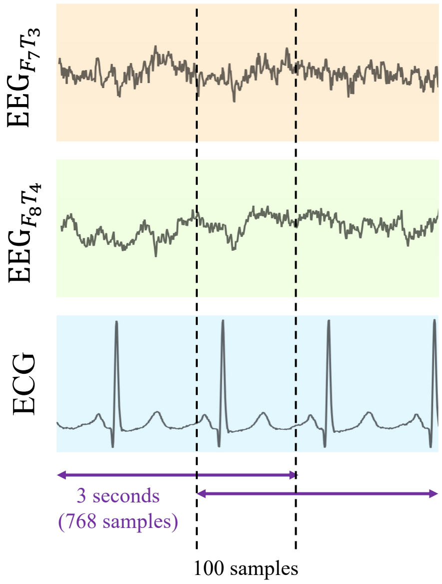


Fig. 4: Segmentation of ECG, EEG$_{F_7T_3}$ and EEG$_{F_8T_4}$ signals using slots of 3-seconds with 100 samples overlap. In the decentralized FL framework, ECG, EEG$_{F_7T_3}$ and EEG$_{F_8T_4}$ signals are synchronized to train the teacher model for each hospital. In other words, ECG, EEG$_{F_7T_3}$ and EEG$_{F_8T_4}$ signals are acquired and measured in parallel.

patient is generally limited. Thus, data augmentation techniques have been used in recent works [63, 64] to increase the amount and diversity of epileptic data. This work exploits data augmentation by segmenting and overlapping synchronized ECG and EEG signals. Since interpreting the QRS complexes and obtaining their characteristics is one of the essential parts of ECG signal processing, we consider 3-second slots to ensure a minimum of two QRS complexes. These slots are obtained by sliding a fixed-length window, with 100 samples overlapping, through the entire signal. Figure 4 shows how ECG and EEG signals are segmented in our procedure.

Pre-processing techniques are required and have been utilized in different applications to train DNN models effectively and efficiently [65]. Thus, we propose a simple method for pre-processing each segment after signal segmentation. Pre-processing steps of an ECG and EEG segments are presented in Algorithm 1 and 2.

The standardization is done on each 3-second segment separately to reduce the number of false positives mostly related to artifacts causing sudden amplitude variations. Consequently, a potentially significant difference in the magnitude of one segment does not cause a high degradation in other segments.

Algorithm 1 Pre-processing of ECG Segment

Require: ECG segment x
Ensure: Standardized ECG segment $x_{\text{standardized}}$

Step 1: Apply 10th-order Low-pass Butterworth Filter
$x_{\text{filtered}} \leftarrow \text{ButterworthFilter}(x, \text{order} = 10, \text{cutoff} = 50)$

Step 2: Perform Linear Detrending
Fit a linear model $y = ax + b$ to x_{filtered}.
Subtract the fit from the initial data: $x_{\text{detrended}} \leftarrow x_{\text{filtered}} - (ax + b)$.

Step 3: Standardize the Detrended ECG Segment
Compute the mean μ and standard deviation σ of $x_{\text{detrended}}$.
$x_{\text{standardized}} \leftarrow \frac{x_{\text{detrended}} - \mu}{\sigma}$

Algorithm 2 Pre-processing of EEG Segment

Require: EEG segment x
Ensure: Standardized EEG segment $x_{\text{standardized}}$

Step 1: Apply 10th-order Low-pass Butterworth Filter
$x_{\text{filtered}} \leftarrow \text{ButterworthFilter}(x, \text{order} = 10, \text{cutoff} = 50)$

Step 2: Standardize the Filtered EEG Segment
Compute the mean μ and standard deviation σ of x_{filtered}.
$x_{\text{standardized}} \leftarrow \frac{x_{\text{filtered}} - \mu}{\sigma}$

B. Epileptic Seizure Detection DNN

This section describes the teacher DNN used in this work for each hospital for seizure detection. We considered our previous work's DNN named Res1DCNN [48] shown in Figure 5. As shown in Figure 2, the teacher network takes synchronized ECG and 2-channel EEG segments as inputs and trains three analogous feature extractors of Res1DCNN models for every input. This process is an ensemble learning technique where various DNN models are combined and trained to solve the same problem [66]. We could use all EEG channels; however, since we studied and experimented with a wearable setup, we considered only two EEG channels where the signals can be acquired from e-Glass [62] shown in Figure 3. In this network, the inputs are EEG and ECG segments $\{x_i^{\text{ECG}}, x_i^{\text{EEG1}}, x_i^{\text{EEG2}}\}$. We extract features from the inputs by passing each of them through Res1DCNN's feature extractor to obtain $\{\mathbf{z}_i^{\text{ECG}}, \mathbf{z}_i^{\text{EEG1}}, \mathbf{z}_i^{\text{EEG2}}\}$. We merge these feature maps into a single $\mathbf{z}_i^T \in \mathbb{R}^L$ using a linear combination of the features. More formally, $\mathbf{z}_i^T = f(\mathbf{z}_i^{\text{ECG}}, \mathbf{z}_i^{\text{EEG1}}, \mathbf{z}_i^{\text{EEG2}}; \theta)$, where θ is the trainable weight for the linear combination. When $\mathbf{z}_i^T$ is obtained, we train a simple fully-connected layer that predicts the output $\hat{y}_i$ from $\mathbf{z}_i^T$. Finally, a softmax layer outputs the predicted value. We present the algorithm for the teacher DNN in Algorithm 3.

C. Personalized Decentralized Federated Learning

A significant amount of data are required to train DNNs and fit the parameters (often in the order of millions). Therefore, transferring the raw data to a server consumes a significant amount of energy [67, 37] and puts the patient's privacy at risk. A promising solution to address this problem is to distribute the computation across several hospitals, which is known as federated learning (FL), when a central hub coordinates the learning process [56, 58]. A typical FL framework consists of a central server and multiple hospitals. The server maintains

Algorithm 3 Teacher DNN

Require: Synchronized pre-processed ECG signal x_i^{ECG}, EEG signals x_i^{EEG1} and x_i^{EEG2}
Ensure: $\hat{y}_i$

Initialize the weights of the teacher model randomly.
for x_i^{feature} in $\{x_i^{\text{ECG}}, x_i^{\text{EEG1}}, x_i^{\text{EEG2}}\}$ **do**
    $\boldsymbol{z}_i^{\text{feature}} \leftarrow \text{Res1DCNN}(x_i^{\text{feature}})$

Merge these feature maps into a single $\boldsymbol{z}_i^T$ using a linear combination of the features.
Initialize the trainable weights θ randomly.
$\boldsymbol{z}_i^T \leftarrow \theta_1 \boldsymbol{z}_i^{\text{ECG}} + \theta_2 \boldsymbol{z}_i^{\text{EEG1}} + \theta_3 \boldsymbol{z}_i^{\text{EEG2}}$

Train a fully-connected layer that predicts the output $\hat{y}_i$ from $\boldsymbol{z}_i^T$.
$\hat{y}_i \leftarrow \text{FullyConnected}(\boldsymbol{z}_i^T)$
$\hat{y}_i \leftarrow \text{Softmax}(\hat{y}_i)$

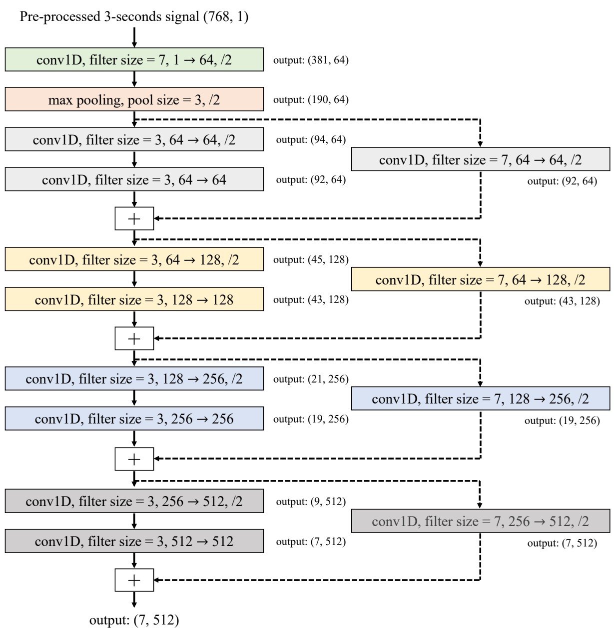


Fig. 5: The architecture of the feature extractor of Res1DCNN [48]. It contains 13 convolutional layers with skip connections followed by a dense layer. Here, '/2' denotes the downsampling operator using a strided convolution with a factor of 2. '→' denotes the transition from the input to output channels. The teacher model includes three feature extractors for ECG, $\text{EEG}_{F_7T_3}$ and $\text{EEG}_{F_8T_4}$ signals. The student model retains one feature extractor for the ECG signal.

