

参赛学生姓名: Simon Leonardo Liu

中学: 上海中学国际部

省份: 上海

国家/地区: 中国

指导老师姓名: 姚艳婕, 陈智能

指导老师单位: 上海中学国际部, 复旦大学

论文题目: Beyond Reactive Assistance:

PV-Care Using Low-Density EEG and AI to

Provide Proactive, Context-Aware Help for MCI

Beyond Reactive Assistance: PV-Care Using Low-Density EEG and AI to Provide Proactive, Context-Aware Help for MCI

Simon L Liu

Abstract—The growing elderly population gives rise to an urgent need for intelligent support systems, particularly for individuals with Mild Cognitive Impairment (MCI). This paper presents PV-Care, a proactive AI-driven assistance scheme that integrates wearable electroencephalogram (EEG) sensing with visual environmental perception to provide real-time, context-aware voice assistance for MCI users. Unlike traditional assistant systems that passively wait for user commands, PV-Care actively initiates helpful interactions based on the user’s detected brain states— Learning, Memory Recall, or Resting, using a novel deep neural architecture named Spatial and Frequency Refinement Network (SFR-Net). By combining EEG-based cognitive state recognition with AI-based analysis of visual data, PV-Care generates structured “4W-UT” prompts to control the output of large language models (LLMs) such as ChatGPT. Simulation results and user studies validate the high accuracy of the proposed SFR-Net and the effectiveness of PV-Care’s naturalistic, context-aware assistance, demonstrating that PV-Care is a feasible solution for supporting MCI patients in daily living scenarios.

Index Terms—Proactive Service, Mild Cognitive Impairment (MCI), Electroencephalogram (EEG) Signals, Visual Perception, LLM

I. INTRODUCTION

THE growing elderly population necessitates innovative approaches to assist senior individuals, especially those experiencing mild cognitive impairment (MCI). MCI patients often encounter problems such as recognizing people or forgetting back to home. Thus, caring for those with MCI always demands personal companions, which requires considerable effort in terms of time, money, and patience [1].

Human-machine collaboration techniques have garnered significant attention, especially with the advancements in Artificial Intelligence (AI) technologies. This collaboration promotes the idea that machines can serve as embodied AI assistants that meet human needs and assist them with various tasks [2]. Silver et al. [3] developed a learning algorithm capable of mastering complex tasks through self-learning, further improving the ability to adapt to users’ needs. AI companion robots, as explored by Clara et al. [4], have been shown to mitigate loneliness among elders, offering emotional support and improving their mental well-being. Additionally, Cantone et al. [5] proposed integrating AI with autonomous robots and sensors to achieve secure and independent living for elderly individuals, which has been particularly beneficial for MCI patients. Zhou et al. [6] demonstrated how an assistant

robot could enhance the perceived communication quality for people with MCI. Saunders et al. [7] introduced a user-personalized companion robot system, and the robot could learn the user’s preferences with AI methods.

Despite these advancements, existing solutions for elder care typically rely on users voluntarily issuing commands for interaction with AI robots or services. Due to cognitive decline, MCI patients often cannot correctly command assistant systems when they need assistance. Therefore, it is crucial to monitor the needs of MCI patients and proactively provide suitable assistance [8].

Electroencephalography (EEG) signals have been widely adopted to analyze users’ brain activities, which can be used to identify whether an MCI patient needs help from the assistant systems. According to Schumacher et al. [9], the EEG features associated with key brain states, such as the attentive state and recall state, are consistent for both healthy individuals and MCI patients. Hence, utilizing EEG analysis to determine the brain state is a promising approach to analyzing the needs of MCI patients. The results of such analysis can be integrated into MCI assistance systems to enable proactive service delivery, such as proactively prompting MCI patients about tasks they may struggle to remember or complete.

Various AI methods have been applied to the processing of EEG data. For example, Yue et al. [10] proposed a temporal-frequency hierarchical transformer network to capture key information from EEG signals. Yao et al. [11] introduced TCNN, which leverages positional encoding and multi-head attention to extract channel information. Wang et al. [12] utilized a CNN-LSTM architecture to progressively extract temporal-spatial features through temporal and spatial convolutions.

However, these methods typically rely on EEG headsets with 32 or more channels. A major practical challenge is the inconvenience of wearing such complex equipment. In practical applications, convenient wearable EEG devices can often only sense a few channels, such as 4 or 6. To enable EEG-based MCI assistive systems under these constraints, more advanced solutions are demanded for robust recognition of brain states using wearable EEG devices.

Beyond the limitations of EEG analysis, existing assistive solutions for individuals with MCI remain constrained in their ability to proactively initiate supportive services based on environmental context for context-aware interaction. With the rapid advancement of electronic hardware, such as compact EEG and visual sensing devices, and breakthroughs in large language models (LLMs), the development of more intelligent

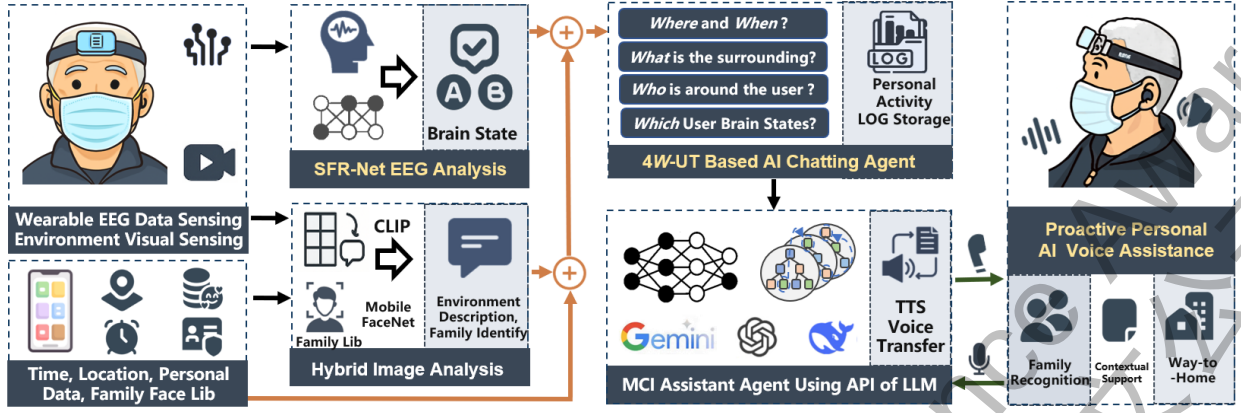


Fig. 1. The Proposed Proactive MCI Care Scheme: PV-Care

MCI support systems has become increasingly feasible. This paper introduces PV-Care, an AI-powered assistance scheme for individuals with MCI, featuring real-time brain state monitoring to enable proactive, context-aware conversational support—particularly valuable when users are unable to explicitly request help. The main contributions of this work are summarized as follows:

- We propose PV-Care, a proactive assistance scheme designed to support the daily lives of individuals with MCI. The scheme integrates EEG signal analysis, visual sensing, and an AI-based conversational agent that emulates a caring family voice. Specifically, PV-Care proactively delivers assistance based on the user's detected brain activity.
- An innovative method for limited-channel EEG analysis is introduced. We propose Spatial and Frequency Refinement Network (SFR-Net) to enable accurate identification of an individual's *Resting/Learning/Memory-Recall* states. This SFR-Net effectively detects brain activity states using only a few EEG channels (e.g., 4 channels) from wearable devices, providing a solid basis for proactively delivering services in our assistance scheme.
- We propose an MCI-oriented conversational agent powered by a large language model (LLM). It integrates environmental information, EEG signals, and other contextual data to generate a structured "4W-UT" prompt, which guides the LLM-based agent in providing voice-based assistance to individuals with MCI. Extensive simulations and user studies demonstrate the high accuracy of the proposed brain activity recognition method and the overall effectiveness of the PV-Care scheme.

II. OUR PROPOSAL

A. The Overall Architecture of Proposed Proactive MCI Care Scheme

The proposed scheme, PV-Care, integrates multiple key modules. Figure 1 illustrates the overall architecture. By utilizing wearable EEG sensing and visual sensing modules, PV-Care enables proactive, context-aware interactions tailored to the user's brain activity.

The Environmental Visual Sensing module captures real-time images of the user's surroundings and employs image

understanding models such as CLIP [13] to generate descriptive textual representations of the environment. Additionally, if a person is present in the environmental images, we utilize a family member face database stored on the user's smartphone and employ a lightweight model, Mobile FaceNet [14] to accurately recognize family members. Meanwhile, the wearable EEG sensing module collects EEG data for SFR-Net to recognize the user's cognitive state, such as *Learning*, *Resting*, or *Memory Recall*. Textual descriptions from the visual module and cognitive state information analyzed from EEG data are integrated with additional contextual details, including time, location, and personal activity logs. These combined descriptions are then provided to a chatting agent powered by LLMs, such as ChatGPT, enabling contextually relevant and personalized interactions. By organizing the different functional modules in this way, PV-Care can proactively provide timely and precise assistance to MCI users based on their identified brain states.

