Journal of Artificial Intelligence and Technology, 2025, 5, 84-92
https://doi.org/10.37965/jait.2025.0661

ISTP
RESEARCH ARTICLE

A Privacy-Preserving System for Alzheimer's Disease Detection Based on Federated Learnings

Wenjun Zhang,[1] Shenghui Zhao,[2] and Huibin Wang[2]
[1]School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, 232000, China
[2]School of Computer and Information Engineering, Chuzhou University, Chuzhou, 239000, China

(Received 06 October 2024; Revised 07 November 2024; Accepted 09 November 2024; Published online 02 February 2025)

Abstract: Alzheimer's disease (AD) is a severe neurodegenerative disorder that primarily affects the elderly. Early detection is crucial for enabling timely interventions and slowing the disease's progression. A promising approach for early detection involves analyzing audio data collected from elderly individuals in their homes using Internet of Things (IoT) devices. However, this method presents significant challenges concerning privacy and data security. This paper introduces Efficient Differential Privacy-Alzheimer's Detection (EDP-AD), an efficient and privacy-preserving system. The system utilizes small IoT devices to collect audio data from elderly individuals. By integrating machine learning, federated learning (FL), and differential privacy (DP) techniques, EDP-AD ensures that raw audio data remains local while breaking down data silos, thereby preventing potential privacy leaks during the disease detection process. To enhance system efficiency, a sparse mask updating algorithm based on Top-k is proposed. This algorithm reduces communication overhead by sparsifying the model parameters uploaded by clients within the FL framework. Evaluation on a real-world dataset shows that the system achieves an accuracy rate of 84.48%, reduces communication costs by two-thirds, and provides robust privacy protection while maintaining high efficiency.

Keywords: Alzheimer's disease; differential privacy; federated learning; privacy-preserving

I. INTRODUCTION

In the current societal context, the phenomenon of population aging is becoming increasingly severe [1,2]. The proportion of elderly individuals is steadily rising, along with the prevalence of chronic diseases like Alzheimer's. Alzheimer's disease (AD) is a neurodegenerative disorder that affects memory, cognition, and daily living skills. Early detection and intervention are crucial for slowing disease progression and improving patients' quality of life. Therefore, the early diagnosis and intervention of AD in the elderly have become key research topics in medicine and smart elderly care.

With the rapid development of Internet of Things (IoT) technology, smart elderly care and healthcare have entered a new era [3]. Integrating IoT smart devices with cloud-based data platforms facilitates remote healthcare services for elderly individuals living alone, particularly in the early detection of chronic conditions like AD. IoT devices, such as wearables, can continuously track patients' vital signs and physiological data, enabling more accurate health assessments and supporting clinical decisions [4].

Currently, IoT devices for Alzheimer's detection mostly rely on expensive, specialized medical equipment or sensors, such as wearable sensors. These sensors are not only difficult to deploy but also hard for elderly individuals to accept. Using smaller sensors like smart speakers and microphones to collect voice and text data from seniors makes Alzheimer's detection more efficient. This approach reduces the need for specialized equipment, increases both convenience and acceptance. However, as voice data is highly sensitive personal information, it poses considerable challenges regarding privacy protection during collection, transmission, and processing. Unprotected data may result in privacy breaches or be maliciously exploited. Therefore, effectively utilizing data while ensuring privacy protection has become a critical issue that IoT healthcare systems must address [5].

To address these issues, this paper proposes a privacy-preserving method for AD detection based on IoT technology. The system utilizes an architecture based on federated learning (FL) [6] and differential privacy (DP) [7] techniques. Raspberry Pi or similar computational devices are installed in the homes of elderly individuals living alone. These devices are equipped with microphones and speakers to collect voice data during the elderly's daily activities. By playing prerecorded greetings from their children, the system simulates natural conversation scenarios. The terminal devices capture and process voice data in real time, converting it into spectrograms. Edge devices use lightweight convolutional neural networks (CNNs) to perform localized training on the spectrograms. During the training process, DP techniques protect model parameters to ensure data privacy. The trained model parameters are then securely transmitted to a cloud server, where data from different terminals are aggregated and further trained to form a more accurate global model. The optimized model is subsequently sent back to the edge devices for continuous personalized adjustments and learning, enabling early and precise detection of AD while ensuring privacy protection.

The rest of the paper is organized as follows. Section II presents the related work in AD detection and FL. Section III provides an overview of the system, while Section IV elaborates on the design. Section V outlines the experiment design and results analysis, and Section VI concludes the paper.

Corresponding author: Huibin Wang (e-mail: wanghuibin@chzu.edu.cn).

II. RELATED WORK

A. IoT HEALTHCARE

As AD is a severe chronic condition, IoT-based healthcare offers significant advantages over traditional screening methods, such as professional medical examinations and neuropsychological assessments [8]. It does not require complex medical equipments and can remotely detect AD in elderly individuals, enabling timely intervention by family members or healthcare providers. Individuals can use IoT devices at home for continuous health monitoring and timely intervention [9]. Currently, most research focuses on wearable devices for continuous tracking of health data and behavior [10]. For example, [11] developed a prototype system using wearable IoT devices to provide psychological support for AD patients and ensure secure information transmission for family members to review. In [12], mobile health applications and IoT-based wearable devices were used to assist in the continuous health screening of AD patients. Ebrahem *et al*. [13] developed a small, lightweight, portable IoT prototype to track AD patients in real time and remind them to take their medication through timely alerts. However, wearable sensor devices still face challenges in deployment and acceptance by elderly individuals. As a result, researchers have begun exploring smaller, nonintrusive IoT devices for AD detection and medical services for AD patients. For instance, [14] collected data from Alzheimer's patients using sensors and smartwatches installed in their homes, enabling timely treatment. Li *et al.* [15] improved AD detection accuracy by deploying IoT devices in smart home environments to collect users' audio data and analyze classical speech features. IoT-based AD detection offers more efficient, cost-effective, and convenient solutions, although its accuracy is slightly lower than traditional hospital-based screenings.

