The EasyCog Dataset: Towards Easier Cognitive Assessment With Passive Video Watching

QINGYONG HU*, The Hong Kong University of Science and Technology, Hong Kong SAR
YUXUAN ZHOU*, The Hong Kong University of Science and Technology, Hong Kong SAR
JINJIAN WANG, The Hong Kong University of Science and Technology, Hong Kong SAR
YANBIN GONG, The Hong Kong University of Science and Technology, Hong Kong SAR
YIZHEN ZHANG, The Hong Kong University of Science and Technology, Hong Kong SAR
JINGNAN SUN, Tsinghua University, China
JIAN YAO, Jinan University Affiliated Guangdong Second Provincial General Hospital , China
QIJIA SHAO, The Hong Kong University of Science and Technology, Hong Kong SAR
LILI QIU, Microsoft Research Asia, China
QIAN ZHANG†, The Hong Kong University of Science and Technology Hospital, Hong Kong SAR
GUIHUA LI†, Jinan University Affiliated Guangdong Second Provincial General Hospital; Second Clinical College, Southern Medical University, China

As the global population ages, the prevalence of cognitive impairment continues to rise, highlighting the urgent need for accessible and low-burden cognitive assessment. Current assessments based on clinical scales are often hindered by subjectivity, significant user burden, and practice effects, limiting their applicability. We observe that passive visual stimuli can engage multiple cognitive domains while minimizing the need for active participation. Considering the lack of related datasets, we establish EasyCog, the first large-scale multimodal dataset for low-burden cognitive assessments. EasyCog collects synchronized forehead/ear EEG and contactless eye tracking data while participants passively view a short, cognitively structured video followed by an eyes-closed rest. The dataset includes 101 participants spanning healthy controls and patients with PD, AD, and VaD, with clinician-administered MoCA/MMSE scores collected in daily settings. We provide detailed collection procedures, quality validation, implementation, and benchmark baselines. Results indicate assessment feasibility while highlighting generalization challenges. By integrating passive visual stimuli with affordable sensing, EasyCog provides a foundation for future research in accessible and scalable cognitive monitoring in both clinical and community settings.

*Both authors contributed equally to this research.
†Both authors are co-corresponding authors.

Authors' Contact Information: Qingyong Hu, qhuag@cse.ust.hk, The Hong Kong University of Science and Technology, Hong Kong SAR; Yuxuan Zhou, yzhoudo@connect.ust.hk, The Hong Kong University of Science and Technology, Hong Kong SAR; Jinjian Wang, jwangjx@connect.ust.hk, The Hong Kong University of Science and Technology, Hong Kong SAR; Yanbin Gong, ygongae@connect.ust.hk, The Hong Kong University of Science and Technology, Hong Kong SAR; Yizhen Zhang, yzhangtf@connect.ust.hk, The Hong Kong University of Science and Technology, Hong Kong SAR; Jingnan Sun, sunbme@foxmail.com, Tsinghua University, China; Jian Yao, yj1959190042@126.com, Jinan University Affiliated Guangdong Second Provincial General Hospital , China; Qijia Shao, qijiashao@ust.hk, The Hong Kong University of Science and Technology, Hong Kong SAR; Lili Qiu, lili@cs.utexas.edu, Microsoft Research Asia, China; Qian Zhang, qianzh@cse.ust.hk, The Hong Kong University of Science and Technology Hospital, Hong Kong SAR; Guihua Li, guihuali19790302@sina.com, Jinan University Affiliated Guangdong Second Provincial General Hospital; Second Clinical College, Southern Medical University, China.

ACM 2474-9567/2026/3-ART4
https://doi.org/10.1145/3789682

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; • **Applied computing → Health informatics**.

Additional Key Words and Phrases: Datasets, Cognitive Assessment, Human Electrical Signal

ACM Reference Format:
Qingyong Hu, Yuxuan Zhou, Jinjian Wang, Yanbin Gong, Yizhen Zhang, Jingnan Sun, Jian Yao, Qijia Shao, Lili Qiu, Qian Zhang, and Guihua Li. 2026. The EasyCog Dataset: Towards Easier Cognitive Assessment With Passive Video Watching. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 1, Article 4 (March 2026), 28 pages. https://doi.org/10.1145/3789682

1 Introduction

The global trend of population aging is driving a significant increase in the prevalence of cognitive impairment such as dementia. Cognitive impairment, as a progressive and heterogeneous condition often caused by neurodegenerative diseases, poses a major public health challenge and significantly impacts quality of life. In 2020, over 50 million people worldwide were living with cognitive impairment, a number projected to reach 152 million in 2050 [6]. Since cognitive impairment is hard to fully cure, early detection and regular cognitive assessment are crucial for timely intervention and disease stabilization [23].

Current clinical assessments typically utilize well-designed scales, including the Mini-Mental State Examination (MMSE) [8] and the Montreal Cognitive Assessment (MoCA) [33], which require patients to complete a series of predefined questions to evaluate cognitive functions. However, their effectiveness is hindered by several limitations, including subjectivity, substantial user burden, and reduced suitability across different stages of cognitive decline [33]. Specifically, (1) **subjectivity:** assessment outcomes depend heavily on both patient and clinician subjective factors. Patients' willingness to cooperate may be compromised by emotional distress or stigma [41], while scoring can be influenced by clinicians' subjective judgment and biases related to patients' backgrounds (e.g., education level and language). (2) **user burden:** these assessments require active participation, which can be challenging for patients with moderate to severe impairment due to attention deficits and communication difficulties [77]. (3) **practice effects:** for patients undergoing repeated assessments, fixed test content may lead to familiarity with the tasks, resulting in distorted scores that do not accurately reflect true cognitive change [43, 62]. These limitations motivate the need for a more objective, low-burden, and easily deployable cognitive assessment method suitable for both clinical and community environments.

Recent EEG-based approaches show promise for cognitive impairment detection and continuous assessment [7, 66], yet require a 50-minute wear time, controlled laboratory settings, and professional operators [27]. Eye-tracking reduces hardware burden but typically relies on instruction-heavy protocols (e.g., anti-saccade) and calibration challenges [18, 24, 46, 60]. Consequently, neither modality alone readily supports low-cost, low-effort screening outside clinics. A promising alternative is passive visual engagement: many cognitive domains assessed by MoCA and MMSE, including attention, memory, executive function, language, and visuospatial processing, can be stimulated by structured visual content (Sec. 3). Prior work shows naturalistic scenes, semantic anomalies, and social imagery elicit measurable responses even without verbal/motor tasks [17, 47, 52]. When combined with easy-to-wear, low-cost EEG [11], this could support a passive and low-burden assessment paradigm. However, publicly available datasets that support low-burden, low-cost cognitive assessment in this setting remain limited, which slows progress toward practical and scalable approaches.

To bridge the gap, we introduce EasyCog, a public multimodal dataset for passive, low-burden cognitive assessment. EasyCog provides synchronized forehead- and ear- EEG together with eye-tracking data from 101 participants collected in real-world clinical settings. During collection, participants simply watch a 7-minute video consisting of carefully designed, cognitively targeted tasks, followed by a 3-minute rest, with no active responses required, as shown in Figure 1. Based on MoCA scores [33], the dataset spans cognitive levels from cognitively healthy (13) to mild cognitive impairment (37) and dementia (51). By clinical diagnosis, it includes 21 controls

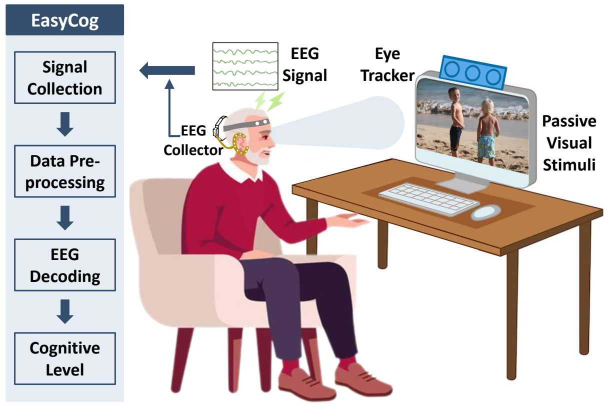


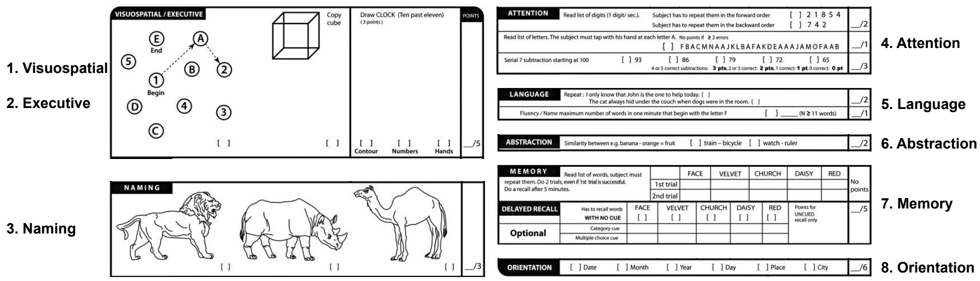
Fig. 1. Illustration of an application scenario of EasyCog. Patients watch a structured video composed of cognitively rich visual stimuli designed to passively engage multiple cognitive domains. During the viewing session, lightweight EEG and eye-tracking devices record neural and behavioral responses without requiring active task participation. This setup enables objective, scalable, and low-cost cognitive evaluation in both clinical and community settings.

without any diagnosed neurological disease and patients with PD (31), AD (19), VaD (29), and neurosyphilis (1). Notably, the diagnosis groups and MoCA-based cognitive levels are not mutually exclusive, and controls may also exhibit cognitive impairment. Data were collected across four daily settings, including outpatient rooms, wards, nursing homes, and public hospital areas, with typical real-world distractions such as noise and bystanders.