For user convenience and comfort, EEG and visual sensing capabilities are integrated into a single wearable device paired with a smartphone. The smartphone's earphones facilitate voice-based interactions between the user and the AI assistant. Using Text-to-Speech (TTS) technology, conversational texts generated by the assistant are converted to voice responses and delivered through the earphones. Furthermore, since MCI patients often place greater trust in familiar family members, voice style transfer techniques [15] can be applied to simulate their voices, thereby enhancing users' emotional engagement and trust in PV-Care.

The detailed methods and implementations of each module of the proposed PV-Care scheme are presented in the following sections.

B. Visual Sensing Module and Image-to-Environment Description

1) *Visual to Textual Environment Description*: In the proposed PV-Care scheme, the Visual Sensing Module captures real-time environmental images in front of the user, and the images are sent to the user's smartphone and further forwarded for cloud-based image AI analysis. Recent advances in AI models have demonstrated remarkable performance in interpreting image content and converting it into textual

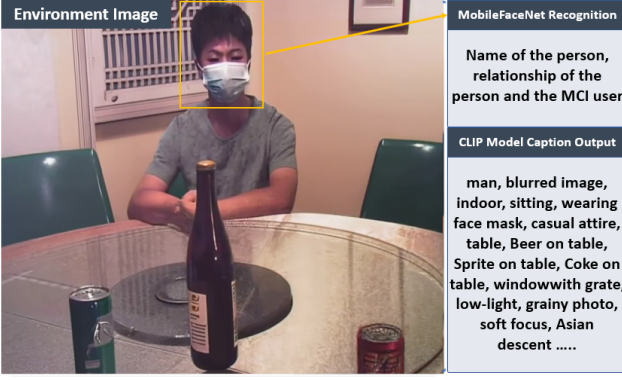


Fig. 2. Example of a sensed image and the corresponding environmental description generated by the cloud AI model.

descriptions. In our scheme, OpenAI's GPT-4 Vision model [16] is employed to analyze the environmental images, and the captured images are uploaded using the method described in [17], which processes the images and generates descriptive captions. For example, as illustrated in Figure 2, a captured image is converted into the caption: "man, blurred image, indoor, sitting, wearing face mask, casual attire, table, beer on table, Sprite on table, Coke on table..."

2) *Smartphone-Based Face Recognition*: As shown in Figure 2, when people appear in the environmental images, PV-Care should identify them. While cloud-based models can generate detailed environmental descriptions, they cannot reliably recognize specific individuals, such as family members, which MCI patients often need.

Typically, an MCI patient only needs to recognize a few dozen familiar individuals. To achieve this, PV-Care employs a smartphone-optimized lightweight face recognition model [14]. Users can configure a personal face database, enabling the smartphone to accurately and efficiently identify them. This local processing ensures both low latency and strong privacy protection, while providing a personalized recognition experience tailored to each user.

Overall, the Visual Sensing Module captures the user's surrounding environment. General objects and scenes are processed by the cloud-based model to generate descriptive captions, while the smartphone-based face recognition model accurately identifies familiar individuals in the scene. This hybrid image analysis leverages the advantages of cloud-based AI and on-device recognition to achieve both comprehensive environmental understanding and personalized assistance.

C. Spatial and Frequency Refinement Network (SFR-Net) for EEG Analysis

The PV-Care is designed to provide proactive services based on the user's brain activity. For instance, when a user attempts to recall information, PV-Care can offer appropriate suggestions without the user's explicit command. Therefore, accurate analysis of brain activity is essential for enabling proactive assistance. In our framework, we use EEG signals to analyze three brain activity states: learning, memory recall, and resting.



Fig. 3. Experiment for collecting EEG during learning activities.

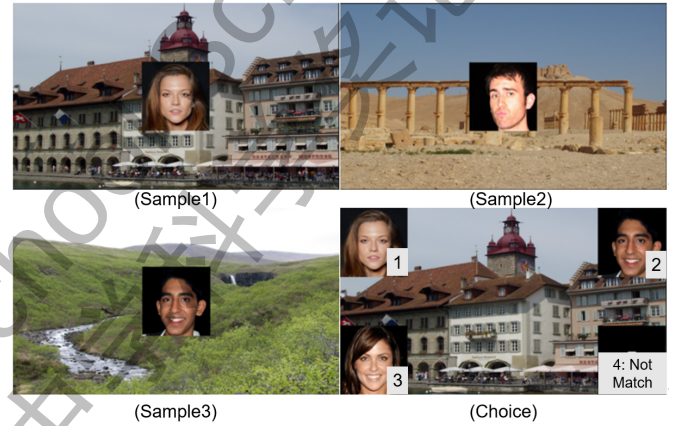


Fig. 4. Experiment for collecting EEG during memory recall activities (participants matched background images with memorized faces).

1) *EEG Signal Acquisition and Experimental Dataset*: EEG signal processing is inherently influenced by both the sensing device and individual variability. To support the development and validation of PV-Care, we designed a dedicated EEG data acquisition experiment incorporating both wearable and high-density EEG devices. Specifically, we employed the portable Muse EEG headband [18], which provides 4-channels (AF7, AF8, TP9, and TP10), alongside a standard 64-channel Neuracle EEG system [19]. By simultaneously recording data from both devices and aligning their signals, we constructed a dataset to enhance the performance of our proposed SFR-Net EEG recognition model. To ensure temporal synchronization, both devices were sampled at 256 Hz.

More than 30 volunteers participated in our EEG sensing experiments to establish a comprehensive dataset. Each participant wore both the wearable 4-channel EEG device and the standard 64-channel EEG device throughout the entire data collection process, as illustrated in the lower-right part of Figure 3. The 64-channel EEG data is primarily used for validation. Classical EEG analysis methods were applied to the 64-channel data to confirm its reliability and ensure the accuracy of the experimental process.

The EEG data acquisition process was designed as follows: 3-Minute Resting State Session → 5-Minute Learning Task →

3-Minute Resting State Session → 5-Minute Memory Recall Task → 3-Minute Resting. This experimental protocol ensures that the necessary data for all three states are collected in a single session, adhering to standard EEG data collection procedures. The specific tasks for each state are described as follows:

- *Resting State*: Participants were instructed to remain seated and relaxed with their eyes open for approximately 3 minutes to establish a baseline EEG signal. This resting state data serves as a reference for each subject.
- *Learning State*: Participants were presented with a visual learning task. They watched a short news video with highlighted text (Figure 3), and participants were required to understand and learn the highlighted content. EEG data collected during this period were labeled as the learning state.
- *Memory-Recall State*: After a 3-minute resting interval, participants were asked to memorize a series of face-background pairs. During the recall phase, background images were presented, and participants were instructed to select the corresponding memorized face. For example, as shown in the lower-right “Choice” panel of Figure 4, the correct selection is label 1, which matches the face-background pair in the upper-left “Sample 1” panel. When the participant correctly selected the face corresponding to the original sample image, the participant’s EEG data were considered valid memory recall data.

After completing the EEG data collection, we first confirmed the validity of the data. Since traditional EEG analysis methods are well-established and reliable, we initially applied classical EEG analysis techniques (methods of Ref. [20] [21]) to process the 64-channel EEG data. If these traditional methods can accurately distinguish the three brain states, it indicates that the participants have properly and diligently completed the tasks as required.

2) *Spatial and Frequency Refinement Network for EEG Analysis*: After validating the EEG signal dataset, we proposed a deep learning architecture named Spatial and Frequency Refinement Network (SFR-Net). This model analyzes 4-channel EEG signals and classifies three brain activity states—*Resting*, *Learning*, and *Memory Recall*. The SFR-Net operates on 4-second EEG signal segments, with a preprocessing pipeline consisting of baseline correction, bandpass filtering (0.5–60 Hz), and artifact removal.

As shown in Figure 5, our SFR-Net model integrates multi-resolution temporal EEG signals with frequency-domain features through a multi-branch neural architecture, enabling robust and hierarchical fusion for enhanced brain state recognition. Let C be the number of EEG channels (electrodes) and L the number of time samples in a 4 s segment sampled at $f_s = 256$ Hz, so $L = 1024$. The proposed SFR-Net processes three complementary EEG representations as inputs:

- High-resolution temporal signal $\mathbf{X}_{HT} \in \mathbb{R}^{C \times L}$ (sampled at 256 Hz to capture fine-grained temporal dynamics).
- Frequency-domain features $\mathbf{X}_F \in \mathbb{R}^{C \times L}$ are derived from the High-resolution temporal signal $\mathbf{X}_{HT}$ via the Short-Time Fourier Transform (STFT). The STFT is

applied to each channel of the input signal $x(n)$ with a sampling rate of 256 Hz, employing a Hann window of length $N = 256$ samples (corresponding to 1 second) and an overlap of 147 samples, resulting in a hop length of 109 samples. This yields 8 time frames for a 4-second EEG segment of length $L = 1024$.