a global model, and each hospital maintains a local model. At the beginning of training, all local models are initialized randomly. Each hospital performs Stochastic Gradient Descent and computes the local gradient for a certain number of iterations [13, 68]. After a certain number of iterations, a synchronization stage happens where the hospitals send their local weights to the server. The server performs the Federated Averaging (FedAvg) algorithm [9] to form a global weight to update the global model. These steps will be iterated until the

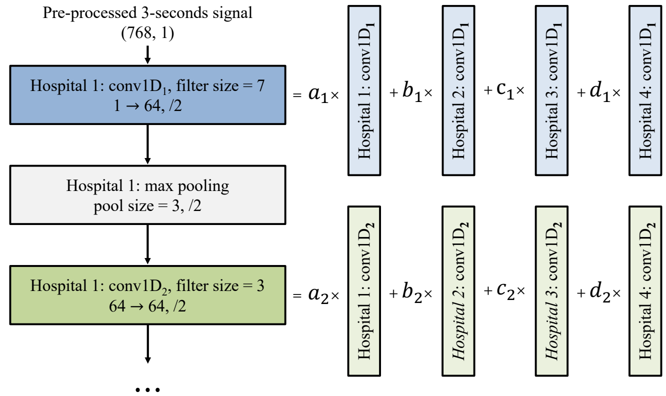


Fig. 6: Example of ensemble learning process of the first and second convolution layers of hospital 1. In the synchronization stage, hospital 1 receives the local weights of other hospitals. Hospital 1 recreates the teacher layers, where each layer is a weighted (a_1, b_1, c_1, d_1, e_1, a_2, b_2, c_2, ...) combination of that precise layer from other hospitals. The weighted combination variables are optimized to improve seizure detection accuracy on the validation set of the hospital 1.

model accuracy meets the requirement.

The typical FL framework has drawbacks. Due to its disability to implement a serverless framework, it is vulnerable to malfunctioning servers. In other words, the FL process stops when the server is down. Another disadvantage is that the common goal of FL is to obtain a global DNN model for all hospitals by assuming the data from all hospitals come from a similar distribution. However, when the hospitals' data come from different distributions and do not necessarily follow the same profile, the global DNN is not adapted for each hospital.

To consider the profile of each hospital while still exploiting the data from other hospitals, we propose a customized variant of the decentralized FL framework. An additional feature of the decentralized FL is the prevention of server failure. Figure 2 illustrates our proposed customized decentralized FL framework with n devices in the context of epileptic seizure monitoring. Decentralized FL is a strategy where the training is distributed among hospitals, and each hospital communicates with other available hospitals. We explore a clustering algorithm based on seizure type, by which model transfer costs can be reduced by limiting the hospitals to which teacher models are transferred. Thus the communication bottleneck (server) is removed.

Due to the heterogeneous computation resources of different hospitals, the local training time varies across hospitals. Therefore, after the hospitals trained their local teacher model for a fixed number of iterations, those who are online send their current weights to other available hospitals ($W_{i,1}$, $W_{i,2}$, ..., $W_{i,n}$). We propose a model ensemble learning process after the synchronization stage to improve the seizure detection accuracy for each hospital separately. The purpose is to obtain a model tailored to each hospital. As shown in Figure 6, each hospital recreates the teacher model, where each layer is obtained by a weighted combination of that precise layer from all hospitals. Each hospital gets these combination weights by enhancing the seizure detection accuracy on its local validation set. We observe in Section VI that by using our proposed ensemble learning process, each hospital will take the highest advantage of the information from the other hospitals' models. When the local distribution is highly correlated with a combination of global and local distribution, the latter is preferable; therefore, the hospital continues the training with the combined DNN; otherwise, the hospital keeps the local DNN. Our proposed framework also contributes to preserving the privacy of the hospitals' patient data involved in the training process by keeping sensitive medical data in the hospitals. We present personalized decentralized federated learning framework in Algorithm 4.

Algorithm 4 Personalized Decentralized Federated Learning Framework for Hospitals

Require: Hospitals $\mathcal{H} = \{H_1, H_2, \ldots, H_n\}$ with local teacher model weights $W_{1,i}, W_{2,i}, \ldots, W_{k,i}$ for each hospital H_i.
Ensure: Personalized teacher model weights $W'_{1,i}, W'_{2,i}, \ldots, W'_{k,i}$ for each hospital H_i.

Step 1: Local Training
for each hospital H_i in $\mathcal{H}$ **do**
  Train local teacher model using local weights $W_{1,i}, W_{2,i}, \ldots, W_{k,i}$.

Step 2: Synchronization
for each online hospital H_i in $\mathcal{H}$ **do**
  Send local weights $W_{1,i}, W_{2,i}, \ldots, W_{k,i}$ to all other online hospitals.

Step 3: Model Ensemble Learning
for each hospital H_i in $\mathcal{H}$ **do**
  for each layer L_j in model of Hospital H_i **do**
    Recreate weight $W_{j,i}$ of the layer L_j as a weighted combination of weight $W_{j,k}$ from the corresponding layer of the models of all other hospitals.
    This weight $W'_{j,i}$ is obtained by: $W'_{j,i} = a_j \cdot W_{j,1} + b_j \cdot W_{j,2} + c_j \cdot W_{j,3} + \ldots$.
    Optimize $a_j, b_j, \ldots$ for accuracy on H_i validation set.

Step 4: Select Best Model
for each hospital H_i in $\mathcal{H}$ **do**
  Compute the seizure detection accuracy of the local DNN model with original weights $W_{1,i}, W_{2,i}, \ldots, W_{k,i}$ and the model with combined weights $W'_{1,i}, W'_{2,i}, \ldots, W'_{k,i}$ on the local validation set.
  if accuracy($W_{j,i}$) $<$ accuracy($W'_{j,i}$) **then**
    $W_{1,i}, W_{2,i}, \ldots, W_{k,i} \leftarrow W'_{1,i}, W'_{2,i}, \ldots, W'_{k,i}$

IV. Seizure Detection DNN Deployment to Low-Power Wearable IoT Systems

This section describes the patient monitoring stage, where a patient is discharged from a hospital. Running the teacher

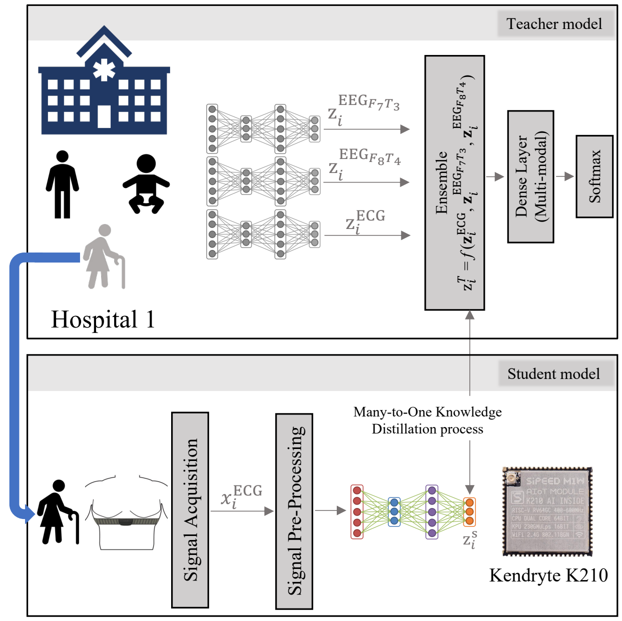


Fig. 7: Deployment phase: The stage where the hospital discharges a patient. Each hospital performs the knowledge distillation process from the enriched teacher model to the smaller student model. The student model is a single DNN that requires only an ECG signal and can be implemented on wearable devices with limited memory and computational resource for long-term patient monitoring.

DNN on resource-limited devices is often not feasible due to its demand for ECG, $\text{EEG}_{F_7T_3}$ and $\text{EEG}_{F_8T_4}$ signals. Using the Many-to-One Knowledge Distillation method, we can compact the knowledge from the teacher DNN, which requires these multiple signals, into a student DNN that needs only the ECG signal.