B. AD DETECTION METHODS

Currently, AD detection methods are primarily categorized into neuroimaging-based analysis and voice-based or text-based analysis. Neuroimaging-based detection methods diagnose AD by analyzing imaging data such as magnetic resonance imaging (MRI), positron emission tomography (PET), and functional MRI (fMRI). These methods have played a significant role in early AD detection. For example, [16] proposed a deep learning method for predicting AD, achieving excellent results on the fMRI AD imaging dataset. In [17], a depthwise separable CNN model was introduction for AD classification, significantly reducing parameters and computational costs compared to traditional neural networks. Ebrahimi *et al*. [18] proposed a method using transfer learning in 3D-CNN, enabling knowledge transfer from 2D image datasets to 3D image datasets. All the aforementioned methods predict AD stages using single data modalities. In contrast, [19] proposed a holistic approach that integrates multiple data modalities using deep learning, proving superior to traditional machine learning models.

Although image-based methods are effective, collecting such data is costly and often difficult. As a result, many researchers have focused on detecting AD through speech or text analysis. For example, [20] conducted a systematic evaluation of methods for detecting AD in elderly individuals through speech analysis, examining relevant features and diagnostic accuracy. In [21], speech recognition and natural language processing techniques were used to detect AD and assess its severity.

Since most AD detection methods rely on highly sensitive data, such as images or speech, privacy protection has become a critical consideration in this research. Some scholars have explored using FL for AD detection. For example, [22] proposed a hybrid FL framework that utilizes unlabeled data to train deep learning networks while ensuring data privacy protection. The authors also introduced a novel brain region attention network (BANet) that highlights important regions of interest using attention mechanisms. In [23], a hierarchical FL model with adaptive model parameters aggregation were studied to improve learning efficiency. Lakhan *et al*. [24] proposed a novel scheme called Evolutionary Deep Convolutional Neural Networks (EDCNNS), which focuses on convex optimization problems, aiming to minimize computation time while maximizing prediction accuracy for AD. Ouyang *et al.* [25] proposed an end-to-end system integrating multimodal sensors and a novel FL algorithm to detect multidimensional AD digital biomarkers in natural living environments.

III. SYSTEM OVERVIEW

In the Efficient Differential Privacy-Alzheimer's Detection (EDP-AD) framework, there is a global server and n clients participating in training, as illustrated in Fig. 1. The framework consists of the following steps: (1) At the current round t, the server of the community or medical institution initializes a global model w_g and distributes this model to all participating clients. (2) Clients process the collected data by converting the voice data into spectrogram representations and extracting features. DP noise is then added to these features. (3) Each client downloads the global model w_g from the server and trains it using their local data. Upon completion of training, each client obtains updated local model parameters w_i^t, resulting in model parameter updates Δw_i^t. (4) Each client employs a Top-k parameter selection method and adds a mask to perturb the parameter updates, thereby preventing adversaries from accessing the original model and compromising client privacy. The perturbed model parameter updates $\Delta \tilde{w}_i^t$ are then uploaded to the server for aggregation. (5) The server utilizes an aggregation algorithm to aggregate all model updates $\Delta \tilde{w}_i^t$ uploaded by the clients, yielding a new global model. In the next communication round $t + 1$, this updated model w_{t+1} is redistributed to the clients for further training, iterating through the aforementioned steps for t training rounds.

A. DESIGN GOALS

1. **Easy Deployment:** A key design goal of our system is ease of deployment in various home-based elderly care environments. For AD detection, elderly individuals only need to interact with IoT devices (such as Raspberry Pi) via voice, without requiring additional analysis or complex equipment setups.
2. **High Performance:** Due to the limited computational resources of client devices, it is essential to optimize communication overhead while maintaining high accuracy, ensuring overall system efficiency and performance.
3. **Privacy Protection:** It is crucial to protect local model parameters, user raw data, and classification features from exposure, both during data collection and AD detection, as well as during transmission over the network.

B. THREAT MODEL

In our proposed system, we assume the central server is honest, meaning both the clients and the server comply with the protocol.