The STFT is computed as:

$$\mathbf{X}(m, \alpha) = \sum_{n=0}^{N-1} x(n + mH)w(n)e^{-j2\pi\alpha n/N}, \quad (1)$$

where $w(n)$ is the Hann window function, m is the frame index, $H = 109$ is the hop length, and $\alpha = 0, 1, \dots, N/2$ indexes the frequency bins (yielding 129 bins from 0 to 128 Hz).

The magnitude spectrogram is computed, the DC component (0 Hz) is discarded, retaining 128 frequency bins. The resulting 128×8 matrix per channel is normalized, transposed and flattened into a 1024-point vector to form $\mathbf{X}_F$ across C channels.

- Low-resolution temporal signal $\mathbf{X}_{LT} \in \mathbb{R}^{C \times L/k}$ (with $k = 4$), obtained by downsampling the 256 Hz original signal to 64 Hz via average pooling to emphasize slower, global temporal trends. Here, C denotes the number of EEG channels, and L denotes the time steps in the high-resolution (HR) signal and the frequency bins.

This multi-input design is motivated by the fusion of different brain activity patterns corresponding to EEG signal features: HR signals excel at detecting rapid neural transients (e.g., event-related potentials during learning), LR signals reduce noise and focus on sustained activities (e.g., resting baselines), and frequency features highlight rhythmic signatures (e.g., the α band of EEG is typically suppressed during memory recall).

To extract core temporal and spectral patterns and promote initial cross-modality alignment. The outputs are fused additively with the frequency branch serving as an anchor to infuse spectral context into temporal branches early on. This additive fusion is inspired by residual connections, enhancing gradient flow while integrating frequency priors to mitigate temporal aliasing. The initial convolution outputs are denoted as:

$$\mathbf{F}_{HT}^{(1)} = f_{HT}^{(1)}(\mathbf{X}_{HT}; \theta_{HT}^{(1)}) + \mathbf{F}^{(1)}, \quad (2)$$

$$\mathbf{F}_{LT}^{(1)} = f_{LT}^{(1)}(\mathbf{X}_{LT}; \theta_{LT}^{(1)}) + \mathbf{F}^{(1)}, \quad (3)$$

$$\mathbf{F}^{(1)} = f^{(1)}(\mathbf{X}_F; \theta^{(1)}), \quad (4)$$

where $f_*^{(i)}(\cdot)$ represents the i -th convolutional block (Conv $d(\cdot)$) with learnable parameters $\theta_*^{(i)}$ for the respective branch, and the subscript $*$ denotes either HT, LT, or is omitted. Each branch begins with a convolutional block, comprising a convolutional layer, batch normalization (BN), ReLU activation, and dropout. Here, $f_{HT}^{(1)}(\cdot)$ and $f^{(1)}(\cdot)$ are configured with 16 output channels, kernel size k_2 , and stride s_2 , whereas $f_{LT}^{(1)}(\cdot)$ uses 16 output channels, kernel size k_1 , and stride s_1 . $\mathbf{F}_{HT}^{(1)}, \mathbf{F}_{LT}^{(1)}, \mathbf{F}^{(1)} \in \mathbb{R}^{16 \times L_1}$ (with L_1 adjusted post-convolution).

To enhance temporal-frequency interactions and capture mid-level features, a second convolutional block is applied,

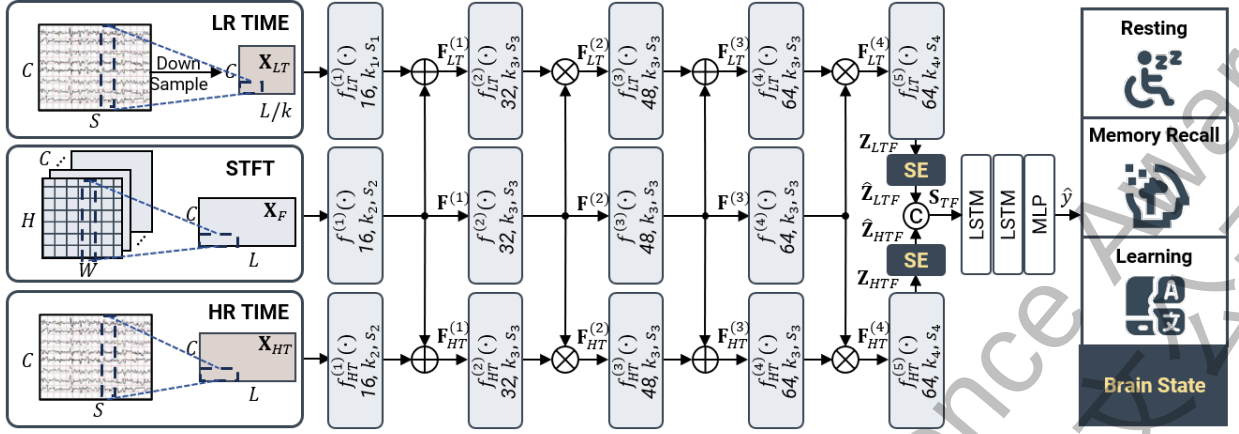


Fig. 5. Overall architecture of the SFR-Net for EEG classification.

followed by multiplicative feature fusion. This gating mechanism, motivated by attention-like modulation, allows the frequency branch to selectively amplify salient temporal patterns, fostering adaptive integration across resolutions. The outputs are:

$$\mathbf{F}_{HT}^{(2)} = f_{HT}^{(2)}(\mathbf{F}_{HT}^{(1)}; \theta_{HT}^{(2)}) \otimes \mathbf{F}^{(2)}, \quad (5)$$

$$\mathbf{F}_{LT}^{(2)} = f_{LT}^{(2)}(\mathbf{F}_{LT}^{(1)}; \theta_{LT}^{(2)}) \otimes \mathbf{F}^{(2)}, \quad (6)$$

$$\mathbf{F}^{(2)} = f^{(2)}(\mathbf{F}^{(1)}; \theta^{(2)}), \quad (7)$$

where $f_{HT}^{(2)}(\cdot)$, $f^{(2)}(\cdot)$, and $f_{LT}^{(2)}(\cdot)$ are configured with 32 output channels, kernel size k_3 , and stride s_3 . The $\otimes$ denotes element-wise multiplication calculation, and $\mathbf{F}_{HT}^{(2)}, \mathbf{F}_{LT}^{(2)}, \mathbf{F}^{(2)} \in \mathbb{R}^{32 \times L_2}$.

Building on this, a third convolutional layer further refines hierarchical representations, with additive fusion to reinforce multi-resolution coherence. This stage is motivated by the progressive abstraction in neural hierarchies, where deeper layers integrate broader contexts. The outputs are:

$$\mathbf{F}_{HT}^{(3)} = f_{HT}^{(3)}(\mathbf{F}_{HT}^{(2)}; \theta_{HT}^{(3)}) + \mathbf{F}^{(3)}, \quad (8)$$

$$\mathbf{F}_{LT}^{(3)} = f_{LT}^{(3)}(\mathbf{F}_{LT}^{(2)}; \theta_{LT}^{(3)}) + \mathbf{F}^{(3)}, \quad (9)$$

$$\mathbf{F}^{(3)} = f^{(3)}(\mathbf{F}^{(2)}; \theta^{(3)}), \quad (10)$$

where $f_{HT}^{(3)}(\cdot)$, $f^{(3)}(\cdot)$, and $f_{LT}^{(3)}(\cdot)$ are configured with 48 output channels, kernel size k_3 , and stride s_3 . $\mathbf{F}_{HT}^{(3)}, \mathbf{F}_{LT}^{(3)}, \mathbf{F}^{(3)} \in \mathbb{R}^{48 \times L_3}$.

Finally, a fourth convolutional block extracts high-level fused features, culminating in multiplicative fusion for fine-grained modulation. This multi-stage alternating fusion (additive-multiplicative) is a key innovation, enabling dynamic recalibration and preventing information loss in heterogeneous EEG modalities. The outputs are:

$$\mathbf{F}_{HT}^{(4)} = f_{HT}^{(4)}(\mathbf{F}_{HT}^{(3)}; \theta_{HT}^{(4)}) \otimes \mathbf{F}^{(4)}, \quad (11)$$

$$\mathbf{F}_{LT}^{(4)} = f_{LT}^{(4)}(\mathbf{F}_{LT}^{(3)}; \theta_{LT}^{(4)}) \otimes \mathbf{F}^{(4)}, \quad (12)$$

where $f_{HT}^{(4)}(\cdot)$, $f^{(4)}(\cdot)$, and $f_{LT}^{(4)}(\cdot)$ are configured with 64 output channels, kernel size k_3 , and stride s_3 . $\mathbf{F}_{HT}^{(4)}, \mathbf{F}_{LT}^{(4)} \in \mathbb{R}^{64 \times L_4}$.