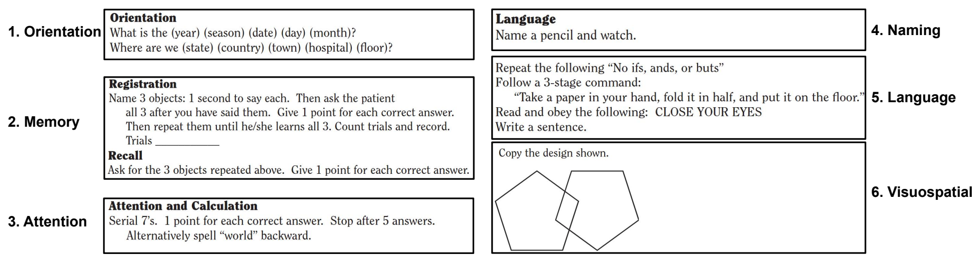
We verified data quality using a rigorous quality control pipeline and targeted stress tests involving motion and electrode displacement, and reported signal quality metrics across settings. To illustrate dataset use, we release implementation and compact benchmarks spanning five feature-extraction families and three training schemes. Across evaluations, EasyCog shows that combining passive visual stimuli with low-cost forehead and ear EEG can capture clinically relevant cognitive information, while highlighting generalization challenges across subjects and settings. Beyond score regression, EasyCog also supports dementia classification and EEG–image response analysis. We expect EasyCog and the accompanying benchmarks to facilitate future research on low-burden, cost-effective cognitive monitoring. Our contributions can be summarized as follows:

- We introduce EasyCog, to our knowledge, the first publicly available multi-modal dataset for low-cost and low-burden cognitive assessments. EasyCog showcases its diversity, comprising 101 subjects spanning 3 types of dementia and all cognitive levels.
- We provide a quality control pipeline and targeted stress tests under motion and electrode displacement, together with signal quality metrics and scripts to support transparency and reuse.
- We construct benchmark baselines with varying sensing modalities, analysis architectures, and optimization strategies for promoting the development of low-effort and low-cost cognitive assessment.
- We present extensive baseline results and analyses that demonstrate feasibility and highlight generalization challenges across subjects and real-world settings. These results highlight the potential value of the EasyCog settings and dedicated designs to fully realize the benefits. Data and code are publicly available [1].

[1]https://github.com/EasyCog/EasyCog-Benchmark



(a) Core cognitive functions of the MoCA scale



(b) Core cognitive functions of the MMSE scale

Fig. 2. Core cognitive functions of MoCA and MMSE scales

2 Related Works

2.1 Cognitive Functions and Standard Assessments

Cognition refers to the set of mental processes involved in acquiring knowledge and understanding through thought, experience, and senses. These processes, while interrelated, are commonly divided into several key functions: complex attention, executive functioning, learning and memory, language, perceptual-motor/visuospatial function, and social cognition [29]. Cognitive decline associated with conditions such as dementia and mild cognitive impairment (MCI) typically manifests as deterioration in one or more of these domains.

Currently, the gold standard for assessing the overall cognitive ability are the MoCA and MMSE scales. These tests assess overlapping cognitive domains and produce a total score out of 30. As shown in Figure 2(a), the MoCA evaluates eight core areas: visuospatial/executive function, naming, attention, memory, abstraction, language, and orientation. A score equal to or above 26 generally indicates normal cognition, while scores between 19 and 25 suggest MCI, and scores equal to or below 18 indicate moderate to severe impairment. Similarly, the MMSE assesses six components spanning comparable domains with simpler questions as shown in Figure 2(b), with score thresholds of >26 for normal cognition, 21–26 for MCI, and <21 for more advanced impairment.

While these tools offer structured, interpretable metrics, they are limited in their applicability across all patient groups. Their reliance on direct verbal responses, culturally specific content, and fixed task formats can reduce fairness and sensitivity, especially when used repeatedly or in diverse populations. This motivates the need for alternative approaches that can probe multiple cognitive functions in a naturalistic and inclusive manner.

2.2 Objective Cognitive Assessment

Recently, some researchers have attempted to simplify cognitive assessment, as shown in Table 1. Many objective methods have been proposed to address the subjective issues of current clinical scales. EEG-based assessments can provide accurate cognitive information. By analyzing the temporal, spectral, and spatial features of EEG signals

Methods	Application	Low-cost Sensor	Wearing Effort	Assess. Effort
EEG [7, 30, 65, 66]	Assessment	×	High	Low
Gamification [31, 45]	Assessment	✓	Low	High
Eye tracking [20, 37, 52, 75]	Classification	✓	Low	High
EasyCog	Assessment	✓	Low	Low

Table 1. Comparison of cognitive assessment methods. Here Assessment indicates continuous cognitive score regression while Classification is for discrete severity classification. "Assess. Effort" denotes the cooperation requirement of the participants.

Datasets	Modalities	#Participants	Diseases	Low-cost Sensor	Wearing Effort	Environment	Cognitive Labels
Anjum et al. [7]	EEG	149	PD	×	High	Controlled	MoCA
Miltiadous et al. [53]	EEG	59	AD, FTD	×	High	Controlled	MMSE
Valdes-Sosa et al. [68]	EEG, MRI	282	-	×	High	Controlled	MMSE, WAIS-III
Cejnek et al. [13]	EEG	168	AD	×	High	Controlled	MMSE
Jesus Jr et al. [32]	EEG, MRI	204	AD	✓	Medium	N/A	MMSE
EasyCog	EEG, ET	101	AD, PD, VaD	✓	Low	Daily	MoCA, MMSE

Table 2. Comparison of cognitive assessment datasets. 'MRI':'Magnetic Resonance Imaging', 'ET': 'Eye Tracking', 'PD': 'Parkinson's Disease', 'AD':' Alzheimer's Disease', 'FTD': 'Frontotemporal Dementia', 'VaD': 'Vascular Dementia', 'N/A':'Not Available'. Valdes-Sosa et al. [68] focus on functionally healthy participants.

during closed-eye resting, it is possible to evaluate cognitive decline at various levels [7, 30]. Furthermore, Sun et al. [65, 66] developed machine learning and deep learning methods based on various EEG computational features to enable continuous mapping of cognitive states, achieving a more refined evaluation. However, current EEG-based assessments typically rely on classic electrode arrangements, such as the 10-20 EEG electrode placement standard, with electrodes uniformly distributed across the scalp. This setup requires professional clinicians' assistance and places a significant burden on both patients and clinicians due to an average wear time of about 50 minutes [27], making it unsuitable for early screening or regular assessments.

There are efforts to develop gamified assessments [31, 45] that reduce system requirements and simplify the assessment process. Moreover, building on previous findings regarding eye movement behaviors in dementia, some researchers have developed a series of eye movement tasks to capture abnormal eye movements (e.g., gaze, saccade, pursuit) for detecting dementia or assessing cognitive levels [20, 37, 52, 75].

However, these attempts often involve complicated calibration processes or the difficulty of learning new tasks. These factors hinder testing for patients with severe cognitive impairments or lower education levels [18, 24, 60]. Towards a low-burden cognitive assessment, the EasyCog dataset collects EEG signals and eye tracking data without complex task descriptions, significantly reducing the likelihood of misinterpreting tasks for severely impaired patients. Participants only need to watch a video and rest with their eyes closed during the assessment.

2.3 Advanced Sensor for Brain Information

Around-the-ear EEG devices offer low-burden wearability and provide extensive information about cognitive functions in the temporal lobe, enabling applications such as imagery and monitoring [35]. There has been research [36] to integrate ear-EEG into the headphone for practical cognitive load monitoring. cEEGrid sensors [11] have been demonstrated to be effective in various investigations, and they have recently become commercially available [2]. Forehead electrodes provide abundant information about the frontal lobe [12, 55] and moderate information from the parietal lobe [12]. This makes them suitable for evaluating cognitive functions such as execution and organization, and complementing around-the-ear electrodes to form comprehensive brain activity sensors. Given that both forehead EEG and ear-EEG share the properties of low cost and easy wearability, we have chosen to adopt both to create a user-friendly and comprehensive EEG device.