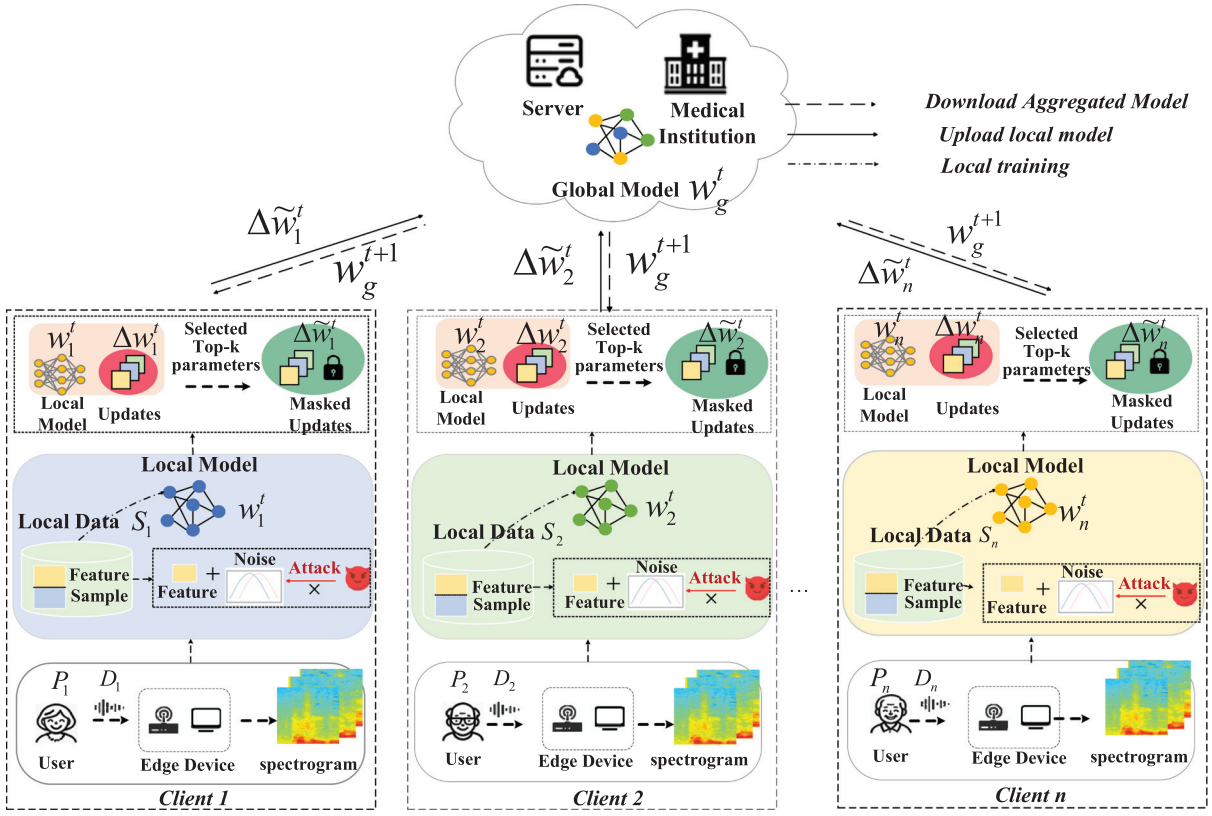

Fig. 1. Figure system architecture diagram.

While clients enhance privacy by training data locally, there remains a potential risk of privacy leakage when uploading trained model parameters to the central server. Specifically, attackers could infer sensitive data from clients by analyzing the uploaded model parameters [26]. To mitigate this risk, we apply DP techniques during local data processing and model updates, ensuring that changes to an individual data point have a minimal impact on the overall output. This effectively prevents data reconstruction attacks and privacy breaches. These measures collectively ensure data security and privacy during transmission. In this paper, we assume attackers cannot infiltrate users' networks to extract raw data. Based on these assumptions, our goal is to protect client privacy throughout the training process and prevent adversaries from compromising it.

IV. SYSTEM DESIGN

This section provides a detailed explanation of the overall system design and the functionality of each component. First, we introduce the formal definitions used in this paper. Table I presents the main symbols used in this work.

Table I. Main symbols

Notations	Explanation
M	Random algorithm
S_i,S_i'	Adjacent datasets
ϵ	Total privacy budget
δ	Relaxation factor
Δf	Sensitivity
N	Number of participating clients
T	Collecting time points of audio recordings from the elderly
t	Current communication round t
w	Model parameter vector after server aggregation
$F_i(w)$	Local loss function of the i-th client
Δw_i^t	Model parameter updates of the i-th client in round t
$\Delta\tilde{w}_i^t$	Masked model parameter updates of the i-th client in round t
$\Delta w_{i,j}^t$	Client i's j-th component of the local model updates in round t
w_g^t	Global model parameters in round t
η	Learning rate

A. DATA COLLECTION AND PROCESSING MODULE

This subsection explains how user voice data is collected and processed. The data collection devices consist of Raspberry Pi units deployed in the homes of elderly individuals living alone. Raspberry Pi [27] is a widely used IoT device in smart elderly care systems due to its small size, easy deployment, and low cost. These devices efficiently meet the requirements for intelligent data collection and provide reliable technical support. As the system's terminal device, the Raspberry Pi, automatically initiates conversations with the elderly at preset times each day to systematically collect data. The device plays prerecorded personalized audio messages from the elderly's children, engaging in daily communication that includes

greetings, reminders, and emotional support. This nonintrusive interaction is easily accepted by the elderly, as it does not disrupt their routines while simplifying data collection and ensuring privacy and comfort. We adopted the audio data processing method proposed by Liu *et al.* [28]. The following section will provide a detailed description of the data collection and processing procedures.

Let $D = \{D_1, D_2, \ldots, D_n\}$ represents all the audio collected from each elderly individual, and $P = \{P_1, P_2, \ldots, P_n\}$ represents the total number of elderly individuals from whom data are collected. That is, P_i represents the i-th elderly individual and D_i represents all the audio data collected from the i-th elderly individual. That is, $\forall D_i \in P_i, 0 < i \leq n$. Here, n represents the total number of elderly individuals from whom data are collected, and D_i represents the audio data collected from the i-th elderly individual at the fixed time point T. It is represented as: $D_i = \{A_1, A_2, \ldots, A_n\}$, an $0 < i \leq n$, $\sum_{i=1}^{m} D_i^{time} \geq 2s$. Where $\forall Group\{A_1, A_2, \ldots, A_m\}$ represents a set of audio segments from D_i as shown. D_i^{time}represents the total duration of audio collected from the i-th elderly individual. Since the collected audio $D_i = \{A_1, A_2, \ldots, A_m\}$ in each session is a continuous segment with varying duration, it is necessary to divide it into multiple subsegments. Where $D_j^{A_i} = \{d_1, d_2, \ldots, d_k\}$ represents the j-th subsegment of the audio segment D_j. That is,

$$\forall Group_j\{d_1, d_2, \ldots, d_m\} \in A_j, \quad \sum_{j=1}^{m}\sum_{i=1}^{k} d_i \geq 2, \quad 0 < j \leq m.$$

where $\forall Group_j\{d_1, d_2, \ldots, d_m\}$ represents a subset of the audio segment collection, and $\sum_{j=1}^{m}\sum_{i=1}^{k} d_i$ represents the total number of audio segments for the elderly individual. The value of k depends on the duration of the audio segments.

B. FEATURE EXTRACTION AND NOISE ADDITION MODULE

This subsection explains how the client extracts features from the data and applies DP to prevent the leakage of classification model parameters. In the AD system, after the terminal collects the audio data, it undergoes preprocessing, where spectrogram features are extracted from each audio subsegment. Random noise is added using DP to obscure internal feature correlations, thereby protecting information privacy. The detailed process is shown in Fig. 2.