To achieve a deeper integration of the multi-resolution temporal and frequency features beyond the stage-wise fusions, we introduce a final fusion step by applying a fifth convolutional block. The fused features are defined as:

$$\mathbf{Z}_{LTF} = f_{LT}^{(5)}(\mathbf{F}_{LT}^{(4)}; \theta_{LT}^{(5)}), \quad (13)$$

$$\mathbf{Z}_{HTF} = f_{HT}^{(5)}(\mathbf{F}_{HT}^{(4)}; \theta_{HT}^{(5)}), \quad (14)$$

where $f_{HT}^{(5)}(\cdot)$ and $f_{LT}^{(5)}(\cdot)$ are configured with 64 output channels, kernel size k_4 , and stride s_4 , and $\mathbf{Z}_{LTF}, \mathbf{Z}_{HTF} \in \mathbb{R}^{64 \times L_5}$.

The fused branch outputs are then processed through a Squeeze-and-Excitation (SE) block to emphasize channel-wise dependencies, motivated by the need to prioritize informative EEG channels. The SE-enhanced features are denoted as:

$$\hat{\mathbf{Z}}_{LTF} = \text{SE}(\mathbf{Z}_{LTF}), \quad (15)$$

$$\hat{\mathbf{Z}}_{HTF} = \text{SE}(\mathbf{Z}_{HTF}), \quad (16)$$

where $\text{SE}(\cdot)$ applies global average pooling followed by a two-layer MLP for excitation weights, and $\hat{\mathbf{Z}}_{LTF}, \hat{\mathbf{Z}}_{HTF} \in \mathbb{R}^{64 \times L_5}$.

To model long-range temporal dependencies across the fused multi-resolution features—essential for distinguishing sequential brain states like transitions from resting to learning—we employ a bidirectional two-layer Long Short-Term Memory (LSTM) network. This is motivated by LSTM's ability to capture bidirectional context in time-series EEG data, outperforming vanilla RNNs in handling vanishing gradients. The LSTM processes flattened features from $\mathbf{Z}_{LTF}$ and $\mathbf{Z}_{HTF}$ (after max-pooling to reduce dimensionality), yielding hidden states:

$$\mathbf{S}_{TF} = \text{Concat}(\hat{\mathbf{Z}}_{HTF}, \hat{\mathbf{Z}}_{LTF}), \quad (17)$$

$$\mathbf{H} = \text{LSTM}(\text{Flatten}(\mathbf{S}_{TF}); \phi), \quad (18)$$

where $\text{Concat}(\cdot)$ is channel-wise concatenation, $\text{Flatten}(\cdot)$ vectorizes the tensor, $\text{LSTM}(\cdot; \phi)$ is the two-layer bidirectional LSTM with parameters ϕ , and $\mathbf{H} \in \mathbb{R}^{256}$.

Finally, a Multi-Layer Perceptron (MLP) with two fully connected layers (hidden size 256, ReLU activation, dropout 0.3) performs the classification, outputting probabilities for the

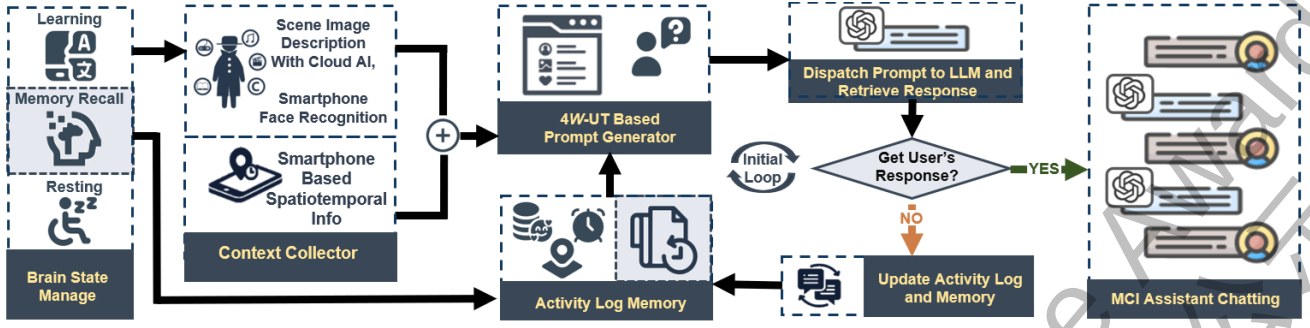


Fig. 6. Working flowchart ‘4W-UT’ and Personal Data Combined Prompts for LLM-Based MCI Chatting Agent.

three brain states. This MLP is motivated by its efficiency in non-linear decision boundaries for multi-class tasks. The prediction is:

$$\hat{y} = \text{Softmax}(\text{MLP}(\mathbf{H}; \psi)), \quad (19)$$

where $\text{MLP}(\cdot; \psi)$ denotes the MLP with parameters ψ , and $\hat{y} \in \mathbb{R}^3$.

Using the EEG signal dataset, which was accurately labeled and established in the previous section, our SFR-Net model was trained end-to-end using cross-entropy loss and incorporates a multi-branch, multi-stage fusion approach. This innovative SFR-Net model effectively captures the spatial, spectral, and temporal dependencies within the EEG signals, enabling robust classification of the three cognitive states—*Resting*, *Learning*, and *Memory Recall*—required for the PV-Care scheme, which uses a wearable 4-channel EEG device.

D. “4W-UT” and Personal Data Combined Prompt for LLM-Based MCI Chatting Agent

Cloud-based AI systems, such as ChatGPT, have demonstrated remarkable capabilities in conversational tasks and problem-solving. However, a significant challenge in assisting MCI patients is the variability in LLM responses.

1) “4W-UT” Prompt for Chatting Setup: In the PV-Care scheme, we designed a chatting agent, the pipeline of which is shown in Figure 6. Prompt-based control, such as “*role: system, content: descriptions*”, is used in our agent to initialize the cloud LLM. Specifically, we define the “4W-UT” prompt, which integrates multi-source context to drive personalized conversations. The key information of “4W-UT” prompt are:

- *Where*: the user’s location, derived from the smartphone GPS (position format: dddmm,N/S, dddmm,E/W).
- *When*: the current time, also from the smartphone (time format: year-month-day-time).
- *Who*: the people present, recognized locally using the smartphone-based face recognition system.
- *What*: the surrounding environment description, generated by the cloud-based visual AI model (e.g., “man, table, bottle on table, can on table, etc.” as in Figure 2).
- *UserState*: the user’s cognitive state and personal log. In addition to *Resting*, *Learning*, or *Memory Recall*, recent activity history and personalized information of the MCI patient are incorporated.

By combining these information, chatting agent can proactively initialize chatting with “4W-UT” prompt: For example

shown in Figure 2, and assume the person in the image is named Leonardo. Based on the collected contextual information, a generated “4W-UT” prompt that is sent to the OpenAI API might be:

role: system, content: “I am in a room with GPS dddmm,N/S, dddmm,E/W. Current time is year-month-day-time. Leonardo is here. In front of me: a man, indoors, sitting, wearing a face mask, casual attire, and a table with bottles and cans. The user is currently in a memory state.”

Additionally, our scheme incorporates the user’s recent weekly activity data and planned next-step actions. For example, if an MCI patient leaves home in the morning, they may forget to return. By maintaining an activity log and integrating it into the prompt, PV-Care chatting can provide reminders during memory recall states, such as suggesting the user should return home and offering the navigation route.

2) *The Workflow of Our Assistant Chatting Agent*: Figure 6 presents the overall workflow of our chatting agent, and the details of each module in the chatting agent pipeline are as follows:

- **Brain State Management**: This module receives the brain activity state outputs from the SFR-Net, as described in the previous section. The *Memory Recall* and *Learning* states activate subsequent modules to proactively initiate user interactions, while the *Resting* state does not trigger any active engagement.
- **Context Collection**: Multi-source context is collected, including environmental descriptions generated by the cloud model from the user’s surroundings, as well as the identification of potential people using the user’s smartphone-based face recognition. This completes the gathering of necessary environmental information for building the chatting agent’s context.
- **Activity Log Memory**: This module logs the user’s recent activities, such as “left home at 9 AM”, which provides contextual information for subsequent interactions. This log is particularly useful when the user is in the *Memory Recall* state, enabling the system to create relevant prompts based on past activities.
- **4W-UT Based Prompt Generator**: If the user is in a *Memory Recall* state, the activity log memory information is fused with environmental data to form the “4W-UT” prompt. If the user is not in a *Memory Recall* state, the prompt is constructed solely from the environmental

image recognition descriptions used for the *Learning* state.

- **Dispatch Prompt and Feedback Loop:** The constructed prompts are proactively sent to the user, using ChatGPT as the “role: system. Based on the user’s feedback, the system determines whether assistant chatting has been established. If no feedback is provided, the system updates the Activity Log and Memory and re-enters the prompt generation process. This loop continues until a meaningful conversation is established, at which point the MCI Assistant Chatting module engages with the user.

Using this conversational agent, PV-Care can provide a highly personalized and context-aware conversational experience, and the agent, which uses the ChatGPT API, is available at [22]. Additionally, the textual responses generated by LLM are transformed into spoken feedback through Text-to-Speech (TTS) technology, and delivered via the user’s paired smartphone earphones, as illustrated in the lower-right part of Figure 1.