Knowledge distillation allows us to reduce the size of the teacher DNN to make it compatible with devices having limited memory and power. As established in [69, 45, 70, 71], one can successfully transfer knowledge from an ensemble model (teacher model) to a singular student model. As depicted in Figure 7, each hospital houses a student network that solely relies on ECG signals. Every hospital carries out the Many-to-One Knowledge Distillation, moving insights from the larger teacher network to the more lightweight student network.

To impart the knowledge from the pre-trained teacher to the student, we use an L2 distance between the output of the teacher DNN ($\mathbf{z}_i^T$) and the output of the student DNN ($\mathbf{z}_i^S$) as our loss function. Training the model with this loss function steers the feature map of the student to mirror that of the teacher. We select the feature map z_i^T of the teacher model due to its enhanced signal intensity and spatial correlation information. The student network's reduced complexity renders it apt for wearable devices running for prolonged periods. Notably, despite using only the ECG signal, the student network reaches a detection performance on par with the teacher network [61]. Acquiring the ECG signal consumes less energy, and the devices for this are readily available. This makes long-term patient monitoring more accessible.

The Many-to-One Knowledge Distillation process is outlined in Algorithm 5.

Algorithm 5 Many-to-One Knowledge Distillation for Student DNN

Require: $\mathcal{H} = \{H_1, H_2, \ldots, H_n\}$, personalized teacher DNN outputs $\boldsymbol{z}_i^T$
Ensure: Student DNN outputs $\boldsymbol{z}_i^S$

for H_i in $\mathcal{H}$ **do**
    Train student DNN using only ECG signals.
    Define loss function as $L = ||\boldsymbol{z}_i^T - \boldsymbol{z}_i^S||_2^2$.
    Update student DNN by minimizing the loss function.

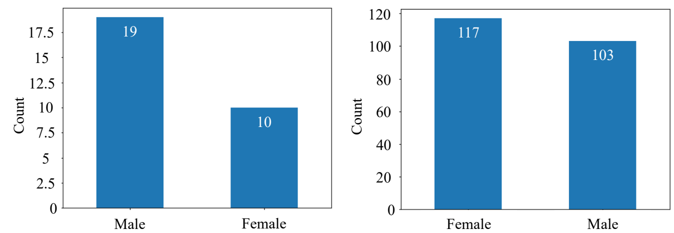
V. Experimental Setup

This section presents the experimental setup to evaluate our proposed decentralized FL framework regarding epileptic seizure detection performance and energy consumption.

A. Epileptic Seizures Datasets

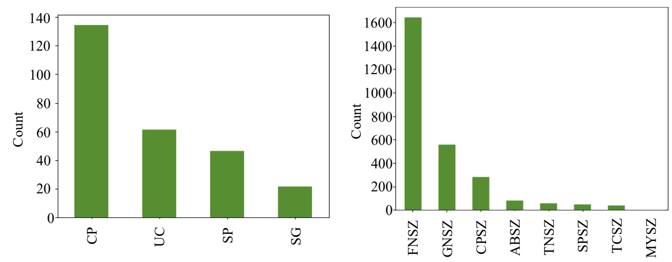
1) EPILEPSIAE dataset: We utilize the EPILEPSIAE dataset [72], an extensive manually annotated epilepsy dataset containing one-lead ECG and 19-channel EEG data from 30 patients. The dataset includes diverse seizure types: complex partial (CP), simple partial (SP), secondarily generalized (SG), and unclassified (UC) (Figure 9a). Due to synchronization issues in ECG and EEG signals for one patient, we analyze data from 29 patients, totaling 4603 hours of recordings with 266 seizures. The patient demographics consist of 19 males and 10 females (Figure 8a), predominantly aged between 40 and 50 years (Figure 10a). The signals were recorded at a 256 Hz sampling rate with 16-bit resolution. While the teacher network utilizes both EEG and ECG data, the student network only employs ECG signals. The teacher network's EEG data is acquired from four electrodes (two channels): F_7T_3 and F_8T_4 (Figure 3), selected for compatibility with the e-Glass wearable device, which avoids social stigma associated with traditional EEG equipment.

2) TUSZ dataset: We also use the TUH EEG Seizure Corpus (TUSZ) dataset [73], the world's most extensive corpus of annotated data for seizure detection, which includes records from 220 patients. As shown in Figure 8b, there are 117 female and 103 male patients. The age distribution (Figure 10b) reveals that most patients are between 40 and 70 years old. The dataset contains EEG and ECG data, with the most common number of EEG channels being 31. As described in Sec. III-A, we only require four EEG channels to train the teacher network. The majority of EEG signals are sampled at 250Hz (87%), with the remaining signals sampled at 256Hz (8.3%), 400Hz (3.8%), and 512Hz (1%). TUSZ encompasses various seizure morphologies, including Focal Non-Specific Seizure (FNSZ), Generalized Non-Specific Seizure (GNSZ), Complex Partial Seizure (CPSZ), Absence Seizure (ABSZ), tonic seizure (TNSZ), simple partial seizure (SPSZ), tonic-clonic seizure (TCSZ), and Myoclonic Seizure (MYSZ), as shown in Figure 9b. The TUSZ dataset offers event-based and term-based annotations, capturing seizure evolution and duration. Event-based annotations provide seizure origin details per channel. Both multi-class and bi-class annotations are included for machine learning research, with multi-class annotations detailing seizure types and bi-class annotations indicating seizure presence.



(a) EPILEPSIAE database   **(b)** TUSZ database

Fig. 8: Gender distribution



(a) EPILEPSIAE database   **(b)** TUSZ database

Fig. 9: Seizure distribution

B. Detection Performance Metrics

We considered three different metrics to evaluate the detection performance of our proposed framework. Sensitivity (Sen) (Eq. (1)) represents the percentage of ictal samples that are labeled correctly. Specificity (Spe) (Eq. (2)) shows the percentage of inter-ictal samples that are labeled correctly. These metrics are defined as follows:

$$Sen = \frac{TP}{TP + FN}, \tag{1}$$

$$Spe = \frac{TN}{FP + TN}, \tag{2}$$

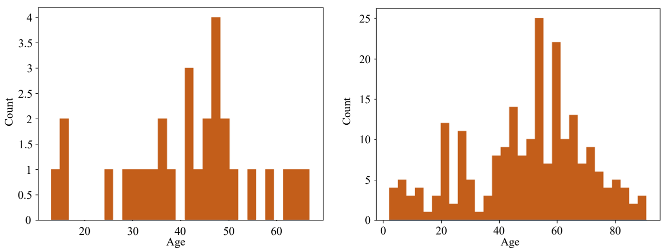
where TP, TN, FP and FN are true positive, true negative, false positive and false negative, respectively.

Finally, we evaluate the geometric mean (Gmean) (Eq. (3)) [74], which reflects both sensitivity and specificity and measures the balance between classification performance in both classes. A low geometric mean indicates poor performance in classifying the seizure cases, even if the non-seizures cases are correctly classified or vice versa.

$$Gmean = \sqrt{Sensitivity \times Specificity}. \tag{3}$$

C. Decentralized Training Setup

In our evaluation of the proposed decentralized Federated Learning (FL) system, we paid special attention to aspects



(a) EPILEPSIAE database

(b) TUSZ database

Fig. 10: Age distribution

Fig. 11: Hardware setup for the energy consumption measurement. Otii Arc is connected to the computer using a USB cable. The main output of the Otii Arc is connected to the voltage supply input of the IoT wearable device. The computer's USB port's current is used to power the Otii Arc and edge AI platforms. The Otii desktop application enables us to measure and analyze the energy consumption of the edge AI platform. In our evaluation, we considered a development system that uses AI technologies embedded on a Kendryte K210 chip in different IoT wearable systems [77] or Raspberry Pi Zero.

such as energy consumption and real-time performance. To this end, we envisioned a scenario involving four hospitals, each with their unique subset of patient data, randomly distributed among them. All the hospitals operated on the same teacher and student Deep Neural Network (DNN) architecture.

The first phase saw each hospital conduct the proposed decentralized FL, where the local teacher model underwent training for 5000 iterations. Following this, the model's weights were transmitted to all other participating hospitals. Each hospital then initiated the model ensemble learning strategy detailed in Section III-C, comparing epileptic seizure detection accuracy between local weights and a blend of global and local weights on their respective local validation data. The weights that yielded higher detection accuracy were accepted for further use.