The system extracts spectrogram features from the audio subsegments $D_j^{A_i} = \{d_1, d_2, \ldots, d_k\}$, which more clearly represent the variation in speech frequency energy for each segment. $S_j^{A_j} = \{S_1, S_2, \ldots, S_k\}$ represents the spectrogram features of the j-th audio segment from the i-th elderly individual's audio data, where

$$S_j^{A_i} = \{S_1, S_2, \ldots, S_k\} = \{F(d_1), F(d_2), \ldots, F(d_k)\}$$

Where $F(\cdot)$ represents the function used to extract the spectrogram features. For the extracted spectrogram feature dataset, DP is applied to add noise, obscuring the feature information and ensuring privacy protection during the feature extraction phase. In our system, S_i refers to the spectrogram feature dataset of all audio subsegments from the i-th elderly individual. S_i' is the adjacent dataset of S_i, where $\|S_i - S_i'\| \leq 1$. The random algorithm M is used to train the deep neural network, and the parameter space of the network (also referred to as weights or coefficients) is denoted as $Range(M)$. The definition of DP is as follows:

Definition 1(DP). If S and $S_i{}'$ represent adjacent datasets differing by only a single record, and the output space of the random algorithm M is $S \subseteq Range(M)$, then for all output results, if the algorithm satisfies (ϵ,δ)-DP on datasets S and S', the algorithm M satisfies the following equation:

$$\Pr[M(D) \in S] \leq \exp(\epsilon)\Pr[M(D') \in S] + \delta$$

Here, ϵ (privacy budget) controls the level of privacy protection, and δ is the failure probability. If $\delta = 0$, the random algorithm is M said to have strict DP. Since ϵ controls the similarity of the random algorithm's output between two different input datasets, a smaller ϵ indicates a higher level of privacy protection, and vice versa. Common methods to achieve (ϵ,δ)-DP include the Gaussian and Laplace mechanisms, which calculate noise based on the sensitivity of the query function and add it to the output. The definition of sensitivity is provided below:

Definition 2 (Sensitivity). Given two adjacent datasets S and $S' \in S$, which differ by at most one record, and a query function $f: S \in R^d$, the sensitivity of the query function is defined as:

$$\Delta f = \max_{S_i, S_i'} \|f(S) - f(S')\|_2$$

When selecting the DP noise mechanism, the system considered the following two options:

Gaussian Mechanism: For a query function $f: S \to R^d$with sensitivity Δf, noise generated from a Gaussian distribution $N(0,\sigma^2)$ is added to the output of f to satisfy (ϵ,δ)-DP if and only if $\sigma \geq \frac{\sqrt{2\ln(1.25/\delta)}\Delta f}{\epsilon}$. Here, $\epsilon,\delta \in (0,1)$ is the sensitivity of the function f.

Laplace Mechanism: For a query function $f: S \to R^d$ with sensitivity Δf on dataset S, it satisfies $(\epsilon,0)$-DP if the following condition holds:

$$M(S) = f(S) + Laplace(\Delta f/\epsilon)$$

Where the added noise follows the Laplace distribution with a probability density function (PDF) given by:

$$P(x|\lambda) = \frac{1}{2\lambda} e^{-|x|/\lambda}$$

In the proposed EDP-AD system, we utilized both of these methods and conducted corresponding evaluations.

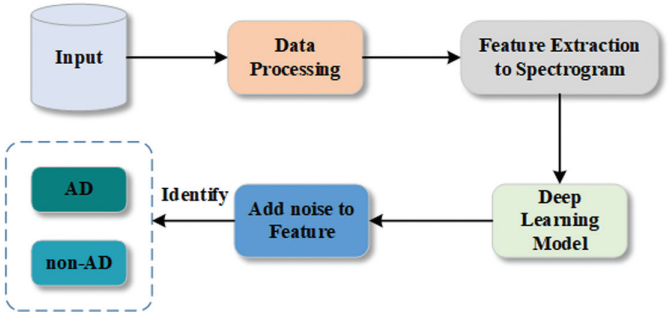

Fig. 2. Data feature extraction and noise addition process.

C. FL-BASED MODULE

In this paper, FL is used to address the issues of data silos and privacy. A FL system consists of a cloud server S, which maintains

the global model w_g^t and multiple clients $C = \{Client_1, Client_2, \ldots, Client_n\}$.

Each client performs multiple iterations using Stochastic Gradient Descent (SGD) to train local model updates Δw_i^t on their spectrogram dataset S_i. The server aggregates all local model updates submitted by the N clients into a global model using a weighted averaging algorithm, based on the size of each client's local dataset, thereby making accurate AD decisions. By keeping client data locally and only transmitting model parameter updates, deploying FL prevents the leakage of users' sensitive information. During each training round, clients provide the server with model parameter updates of the local model instead of directly transmitting user data. The server updates the global model from w_g^t to w_g^{t+1}, where t represents the current round. The following formula outlines the training process of FL from round t to round t+1:

$$w_g^{t+1} = w_g^t + \frac{1}{N}\sum_{i=1}^{N}\frac{S_i}{S}w_i^t$$

where t represents the current training round, N denotes the total number of participating clients, w_i^t indicates the local training parameters of the i-th client in round t, and S_i represents the local spectrogram data. S denotes the total size of the spectrogram data held by all clients.

Client Training Process: For the clients participating in FL, they first need to utilize the collected dataset $D_i = \{(p_1,y_1), (p_2,y_2), \ldots, (p_m,y_m)\}$, $P = \{P_1,P_2,\ldots,P_m\}$ where represents the spectrogram data from all clients. $Y = [y_1,y_2,\ldots,y_m] \in \{0,1\}$ (0 represents the healthy population and 1 represents AD patients). For each client i, its loss function $L(w_i)$ is defined as:

$$L(w_i) = -\frac{1}{m}\sum_{j=1}^{m_i}[y_i \log(\widehat{y}_j + (1-y_j)\log(1-\widehat{y}_j)]$$

where $\widehat{y}_j = \sigma(w_i * p_j)$ is the predicted value of the j-th sample, σ is the activation function, y_j is the corresponding label, and m_i is the data size of client i.