III. SIMULATION FOR OUR PROPOSALS

In the experimental evaluation, we conducted functional simulations of the proposed proactive MCI assistance scheme. As illustrated in Figure 1, the PV-Care scheme integrates three core components: scene image analysis, EEG-based brain state analysis, and LLM-based chatting agent. The image analysis is a relatively mature technique, and in our implementation, we directly utilized OpenAI’s API. Hence, our experiment primarily focused on the EEG-based brain state recognition, the usability of PV-Care prototype for proactive assistance.

A. Evaluation of SFR-Net for EEG-Based Brain State Classification

To evaluate the effectiveness of the proposed SFR-Net model, we conducted experiments on the 30-subject EEG dataset described in Section II.C. As supported by [9], EEG patterns corresponding to cognitive states: *Resting*, *Learning*, and *Memory Recall* are physiologically consistent between healthy participants and MCI patients, making this dataset suitable for validating SFR-Net, although the data were collected from healthy participants.

1) *Experimental Setup:* The EEG signals were preprocessed and segmented into 4-second windows, with both temporal and frequency-domain features. We adopted 5-fold cross-validation to improve the reliability of the evaluation and reduce the impact of data partitioning. For each fold, the SFR-Net model was trained from scratch and evaluated on the corresponding held-out set. Our model was implemented in PyTorch and trained on an NVIDIA GeForce RTX 4090 GPU using the Adam optimizer (batch size=32, learning rate=0.001) for up to 300 epochs. Final performance, evaluated on the test set, reports the average precision, recall, accuracy, and F1 score.

2) *Comparison with State-of-the-Art Methods:* We compared SFR-Net against several established EEG classification models:

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT METHODS

Method	Acc	Precision	Recall	F1 Score
EEGNet	0.669±0.004	0.675±0.002	0.667±0.008	0.668±0.005
Tsception	0.720±0.018	0.751±0.014	0.715±0.022	0.713±0.024
Conformer	0.726±0.011	0.738±0.012	0.728±0.010	0.728±0.011
MSTCNN	0.754±0.014	0.754±0.015	0.768±0.017	0.754±0.014
CNNLSTM	0.781±0.013	0.800±0.020	0.789±0.016	0.782±0.013
BMFCNet	0.786±0.019	0.779±0.018	0.799±0.019	0.785±0.020
SFR-Net	0.804±0.016	0.812±0.020	0.813±0.011	0.804±0.017

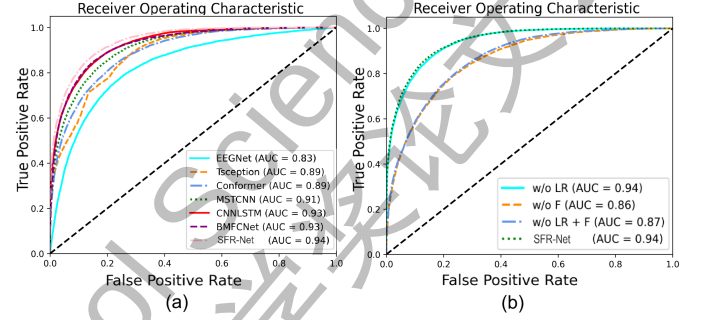


Fig. 7. (a) ROC of SFR-Net for three-state brain activity classification using 5-fold cross-validation, (b) ROC analysis of ablation study for SFR-Net

- **EEGNet [23]:** A lightweight CNN using depthwise and separable convolutions for efficient EEG spatial-temporal feature extraction.
- **Tsception [24]:** Captures hemispheric asymmetry through bilateral kernels in its multi-scale temporal-spatial architecture.
- **Conformer [25]:** Integrates convolutional layers for local patterns with self-attention mechanisms to model global dependencies in EEG signals.
- **MSTCNN [26]:** Uses parallel temporal convolutions with varied kernel sizes to extract multi-scale features from EEG time series.
- **CNNLSTM [27]:** Combines CNNs for spatial-spectral feature learning with LSTMs to capture temporal dynamics in EEG sequences.
- **BMFCNet [28]:** Fuses temporal and spectral EEG representations via dedicated bimodal modules for enriched feature learning.

As shown in Table I, SFR-Net outperforms all baselines across all metrics, achieving the highest accuracy of 0.804 ± 0.016 . Notably, SFR-Net improves upon the closest competitor, BMFCNet, by approximately 1.8% in accuracy, demonstrating the advantages of its hierarchical fusion of multi-resolution temporal and frequency-domain features. Lower-performing models rely primarily on single-resolution inputs, underscoring the limitations of ignoring complementary dynamics across time scales.

The superiority of SFR-Net is further evidenced by the ROC curves in Figure 7 (a), where it achieves the highest AUC of 0.94. This indicates robust separability, with SFR-Net’s curve being more closer to the top-left corner than baselines such as CNNLSTM and BMFCNet (both AUC 0.93) or EEGNet

TABLE II
CHAT PROMPT EVALUATION METRICS FOR PV-CARE, INCLUDING SUBJECTIVE (M1–M4) AND OBJECTIVE (M5) INDICATORS.

No.	Metric Name	Description	Evaluation Method	Score
M1	Appropriateness of Proactive Guidance	Whether proactive questions or guidance match the user’s context and needs	Human rating (1–5 scale)	4.7
M2	Effectiveness of Explanations	Whether explanations help the user understand and reduce confusion	Human rating (1–5 scale)	4.8
M3	Accuracy of Fact Verification	Whether dialogue content aligns with factual information	Human rating (1–5 scale)	4.6
M4	Semantic Relevance	Semantic similarity between generated content and the user’s context/input	Human rating (1–5 scale)	4.7
M5	Discourse Coherence Score	Logical and semantic coherence of the dialogue	Coh-Matrix [29] (-1–1 scale)	0.92

TABLE III
ABLATION STUDY OF SFR-NET COMPONENTS

Method	Acc	Prec.	Recall	F1
w/o LR	0.797±0.010	0.809±0.015	0.803±0.007	0.797±0.009
w/o F	0.685±0.007	0.688±0.007	0.700±0.007	0.684±0.008
w/o LR&F	0.688±0.012	0.697±0.012	0.701±0.010	0.688±0.012
SFR-Net	0.804±0.016	0.812±0.020	0.813±0.011	0.804±0.017

(AUC 0.83). The marginal AUC gains highlight SFR-Net’s enhanced ability to minimize false positives while maintaining high true positive rates, which is critical for real-world brain-computer interface applications where misclassification of cognitive states (e.g., confusing Learning with Resting) could impair usability.

3) *Ablation Study*: To validate the contributions of key components in SFR-Net—specifically the low-resolution (LR) temporal branch and the frequency-domain (F) branch—we performed an ablation study by systematically removing these elements and retraining the model.

Results in Table III reveal that both branches are essential. Removing the LR branch (w/o LR) results in a modest drop in accuracy to 0.797 ± 0.010 , suggesting that while high-resolution inputs capture fine details, the LR branch provides complementary stability for slower neural dynamics. More critically, ablating the F branch (w/o F) causes a substantial decline to 0.685 ± 0.007 accuracy, emphasizing the importance of spectral features in encoding oscillatory patterns like theta-band activity during Memory Recall. The combined removal (w/o LR&F) yields performance comparable to w/o F (accuracy 0.688 ± 0.012), confirming that frequency-domain integration is the dominant factor, though LR enhances it further in the full model.

The ROC analysis in Figure 7 (b) corroborates these findings, with the full SFR-Net achieving an AUC of 0.94. Interestingly, w/o LR maintains the same AUC (0.94), but the table metrics indicate reduced consistency, implying that LR aids in balanced multi-class performance rather than overall separability. In contrast, w/o F and w/o LR&F drop to AUCs of 0.86 and 0.87, respectively, with curves deviating further from the ideal, highlighting how frequency fusion mitigates trade-offs in false positive rates. These results affirm the innovative multi-branch, multi-stage fusion in SFR-Net as pivotal for superior EEG decoding.

To promote reproducibility and further research, the source code and experimental data for the proposed SFR-Net have been made publicly available at our GitHub repository [30].

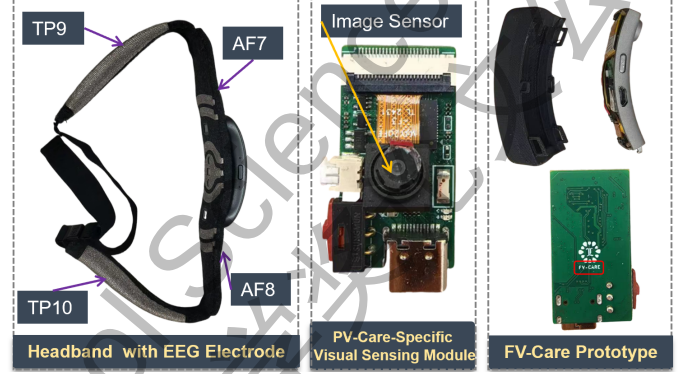


Fig. 8. Wearable PV-Care prototype.