2.4 Existing Dataset for Cognitive Assessment

Many datasets have been proposed for cognition-related tasks utilizing different modalities, primarily including EEG [7, 13, 32, 53, 68, 78] and eye movement [38, 42, 61, 64], and the comparison is shown in Table 2.

For EEG datasets, Anjum et al.[7] propose a large dataset comprising 100 patients with Parkinson's Disease (PD) and 49 healthy controls. The eyes-open resting EEG is collected in a quiet room using high-density 64-channel EEG devices, with labels provided by MoCA scores. Miltiadous et al.[53] construct a dataset featuring multiple types of dementia, including 36 patients with Alzheimer's Disease (AD) and 23 patients with frontotemporal dementia (FTD). They utilize a 19-channel scalp EEG device to collect resting EEG, with the ground truth determined by the MMSE scale. Valdes et al.[68] collect a large dataset of 282 healthy controls with multimodal data, including magnetic resonance imaging (MRI), 30 minutes of resting EEG, and MMSE scale scores. Additionally, another dataset collects 21-channel resting EEG data from 102 healthy controls, 7 patients with Mild Cognitive Impairment (MCI), and 59 patients with dementia [13]. Jesus et al. [32] gather data from 103 participants with probable AD and 101 healthy controls using a 7-channel scalp EEG device, with labels based on MMSE scores. Zhang et al. [78] recruit 104 participants to investigate flow rather than cognitive assessment. Most of these datasets are collected using professional high-density EEG devices, which can cause discomfort and distress for participants. Although some datasets [32] utilize consumer-grade EEG devices, these scalp EEG setups still require hair washing in advance, making them less convenient. Besides, they only capture part of the clinical scales, which limits applicability across patient groups [63].

Several eye-tracking datasets have also been proposed to investigate the correlation between cognitive activities and eye movements [38, 42, 61, 64], but they remain limited to shallow cognitive behaviors or specific diseases. Ktistakis et al. [38] and Langner et al. [42] focus on cognitive load and flow, Skaramagkas et al. [64] investigate emotional states triggered by videos, and Rojas et al. [61] study the behavior of children with Attention Deficit/Hyperactivity Disorder (ADHD).

In summary, while low-burden and easy-to-follow cognitive assessment systems are urgently needed, existing EEG datasets are still limited to cumbersome and inconvenient devices in a controlled environment, and eye-tracking datasets are restricted to shallow cognitive characteristics. This limitation hinders the development of practical and effortless cognitive assessments. To bridge these gaps, we introduce EasyCog, the first publicly available dataset to simultaneously provide EEG and gaze data captured using consumer-grade, easy-to-wear hardware during passive visual stimulation. This dataset provides: (1) **multi-modal** data captured using low-cost, easy-to-wear sensors (forehead and cEEGrid electrodes), and eye tracking data from a commercial eye tracker device [4]; (2) **both MoCA and MMSE scores** for a comprehensive understanding of the cognitive levels; (3) **participants with various dementia types** (PD, AD, and Vascular Dementia (VaD). Crucially, the EasyCog dataset contains **everyday settings** like nursing homes to facilitate practical real-world deployment.

3 Design of The Data Collection System

In this section, we introduce the EasyCog collection rationale. We firstly present the basic insights, followed by our observations and design rationale for the passive video stimuli and signal collection, respectively.

3.1 Motivation

Designing an effective cognitive assessment tool requires careful consideration of the diverse needs and capabilities of the target population while traditional task-based methods present significant barriers:

- Patient Burden and Stigma: Individuals with early cognitive changes or normal cognition often feel "reluctant" to participate due to stigma or the anxiety of formal testing [41]. To increase acceptance and participation, assessments should minimize perceived pressure and cognitive load, like free-viewing tasks [72] which offer a low-stress alternative.

- Practice Effects: For patients with mild to moderate impairment requiring regular monitoring, repeated exposure to fixed-content tests (like MoCA/MMSE) leads to "practice effects" that mask true cognitive changes [43, 62]. To counter this, assessments must be dynamic and complex enough to capture real variation, while remaining cost-effective and practical for routine use [34].
- Instruction Complexity: Patients with severe impairment may be "unable to comprehend instructions" [24, 60] or "sustain attention"s [77], rendering assessments that require strict task compliance unusable. In such cases, methods that rely on passive interaction and intuitive responses—rather than explicit instructions—are essential [18].

Existing approaches that attempt to reduce assessment burden—such as gamified tasks or eye-tracking-based evaluations typically require complex calibration procedures, strict task compliance, or abstract cognitive processing [37, 45, 60]. While prior research suggests that naturalistic tasks can reduce task demands [24, 76], such tasks are too coarse to probe specific cognitive functions. Eye movement patterns alone (e.g., gaze duration or saccade frequency) are insufficient to capture the multidimensional deficits seen in early-stage dementia.

By recognizing and addressing the distinct needs of these patient subgroups, the EasyCog dataset is specifically designed to support inclusive, low-burden cognitive assessment. Driven by the insight that comprehensive cognitive assessment can be achieved through passive interaction when supported by structured stimuli and response sensing, the design of EasyCog is grounded in two key principles: (1) crafting structured visual stimuli that are intuitively processed but systematically target specific cognitive domains; and (2) collecting deeper neural and behavioral signals with low-cost, wearable EEG and eye-tracking hardware, without requiring active task execution or complex calibration.

3.2 Passive Visual Stimuli Protocol

Our assessment is based on a structured, 7-minute video protocol designed to probe the cognitive domains of the MoCA scale passively but systematically. While no explicit responses are required, the protocol is composed of nine distinct tasks (Figure 3) that elicit measurable neural and behavioral correlates of attention, memory, executive function, language, and visuospatial processing. The tasks implicitly probe high-level semantic processing (abstraction), visual-semantic linkage (naming), and logical inference (orientation) through carefully selected visual content. The protocol is complemented by a 3-minute eye-closed resting state, a standard paradigm for assessing baseline brain activity in cognitive studies [7, 65, 66]. The nine visual tasks, grounded in prior cognitive research, are as follows:

Naturalistic Scene Exploration. Participants freely view complex scenes. This task probes attention and executive functions (e.g., curiosity), where cognitive decline is associated with reduced visual exploration [40, 54].

Pictures with Multiple Objects. Viewing images with multiple distinct objects tests visual search strategies, attention, and visuospatial organization [76].

Social vs. Non-social Scenes. Images with and without people are presented to evaluate how social cues (e.g., faces) modulate attention and exploratory behavior, which can be altered in certain dementias [52, 58, 76].

Semantically Congruent vs. Incongruent Images. Participants view images containing semantic anomalies [52, 54], e.g., a lion in a classroom [17, 44]. A diminished response to incongruity can indicate deficits in attention, executive function, or semantic processing.

Paired Social Scenes. Side-by-side images, one with a person and one without, are shown to more directly assess social attention and apathy, a component of executive function [40, 52, 58].

Moving Target Pursuit. A single object moves across a static background. This task assesses sustained attention and visuospatial tracking, which can be impaired in dementia [48].

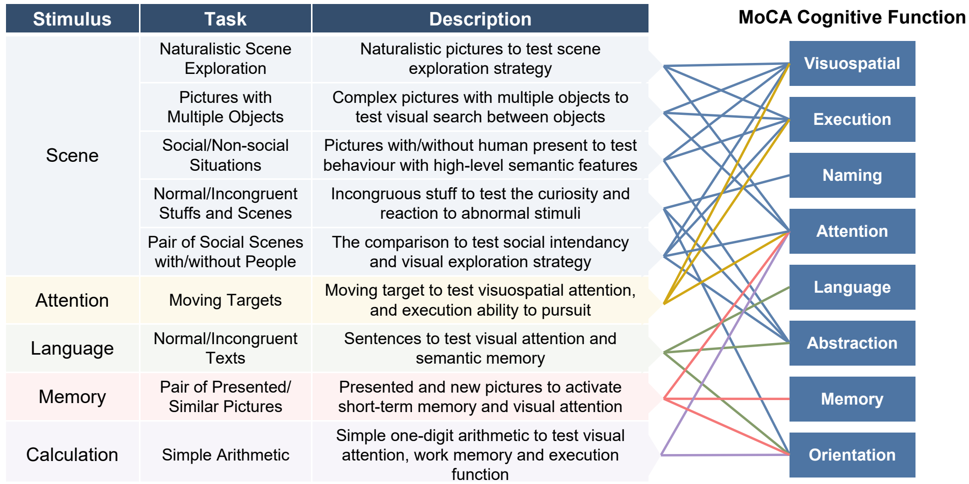


Fig. 3. Correlation between passive visual stimuli and MoCA cognitive functions

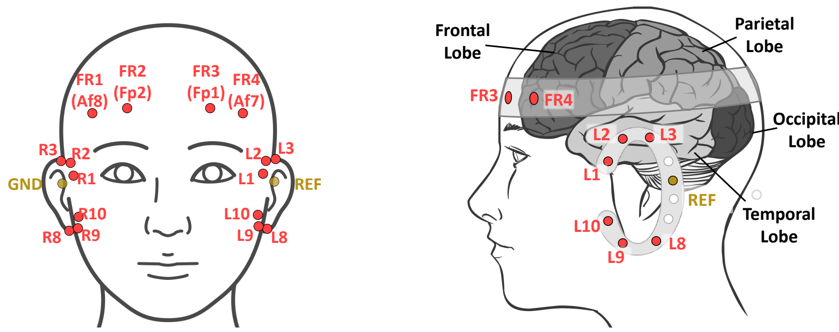
Fig. 4. Task examples in the visual stimuli. The text in the Task 7 example means "The old man is swimming in the swimming pool on his cane.". Here we select the language fits best for the patients but that can be transferrable to other languages.