In order to assess the detection accuracy of our proposed decentralized FL more effectively, we considered the division of patient data across hospitals based on similarities in seizure type and age range. Importantly, no two hospitals shared data from the same patients. Moreover, as part of our data handling strategy, each hospital's data was split into a training set, comprising 80% of the data, and a test set with the remaining 20%. This separation ensured an ample dataset for training our DNN models, while also reserving a substantial amount for an unbiased evaluation of the models' performance. Further refining our strategy, we carved out a validation set from the initial training set. This validation set was used to guide hyperparameter tuning and inform decisions regarding the training process, such as implementing early stopping to prevent overfitting.

D. Learning Parameters

We trained the teacher and student DNNs from scratch in our proposed framework using pre-processed 3-second ECG and EEG segments. The weights initialization of the layers follows a normal distribution with zero mean and 0.01 as the standard deviation. We initialize all the biases to zero. During the training, the network adjusts the model's parameters to minimize the cross-entropy loss. We performed the binary classification with two nodes where each class gets its output neuron and can be easily extended to multi-class classification in future work. Finally, we use the Adam optimizer [75] with a base learning rate of 10^{-4} and implement the DNNs on Tensorflow 1.14.0 [76].

E. Edge AI Evaluation Platforms

Wearable devices have small batteries and low-power processors compared to desktop processors. After a hospital discharges a patient and in the patient monitoring stage, this work uses two platforms, the Kendryte K210 [77] and Raspberry Pi Zero [78] with different features such as computing power and memory size to analyze and compare the energy consumption and timing requirements for the continuous execution of the proposed DNNs. Note that the proposed process must be executed repeatedly in real-time and have a satisfying detection performance.

The Raspberry Pi Zero includes an ARM11 CPU running at 1 GHz, has 512MB RAM, and performs the inference process of a given DNN with power supplied via a micro USB connector. The Kendryte K210 is a chip system with specific circuits/components for machine vision and ML. This chip system employs advanced ultra-low processing with the help of a 64-bit dual-core processor equipped with a high-performance hardware accelerator of the CNN. It supports convolution kernels, various type of activation function, and neural network parameter size up to 6 MB for real-time application.

We used Otii Arc [79] as a power analyzer and power supply for the inference process of our proposed DNNs. Otii Arc is a measurement tool for designing highly energy-efficient algorithms. It is powered via USB from the laptop and records both current and voltage, and it displays them in real-time for analysis and comparison. It provides up to 5 V output voltage and runs high-resolution current measurements with a sample rate of up to 4 kHz for the range of 1 μA-5 A. Figure 11 shows the hardware setup of our energy consumption measurement. We considered the Kendryte K210 chip and Raspberry Pi Zero as they have comparable processing capabilities to modern wearable architectures [80].

F. Baselines Description

- Federated Averaging (FedAvg) [9]: FedAvg is traditionally known for its approach to Federated Learning where devices train models locally on their data and send model updates to a central server for aggregation. However, in our experimental setup, we've adapted FedAvg to fit a decentralized federated learning framework. In this context, hospitals still compute local models, but instead of communicating with a central server, they share updates directly with other hospitals. This decentralized approach facilitates iterative averaging of model updates without the need for a centralized authority.

- Decentralized Federated Learning via Mutual Knowledge Transfer (Def-KT) [53]: In the context of IoT systems, [53] explores decentralized federated learning (DFL) where multiple IoT clients collaboratively train models for a shared task without disseminating their private training data and devoid of a central server. A prominent challenge that arises in such a setting is client-drift, especially when data is diverse across clients, resulting in slower convergence and diminished learning efficacy. To address this, the "Decentralized Federated Learning via Mutual Knowledge Transfer (Def-KT)" algorithm is introduced. Unlike conventional methods that directly average model parameters, Def-KT facilitates clients to merge models by transferring their individual learned knowledge to one another, effectively tackling the client-drift concern.
- Learning to Collaborate (L2C) [54]: This baseline tackles the challenge of crafting personalized models for distinct tasks given constraints of limited data and computational resources on edge devices. Traditional decentralized learning (DL), while suitable for creating a universal model using distributed data, isn't adept at personalizing models for varied tasks or optimizing network topologies. In DL, the fixed mixing weights can't easily adjust to different nodes, tasks, or stages of learning, making personalization difficult. To overcome these limitations, [54] offers "Learning to Collaborate (L2C)", which refines the models to minimize local validation loss for specific tasks.

VI. Evaluation

This section presents an evaluation of the accuracy, runtime, and energy consumption of seizure detection of our proposed decentralized FL with the knowledge distillation approach on the Kendryte K210 and the Raspberry Pi Zero unit.

A. Detection Performance Analysis

In an effort to assess the proposed ensemble learning framework for decentralized FL, we conducted a simulation using a realistic scenario involving four hospitals with a random distribution of patients. The detection accuracy using TUSZ and EPILEPSIAE datasets is compared in Table I and Table II. Here, we contrasted the results from our proposed ensemble learning approach, which uses a combination of local and other models, with conventional methods like FedAvg, which averages local and other models, and other techniques as proposed in [53] and [54]. Upon analysis of the TUSZ dataset, we noticed that the Gmean average rose to 85.83% with our proposed ensemble learning, from 77.58% with FedAvg, 81.08% with [53], and 84.84% with [54]. Similar improvements were observed with the EPILEPSIAE dataset,

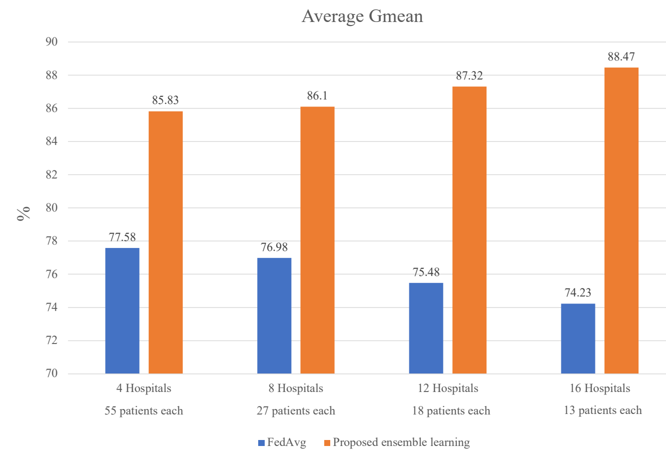


Fig. 12: Proposed ensemble learning versus FedAvg in decentralized FL on the TUSZ dataset with different number of hospitals.

where the Gmean average increased to 88.72% from 80.77% with FedAvg, 85.41% with [53], and 87.79% with [54]. In our ensemble learning methodology, each hospital generates a 'teacher model', with each layer representing a weighted combination of the corresponding layer from all other hospitals. This allows each hospital to maximally leverage the information available from the models of the other hospitals. We also present results from the student model, derived using the Many-to-One Knowledge Distillation technique from the teacher model. Notably, the student model exclusively utilizes the ECG signal, eliminating the need for both ECG and EEG signals, which is a distinct advantage in practical applications.

We deployed the detection algorithm with more hospitals to validate the practicality of the proposed ensemble learning in decentralized FL in large-scale settings. Figure 12, shows and compares the average Gmean of our proposed ensemble learning with FedAvg in FL, where we have 4, 8, 12, and 16 hospitals. We observe that the detection accuracy of FedAvg tends to degrade in large-scale FL settings. In these cases, increasing the number of hospitals makes the data distribution of patients between hospitals more dissimilar. As a result, solely averaging the weights does not create a more accurate model. That is where implementing the proposed ensemble learning demonstrates its importance in large-scale FL settings. When scaling the number of hospitals, each hospital contains fewer patients. As a result, the proposed ensemble learning aids hospitals in obtaining a more personalized DNN for their patients.

To assess how similarities between patients in the same hospital can help to improve the models obtained, we studied different arrangements in which we placed patients manually among hospitals. In the first plan, we divide patients according to their ages. Therefore, patients within an age range are assigned to a hospital. Table III shows the results of our proposed framework for each hospital in the TUSZ database. We note that, once again, our proposed ensemble learning outperforms FedAvg by 15%.

In our second plan, our goal is to group patients experiencing similar seizures and assign them to clusters. Put simply,

TABLE I: Comparative performance analysis of epileptic seizure detection models on the TUSZ dataset distributed across four hospitals. This analysis involves the decentralized Federated Learning (FL) model trained using FedAvg, models presented in Li et al., 2021 [53] and Li et al., 2022 [54], and our newly proposed ensemble learning model (the teacher model). All these models require both ECG and EEG signals for their operation. In addition to these, we present results from our student model, obtained through the Many-to-One Knowledge Distillation technique from the teacher model. Notably, the student model requires only the ECG signal. For the purpose of this experiment, patients were randomly assigned to each hospital.