B. Evaluation of “4W-UT” Prompt Controlled Proactive Chatting

Given the inherent variability and occasional hallucinations in LLM outputs, we evaluated the stability and reliability of conversational responses generated under our proposed “4W-UT” prompt. By setting the LLM’s system role explicitly through structured environmental, temporal, and cognitive context, we aimed to standardize the dialogue generation process and improve consistency.

We assessed dialogue quality using both subjective metrics (M1–M4 in Table II) and objective metrics (M5). Subjective evaluations included the appropriateness of proactive guidance (M1) and clarity of explanations (M2), scored on a 5-point Likert scale by human raters. Objective indicators such as discourse coherence score (M5) was automatically computed using established NLP tools including Coh-Matrix [29]. Some results from the simulation are available at [22].

We used the beverage recognition scenario (Scenario 1) as a representative case. A total of 50 conversation samples were generated using identical “4W-UT” prompts. Ten healthy participants were recruited to review the conversations, each evaluating all 50 samples. This yielded 500 subjective ratings per metric. An automated hallucination detection module was also applied to assess factual consistency and contextual relevance in the generated responses.

Overall, the evaluation results (Table II) indicate that the integration of LLM with our structured “4W-UT” prompting framework enables the PV-Care to deliver reliable, coherent, and contextually appropriate proactive assistance.

C. Overall Performance Assessment for PV-Care Scheme

1) *Hardware Implementation of the PV-Care Prototype*: In addition to evaluating the performance of the proposed SFR-



Fig. 9. MoCA and MMSE cognitive screening for participant selection.

TABLE IV
COGNITIVE ASSESSMENT SUMMARY FOR 60 PARTICIPANTS

Metric	N	Age (Years)	Gender (Female)	MMSE	MoCA
Value	60	68.13 $\pm$ 6.14	37 (61.7%)	26.13 $\pm$ 3.65	23.48 $\pm$ 5.02

Net for EEG analysis, we further assessed the overall system feasibility of the PV-Care scheme through hardware implementation. To validate the practicality of the proposed system integration, we developed a functional prototype focused on environmental sensing and EEG data acquisition.

Given the limited availability of compact and high-quality EEG devices, we adopted a commercially available 4-channel wearable EEG headset (Muse) [18], which is capable of capturing real-time brainwave activity suitable for cognitive state monitoring. To meet the integrated and wearable design requirements of PV-Care, we specially developed a lightweight camera module capable of capturing 640 $\times$ 480 resolution images for real-time environmental sensing. This visual sensing unit was mechanically integrated with the EEG headset via a custom-designed 3D-printed interface and assembly mechanism, ensuring stable alignment during operation while maintaining user comfort. Furthermore, by equipping the system with a compact rechargeable power supply, we constructed a fully functional prototype of the PV-Care sensing hardware, as illustrated in Figure 8.

2) *Participant Selection for Subjective Evaluation*: In this section, we present the subjective usability evaluation of the PV-Care scheme conducted with MCI patients, along with a simulated case study to demonstrate its practical functionality and user experience.

Since PV-Care is designed primarily for individuals with MCI, we recruited suitable participants in collaboration with the Brain and Behavior Research Institute (BABRI) [31], following established ethical guidelines. Cognitive screening was conducted using two standardized tools: the Montreal Cognitive Assessment (MoCA) and the Mini-Mental State Examination (MMSE). The MoCA is known for its high sensitivity in detecting MCI by assessing memory, language, attention, and executive functions [32], while the MMSE provides a general evaluation of cognitive status and is widely used in dementia screening [33].

60 volunteers from local communities completed the MoCA and MMSE assessments (Figure 9). Table IV summarizes the cognitive and demographic profiles of the participants. The average age was 68.13 years (SD = 6.14), with a majority being female (61.7%). The mean MMSE and MoCA scores were 26.13 and 23.48, respectively, confirming mild cognitive

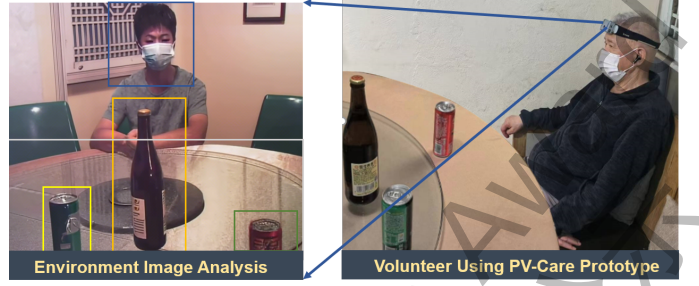


Fig. 10. (a) Environmental image captured by PV-Care. (b) Volunteer wearing the PV-Care prototype during testing.

TABLE V
SUBJECTIVE EVALUATION OF PV-CARE PROTOTYPE (1 = LOW, 5 = HIGH).

Metric	Remarks	Avg. Score
Wearing comfort	Comfort while wearing the device	4.7
Ease of use	User-friendliness and interaction simplicity	4.8
Prompt clarity	Clarity of system prompts and feedback	4.6
Overall rating	General assessment of system performance	4.8
Willingness to use	Interest in continued usage in daily life	4.9

decline in the group.

To ensure effective participation in the usability study, 20 individuals with MoCA scores between 18 and 25 were selected in our subjective evaluation of PV-Care. These participants demonstrated sufficient cognitive capacity to follow instructions and engage with the system.

3) *Prototype Testing and Usability Evaluation*: Selected participants were invited to wear the PV-Care prototype and complete a series of guided assistance tasks in a controlled environment. During these simulations, the system provided real-time interaction and context-aware navigation support, based on both environmental perception and EEG-based brain state recognition. One of the participant scenarios is illustrated in Figure 10.

To assess the system's usability, participants completed a subjective evaluation form (Table V) that measured comfort, usability, prompt clarity, overall satisfaction, and willingness to use the system long-term.

The subjective evaluation results indicated strong positive responses across all metrics. Participants rated the device with an average score of 4.7 for wearing comfort, 4.8 for ease of use, 4.6 for prompt clarity, 4.8 for overall satisfaction, and 4.9 for willingness to continue using the system. Most users described the prototype as intuitive, comfortable, and helpful in managing daily tasks.

4) *Case Study: Assistant Chatting Powered by 4W-UT Prompts*: To evaluate the functionality of the PV-Care chatting agent (In this experiment, we used ChatGPT APIs of OpenAI as the LLM to support our chatting agent.), we conducted two representative simulation scenarios that tested the system's ability to detect users' cognitive states, construct contextual prompts, and deliver proactive, context-aware assistance. Volunteers were invited to participate and provide feedback. Each simulation involved real-time EEG signal acquisition, environmental sensing via a wearable camera, and LLM-based conversational interaction.

a) *Scenario 1: Assisting with Beverage Recognition and Learning:* As shown in Figure 10, a volunteer wearing the PV-Care prototype was seated at a round table with three beverages: a bottle of beer, a can of Coke, and a can of Sprite. To simulate a realistic home scenario, another participant—a registered family member named Leonardo—was seated opposite the volunteer.

Before initiating any conversation, PV-Care performed a multi-stage perception process. The SFR-Net identified that the user was in a *Learning* state based on EEG signals, indicating a focus on the objects in front. Meanwhile, the wearable camera captured the surrounding scene, and visual analysis recognized Leonardo. The system then constructed a “4W-UT” contextual prompt, which was transmitted to ChatGPT using the *system* role (see Section II.D for details). Based on the generated response, PV-Care proactively initiated a context-aware conversation, as shown below:

- 1) **PV-Care:** *I see you're looking at the three drinks in front of you, and Leonardo is here with you. Would you like some help understanding what they are?*
- 2) **PV-Care:** *[Waiting for user response...]*
- 3) **PV-Care** (if no response within 10 seconds, re-initiates the prompt): *I can help you learn more about the drinks you're looking at—just let me know when you're ready.*
- 4) **Loop:** repeat 2), 3) processes until a response is received from the user.
- 5) **PV-Care** (upon receiving user response, continues with contextual explanation): *The can with the red label is Coke. It's a sweet, caffeinated soft drink....*
- 6) **User:** *I want....*

This case demonstrates that PV-Care can recognize when the user is in a learning state and initiate intelligent, personalized support. The system integrates EEG-based cognitive monitoring, real-world visual context, and identity recognition to help users better understand and interact with their environment.

b) *Scenario 2: Reminding and Guiding the User to Return Home:* The second scenario simulated a safety-critical situation for an MCI user. A volunteer equipped with the PV-Care prototype walked outdoors and paused at a crossroads, attempting to recall the correct direction home. At this moment, SFR-Net detected a *Memory Recall* cognitive state. Simultaneously, the wearable camera captured nearby visual landmarks, and the GPS module determined the user's exact location. Combining this information with stored home address data, PV-Care constructed a safe navigation route and generated a contextual 4W-UT prompt for the language model:

“role: system, content: I am on Livernois Road with GPS dmm,N/S, dddmm,E/W. Current time is year-month-day-time. In front of me: road, tree, building. The user is currently in a memory recall state.”