Semantically Congruent vs. Incongruent Text. Sentences with or without semantic errors (e.g., "The seawater is hard") are displayed. This task specifically targets language comprehension and semantic memory, known to elicit distinct neural signatures [39, 52].

Presented vs. Similar Paired Pictures. A previously seen image is paired with a novel but similar one. Preferential viewing patterns probe short-term recognition memory, as healthy individuals typically focus more on the novel image [16, 47, 52].

Simple Arithmetic. Single-digit arithmetic problems are displayed to passively engage working memory and executive functions related to calculation [49].

To construct the 430-second video, stimuli were presented for 3-5 seconds each, a duration validated in prior work [52, 76]. The task examples are shown in Figure. 4. Stimuli were from established cognitive research datasets [73], generated by DALLE-3 according to clinical references [17, 44], or filmed for moving targets. The final protocol was reviewed by three experienced clinicians who confirmed its comprehensive coverage of the core cognitive domains assessed by the MoCA and MMSE scales. More details about the stimuli design are in the attached Appendix A.



(a) Electrode position (front view)

(b) Electrode position (side view)

Fig. 5. Electrode positions: (a) from the front view and (b) from the left side view. The electrodes in white are not enabled.

3.3 EEG Sensing Hardware Design

To assess cognitive functions comprehensively, we need detailed information to capture neural activities with a low-cost and non-intrusive sensing system. This goal precluded traditional, high-density scalp EEG systems, which are costly, cumbersome, and require expert setup.

3.3.1 Electrode Position Design. To solve this, we design a hybrid, low-cost wearable collection system that integrates two non-intrusive, commercially available sensor types to achieve comprehensive brain coverage:

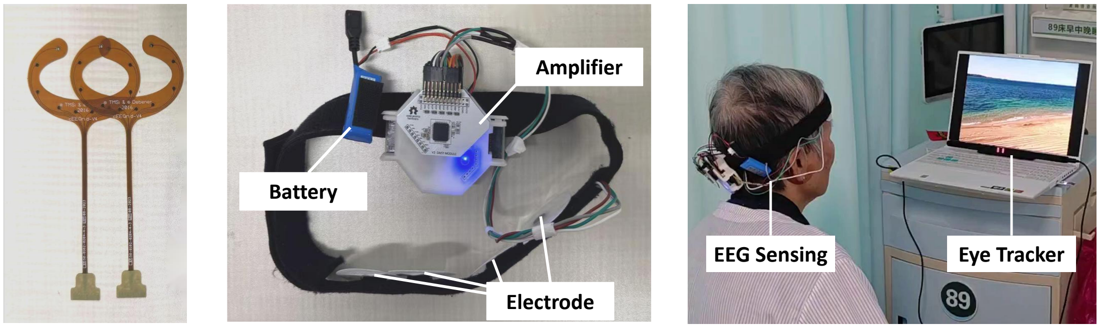
Forehead EEG. We use four Ag-AgCl electrodes integrated into a simple headband [55]. This form factor is easy to wear, bypasses interference from hair, and provides robust signals corresponding to the 10-20 standard positions (Af7, Af8, Fp1, Fp2). These channels are optimally positioned to capture signals from the frontal lobe as shown in Figure 5(a), which is critical for functions like execution and organization.

Ear-EEG. We complement the forehead sensors with ear-EEG sensors [11], which have been widely used in cognitive applications such as speech attention [28] and sleep monitoring [19]. This effectiveness stems from its high sensitivity to signals from the temporal lobe and its ability to obtain partial signals from the parietal lobe [51]. As shown in Figure 5(b), we select 6 electrodes on the cEEGrid as input channels and 1 electrode as the reference or the ground channel.

Neither of these "easy-to-wear" sensor types alone can provide comprehensive cognitive coverage. By combining both ear electrodes and forehead electrodes, it is possible to approximate full-brain sensing: (1) **Frontal and Temporal Lobes** are directly captured by the forehead and ear electrodes, respectively. (2) **Parietal Lobe** activity can be captured in the combined signals as all electrodes capture signals with moderate strength in different directions [12, 25]. (3) **Occipital Lobe** is the furthest area from both the ear and forehead but can be inferred by eye movements as this brain region processes visual information. As all electrodes are placed near the eyes, it is feasible to extract electrooculography (EOG) signals that indirectly indicates the visual cognitive processes occurring in the occipital lobe.

3.3.2 Hardware Implementation. The hardware platform and the data collection scenario of the EasyCog are shown in Figure 6.

Electrodes. We adopt the commercial TMSi cEEGrid electrodes [2] as the around-the-ear channels in Figure 6(a). For the forehead, we integrate four Ag-AgCl electrodes inside a headband. To ensure data quality, all electrodes are wet electrodes with electrolyte gel. The overall cost is within 3 USD and can be quickly set up within 10 minutes for each assessment.



(a) Ear electrode (b) Headband with forehead electrodes (c) Data collection scenario

Fig. 6. Illustration of our hardware platform and the patients conducting the cognitive assessment

Acquisition Board. All channels are connected to the OpenBCI Cyton + Daisy Biosensing Boards with 16 channels [57], as shown in Figure 6(b) with a sampling rate of 125 Hz.

Data Transmission. Data are streamed wirelessly via Bluetooth to a collection laptop.

To ensure user comfort, the ear and headband electrodes are implemented with flexible circuits that can accommodate different user conditions. The weight of the whole collection system is 136 grams, lighter than the commercial headphone [5]. Thus, users can wear the system with natural behaviors watching the video stimuli.

3.4 Eye Tracking Hardware

Since eye movement can also reflect the cognition condition, and it can be monitored contactlessly, which is complementary to the wearable platform, we integrate a commercial eye tracker into our data collection setup. This non-contact approach aligns with our "low-burden" design goal, as it only requires participants to sit in front of a monitor as shown in Figure 6(c).

Hardware Setup. We use a Tobii 4c [4] eye tracker mounted to the laptop monitor used for stimuli presentation. The 16-inch laptop monitor has a resolution of 2560×1600. The eye tracker works at an operating distance from 50 to 95 centimeters with a high sampling rate of 90 Hz [4]. It collects the gaze positions on the screen and we use an open-source software, GazeTrack [1] to collect the streamed gaze coordinations on the screen.

Calibration Protocol. We adopt the standard calibration protocol of Tobii 4c. The participant is suggested to sit in front of the sensor at a distance of around 70 centimeters to make sure the eyes are well captured. Then the participant is instructed to conduct six-point calibration, which requires to fix the gaze on six points in the screen (upper left, upper middle, upper right, lower left, lower middle, lower right) for around 3 seconds.

Accuracy and Dropout Rates. Tobii 4c is a widely-used commercial eye tracker that has enabled various applications [9, 56] with a low error of around 2 degrees [3]. We calculate the dropout rates by counting the overall ratio of zero reports. The dropout distribution across the overall 101 participants is 9.72% (median), 6.23% (25% percentile), and 19.12% (75% percentile).

Data Synchronization. The eye-tracking data stream from the Tobii (90 Hz) and the EEG data stream from the OpenBCI (125 Hz) were synchronized using the software trigger from the control laptop with timestamp records. Then we align the start timestamps from the two modalities.