		Hospital 1			Hospital 2			Hospital 3			Hospital 4					
	Number of patients	55			55			55			55			Average		
		Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean
TUSZ	FedAvg ECG + EEG	75.51%	78.89%	77.18%	71.26%	76.71%	73.93%	79.52%	78.20%	78.85%	80.85%	79.91%	80.37%	76.78%	78.42%	77.58%
	Results of [53] ECG + EEG	80.92%	79.61%	80.26%	79.12%	80.33%	79.72%	81.99%	82.78%	82.38%	81.33%	82.67%	81.99%	80.84%	81.34%	81.08%
	Results of [54] ECG + EEG	86.77%	82.51%	84.61%	89.02%	84.20%	86.57%	84.55%	83.07%	83.80%	85.00%	83.84%	84.41%	86.33%	83.40%	84.84%
	Proposed ensemble learning (Teacher model) ECG + EEG	87.81%	82.06%	84.88%	91.45%	84.36%	87.83%	85.88%	83.77%	84.81%	86.63%	85.05%	85.83%	87.94%	83.81%	85.83%
	Many-to-One Knowledge distillation (Student model) ECG	80.11%	79.23%	79.66%	82.04%	83.15%	82.59%	79.09%	80.43%	79.75%	78.85%	80.63%	79.73%	80.02%	80.86%	80.43%

TABLE II: Comparative performance analysis of epileptic seizure detection models on the EPILEPSIAE dataset distributed across four hospitals. This analysis includes the decentralized Federated Learning (FL) model trained using FedAvg, models presented in Li et al., 2021 [53] and Li et al., 2022 [54], and our newly proposed ensemble learning model (the teacher model). Each of these models necessitate the usage of both ECG and EEG signals. Additionally, we include results from our student model, which was obtained via the Many-to-One Knowledge Distillation technique from the teacher model. Remarkably, the student model operates using only the ECG signal. For this experiment, patients were randomly assigned to each hospital.

		Hospital 1			Hospital 2			Hospital 3			Hospital 4					
	Number of patients	7			7			8			7			Average		
		Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean
EPILEPSIAE	FedAvg ECG + EEG	84.22%	86.19%	85.19%	77.97%	79.78%	78.86%	80.49%	82.01%	81.24%	77.33%	78.29%	77.80%	80.00%	81.56%	80.77%
	Results of [53] ECG + EEG	89.27%	90.71%	89.98%	83.42%	84.89%	84.15%	86.63%	86.17%	86.39%	80.51%	81.75%	81.12%	84.95%	85.88%	85.41%
	Results of [54] ECG + EEG	93.38%	90.26%	91.80%	84.73%	88.57%	86.62%	90.47%	89.67%	90.06%	83.58%	81.82%	82.69%	88.04%	87.58%	87.79%
	Proposed ensemble learning (Teacher model) ECG + EEG	94.34%	91.79%	93.05%	85.71%	88.43%	87.05%	91.14%	91.64%	91.38%	84.91%	81.97%	83.42%	89.02%	88.45%	88.72%
	Many-to-One Knowledge distillation (Student model) ECG	89.09%	89.86%	89.47%	83.39%	86.75%	85.05%	86.48%	86.91%	86.69%	82.60%	78.59%	80.57%	85.39%	85.52%	85.44%

we segregate patients across multiple hospitals using the k-means clustering method. The intent here is to ensure that seizure types within a given hospital are more similar to each other than to those in other hospitals. The outcomes of our proposed decentralized federated learning for each hospital, utilizing the TUSZ and EPILEPSIAE databases, are illustrated in Tables IV and V. Comparing our proposed ensemble learning to FedAvg, we noted that the average Gmean rises from 83.92% to 89.52% on the TUSZ dataset, and from 88.69% to 94.64% on the EPILEPSIAE dataset. We deduced

TABLE III: Performance comparison between a decentralized Federated Learning (FL) model trained using FedAvg and our novel ensemble learning model for epileptic seizure detection. The analysis uses the TUSZ dataset distributed across four hospitals, with patient assignment to each hospital based on a specific age range.

		Hospital 1			Hospital 2			Hospital 3			Hospital 4					
	Number of patients	32			39			116			33					
	Age intervals	2-25			26-48			49-71			72-93			Average		
		Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean
TUSZ	FedAvg	69.91%	74.71%	72.27%	78.18%	62.51%	69.90%	78.77%	75.82%	77.28%	63.71%	65.78%	64.73%	72.64%	69.70%	71.04%
	Proposed ensemble learning	86.37%	84.00%	85.17%	86.02%	84.41%	85.21%	84.88%	83.69%	84.28%	83.52%	84.18%	83.84%	85.19%	84.07%	85.85%

that clustering patients based on their seizure type results in superior detection performance from each hospital's DNN. However, the sharing of cluster identities amongst hospitals is necessary, leading to potential privacy issues. Nevertheless, our proposed ensemble learning technique offers a better means of safeguarding patient privacy within a decentralized federated learning system, while simultaneously providing a more tailored DNN that leads to improved seizure detection for each hospital.

Federated learning, although advantageous, faces challenges. In particular, when the number of hospitals grows, the communication cost can escalate significantly. These challenges can broadly be categorized into three areas: heterogeneous data, asynchronous updates, and client participation. Heterogeneity in data distribution, particularly non-IID data among clients, can affect model convergence and accuracy. Similarly, with a larger number of clients, the time it takes for each client to compute and send their model updates can vary, thus necessitating asynchronous updates. Client participation, referring to the involvement of each client in the collaborative training process, is also crucial for the success of the federated learning system due to variations in availability, resources, and data distributions. Our research is dedicated to addressing these challenges.

To mitigate the impact of data heterogeneity on model convergence and accuracy, we introduce an adaptive ensemble learning approach. This technique is formulated to effectively manage the variability in data across different clients. Additionally, we have devised a strategy, as displayed in Tables IV and V, to cluster patients based on their seizure type, enhancing the DNN's detection performance for each hospital. This method is analogous to client sampling in federated learning, where a subset of clients with similar seizure types is selected for each training round. This strategy not only helps reduce communication overhead but also accommodates clients with limited resources or availability. Through these solutions, we effectively address the challenge of high communication costs while enhancing the performance and efficiency of federated learning in our healthcare context.

Further supplementing our evaluation, we have incorporated Table VI, showcasing the potency of our proposed ensemble learning methodology on various other medical datasets, including the MIT-BIH Arrhythmia Dataset [81]. This addition solidifies the robustness and versatility of our approach, highlighting its potential value not just within the confines of the datasets initially considered, but also extending its utility across the broader healthcare domain and patient monitoring scenarios. These results, thereby, emphasize the importance and wide ranging applicability of our proposed ensemble learning approach in the landscape of decentralized Federated Learning within healthcare.

B. Energy Consumption Analysis

In this work, we also want to perform long-term patient monitoring to perform the epileptic seizure detection algorithm on a low-energy embedded medical platform with limited computational resources that operates on a battery. For instance, the e-Glass wearable system [62] shown in Figure 3 contains a 570 mAh battery and features an ultra-low-power 32-bit microcontroller STM32L151 [82] with an ARM® Cortex®-M3 with 48 KB RAM and 384 KB Flash. Therefore, e-Glass communicates with a smartphone or smartwatch using Bluetooth low energy (nRF8001) [83].

We study the Kendryte K210 and Raspberry Pi Zero platforms to examine the complexity, lifetime, and energy efficiency of our approach. In the implementation code, all the computations and storage are in a 16-bit fixed point. We chose 13 bits for the fractional part using the results of the validation set. We observed that dedicating more bits to the fractional part causes overflows in the computations. On the other hand, reducing the number of fraction bits gives rise to a considerable accuracy drop. Since we use the fixed-point representation of numbers, we save the amount of storage by a factor of 4 compared to 64-bit floating-point operations. This compression is crucial because it enables our student network to be applicable on various memory-limited embedded devices.