PV-Care then initiated a proactive navigation dialogue:

- 1) **PV-Care:** *You seem to be near the park. Are you trying to find your way home?*
- 2) **PV-Care:** *[Waiting for user response...]*
- 3) **PV-Care** (if no response within 10 seconds, re-initiates prompt): *If you're unsure which direction to go, I can guide you back home. Just let me know when you're ready.*

- 4) **Loop:** repeat 2), 3) until a response is received from the user.
- 5) **PV-Care** (upon receiving user response, proceeds with personalized navigation instructions): *From your current location, turn right at the next intersection. Then continue straight for about 300 meters until you reach Timberview Street. Turn left there—your home is just around the corner.*
- 6) **User:** *Which way...*

This scenario demonstrates PV-Care's capability to identify memory recall states, contextualize the surrounding environment, and provide timely, safety-critical navigation support through its LLM-powered conversational interface.

IV. CONCLUSION

We proposed PV-Care, a proactive assistance scheme for individuals with MCI. By integrating wearable 4-channel EEG sensing, environmental perception, and large language model-based dialogue, PV-Care enables real-time, personalized support. The proposed SFR-Net accurately recognizes brain states, while structured “4W-UT” prompts guide LLM to generate context-aware responses. Simulation experiments and user evaluations confirm the system's usability, cognitive recognition performance, and conversational effectiveness. PV-Care shows strong potential as a practical solution for daily cognitive support in aging populations.

REFERENCES

- [1] F. H. Frech *et al.*, “Economic impact of progression from mild cognitive impairment to alzheimer disease in the united states,” *Journal of Prevention of Alzheimer's Disease*, vol. 11, no. 6, pp. 983–991, Dec 2024.
- [2] S. Li *et al.*, “Proactive human–robot collaboration: Mutual-cognitive, predictable, and self-organising perspectives,” *Robotics and Computer-Integrated Manufacturing*, vol. 81, p. 102510, Jan 2023.
- [3] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, and T. Graepel, “A general reinforcement learning algorithm that masters chess, shogi, and go through self-play,” *Science*, vol. 362, pp. 1140–1144, 2018.
- [4] B. Clara, Z. Yuanjin, R. Julie M, and J. Kaye, “Companion robots to mitigate loneliness among older adults: Perceptions of benefit and possible deception,” *Front. Psychol*, vol. 14, Feb 2023.
- [5] A. A. Cantone, M. Esposito, F. P. Perillo, M. Romano, M. Sebillio, and G. Vitiello, “Enhancing elderly health monitoring: Achieving autonomous and secure living through the integration of artificial intelligence autonomous robots and sensors,” *Electronics*, vol. 12, p. 3918, 2023.
- [6] D. Zhou, E. I. Barakova, P. An, and M. Rauterberg, “Assistant robot enhances the perceived communication quality of people with dementia: A proof of concept,” *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 3, 2022.
- [7] J. Saunders, D. S. Syrdal, K. L. Koay, N. Burke, and K. Dautenhahn, ““teach me—show me”—end-user personalization of a smart home and companion robot,” *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, 2016.
- [8] S. L. Liu, “Personalized caring: Integrating eeg/visual analysis with chatgpt for mci assistance,” *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2025, DOI: 10.1109/HRI61500.2025.10973826.
- [9] J. Schumacher, J.-P. Taylor, C. A. Hamilton, M. Firbank, R. A. Cromarty, P. C. Donaghy, G. Roberts, L. Allan, J. Lloyd, R. Durcan *et al.*, “Quantitative eeg as a biomarker in mild cognitive impairment with lewy bodies,” *Alzheimer's Research & Therapy*, vol. 12, no. 1, pp. 1–12, 2020.
- [10] C. Yue, H. Hu, E. Shi, and S. Zhang, “Tf-hitnet: A temporal-frequency hierarchical transformer network for eeg motor imagery classification,” in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2024, pp. 3961–3965.

- [11] X. Yao, Z. Luo, C. Lu, and Y. Yan, "Emotion classification based on transformer and cnn for eeg spatial-temporal feature learning," *Brain Sciences*, vol. 14, no. 3, p. 268, 2024.
- [12] C. Wang, X. Zhang, R. Li, G. Wei, M. He, X. Li, and R. Wang, "Towards high-accuracy classifying attention-deficit/hyperactivity disorders using cnn-lstm model," *Journal of Neural Engineering*, vol. 19, no. 4, p. 046015, 2022.
- [13] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021, PMLR 139:8748-8763.
- [14] S. Panchal, "Face recognition with facenet on android," https://github.com/shubham0204/FaceRecognition_With_FaceNet_Android, 2022, accessed: 2024-11-26.
- [15] RVC-Boss, "Voice-transfer," <https://github.com/RVC-Boss/GPT-SoVITS>, 2024, accessed: November 11, 2024.
- [16] OpenAI, "Openai vision api documentation," <https://platform.openai.com/docs/guides/vision>, 2024, accessed: November 29, 2024.
- [17] Jiaye, Fok, and GPT4, "Gpt4v-image-captioner," <https://github.com/jiayev/GPT4V-Image-Captioner>, 2023, accessed: November 5, 2024.
- [18] Muse by Interaxon Inc., "Muse s athena," 2024, available at: <https://choosemuse.com/> [Accessed 01-Aug-2025].
- [19] Neuracle Technology Inc., "Neuracle," 2024, available at: <http://www.neuracle.cn/productinfo/148706.html> [Accessed 02-Aug-2025].
- [20] D. Friedman, D. Nessler, and R. J. Jr., "An erp study of memory encoding and retrieval in younger and older adults," *Neuropsychologia*, vol. 42, no. 10, pp. 1423-1437, 2004.
- [21] I. Neuner, J. Arrubla, C. J. Werner, K. Hitz, F. Boers, W. Kawohl, and N. J. Shah, "The default mode network and eeg regional spectral power: A simultaneous fmri-eeg study," *PLOS ONE*, vol. 9, no. 2, p. e88214, 2014.
- [22] S.L.Liu, "Pv_care_assistant-llm-agent," https://github.com/Simon-0523/PV_Care_Assistant-LLM-Agent, 2025, accessed: September 8, 2025.
- [23] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, p. 056013, 2018.
- [24] Y. Ding, N. Robinson, S. Zhang, Q. Zeng, and C. Guan, "TSception: Capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 2238-2250, 2023.
- [25] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG conformer: Convolutional transformer for EEG decoding and visualization," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 710-719, 2023.
- [26] W. Liu, G. Li, Z. Huang, W. Jiang, X. Luo, and X. Xu, "Enhancing generalized anxiety disorder diagnosis precision: MSTCNN model utilizing high-frequency EEG signals," *Front. Psychiatry*, vol. 14, p. 1310323, 2023.
- [27] N. Masuda and I. E. Yairi, "Multi-Input CNN-LSTM deep learning model for fear level classification based on EEG and peripheral physiological signals," *Front. Psychol.*, vol. 14, p. 1141801, 2023.
- [28] M. Karnati, G. Sahu, G. Verma, A. Seal, M. Kishore Dutta, and J. Jaworek-Korjakowska, "Bmfnet: Blended multilevel features with constraint fusion network for depression detection from eeg signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1-14, 2025.
- [29] D. S. McNamara, A. C. Graesser, P. M. McCarthy, and Z. Cai, *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press, 2014.
- [30] S. L. Liu, "Eeg 3-state classification with sfr-net," https://github.com/Simon-0523/EEG_3StateClassification, 2025, accessed: 2025-08-28.
- [31] Y. Yang *et al.*, "Community-based model for dementia risk screening: The beijing aging brain rejuvenation initiative (babri) brain health system," *Journal of the American Medical Directors Association*, vol. 22, no. 4, pp. 766-772, 2021. [Online]. Available: [https://www.jamda.com/article/S1525-8610\(20\)31091-4/fulltext](https://www.jamda.com/article/S1525-8610(20)31091-4/fulltext)
- [32] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment (moca): A brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695-699, 2005.
- [33] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189-198, 1975.