AD Detection Process: After training, the client's model is optimized to minimize the loss function. When voice data is received from the user, the client can predict the result as either AD or healthy (HC).

D. TOP-K-BASED UPDATES SELECTION AND MASKING MODULE

Uploading local model parameters from each client to the server in FL results in significant communication overhead [29]. We propose a mask-based sparse updates strategy to reduce communication costs. Sparsification is a widely used technique in deep learning to improve communication efficiency in distributed training [30–32]. Inspired by previous work, this paper aims to reduce communication overhead in FL by eliminating certain parameter updates, selecting them based on importance while minimizing the impact on model performance. Therefore, we select the model parameter updates with the largest absolute values for model aggregation and updating.

Assume that during the client's updates process, the initial model weights are $w_g(0)$, and the model weights for the i-th client after training with its private dataset in round t are w_i^t, with the corresponding updates being Δw_i^t:

$$\Delta w_i^t = w_i^t - w_g^{t-1}$$

where w_g^{t-1} represents the global model parameters from round $t-1$ and w_i^t represents the locally trained model parameters of client i. The specific parameters in the model are denoted as w, and the absolute value of each component in the updates are calculated as $|\Delta w_{i,j}^t|$:

$$|\Delta w_{i,j}^t| = |w_{i,j}^t - w_{g,j}^t|$$

where j represents the j-th component of the updates vectors. Let Λ_i^k represent the k maximum values from the parameters updates set$\{Top_k(|\Delta w_{i,j}^t|)|w_{i,j}^t \in w_i^t\}$ of the i-th client. To sparsify the client's updates parameters Δw_i^t, we use a masking function to generate a 0–1 mask matrix M_i, which is a binary vector of the same dimension as Δw_i^t:

$$\mathrm{M}_i = \begin{cases} 1, & if\ j \in \Lambda_i^k \\ 0, & otherwise \end{cases}$$

Thus, after applying the mask matrix M_i to the updated vectors Δw_i^t, the sparsified updated vectors $\Delta\tilde{w}_i^t$ are obtained, that is,

$$\Delta\tilde{w}_i^t = \mathrm{M}_i \odot \Delta w_i^t$$

where $\odot$ represents the Hadamard product. After sparsification, the top k largest updates components $Top_k(\Delta w_{i,j}^t)$ are retained, while the other values are set to zero. As a result, the sparsified updates vectors $\|\Delta\tilde{w}_i^t\|$ will always have fewer nonzero elements than the original vectors $\|\Delta w_{i,j}^t\|$. By adjusting the value of k, the sparsity of the local updates can be controlled, improving the efficiency of uploading model updates. Table II presents Algorithm 1, which is based on the Top-K sparsification mask.

Table II. Top-K-based updates and mask selection algorithm

Algorithm 1 Federated learning with Top-*k* updates selection and masking

Input: Initial global model parameters $\mathrm{w}_g(0)$, number of clients N, number of local epochs E, selection threshold k
Server initializes: $\mathrm{w}_g(0)$
Server sends $\mathrm{w}_g(0)$ **to all clients**
for each communication round $t = 1, 2, \ldots$ **do**
  Client side:
  for each client $i \in [N]$ in parallel do
    Client receives w_g^t
    Client trains on local data:
    $w_i^t \leftarrow DeviceLocalUpdates(w_g^t, S_i)$
    Client computes local updates:
    $\Delta w_i^t \leftarrow w_i^t - w_g^t$
    Client selects Top-*k* updates:
    $\Lambda_i^k \leftarrow Top(|\Delta w_{i,j}^t|)$
    Client generates mask:
    $\mathrm{M}_i = \begin{cases} 1, & if\ j \in \Lambda_n^k \\ 0, & otherwise \end{cases}$
    Client applies mask to the updates:
    $\Delta\tilde{w}_i^t \leftarrow M_i^{\circ}\Delta w_i^t$
    Client sends $\Delta\tilde{w}_i^t$ **to the server**
  end for
  Server side:
  Server aggregates masked updates:
  $w_g^{t+1} \leftarrow w_g^t + \eta\sum_{i=1}^{N}\Delta\tilde{w}_i^t$
end for

V. EXPERIMENTS AND PERFORMANCE ANALYSIS

In this section, we first introduce the experimental setup, including the dataset, model architecture, and algorithm. Then, we evaluate the system's performance in terms of privacy levels, number of participants, and communication overhead.

A. DATASET AND SETTINGS

The dataset used in this paper is the Voice-Based Spectrogram Dataset (VBSD), a real-world dataset collected by our team and published in [28]. It comes from wearable IoT devices, with each audio sample having a frequency of 44.1 kHz and a duration of 1 second. Spectrogram features are extracted from the audio data and fed into the neural network model. The authors of [28] extracted 254 speech samples from AD patients and 250 from healthy controls (HCs), collected from 36 participants, resulting in a total of 504 speech samples and corresponding spectrogram features.

Table III shows the age and gender distribution of the collected data. The ages of AD patients range from 65 to 94 years, with 23 speech samples collected from AD patients, from which 254 spectrogram features were extracted. Additionally, speech data were collected from 13 healthy elderly individuals aged 65 to 92 years, resulting in 250 spectrogram features.

In this paper, classification accuracy, precision, recall, F1 score, communication overhead, and privacy protection strength are used as evaluation metrics for the model. We build a neural network model using the PyTorch deep learning framework with Python version 3.10.14. Experiments are conducted on a 64-bit Ubuntu 22.04.3 system using a single NVIDIA GeForce RTX 3080 Ti GPU with 64 GB of memory.