Table VII shows the seizure detection execution time per 3-second segment for teacher and student DNNs. We obtain the represented numbers by running the experiments for the whole test set. We observe that we can continuously monitor the patients in real time using teacher and student DNNs in the proposed decentralized FL. One of the benefits of the

TABLE IV: Performance comparison of epileptic seizure detection models on the TUSZ dataset across four hospitals. The models under comparison are a decentralized Federated Learning (FL) model trained using FedAvg and our proposed ensemble learning model. For the purpose of this study, patients exhibiting similar seizure types were grouped and assigned to the same hospital.

		Hospital 1			Hospital 2			Hospital 3			Hospital 4			Average		
	Number of patients	63			67			41			49					
	Type and number of seizure	FNSZ:139 GNSZ:93 CPSZ:367 ABSZ:87 TNSZ:62 SPSZ:52 TCSZ:48 MYSZ:3			FNSZ:487 GNSZ:489 CPSZ:0 ABSZ:0 TNSZ:0 SPSZ:0 TCSZ:0 MYSZ:0			FNSZ:379 GNSZ:0 CPSZ:0 ABSZ:0 TNSZ:0 SPSZ:0 TCSZ:0 MYSZ:0			FNSZ:645 GNSZ:0 CPSZ:0 ABSZ:0 TNSZ:0 SPSZ:0 TCSZ:0 MYSZ:0					
		Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean
TUSZ	FedAvg	85.31%	77.60%	81.36%	79.59%	77.98%	78.78%	90.22%	93.97%	92.07%	85.40%	81.65%	83.50%	85.13%	82.80%	83.92%
	Proposed ensemble learning	87.94%	84.90%	86.40%	86.58%	85.94%	86.25%	98.57%	96.79%	97.67%	91.92%	83.84%	87.78%	91.25%	87.86%	89.52%

TABLE V: Comparative analysis of epileptic seizure detection performance between a decentralized Federated Learning (FL) model trained with FedAvg and our proposed ensemble learning model. The evaluation is conducted on the EPILEPSIAE dataset distributed across four hospitals. In this study, patients experiencing similar types of seizures were assigned to the same hospital.

		Hospital 1			Hospital 2			Hospital 3			Hospital 4			Average		
	Number of patients	6			2			2			19					
	Type and number of seizure	UC:4 CP:3 SG:6 SP:33			UC:4 CP:29 SG:0 SP:3			UC:26 CP:103 SG:14 SP:9			UC:28 CP:0 SG:2 SP:2					
		Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean
EPILEPSIAE	FedAvg	86.72%	89.41%	88.05%	88.75%	88.00%	88.37%	95.01%	96.63%	95.81%	80.50%	84.64%	82.54%	87.74%	89.67%	88.69%
	Proposed ensemble learning	89.70%	96.99%	93.27%	94.33%	96.58%	95.44%	100%	100%	100%	83.89%	96.27%	89.86%	91.98%	97.46%	94.64%

TABLE VI: Performance comparison in terms of sensitivity, specificity, and geometric mean between a decentralized Federated Learning (FL) model trained using FedAvg and our proposed ensemble learning model. The models are evaluated on the MIT-BIH Arrhythmia Database distributed across four hospitals, with patients being randomly assigned to each hospital.

		Hospital 1			Hospital 2			Hospital 3			Hospital 4					
	Number of patients	12			12			12			12					
		Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean	Sen	Spe	Gmean
MIT-BIH	FedAvg	100%	96.00%	97.97%	97.10%	95.97%	96.53%	97.00%	96.53%	96.76%	96.22%	96.87%	96.54%	97.58%	96.34%	96.95%
	Proposed ensemble learning	100%	98.66%	99.33%	100%	100%	100%	100%	98.00%	98.99%	99.33%	100%	99.66%	99.83%	99.16%	99.49%

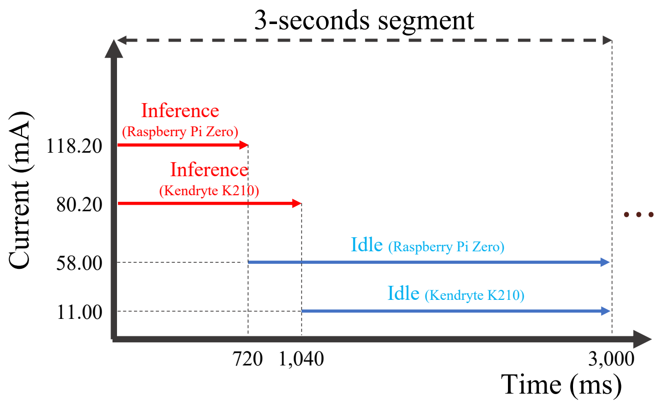


Fig. 13: Energy consumption monitoring on the Raspberry Pi Zero and Kendryte K210 during the inference of student DNN and Idle state. Both platforms accomplish seizure detection in real-time. The Raspberry Pi Zero executes the inference quicker than the Kendryte K210 but consumes more energy. The energy consumption of the Idle state of Kendryte K210 is significantly less than the Raspberry Pi Zero.

student DNN is that it runs faster than the teacher DNN while achieving detection accuracy comparable to the teacher DNN. It avoids EEG signal acquisition and, as a result, provides more comfort to patients. This distillation (compression) from the teacher to student DNN is crucial because it enables our network to be applicable on various memory-limited embedded devices without affecting detection accuracy.

TABLE VII: Results of the quantitative evaluation of the run time for every 3-second segment in the student and teacher network architectures in the Raspberry Pi Zero and Kendryte K210.

Network	Platform	Run time (millisec.)
Teacher DNN (EEG + ECG)	Raspberry Pi Zero	2,080.56 ± 12.56
Distilled Student DNN (ECG)	Raspberry Pi Zero	720.15 ± 32.46
Teacher DNN (EEG + ECG)	Kendryte K210	2,998.08 ± 4.09
Distilled Student DNN (ECG)	Kendryte K210	1,040.64 ± 5.67

Epilepsy is an unpredictable disorder that can produce another health complication; therefore, patients must be continuously monitored. Figure 13 shows the inference energy consumption of a 3-second segment utilizing the student DNN. We realize that Raspberry Pi Zero and Kendryte K210 achieve the inference rapidly and often remain in an idle state, resulting in longer battery life and longer patient monitoring. Table VIII evaluates the battery lifetime of the student network in our desensitized FL using the 570 mAh battery of the e-Glass [62]. We understand that the proposed framework can monitor a patient in real-time for 7.86 hours on the Raspberry Pi Zero and 16.29 hours on the Kendryte K210 on a single charge.

TABLE VIII: Battery life of an edge device using the e-Glass [62] battery to run a student network to monitor patients and detect epileptic seizures.

Method	Platform	Battery life (hours)
Distilled Student DNN (ECG)	Raspberry Pi Zero	7.86 ± 0.09
Distilled Student DNN (ECG)	Kendryte K210	16.29 ± 0.06

VII. Conclusion

In this paper, we have proposed a serverless FL framework consisting of a training phase with an adaptive ensembling stage and a deployment phase using a knowledge distillation technique. The adaptive ensembling stage leads to learning a specific DNN for each medical center by discovering the optimal combination of local models and models from other available hospitals. It demonstrates its benefit in an actual scenario when we scale up the number of hospitals and where the patients' data distribution in different hospitals is usually non-IID. We adjusted the DNN complexity by using a knowledge distillation technique to reduce the computation requirements of the model (memory usage and energy consumption) to meet the resource constraints of wearable systems. Thanks to this solution, we can leave the training phase as complex as we need to obtain a high detection accuracy. We conducted extensive experiments with the TUSZ and EPILEPSIAE datasets to verify our analysis.

References

[1] Epilepsy. [Online]. Available: URL: http://www.who.int/mental_health/neurology/epilepsy/en/.

[2] Simon Shorvon and Torbjorn Tomson. "Sudden unexpected death in epilepsy". In: *The Lancet* 378.9808 (2011), pp. 2028–2038.

[3] Orrin Devinsky. "Sudden, unexpected death in epilepsy". In: *New England Journal of Medicine* 365.19 (2011), pp. 1801–1811.

[4] Anouk Van de Vel et al. "Non-EEG seizure detection systems and potential SUDEP prevention: State of the art Review and update". In: *Seizure* 41 (2016), pp. 141–153.

[5] Yang Li et al. "Automatic seizure detection using fully convolutional nested LSTM". In: *International journal of neural systems* 30.04 (2020), p. 2050019.

[6] Mustafa Talha Avcu, Zhuo Zhang, and Derrick Wei Shih Chan. "Seizure detection using least EEG channels by deep convolutional neural network". In: *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2019, pp. 1120–1124.