致谢

本项目旨在为患有轻度认知障碍（MCI）的老年人提供个性化的辅助支持。我负责执行研究框架，包括构思研究思路、收集数据、进行计算机模拟和撰写论文。

本研究涉及多个关键组成部分，如通过视觉传感进行场景识别和通过脑电信号处理进行认知状态分析，这些工作是在复旦大学陈智能教授的指导下完成的。陈教授在系统的技术和计算方面提供了宝贵的见解，帮助优化实验的设计和实现。作为视觉感知技术领域的专家，陈教授在人工智能方法和工具的使用上提供了极大的帮助。当需要使用视觉分析系统时，陈教授引导并教会我如何使用基于 AI 的在线 API 来分析视觉数据，并将这些数据应用于进一步的分析。

项目的总体设计和组织是在上海高中的姚艳婕博士的精心指导下完成的。姚博士在项目过程中提供了重要的指导，确保研究遵循了严谨的逻辑过程。作为我在学校的主要导师，姚博士在选题和研究方向的确定上给予了我很多帮助。我们共同制定了研究项目的进度安排，她为项目设定了具体的时间节点，并在研究过程中始终作为负责的监督者，确保项目按时推进。

2024 年，我们对系统进行了进一步的改进，尤其是在深度学习模型的整合和脑电数据处理的实际应用方面。此阶段还包括对 PV-Care 原型的广泛测试，并根据多个用户研究的反馈对系统的响应和互动性进行了优化。

导师指导过程

• 2024 年 6 月 - 2024 年 8 月：项目启动与文献回顾

- 在陈智能教授的指导下，开展了关于轻度认知障碍（MCI）和基于脑电信号的脑波分析的文献综述。
- 在姚艳婕博士的协助下，确定了研究目标和范围，重点是开发面向 MCI 患者的 AI 辅助系统。

• 2024 年 9 月 - 2024 年 12 月：系统设计与原型开发

- 设计了 AI 辅助系统的架构，整合脑电信号处理与视觉识别技术。
- 在陈智能教授的指导下，开发了 EEG-Visual 设备原型，重点开发信号采集和处理算法，包括利用先进的 AI 技术提升视觉信息处理。
- 陈教授提供了关于如何将 AI 方法应用于系统的视觉识别和分析的详细指导。
- 姚博士在项目管理上提供了全程监督，确保设计目标与总体研究目标的对齐。

• 2025 年 1 月 - 2025 年 4 月：数据收集与实验设置

- 在陈智能教授的指导下，设立了脑电数据采集和视觉信号处理的实验，重点开发基于 SFR-Net 的脑电信号分析方法进行认知状态检测。
- 与姚博士合作，制定了详细的实验进度和数据收集协议，确保数据收集过程的系统化。

• 2025 年 5 月 - 2025 年 7 月：数据分析与 SFR-Net 验证

- 在陈智能教授的指导下，分析了收集的脑电数据，重点区分认知状态和记忆回忆状态。
- 使用先进的 AI 方法，特别是 SFR-Net（空间增强与重构网络），增强了系统处理和分类脑活动的能力。
- 进行系统验证实验，测试 AI 驱动对话系统和基于脑电的认知状态识别系统的有效性。

– 姚博士在这一阶段对分析方法提供了重要反馈，确保了项目里程碑的达成。

• **2025 年 5 月 - 2025 年 8 月：论文撰写与审阅**

- 撰写了总结项目成果的研究论文，特别是脑电和 AI 处理部分的技术性章节，经过陈智能教授的详细审阅。
- 姚博士在论文的结构和学术写作规范上提供了指导。
- 陈教授和姚博士参与了三轮论文修改，提供了修改意见，提升了论文的整体质量。

尽管他们的贡献巨大，姚博士和陈教授依然无偿提供了帮助。感谢姚博士和陈教授的帮助！

导师简历

姚艳婕博士是上海高中国际部的二级物理教师，持有上海交通大学物理学博士学位。姚博士在教学和研究方面有着丰富的经验，曾主讲多门高级课程，包括 IB 和 A-Level 物理，开展了关于脑电信息处理方面的研究。姚博士还积极参与开发创新的教学方法。

陈智能教授于 2011 年获得中国科学院计算技术研究所的博士学位。他曾担任中国科学院自动化研究所副教授，并曾在香港城市大学计算机科学系担任高级研究员。目前，他是复旦大学计算机科学与技术学院的预聘教授。陈教授的研究兴趣包括多媒体分析、人机交互，医学图像处理和计算机视觉。陈教授是多个国家级项目的负责人，包括由国家自然科学基金资助等，并与百度和腾讯等公司有过合作。他在国内外学术期刊和会议上发表了 60 多篇学术文章。

Acknowledgments

This project addresses the pressing need for personalized assistance for elderly individuals with Mild Cognitive Impairment (MCI). I was responsible for executing the research framework, including conceptualizing the research idea, collecting data, performing computational simulations, and writing the paper.

The research involved several key components, such as scene recognition via visual sensing and cognitive state analysis through EEG signal processing, which were completed under the guidance of Prof. Zhineng Chen from Fudan University. Prof. Chen provided invaluable insights into the technical and computational aspects of the system, helping to refine and optimize the experiment's design and implementation. As a professor in the field of visual sensing technology, Dr. Chen offered great help with the use of AI methods and tools. When there was a need for a visual analysis system, Dr. Chen introduced and taught me how to use AI-based online APIs to analyze the visual data and how to apply those data in further analysis.

The overall design and organization of the project were carried out under the careful supervision of Dr. Yanjie Yao from Shanghai High School International Division. Dr. Yao offered crucial guidance in the project, ensuring the research followed a logical and rigorous process. As my main counselor in school, Dr. Yao offered me a lot of help by choosing the research topic with me and pointing the research direction. Together, we also made schedules for the research project. She set deadlines for the projects, and during the research, she was always a responsible supervisor of the progress.

In 2024, further improvements were made to the system, especially with regard to the integration of advanced deep learning models and the practical application of EEG data processing. Significant adjustments were made to the system design, which involved a more comprehensive use of wearable EEG sensing devices and further validation of the system's effectiveness. This period also included extensive testing of the PV-Care prototype in real-world scenarios, refining its response and interactivity based on feedback from multiple user studies.

Supervisor Guidance

- **June 2024 - August 2024: Project Initialization and Literature Review**

- Conducted an extensive literature review on Mild Cognitive Impairment (MCI) and EEG-based brain signal analysis under the guidance of Prof. Zhineng Chen.
- Defined the research objectives and scope with input from Dr. Yanjie Yao, focusing on the development of an AI-assisted system for elderly individuals with MCI.

- **September 2024 - December 2024: System Design and Prototype Development**

- Designed the architecture of the AI-assisted system, integrating EEG signal processing with visual recognition technologies.
- Developed a prototype for the EEG-Visual device under the supervision of Prof. Zhineng Chen, focusing on both signal acquisition and processing algorithms, including advanced AI techniques for visual information enhancement.
- Prof. Chen provided detailed guidance on incorporating AI methods for improved visual recognition and analysis within the system.

- Dr. Yao provided oversight on project management, ensuring the alignment of design objectives with overall research goals.

- **January 2025 - April 2025: Data Collection and Experimental Setup**

- Set up experiments for EEG data collection and visual signal processing, with an emphasis on visual information processing guided by Prof. Chen. This included implementing advanced AI methods for EEG signal analysis and cognitive state detection using the custom-designed SFR-Net.
- Collaborated with Dr. Yao to develop a detailed experiment schedule and data collection protocol, ensuring a systematic approach to data gathering.

- **May 2025 - July 2025: Data Analysis and SFR-Net Validation**

- Analyzed the collected EEG data using computational methods introduced by Prof. Chen, focusing on differentiating cognitive and memory recall states.
- Applied advanced AI methods, particularly SFR-Net (Spatial Augmentation and Refinement Network), to enhance the system's ability to process and classify brain activity.
- Conducted system validation experiments to test the effectiveness of both the AI-driven dialogue and EEG-based cognitive state recognition systems.
- During this period, Dr. Yao provided critical feedback on the analysis methods and ensured that the project milestones were met.

- **May 2025 - August 2025: Paper Writing and Review**

- Drafted the research paper summarizing the project outcomes, with specific technical sections, particularly those on EEG and AI-based processing, reviewed by Prof. Chen.
- Dr. Yao provided guidance on structuring the paper and aligning it with scientific writing standards.
- Both Prof. Chen and Dr. Yao participated in multiple rounds of paper revisions between July and August, offering corrections and suggestions to enhance the overall quality of the manuscript.

Despite their significant contributions, Dr. Yao and Prof. Chen offered their support without any financial payment. Thanks for the help of Dr. Yao and Prof. Chen!

Supervisor's CV

Dr. Yanjie Yao Dr. Yanjie Yao is a second-level physics teacher at Shanghai High School International Division. She holds a Ph.D. in physics from Shanghai Jiao Tong University. Dr. Yao has extensive experience in both teaching and research. She has led several advanced courses, including IB and A-level physics, and has hosted public lectures that explore experimental physics concepts. Dr. Yao is also actively involved in developing innovative teaching techniques.

Prof. Zhineng Chen Prof. Zhineng Chen obtained his Ph.D. from the Institute of Computing Technology, Chinese Academy of Sciences in 2011. He was an Associate Professor at the Institute

of Automation, Chinese Academy of Sciences, and was a Senior Research Associate with the Department of Computer Science, City University of Hong Kong. He is currently a Professor at the School of Computer Science, Fudan University, Shanghai, China. His research interests include multimedia analysis, medical image processing, and computer vision. Prof. Chen has been the principal investigator of multiple national projects, including key research initiatives funded by the National Natural Science Foundation of China, and has collaborated with companies such as Baidu and Tencent. He has published over 60 academic articles in prestigious journals and conferences.