B. EXPERIMENTS PERFORMANCE COMPARISON AND ANALYSIS

1. Comparison of System Recognition Accuracy Across Different Models: In the experiments for the AD detection system, we have evaluated three neural network architectures: ResNet18, MobileNet, and EfficientNet on the VBSD dataset. The learning rate is set to 0.001, and distributed training is conducted with three clients. As shown in Fig. 3, among these models, ResNet18 demonstrated significant performance advantages. As the number of training rounds increases, the accuracy approaches 95%, reaching a maximum of 95.1%. In contrast, the accuracy of MobileNet and EfficientNet is lower compared to the ResNet18 model, with maximum accuracies of 77.5% and 81.4%, respectively. The results indicate that the ResNet18 model used in FL performs better on the actual dataset. In FL, the training accuracy curve exhibits certain fluctuations, primarily due to uneven data distribution and differences in client training environments. Different clients may have distinct feature distributions in their data, leading to fluctuations during model updates aggregation. Additionally, variations in computational resources and training epochs among clients can result in inconsistent updates quality, further exacerbating performance fluctuations after model merging. These factors collectively contribute to the fluctuations in the accuracy curve during the FL training process, and the ability of ResNet18 to maintain high performance in such an environment, thanks to its stable architecture, is a key reason for our choice of this model.

Table III. VBSD dataset

Group	Age range	Males	Females
AD	65–94	10	13
HC	65–92	5	8

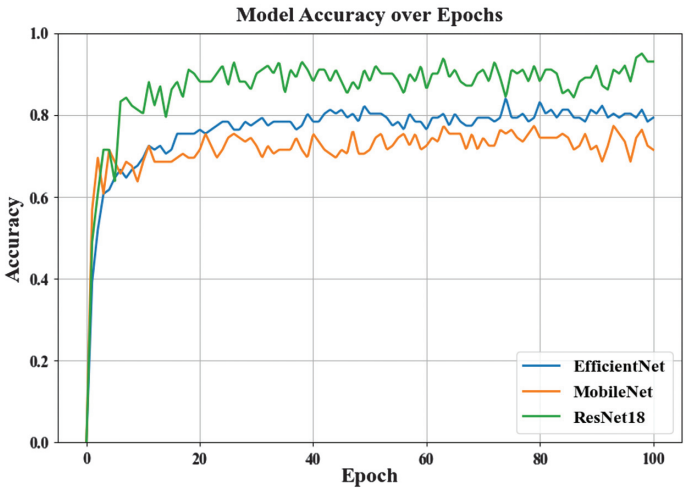

Fig. 3. AD detection accuracy across different models.

2. The Impact of Different Numbers of Clients on System Recognition Accuracy: As shown in Fig. 4, by comparing the performance of ResNet18, MobileNet, and EfficientNet with different numbers of clients, we analyze the impact of client numbers on AD detection accuracy in a FL environment. The study finds that ResNet18 consistently have achieved the highest accuracy across all configurations, particularly when the number of clients is 4, where its accuracy approached 90%, significantly higher than the other networks.

This result is attributed to ResNet18's residual connection structure, which effectively maintains gradient flow stability and addresses the challenges of distributed data training. Additionally, we observed that as the number of clients increased, the accuracy of all networks fluctuated slightly but tended to stabilize overall. This suggests that the non-independent and identically distributed (Non-IID) nature of the data and communication efficiency between clients significantly impact model performance. As the number of clients increases, model aggregation becomes more complex.

3. Evaluation Results for Different Models: As shown in Table IV, we have evaluated three neural network models—EfficientNet b0, MobileNet v2, and ResNet18—on four metrics: accuracy, precision, recall, and F1 score, within the FL system. These metrics are essential for assessing the effectiveness of models in applications requiring reliable classification. Upon observation, it is clear that ResNet18 is the most robust model, performing well across all metrics and proving highly effective in a distributed environment. In contrast, EfficientNet and MobileNet show limitations in classification capabilities, making them less effective when handling complex and highly heterogeneous data.

The Impact of Differential Privacy on the Model Under Different Privacy Budgets: To evaluate the impact of the DP

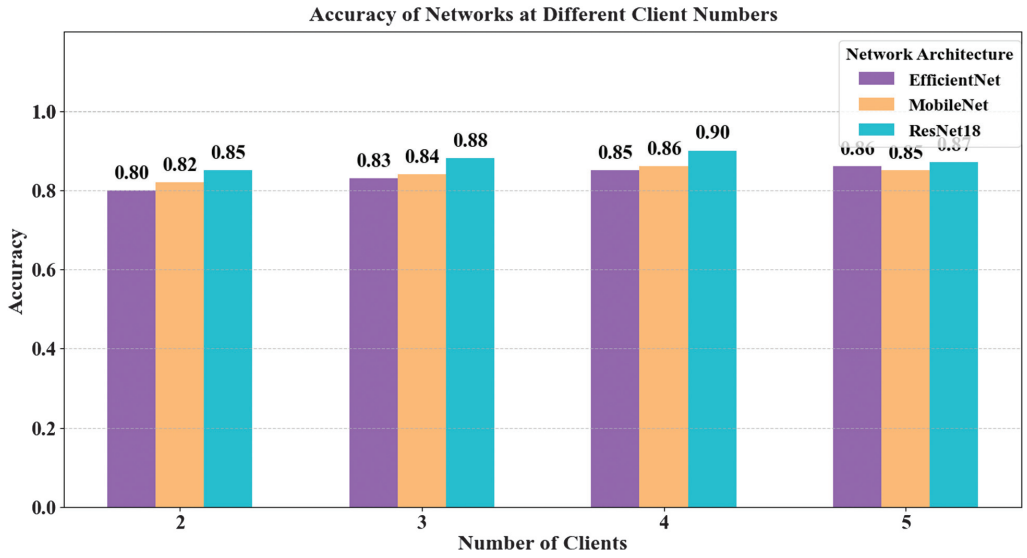

Fig. 4. The impact of client numbers on accuracy across different models.

Table IV. Evaluation results of the system across different models

Model type	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EfficientNet b0	82.35	83.43	84.74	84.43
MobileNet v2	76.47	77.13	87.52	76.43
ResNet18	**95.10**	**92.83**	**93.17**	**94.27**

mechanism on the system, we apply both the Laplace and Gaussian mechanisms with different privacy budgets to the ResNet18 model. As shown in Table V, both mechanisms exhibit a decline in accuracy as the privacy budget decreases, due to the increased noise added to enhance data protection. Additionally, the results show that the Gaussian mechanism achieves better accuracy compared to the Laplace mechanism.