[7] Pierre Thodoroff, Joelle Pineau, and Andrew Lim. "Learning robust features using deep learning for automatic seizure detection". In: *Machine learning for healthcare conference*. PMLR. 2016, pp. 178–190.

[8] Lasitha Vidyaratne et al. "Deep recurrent neural network for seizure detection". In: *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2016, pp. 1202–1207.

[9] Brendan McMahan et al. "Communication-efficient learning of deep networks from decentralized data". In: *Artificial intelligence and statistics*. PMLR. 2017, pp. 1273–1282.

[10] Wenrui Dai et al. "Privacy preserving federated big data analysis". In: *Guide to Big Data Applications*. Springer, 2018, pp. 49–82.

[11] Qiang Yang et al. "Federated machine learning: Concept and applications". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 10.2 (2019), pp. 1–19.

[12] Paul Vanhaesebrouck, Aurélien Bellet, and Marc Tommasi. "Decentralized collaborative learning of personalized models over networks". In: *arXiv preprint arXiv:1610.05202* (2016).

[13] Sebastian U Stich. "Local SGD converges fast and communicates little". In: *arXiv preprint arXiv:1805.09767* (2018).

[14] Sebastian U Stich and Sai Praneeth Karimireddy. "The error-feedback framework: Better rates for SGD with delayed gradients and compressed communication". In: *arXiv preprint arXiv:1909.05350* (2019).

[15] Jianyu Wang and Gauri Joshi. "Cooperative SGD: A unified framework for the design and analysis of communication-efficient SGD algorithms". In: *arXiv preprint arXiv:1808.07576* (2018).

[16] Tzu-Ming Harry Hsu, Hang Qi, and Matthew Brown. "Measuring the effects of non-identical data distribution for federated visual classification". In: *arXiv preprint arXiv:1909.06335* (2019).

[17] Sai Praneeth Karimireddy et al. "SCAFFOLD: Stochastic Controlled Averaging for On-Device Federated Learning." In: (2019).

[18] Xiang Li et al. "On the convergence of fedavg on non-iid data". In: *arXiv preprint arXiv:1907.02189* (2019).

[19] Cristian Buciluundefined, Rich Caruana, and Alexandru Niculescu-Mizil. "Model Compression". In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '06. Philadelphia, PA, USA: Association for Computing Machinery, 2006, pp. 535–541. ISBN: 1595933395. DOI: 10.1145/1150402.1150464. URL: https://doi.org/10.1145/1150402.1150464.

[20] Renée A Shellhaas. "Continuous long-term electroencephalography: the gold standard for neonatal seizure diagnosis". In: *Seminars in Fetal and Neonatal Medicine*. Vol. 20. 3. Elsevier. 2015, pp. 149–153.

[21] D. Wang et al. "Epileptic Seizure Detection in Long-Term EEG Recordings by Using Wavelet-Based Directed Transfer Function". In: *IEEE Transactions on Biomedical Engineering* 65.11 (Nov. 2018), pp. 2591–2599. ISSN: 1558-2531. DOI: 10.1109/TBME.2018.2809798.

[22] Ling Xu Guo, Daniel Rivero, and Alejandro Pazos. "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks". In: *Journal of Neuroscience Methods* 193 (2010), pp. 156–163.

[23] S. Raghu et al. "A Novel Approach for Real-Time Recognition of Epileptic Seizures Using Minimum Variance Modified Fuzzy Entropy". In: *IEEE Transactions on Biomedical Engineering* 65.11 (Nov. 2018), pp. 2612–2621. ISSN: 1558-2531. DOI: 10.1109/TBME.2018.2810942.

[24] Kai Fu et al. "Hilbert marginal spectrum analysis for automatic seizure detection in EEG signals". In: *Biomed. Signal Proc. and Control* 18 (2015), pp. 179–185.

[25] Yissel Rodriguez Aldana et al. "Nonconvulsive epileptic seizure detection in scalp EEG using multiway data analysis". In: *IEEE journal of biomedical and health informatics* 23.2 (2018), pp. 660–671.

[26] Jiang-Ling Song, Wenfeng Hu, and Rui Zhang. "Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine". In: *Neurocomputing* 175 (2016), pp. 383–391.

[27] Ahnaf Rashik Hassan, Siuly Siuly, and Yanchun Zhang. "Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating". In: *Computer methods and programs in biomedicine* 137 (2016), pp. 247–259.

[28] Hisham Daoud and Magdy Bayoumi. "Deep Learning Approach for Epileptic Focus Localization". In: *IEEE Transactions on Biomedical Circuits and Systems* (2019).

[29] Hisham Daoud and Magdy A Bayoumi. "Efficient epileptic seizure prediction based on deep learning". In: *IEEE transactions on biomedical circuits and systems* 13.5 (2019), pp. 804–813.

[30] Ömer Türk and Mehmet Siraç Özerdem. "Epilepsy Detection by Using Scalogram Based Convolutional Neural Network from EEG Signals". In: *Brain sciences*. 2019.

[31] Ye Yuan et al. "A multi-view deep learning framework for EEG seizure detection". In: *IEEE journal of biomedical and health informatics* 23.1 (2018), pp. 83–94.

[32] Yu Liu et al. "Epileptic seizure detection using convolutional neural network: A multi-biosignal study". In: *Proceedings of the Australasian Computer Science Week Multiconference*. 2020, pp. 1–8.

[33] LD Blumhardt, PEM Smith, and Lynne Owen. "Electrocardiographic accompaniments of temporal lobe epileptic seizures". In: *The Lancet* 327.8489 (1986), pp. 1051–1056.

[34] H Mayer et al. "EKG abnormalities in children and adolescents with symptomatic temporal lobe epilepsy". In: *Neurology* 63.2 (2004), pp. 324–328.

[35] Carolina Varon et al. "Detection of epileptic seizures by means of morphological changes in the ECG". In: *Computing in Cardiology 2013*. IEEE. 2013, pp. 863–866.

[36] Maeike Zijlmans, Danny Flanagan, and Jean Gotman. "Heart rate changes and ECG abnormalities during epileptic seizures: prevalence and definition of an objective clinical sign". In: *Epilepsia* 43.8 (2002), pp. 847–854.

[37] Farnaz Forooghifar, Amir Aminifar, and David Atienza. "Resource-aware distributed epilepsy monitoring using self-awareness from edge to cloud". In: *IEEE transactions on biomedical circuits and systems* 13.6 (2019), pp. 1338–1350.

[38] Kaat Vandecasteele et al. "Automated epileptic seizure detection based on wearable ECG and PPG in a hospital environment". In: *Sensors* 17.10 (2017), p. 2338.

[39] OM Doyle et al. "Heart rate based automatic seizure detection in the newborn". In: *Medical engineering & physics* 32.8 (2010), pp. 829–839.

[40] Farnaz Forooghifar et al. "A self-aware epilepsy monitoring system for real-time epileptic seizure detection". In: *Mobile Networks and Applications* (2019), pp. 1–14.

[41] F. Forooghifar, A. Aminifar, and D. Atienza. "Resource-Aware Distributed Epilepsy Monitoring Using Self-Awareness From Edge to Cloud". In: *IEEE Transactions on Biomedical Circuits and Systems* (2019), pp. 1–1. ISSN: 1940-9990. DOI: 10.1109/TBCAS.2019.2951222.

[42] Francesco Onorati et al. "Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors". In: *Epilepsia* 58.11 (2017), pp. 1870–1879.

[43] F Fürbass et al. "Automatic multimodal detection for long-term seizure documentation in epilepsy". In: *Clinical Neurophysiology* 128.8 (2017), pp. 1466–1472.

[44] Marwa Qaraqe et al. "Epileptic seizure onset detection based on EEG and ECG data fusion". In: *Epilepsy & Behavior* 58 (2016), pp. 48–60.

[45] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network". In: *arXiv preprint arXiv:1503.02531* (2015).

[46] Pradip Chitrakar et al. "Social media image retrieval using distilled convolutional neural network for suspicious e-crime and terrorist account detection". In: *2016 IEEE International Symposium on Multimedia (ISM)*. IEEE. 2016, pp. 493–498.

[47] Andrew G Howard et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications". In: *arXiv preprint arXiv:1704.04861* (2017).

[48] Saleh Baghersalimi et al. "Personalized Real-Time Federated Learning for Epileptic Seizure Detection". In: *IEEE Journal of Biomedical and Health Informatics* (2021), pp. 1–1. DOI: 10.1109/JBHI.2021.3096127.