4 Dataset and Preliminary Validation

4.1 Data Collection Protocol

Demographics	
Gender	51 male, 50 female
Age, years	67.5 ± 11
Education	64 (> 9 years)
Number of Sessions	
One-time Session	94
Multiple Sessions	7
Clinical Assessment	
MoCA Score	16.58 ± 7.36
MMSE Score	21.35 ± 6.76
Impairment Severity	**# participants**
Cognitively-Healthy (MoCA ≥ 26)	13
MCI (26 > MoCA ≥ 19)	37
Dementia (19 > MoCA)	51
Disease Diagnosis	**# participants**
PD	31
AD	19
VaD	29
Others	1
Controls	21

Table 3. Participant characteristics

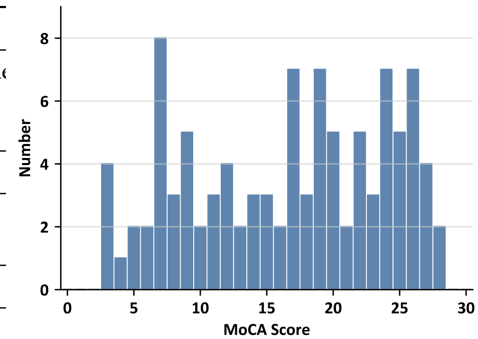


Fig. 7. MoCA score distribution

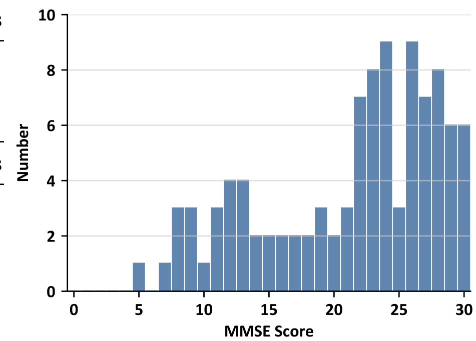
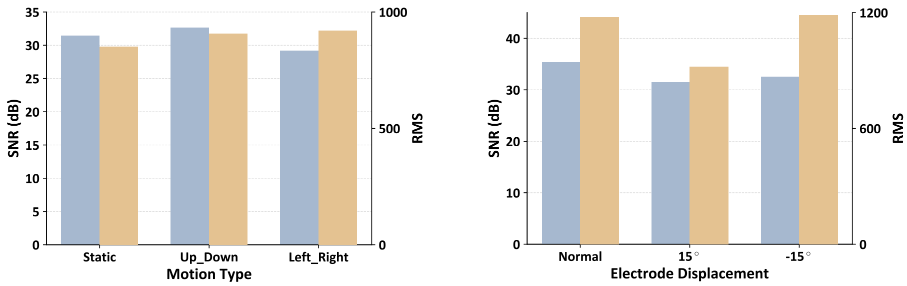


Fig. 8. MMSE score distribution

We collaborate with medical centers under our institute's Institutional Review Board (IRB) approval. To ensure the practicality of the dataset, the assessments are conducted in daily environments with typical real-world factors, such as ambient noises around 50-80 dB, adjacent people, and uncontrolled participant behavior.

Sensors Setup and Labeling. The participant is asked to sit in front of the laptop monitor with a comfortable posture. The data collection process begins with the eye tracker calibration procedure as mentioned in Section 3.4. Specifically, we exclude participants with epilepsy due to potential symptoms caused by near infrared sensing by the eye tracker. For the EEG sensor setup, before presenting the visual stimulus, we firstly clean the skin on the forehead and ear and apply electrolyte gel to the cEEGrid electrodes. Then, we assist the patients in wearing the devices and test the impedance. An impedance of below 150 kΩ is considered acceptable. After all the setup, we ask the participant to watch the video carefully without additional instructions. After the 7-minute video task finishes, we ask the participants to close their eyes for another 3 minutes. Figure 6(c) depicts an example of our data collection in the ward, where real-world interference (e.g., ambient noise) is present. Finally, the clinicians administer the clinical cognitive scales MoCA and MMSE with the participant.

Participant Recruitment. Our study recruited a total of 110 participants and selected 7 individuals to undergo testing 3-4 times over successive days. By screening for data quality, participant adherence, and label availability, we discarded data from 9 participants and data from 3 sessions in successive assessments. A total of 101 participants contributed to our dataset. They were sourced from the outpatient clinic and inpatient department with informed consents, supervised by experienced clinicians. The clinicians identified targeted users of cognitive assessments, like the participants related cognitive diseases including PD, AD, and VaD or controls. Our dataset includes a diverse sample in terms of diseases, cognition conditions, gender, and education levels.



(a) Motion artifact

(b) Electrode displacement

Fig. 9. Signal validity under real-world factors

Ethics. The research received approval from the IRB approval of the host institute. The collaborated clinicians ensure the safety of the participant for the sensors. All participants are informed of their right to withdraw from the collection at any time. We remove all identifying information from the dataset to ensure anonymity. The collected data is stored in a local data repository with downloading logs to maintain an access record.

4.2 Participant Characteristics

As shown in Table 3, the EasyCog dataset is collected from 101 participants, 50 of whom are female and 51 of whom are male. The mean and standard deviation of the age are 67.5 and 11.0, respectively. The cohort includes 21 controls and 80 patients with different diseases, including 31 patients with Parkinson's disease (PD), 19 patients with Alzheimer's Disease (AD), 29 patients with vascular dementia (VaD), and 1 patient with neurosyphilis. The mean and standard deviation for MoCA scores are 16.58 and 7.36, while the mean and standard deviation for MMSE scores are 21.35 and 6.76, respectively. We also summarize the education background of each participant. For participants with fewer than 9 years of education, we add 1 point to the MoCA score, following standard practice [33]. The detailed distributions of MoCA and MMSE scores are shown in Figures 7 and 8.

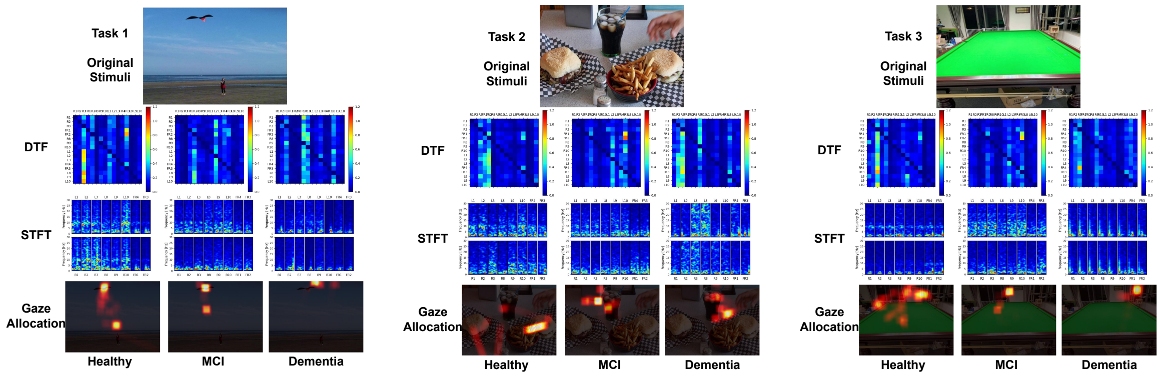
EEG. We first apply a 50-Hz notch filter to remove power noise. Next, we apply a bandpass filter with a frequency range from 1 Hz to 60 Hz, which includes our target $\delta, \theta, \alpha, \beta$, and γ bands. We then use the regular Artificial Subspace Reconstruction (ASR) method [14] from the widely used open-source EEGLab toolbox [21]. ASR is designed to handle non-stationary and non-repeating artifacts, making it suitable for unpredictable motion artifacts caused by patients' unconscious body movements.

Eye tracking. The invalid gaze samples identified by the eye tracker are removed. We perform linear interpolation to align the sampling rate of eye tracking (90 Hz) with the EEG data (125 Hz).

4.3 Signal Validity

To obtain a better trade-off between signal quality and wearability, we select ear-EEG and headband electrodes with wet electrodes with sticky tapes fixed. Despite this consideration, EEG signals can also be influenced by user motion artifacts and electrode displacement. We recruit 4 volunteers to investigate signal validity. After filtering signals as in Section 4.2, we use root mean square (RMS) of the signals and the signal-to-noise ratio (SNR), which is the ratio of the energy of 1-60 Hz over the energy of higher frequencies.

Motion Artifacts. We ask the participants to wear our sensor at the normal position and rest with their eyes closed for one minute with few motions. Then the participants are asked to move their head up and down with multiple rounds for one minute. Later, the head motion from left to right is conducted similarly. As shown in Figure 9(a), with head motions, the signal RMS increase slightly while the SNR values remain stable.



(a) Task 1: naturalistic scenes (b) Task 2: multiple objects (c) Task 3: social/non-social situations

Fig. 10. Differences between HC, MCI, and dementia for Task 1, 2, and 3 in passive visual stimuli

Electrode Displacement. Since our ear-EEG and headband are designed for the ear region and the head region, the space for displacement is relatively limited. We firstly ask the participants to keep resting with eye closed for one minute with few motions, wearing our sensor at the normal position as shown in Figure 5. Then we move the ear-EEG by up to 15 degrees clockwise and the head-EEG with around 1 centimeter right from the participant. Later, the ear-EEG moves from the normal position to around 15 degrees in counter-clockwise (-15°) and the head-EEG with 1 centimeter left from the participant. Figure 9(b) shows the results of the displacement. The clockwise displacement 15° influences the signal RMS as it has more contact with hair. Since our electrodes are gel based, the SNR values remain stable across the displacements.