Furthermore, we observed that even with a privacy budget of 0.1, the system can still maintain an accuracy of 80.02%.

4. Communication Overhead Across Different Models

To evaluate the impact of the Top-k updates and masking selection algorithm on communication, we measured the communication overhead incurred by data uploads. As shown in Fig. 5, the communication overhead of the entire system reached up to 13,426 MB without any sparsification applied to the model. By adjusting different values of k for comparison, we found that when k=0.3, the communication cost for the ResNet18 model was significantly reduced to 4,027 MB, averaging a two-thirds reduction in communication volume while still achieving an accuracy of 84.48%. It is evident from the figure that although the lightweight models EfficientNet b0 and MobileNet v2 have significant advantages in terms of communication overhead, their accuracy in AD recognition is quite poor. The communication volume for the EfficientNet_b0 network reached 4,863 MB, with an accuracy of 82.35%. In contrast, with $k = 0.3$, ResNet not only reduced communication volume but also improved AD detection accuracy by 2.13%. This demonstrates the effectiveness of the Top-k updates and masking selection algorithm in balancing communication efficiency and model performance. This not only enhances the applicability of the model in resource-constrained environments but also provides a feasible technical pathway for developing efficient and accurate machine learning models.

Table V. Evaluation results of the system across different models

Epsilon setting	Type	Accuracy (%)	Type	Accuracy (%)
$\epsilon = 0.1$	Gaussian	80.02	Laplace	76.67
$\epsilon = 0.5$	Gaussian	81.57	Laplace	78.89
$\epsilon = 1.0$	Gaussian	82.19	Laplace	80.14
$\epsilon = 1.5$	Gaussian	85.56	Laplace	82.37
$\epsilon = 2.0$	Gaussian	83.24	Laplace	82.67
$\epsilon = 2.5$	Gaussian	84.41	Laplace	83.32
$\epsilon = 3.0$	Gaussian	88.89	Laplace	84.75

VI. CONCLUSION

In this paper, we proposed a privacy-preserving AD detection system based on FL, designed for low-cost AD detection. The system utilized IoT devices, such as Raspberry Pi, to collect and preprocess audio data, while employing DP mechanisms and a FL framework to prevent raw data and model parameters from leaking during transmission. Additionally, the Top-k-based sparsification strategy reduced communication overhead. Experiments demonstrated that the system was highly efficient, lightweight, and ensured privacy protection.

For future work, we plan to deploy this system in real-world environments for broader field testing to evaluate its performance and practicality in everyday settings. We also aim to explore more advanced machine learning algorithms, to improve diagnostic accuracy, and to consider incorporating additional sensor data, such as video and physiological signals, for more comprehensive

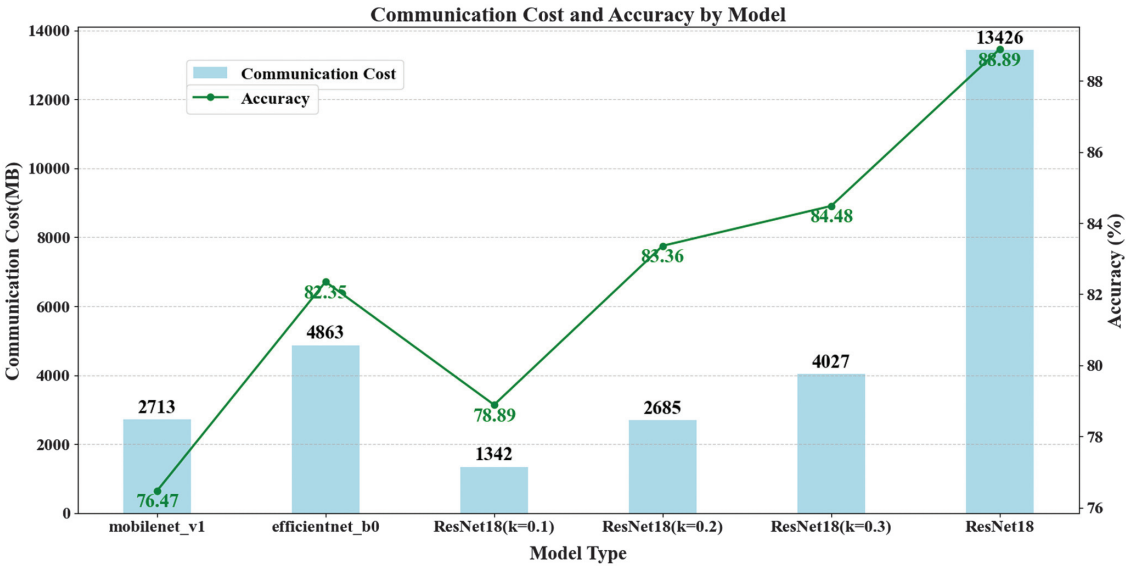

Fig. 5. Communication cost and accuracy by model.

symptom monitoring and analysis. Furthermore, we will focus on optimizing data synchronization and model updates processes to reduce energy consumption and enhance responsiveness, making the system more suitable for resource-constrained devices.

ACKNOWLEDGMENTS

This work is supported by the Major Program of Natural Science Research Foundation of Anhui Provincial Education Department: Research on Key Technologies of Mobile Crowdsensing Privacy Protection in Smart Community Environments (2022AH040148). The authors thank the support from the key project at the school level of Chuzhou University: Research on the Smart Elderly Care Collaborative Care Platform (2022AH051084) and the key scientific research project of the Anhui Provincial Department of Education: Research on Alzheimer's Disease Detection Based on Visual Features (2022XJZD10).

CONFLICT OF INTEREST STATEMENT

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] National Bureau of Statistics of China, "Statistical bulletin of the People's Republic of China on national economic and social development in 2023 (in Chinese)[J]," *China Stat.*, vol. 03, pp. 4–21, 2024.

[2] Z. W. Zhai, J. J. Chen, and L. Li, "China's population and aging trends from 2015 to 2100," *Popul. Stud.*, vol. 41, pp. 60–70, 2017.