[49] Hongzheng Yu et al. "FedHAR: Semi-Supervised Online Learning for Personalized Federated Human Activity Recognition". In: *IEEE Transactions on Mobile Computing* (2021), pp. 1–1. DOI: 10.1109/TMC.2021.3136853.

[50] Weiping Ding et al. "Fed-ESD: Federated learning for efficient epileptic seizure detection in the fog-assisted internet of medical things". In: *Information Sciences* 630 (2023), pp. 403–419. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2023.02.052. URL: https://www.sciencedirect.com/science/article/pii/S0020025523002414.

[51] Suo Chen et al. "Decentralized Federated Learning with Intermediate Results in Mobile Edge Computing". In: *IEEE Transactions on Mobile Computing* (2022), pp. 1–17. DOI: 10.1109/TMC.2022.3221212.

[52] Zhenguo Ma et al. "Like Attracts Like: Personalized Federated Learning in Decentralized Edge Computing". In: *IEEE Transactions on Mobile Computing* (2022), pp. 1–17. DOI: 10.1109/TMC.2022.3230712.

[53] Chengxi Li, Gang Li, and Pramod K Varshney. "Decentralized federated learning via mutual knowledge transfer". In: *IEEE Internet of Things Journal* 9.2 (2021), pp. 1136–1147.

[54] Shuangtong Li et al. "Learning to collaborate in decentralized learning of personalized models". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 9766–9775.

[55] Tao Lin et al. "Ensemble distillation for robust model fusion in federated learning". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 2351–2363.

[56] Ruben Mayer and Hans-Arno Jacobsen. "Scalable Deep Learning on Distributed Infrastructures: Challenges, Techniques and Tools". In: *ArXiv* abs/1903.11314 (2019).

[57] Yutao Jiao et al. "Toward an Automated Auction Framework for Wireless Federated Learning Services Market". In: *IEEE Transactions on Mobile Computing* 20.10 (2021), pp. 3034–3048. DOI: 10.1109/TMC.2020.2994639.

[58] Praneeth Vepakomma et al. "Split learning for health: Distributed deep learning without sharing raw patient data". In: *CoRR* abs/1812.00564 (2018). arXiv: 1812.00564. URL: http://arxiv.org/abs/1812.00564.

[59] Yang Liu et al. "A secure federated transfer learning framework". In: *IEEE Intelligent Systems* 35.4 (2020), pp. 70–82.

[60] Mikhail Yurochkin et al. “Bayesian nonparametric federated learning of neural networks”. In: *International Conference on Machine Learning*. PMLR. 2019, pp. 7252–7261.

[61] Saleh Baghersalimi et al. *Many-to-One Knowledge Distillation of Real-Time Epileptic Seizure Detection for Low-Power Wearable Internet of Things Systems*. 2022. DOI: 10.48550/ARXIV.2208.00885. URL: https://arxiv.org/abs/2208.00885.

[62] Dionisije Sopic, Amir Aminifar, and David Atienza. “e-glass: A wearable system for real-time detection of epileptic seizures”. In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE. 2018, pp. 1–5.

[63] Damian Pascual et al. “Epilepsygan: Synthetic epileptic brain activities with privacy preservation”. In: *IEEE Transactions on Biomedical Engineering* 68.8 (2020), pp. 2435–2446.

[64] Xuyang Zhao et al. “Classification of epileptic IEEG signals by CNN and data augmentation”. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, pp. 926–930.

[65] Lei Huang et al. “Normalization Techniques in Training DNNs: Methodology, Analysis and Application”. In: *arXiv preprint arXiv:2009.12836* (2020).

[66] Cha Zhang and Yunqian Ma. *Ensemble machine learning: methods and applications*. Springer, 2012.

[67] Hossein Mamaghanian et al. “Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes”. In: *IEEE Transactions on Biomedical Engineering* 58.9 (2011), pp. 2456–2466.

[68] Jianyu Wang and Gauri Joshi. “Adaptive communication strategies to achieve the best error-runtime trade-off in local-update SGD”. In: *Proceedings of Machine Learning and Systems* 1 (2019), pp. 212–229.

[69] Cristian Buciluǎ, Rich Caruana, and Alexandru Niculescu-Mizil. “Model compression”. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2006, pp. 535–541.

[70] Jangho Kim, SeoungUK Park, and Nojun Kwak. “Paraphrasing complex network: Network compression via factor transfer”. In: *arXiv preprint arXiv:1802.04977* (2018).

[71] Lei Jimmy Ba and Rich Caruana. “Do deep nets really need to be deep?” In: *arXiv preprint arXiv:1312.6184* (2013).

[72] Matthias Ihle et al. “EPILEPSIAE - A European Epilepsy Database”. In: *Comput. Methods Prog. Biomed.* 106.3 (June 2012), pp. 127–138. ISSN: 0169-2607.

[73] Vinit Shah et al. “The temple university hospital seizure detection corpus”. In: *Frontiers in neuroinformatics* 12 (2018), p. 83.

[74] Philip J Fleming and John J Wallace. “How not to lie with statistics: the correct way to summarize benchmark results”. In: *Communications of the ACM* 29.3 (1986), pp. 218–221.

[75] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[76] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[77] Kendryte K210: URL: https://canaan.io/product/kendryteai.

[78] Warren Gay. *Raspberry Pi hardware reference*. Apress, 2014.

[79] Qoitech AB: Otii arc - otii by qoitech (2020): URL: https://www.qoitech.com/otii/.

[80] Antonio Pullini et al. “Mr. Wolf: An energy-precision scalable parallel ultra low power SoC for IoT edge processing”. In: *IEEE Journal of Solid-State Circuits* 54.7 (2019), pp. 1970–1981.

[81] George B Moody and Roger G Mark. “The impact of the MIT-BIH arrhythmia database”. In: *IEEE engineering in medicine and biology magazine* 20.3 (2001), pp. 45–50.

[82] STM32L1 Series - STMicroelectronics.

[83] [Online]. Available: URL: https://www.nordicsemi.com/Products/Low-power-short-range-wireless/nRF8000-series.

Saleh Baghersalimi earned a Master’s in Electrical & Electronic Engineering from École Polytechnique Fédérale de Lausanne (EPFL) in Switzerland in 2017. He is now pursuing a Ph.D. at the Embedded Systems Laboratory at EPFL. His research centers on biomedical signal processing, machine learning, and computer vision for low-power wearable devices.

Tomas Teijeiro received his Ph.D. in Artificial Intelligence and Computer Science from the University of Santiago de Compostela, Spain, in 2017. From 2018 to 2022, he was a post-doctoral researcher in the Embedded Systems Laboratory at the EPFL. He is currently a tenure-track researcher at the Basque Center for Applied Mathematics (BCAM), Spain. His research interests include biomedical signal processing, non-monotonic temporal reasoning, and efficient machine-learning applied to low-power wearable devices.

Amir Aminifar is Assistant Professor in the Department of Electrical and Information Technology at Lund University, Sweden. He received his Ph.D. degree from the Swedish National Computer Science Graduate School (CUGS), Sweden, in 2016. During 2016-2020, he held a Scientist position at the Swiss Federal Institute of Technology (EPFL), Switzerland. His research interests are centered around federated and edge machine learning for Internet of Things (IoT) systems, intelligent mobile-health and wearable systems, and health informatics.

David Atienza is a Professor of electrical and computer engineering, Heads the Embedded Systems Laboratory (ESL) and he is the Scientific Director of the EcoCloud Center for Sustainable Computing at the Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland. He received his PhD in computer science and engineering from UCM, Spain, and IMEC, Belgium, in 2005. His research interests include system-level design methodologies for high-performance multi-processor system-on-chip (MPSoC) and low power Internet-of-Things (IoT) systems, including new 2-D/3-D thermal-aware design for MPSoCs and many-core servers, and edge AI architectures for wearable systems and smart consumer devices. He is a co-author of more than 400 papers in peer-reviewed international journals and conferences, one book, and 14 patents in these fields. Dr. Atienza has received among other recognitions the ICCAD 2020 10-Year Retrospective Most Influential Paper Award, the DAC Under-40 Innovators Award in 2018, the IEEE TCCPS Mid-Career Award, and an ERC Consolidator Grant in 2017. He served as DATE 2015 Program Chair and DATE 2017 General Chair. He served as IEEE CEDA President (period 2018-2019) and now is serving as Chair of EDAA (period 2022-2023). He is an IEEE Fellow and an ACM Fellow.