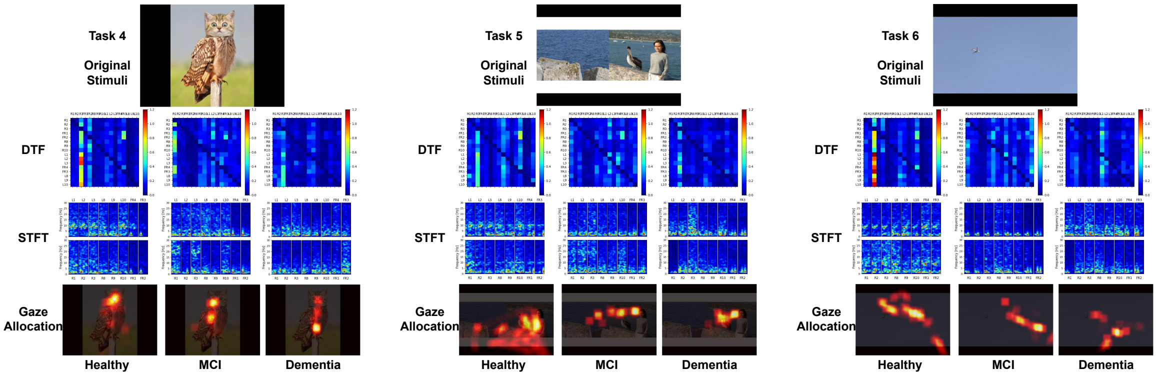
4.4 Neuro-physiological and Behavioral Feature Validation

To validate the feasibility of our collected features, we examined systematic differences in neurophysiological and behavioral responses among cognitively-healthy controls (HC), individuals with MCI, and patients with dementia. Participants were stratified based on MoCA scores: HC (≥26), MCI (19-25), and dementia (<19). We analyzed Short-Time Fourier Transform (STFT) of 16-channel EEG to assess spectral power distribution and Directed Transfer Function (DTF) over 0-30 Hz to quantify inter-regional information flow. Concurrently, gaze allocation was analyzed to reveal attentional patterns.

A consistent pattern emerged across nearly all visual tasks: a gradient of neurophysiological activity corresponding to the severity of cognitive impairment. Both DTF and STFT analyses revealed that HCs consistently demonstrated stronger functional connectivity and higher alpha-band (8-14 Hz) power compared to MCI and dementia groups (Figure 10, 11, 12). This observed decrease in alpha power, and consequently an increased theta-to-alpha ratio (TAR), aligns with established neural correlates of visuospatial, memory, and language dysfunction [10]. Gaze patterns provided complementary behavioral evidence, revealing distinct group-specific attentional strategies.

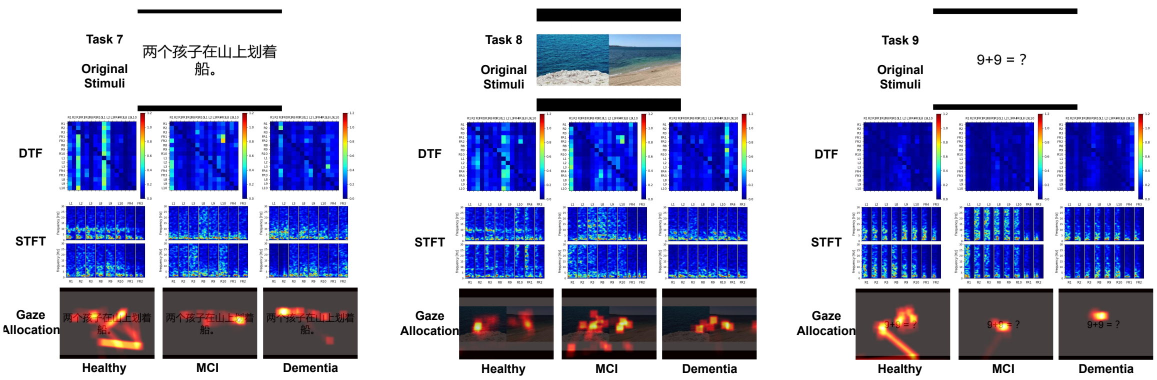
During naturalistic scene exploration (Task 1, Figure 10(a)) and object viewing (Task 2, Figure 10(b)), HCs exhibited broad, exploratory gaze patterns. In contrast, patients with dementia demonstrated pronounced gaze fixation, concentrating on single salient objects (e.g., the eagle in Task 1) and a limited subset of items in Task 2 and Task 3 (Figure 10(c)). This restricted visual search is consistent with prior findings of diminished curiosity [17] or attentional dysfunction [54] in dementia.

Tasks designed to probe the detection of incongruity further highlighted these attentional differences (Figure 11). When presented with an anomalous image (a bird with a cat's head; Task 4, Figure 11(a)), HCs not only



(a) Task 4: normal/incongruent pictures (b) Task 5: social scenes w./w.o people (c) Task 6: moving targets

Fig. 11. Differences between HC, MCI, and dementia for Task 4, 5, and 6 in passive visual stimuli



(a) Task 7: normal/incongruent texts (b) Task 8: presented/similar pictures (c) Task 9: simple arithmetic

Fig. 12. Differences between HC, MCI, and dementia for Visual Task 7-9. The text in (a) means 'Two children are boating on the mountain'. Here we select the language fits best for the patients but that can be transferrable to other languages.

fixated on the incongruous feature but also scanned its surroundings, suggesting a comprehensive evaluation. MCI participants identified the anomaly and showed reduced subsequent exploration, whereas dementia patients showed minimal attention to it, potentially indicating a failure to recognize the incongruity. Similarly, in a social scene task (Task 5, Figure 11(b)), HCs inspected both the present individual and the empty space where a person was absent in a paired image. Dementia patients, however, disproportionately attended to a non-social element (a bird) over the person, particularly in one of the images. This behavior may be indicative of apathy, a common symptom in certain dementias [70]. For the moving target task (Task 6, Figure 11(c)), HCs successfully tracked the object's full trajectory. While MCI participants also tracked the target, their performance was most effective in the latter half of its path. In stark contrast, the gaze patterns of dementia patients failed to form a coherent trajectory, indicating a clear impairment in smooth pursuit and sustained attention.

The final set of tasks targeted higher-order cognitive processes (Figure 12). In a semantic incongruity task (Task 7, Figure 12(a)), HCs and MCI participants focused their gaze on the conflicting words ("boating" and "mountain") within the sentence "Two children are boating on the mountain," demonstrating recognition of the semantic error. Patients with dementia, however, fixated on a non-conflicting word ("children"), failing to engage with the logical inconsistency. The visual recognition memory task (Task 8, Figure 12(b)) also revealed clear deficits. HCs

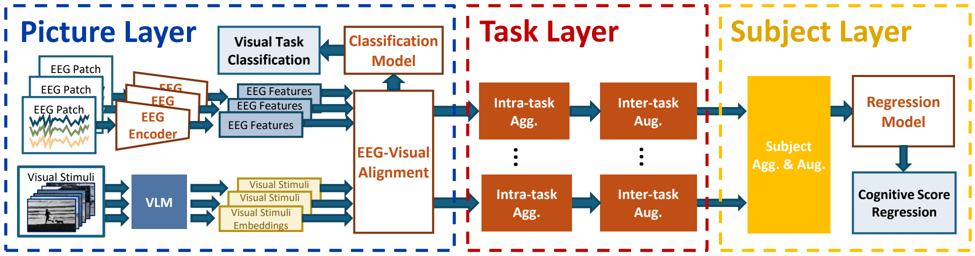


Fig. 13. Overview architecture of our proposed CogAssess model

preferentially explored novel over previously seen images, a hallmark of intact recognition memory [47]. MCI participants showed a less distinct preference, suggesting memory impairment, while dementia patients tended to stare at the screen's center or the previously presented image, a behavior consistent with severe memory dysfunction or apathy [16, 17].

We also observed task-dependent modulation of brain activity. The simple arithmetic task (Task 9, Figure 12(c)) elicited markedly lower overall DTF amplitudes across all groups compared to the complex visual tasks, suggesting lower cognitive resource recruitment. For dementia patients in this task, EEG signals were characterized by stable, low-frequency activity with minimal high-frequency energy. This pattern suggests either a failure to engage with the task or an inability to perform the calculation. We posit that eye-movement information, potentially extractable from forehead electrodes, could be leveraged to disambiguate between non-engagement and cognitive failure by assessing gaze fixation on the equation.

In summary, the consistent neuro-physiological and behavioral differences observed across the cognitive spectrum provide a strong empirical foundation for cognitive analysis.

5 Benchmark

The benchmark task is cognitive score regression (MoCA/MMSE, 0-30) from EEG signals collected during passive viewing of visual stimuli and during rest. We propose CogAssess, a novel deep learning framework designed to address the challenge of regressing a stable, long-term cognitive status from short-term, multi-task signals. CogAssess employs a three-stage hierarchical architecture (Picture, Task, Subject) and is trained with a specialized alternative optimization protocol. The backbone is an EEG foundation model, CBraMod [71], adapted to our data by masked signal reconstruction with 75% mask ratio.

5.1 CogAssess: A Hierarchical Learning Framework

CogAssess is a three-stage hierarchical architecture with a specialized training protocol. More details can be found in Appendix B.