[3] J. Mistry and A. Ganesh, "An analysis of IoT-based solutions for congenital heart disease monitoring and prevention," *J. Xidian Univ.*, vol. 17, pp. 325–334, 2023.

[4] P. Kulurkar, et al., "AI based elderly fall prediction system using wearable sensors: a smart home-care technology with IOT," *Meas. Sens.*, vol. 25, pp. 100614, 2023.

[5] D. Dhinakaran, et al., "Privacy-preserving data in IoT-based cloud systems: a comprehensive survey with AI integration," arXiv preprint arXiv:2401.00794, 2024.

[6] M. Brendan, et al., "Communication-efficient learning of deep networks from decentralized data," *Artif. Intell. Stat.*, vol. 54, pp. 1273–1282, 2017.

[7] C. Dwork, et al., "Differential privacy," in *International Colloquium on Automata, Languages, and Programming*. Berlin, Heidelberg: Springer, 2006, pp. 1–12.

[8] R. Creaney, L. Reid, and M. Currie, "The contribution of healthcare smart homes to older peoples' wellbeing: a new conceptual framework," *Wellbeing, Space Soc.*, vol. 2, p. 100031, 2021.

[9] L. Fang, et al., "A practical model based on anomaly detection for protecting medical IoT control services against external attacks," *IEEE Trans. Ind. Inf.*, vol. 17, no. 6, pp. 4260–4269, 2020.

[10] S. Nasiri, M. R Khosravani, and R. Mohammad, "Progress and challenges in fabrication of wearable sensors for health monitoring," *Sens. Actuat. A: Phys.*, vol. 312, p. 112105, 2020.

[11] N. Surantha, P. Atmaja, and M. Wicaksono, "A review of wearable internet-of-things device for healthcare," *Proc. Comput. Sci.*, vol. 179, pp. 936–943, 2021.

[12] W. Salehi, et al., "IoT-based wearable devices for patients suffering from Alzheimer disease," *Contrast Media Mol. Imaging*, vol. 2022, no. 6, pp. 3224939, 2022.

[13] A. T. Ebrahem, et al., "Using IoT technology for monitoring Alzheimer's and elderly patients," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 31, no. 2, pp. 986–994, 2023.

[14] R. J. Oskouei, et al., "IoT-Based Healthcare Support System for Alzheimer's Patients," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 2020, no. 1, p. 8822598, 2020.

[15] J. Li, et al., "A federated learning based privacy-preserving smart healthcare system," *IEEE Trans. Ind. Inf.*, vol. 18, no. 3, pp. 2021–2031, 2021.

[16] M. Odusami, et al., "Analysis of features of Alzheimer's disease: detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network," *Diagnostics*, vol. 11, no. 6, p. 1071, 2021.

[17] J. Liu, et al., "Alzheimer's disease detection using depthwise separable convolutional neural networks," *Comput. Methods Progr. Biomed.*, vol. 203, p. 106032, 2021.

[18] A. Ebrahimi, S. Luo, and R. Chiong, "Introducing transfer learning to 3D ResNet-18 for Alzheimer's disease detection on MRI images," in

2020 35th Int. Conf. Image Vis. Comput. New Zealand (IVCNZ), IEEE, 2020, pp. 1–6.

[19] H. A. Helaly, M. Badawy, and A. Y. Haikal, "Deep learning approach for early detection of Alzheimer's disease," *Cognit. Comput.*, vol. 14, no. 5, pp. 1711–1727, 2022.

[20] R. P. Filiou, et al., "Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: a scoping review," *Aphasiology*, vol. 34, no. 6, pp. 723–755, 2020.

[21] R. Pappagari, et al., "Using state of the art speaker recognition and natural language processing technologies to detect Alzheimer's disease and assess its severity," in *Interspeech*, ISCA-International Speech Communication Association, 2020, pp. 2177–2181.

[22] B. Lei, et al., "Hybrid federated learning with brain-region attention network for multi-center Alzheimer's disease detection," *Pattern Recognit.*, vol. 153, p. 110423, 2024.

[23] Z. Chen, "A hierarchical federated learning model with adaptive model parameter aggregation," *Comput. Sci. Inf. Syst.*, vol. 12, no. 3, pp. 75–85, 2023.

[24] A. Lakhan, et al., "EDCNNS: federated learning enabled evolutionary deep convolutional neural network for Alzheimer disease detection," *Appl. Soft Comput.*, vol. 147, pp. 110804, 2023.

[25] X. Ouyang, et al., "ADMarker: a multi-modal federated learning system for monitoring digital biomarkers of Alzheimer's disease," in *Proc. 30th Annu. Int. Conf. Mobile Comput. Netwk.*, ACM, 2024, pp. 404–419.

[26] Z. Wu, et al., "Detecting converted speech and natural speech for anti spoofing attack in speaker recognition," in *13th Annu. Conf. Int. Speech Commun. Assoc. (Interspeech 2012)*, ISCA-International Speech Communication Association, 2012, pp. 1–4.

[27] E. Upton, et al., "Raspberry Pi user guide," John Wiley & Sons, 2016.

[28] L. Liu, et al., "A new machine learning method for identifying Alzheimer's disease," *Simul. Model. Pract. Theory*, vol. 99, p. 102023, 2020.

[29] Z. Chen, et al. "A hierarchical federated learning model with adaptive model parameter aggregation," *Comput. Sci. Inf. Syst.*, vol. 12, no. 3, pp. 75–85, 2023.

[30] Y. J. Lin, et al., "Deep gradient compression: reducing the communication bandwidth for distributed training," arXiv preprint arXiv:1712.01887, 2017.

[31] F. Sattler, et al., "Sparse binary compression: Towards distributed deep learning with minimal communication," in *2019 Int. Joint Conf. Neural Netwk (IJCNN)*, IEEE, 2019, pp. 1–8.

[32] Y. Tsuzuku, et al., "Variance-based gradient compression for efficient distributed deep learning," arXiv preprint arXiv:1802.06058, 2018.