Picture Layer: This layer extracts features from 3-second EEG segments corresponding to each picture. To learn rich representations, it jointly optimizes two objectives:

(1) **EEG-Visual Alignment:** It aligns EEG features with the visual embeddings produced by a VLM (Janus-Pro-7B [15]) using cross-attention. This module is guided by a proxy task: visual task classification.
(2) **Sample-wise Contrastive Learning:** To preserve subject-specific cognitive patterns and prevent over-fitting to visual stimuli, it uses contrastive learning [26] with pink noise augmentation [65] to make representations from different subjects discriminative.

Task Layer: This layer aggregates picture-level features into robust task embeddings.

(1) **Intra-task Aggregation:** It uses conditional attention mechanisms—a standard one for isomorphic tasks and a difference-based one for contrastive tasks—to create task-specific representations.

Variants	Baselines	Feature Extractor	Training Scheme	Data
Cognitive Assessment with EEG Solutions	BL-A	Handcraft+ML	End-to-End	R
	BL-B	Handcraft+ML	End-to-End	R+V
	BL-C	GCN-Transformer	End-to-End	R
EEG Foundation Model	BL-D	Foundation Model	End-to-End	R
	BL-E	Foundation Model	Frz. E+Trainable Reg.	R
	BL-F	Foundation Model	Frz. E+Trainable Reg.	R+V
	CogAssess-v1	Foundation Model	Alternative	R+V
Eye Tracking Input	BL-G	Eye tracking Model	Alternative	ET
CogAssess with Visual Stimuli	CogAssess-v2	Ours	Alternative	R
	CogAssess-v3	Ours	Alternative	V
	CogAssess	Ours	Alternative	R+V

Table 4. Benchmark models. 'BL': 'Baseline', 'ML': 'Machine Learning', 'Frz.': 'Frozen', 'E': 'Encoder', 'Reg': 'Regressor', 'R': 'Resting EEG', 'V':'Visual Task EEG', 'ET': 'Eye Tracking'.

(2) **Inter-task Augmentation:** Based on the clinical observation of order invariance in cognitive tests, we apply a permutation-based augmentation by shuffling a random subset of task embeddings. This forces the model to learn order-invariant features and expands the training data.

Subject Layer: The final layer concatenates the augmented task embeddings and feeds them into a 2-layer Transformer to model inter-task correlations. The resulting subject-level embedding is used to regress MoCA/MMSE scores. An auxiliary loss on predicting clinical sub-scores provides additional regularization.

5.2 Alternative Optimization Protocol

Due to sparse supervision, i.e., one score per subject, we employ an Alternative optimization strategy. The training alternates between updating the picture-level feature extractor using the classification and contrastive losses and updating the entire model end-to-end using the final regression loss. This stabilizes training by balancing the general feature learning with task-specific tuning.

5.3 Baseline Methods

To the best of our knowledge, there are currently no methods specifically designed for stimulus-triggered cognitive assessment with low-cost devices. However, to further validate the reliability of the proposed dataset and evaluate the efficiency of our proposed CogAssess, we re-implemented four advanced baseline models from similar applications as shown in Table 4.

- Baseline A: The work [66] proposes a machine learning method based on multi-domain features extracted from resting EEG, achieving impressive performance in predicting cognitive scores using standard 19-channel EEG data. We re-implement this method by extracting the same statistical features from the resting EEG signals and predicting with the AdaBoost regressor.
- Baseline B: We replace the resting input from Baseline A with EEG from both visual and resting tasks.
- Baseline C: Sun et al. [65] designs a graph convolution (GCN) enhanced Transformer to analyze scalp EEG data from over 700 subjects. We reproduce the model using the 3-minute resting EEG signals as input.
- Baseline D: As the SOTA EEG foundation model, CBraMod [71] leverages spatiotemporal attention with extensive EEG knowledge from multiple datasets. We re-implement the end-to-end full-parameters finetuning in CBraMod to the cognition regression task with the 3-minute resting EEG signals.
- Baseline E: To further explore the optimization strategy, we freeze CBraMod as the encoder and use our trainable regressor to predict the cognitive scores with 3-minute resting EEG signals.

- Baseline F: Compared with Baseline E, we replace the input from the resting EEG signals with the signals during both visual tasks and the resting task. The encoder is frozen on the picture layer with average aggregation for task features and concatenates the task features as subject features to train the regressor.
- Baseline G: Mengoudi et al. [52]. propose an eye-tracking-based dementia detection method using visual stimuli. We reproduce the same feature extractor and replace the classifier to our regressor with our eye-tracking data.
- CogAssess-v1: We further replace Baseline F's training strategy with our alternative scheme. We train the foundation model with the cross entropy loss on the picture layer, whose features are processed with our task layer and subject layer as the alternative training.
- CogAssess-v2: We use our CogAssess method but only take the resting EEG signals as the input. Therefore, the picture-layer classifier and EEG-visual alignment are disabled correspondingly.
- CogAssess-v3: To evaluate the importance of our visual stimulus, we use our CogAssess model and only signals during visual tasks as the input.
- Ours: The complete version of CogAssess with EEG signals during both visual and resting tasks as input. It incorporates EEG-visual alignment and sample-wise contrastive learning in the picture layer, and adopts our task layer and subject layer to alternatively train the regressor.

6 Results and Findings

In this section, we will illustrate how we have demonstrated the effectiveness of EasyCog. We will first introduce the evaluation methodology, and then the performance of benchmark methods on our EasyCog dataset. Then we will provide an analysis of the results. Finally, we present the robustness study of our passive stimuli and provide extra insights about electrode configuration optimization and disease diversity.

6.1 Evaluation Methodology

Train / Validation / Test Split. We conduct 10-fold **cross-subject** split to form the training, validation, and testing sets. A total of 101 participants are divided into 10 groups of 10 to 11 individuals. We ensure that each group contains a balanced number of cognitively-healthy individuals (MoCA score ≥ 26), those with MCI ($25 >$ MoCA score ≥ 19), and individuals with moderate or severe cognitive impairment (MoCA score < 19). Each group is rotated as the test set while another group is selected as the validation set. The other left groups are all training sets. In the training phase, we save the model parameters with the best performance on the validation set to demonstrate the optimal feature extraction capability for the unseen group. The model from the last epoch with the early stop on the validation set will be evaluated on the test set to demonstrate the generalization capability.

Metrics. We evaluate the benchmarks' performances compared with the clinical professional's assessment of MoCA and MMSE. The metrics include the mean absolute error (MAE) and Pearson correlation coefficient (PCC) with MoCA and MMSE scores. Since both the validation and test sets contain subjects **unseen** during training, we report the results on the validation set with the best performance on the validation set and the test set with the last model parameters. We regard the best performance on the validation set and the test set as the metrics of cognitive information fitting and generalization ability, respectively.

6.2 Overall Performance

We first conduct the cognitive score prediction task using our benchmark methods, as shown in Table 5.

Performance of SOTA Scalp-EEG solutions: Baseline A uses handcrafted statistical features as input and Baseline B adds visual task signals; however, the performance does not improve. This may be due to a weak regressor. Baseline C is a spatiotemporal transformer using resting EEG, achieving a better validation performance but a poor test performance. This indicates that traditional methods, while effective for clean, lab-grade EEG,

	Validation Set				Test Set			
	MoCA		MMSE		MoCA		MMSE	
	MAE	PCC	MAE	PCC	MAE	PCC	MAE	PCC
BL-A	6.223	0.237	5.677	0.143	**6.183**	**0.218**	**5.536**	0.223
BL-B	6.617	0.105	5.783	0.132	6.533	0.069	5.628	0.197
BL-C	5.830	0.264	5.264	0.296	8.352	-0.106	7.360	-0.081
BL-D	6.336	0.155	5.718	0.190	6.892	0.095	5.892	0.136
BL-E	6.256	0.039	5.339	0.059	6.498	-0.013	5.814	-0.003
BL-F	6.265	0.029	5.381	0.034	7.066	0.011	5.932	0.000
BL-G	**4.792**	**0.596**	**4.238**	**0.559**	7.068	-0.025	5.875	0.021
CogAssess-v1	5.156	0.502	4.663	0.518	7.934	0.068	6.616	0.101
CogAssess-v2	5.834	0.296	4.787	0.366	7.094	0.023	6.025	0.046
CogAssess-v3	5.451	0.421	4.363	0.469	7.445	0.116	6.062	0.195
CogAssess	5.339	0.453	4.291	0.526	6.975	0.213	5.659	**0.310**

Table 5. Overall performance of all subjects. Both the validation set and test set are unseen to the training data.

are insufficient to capture the complex and subtle cognitive patterns present in data from our low-cost sensors in daily environments. On the other side, handcrafted feature methods exhibit more stable patterns in both validation and test sets.

Performance of SOTA EEG Foundation Model: BL-D to BL-F are based on EEG foundation models, which are mainly trained on general cognitive tasks. They have poor performances even in the validation set. In CogAssess-v1, we change the training scheme to our alternative method, obtaining a better validation performance. This indicates the data from cognitive assessment is unique and requires specialized model adaptation considerations.

Performances of CogAssess: Our complete CogAssess improves CogAssess-v1 with a better feature extraction design on the picture layer with a much better generalization performance. This indicates that EasyCog data can reflect cognitive conditions but need careful feature extraction.

Impact of Visual Stimuli: In CogAssess-v2 and CogAssess-v3, we use resting EEG and visual task EEG as input, respectively. Compared to the complete CogAssess, which achieves the most balanced validation and test performance with the best generalizability and effective feature decoding, both resting tasks and visual tasks are essential for obtaining a comprehensive cognitive assessment.

Modality Comparison: Baseline G analyses eye tracking data with an advanced eye tracking feature extractor [52] during the visual tasks. Compared with CogAssess-v3, both modalities can achieve comparable performances during the visual tasks. This indicates that eye tracking can also reflect brain functions which imply the feasibility of a contactless solution. However, it cannot utilize the brain functions during the resting part, which could further complement the cognitive understanding as shown in our complete CogAssess.

Summary: (1) The EEG data under our stimuli can reflect cognitive conditions validly but requires dedicated feature extraction and emphasizes generalization design. (2) Both resting and visual tasks EEG are complementary to obtain a comprehensive cognitive assessment. (3) Both modalities, EEG and eye movement contain cognitive information with different characteristics. Eye movement is closer to visual tasks and requires fewer efforts for feature extraction, which indicates the contactless assessment feasibility. EEG can complement eye tracking with more diverse scenarios while eye tracking may guide EEG efficient feature extraction. (4) The major challenge is the generalizability issue, which is common in EEG-related works [71] for both subject variety and signal changes. Our dataset can reflect this challenge and function as a benchmark to indicate the algorithm capability.

6.3 Impact on EEG Electrodes

We adopt the existing forehead and around-the-ear electrode deployment in the EasyCog dataset, but there is still room to further improve the design of electrode sensors. We check the importance of each electrode to explore

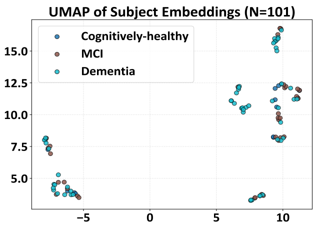


Fig. 14. Feature distribution for subjects with different severities

	Validation Set				Test Set			
	MoCA		MMSE		MoCA		MMSE	
#Channels	MAE	PCC	MAE	PCC	MAE	PCC	MAE	PCC
16	<u>5.339</u>	**0.453**	**4.291**	<u>0.526</u>	**6.975**	**0.213**	**5.659**	**0.310**
10	6.031	0.206	4.859	0.359	<u>7.296</u>	<u>0.062</u>	<u>6.026</u>	<u>0.154</u>
6	**5.254**	<u>0.435</u>	<u>4.465</u>	**0.539**	8.667	-0.030	6.760	0.100

Table 6. Performance of different electrode numbers

the impact of electrodes from a different number of channels. Specifically, for the 10 channels setting, we select R1, R3, R8, R10, FR1, FR4, L1, L3, L8, and L10 to reduce the spatial resolution while maintaining the distribution range, and for the 6 channels setting we further reduce to R3, R8, FR1, FR4, L3, and L8. The results are shown in Table 6. We observe that although the reduced channels setting achieves comparable results on the validation set, the maximal generalizability still occurs in the all channels setting.

Key Insight: This result emphasizes the effectiveness of our sensor in that the increasing channel number brings obvious improvement in generalizability, but it is still promising to further optimize the number and positions of the electrodes.

6.4 Feature Analysis

To further explore the EasyCog dataset for deeper insight, we visualize the feature distribution for subjects across different severity levels in Figure 14, using UMAP [50] to project regression features from the final layer of the regression model on the validation set.

We calculate the inter-class distance to demonstrate our observation. The distance between two classes is defined as the distance between the two centroids of the classes in feature space. The inter-class distance for Healthy-Dementia, Healthy-MCI, and MCI-Dementia are 1.03, 0.15, and 0.95 respectively, where Healthy is short for Cognitively-healthy. The closed distance for Healthy-MCI indicates the difficulties in recognition of MCI. However, the distance for Healthy-Dementia is larger than MCI-Dementia, representing our CogAssess's ability to distinguish different severities.

Key Insight: EasyCog can facilitate impairment severity analysis from the feature understanding. This provides a potentially valuable research opportunity: MCI might not be a homogeneous class [22], and its neural signature may lie on a spectrum. The EasyCog dataset allows investigation into these nuanced boundaries, which are often hidden in categorical diagnostic labels.

6.5 Case Study: Successive Assessments

To examine successive assessments, we select some subjects and conduct both scale-based assessments and our easy cognitive assessments in successive 3 to 5 days, and the results are shown in Table 7. The prediction is given by CogAssess. We first observe that both the MoCA and MMSE scores fluctuate substantially over short intervals for all subjects, proving the significant impact of the practice effect. Recall that 26 is the cutoff for MoCA MCI screening, and we can find that subject D and E change from MCI to Cognitively-healthy persons with continuous scale tests. Here we compute the ground truth standard deviation (GT STD) and prediction standard deviation (Pred. STD), which is computed by the standard deviation (MoCA+MMSE) across different sessions for the same subject. The large ground truth variation demonstrates the significant practice effect of scale-based methods, and the prediction variation with the EasyCog setting is usually much less than the scale. This indicates that it is possible to achieve stable assessment despite repeated tests via the cognitive models. Although current predictions are not accurate enough, we envision that by including more diverse subjects and involving stronger design with deeper understanding, it is possible to achieve an accurate cognitive assessment.

Subject	Session	Ground Truth MoCA	Ground Truth MMSE	Prediction MoCA	Prediction MMSE	GT STD	Pred. STD
A	1	19	23	24.660	28.411	1.5	0.580
	2	21	24	25.775	28.456		
B	1	15	18	27.134	26.116	1.0	1.499
	2	16	19	23.830	26.422		
C	1	26	28	17.169	25.686	1.700	0.702
	2	29	28	17.130	24.989		
	3	30	28	17.110	24.032		
D	1	25	29	14.957	22.816	1.247	1.487
	2	28	28	12.670	21.843		
	3	28	29	16.755	20.741		
E	1	27	25	15.040	24.455	1.0	0.025
	2	28	26	15.575	23.869		
F	1	14	20	13.235	22.447	6.944	0.204
	2	19	24	13.186	22.521		
	3	22	29	13.679	22.448		
G	1	12	18	26.067	27.506	5.0	1.487
	2	19	21	21.810	28.790		

Table 7. Performance of successive assessments

	MoCA		MMSE	
	MAE	PCC	MAE	PCC
Cognitively-healthy	4.855	-0.176	4.041	0.119
MCI	3.430	0.190	2.475	0.205
Moderate Dementia	4.184	0.379	2.748	0.619
Severe Dementia	9.409	-0.098	8.482	0.292

Table 8. Performance across different severity

	MoCA		MMSE	
	MAE	PCC	MAE	PCC
Control	4.900	0.139	3.337	0.082
PD	4.294	0.588	3.951	0.474
AD	6.100	0.404	7.500	0.388
VaD	6.205	0.341	6.189	0.472

Table 9. Performance across different diseases

6.6 Demographic Analysis

We also present performance statistics across different severities and diseases from the validation models, as shown in Table 8 and Table 9, respectively.

For performances across different severities, CogAssess achieves low MAE and low PCC for Cognitively-healthy individuals and those with Mild Cognitive Impairment (MCI), indicating that the model can classify these groups well but is unable to capture fine-grained differences within each type. For subjects with severe dementia, the model obtains high MAE, which likely indicates that the model is not robust enough to decode the common cognitive information from heterogeneous dementia patients,. Another possible reason is that MoCA and MMSE scales may not fit patients with severe dementia [77], resulting in inaccurate ground truth and thus high MAE.

Regarding performances across different diseases, we first observe that the model also achieves a low PCC for the control group, which possibly means that the factors influencing cognitive impairments in the control group are inconsistent and complex. When comparing across diseases, participants with Parkinson's Disease (PD) are the most accurately classified in our dataset, while subjects with Alzheimer's Disease (AD) and Vascular Dementia (VaD) have higher MAE. This discrepancy also reflects differences in EEG signal patterns across diseases. PD is known to produce more consistent EEG alterations (e.g., frontal slowing), whereas AD and VaD often present with more variable and diffuse abnormalities [59, 67]. Another reason is that the AD and VaD participants have lower cognitive scores with lower cooperation and reactions. These findings highlight the potential of the EasyCog dataset not only for general cognitive assessment but also as a foundation for disease-specific biomarker discovery.

Fortunately, we also observe that, regardless of whether it is PD, AD, or VaD, the model can obtain high PCC among each disease, indicating that our system shows promise in capturing intra-disease severity information.

7 Discussion

Our work demonstrates the feasibility of decoupling cognitive assessment from active, instruction-based tasks. By showing that passive viewing of structured visual stimuli can elicit differentiable neuro-physiological and behavioral signatures across the cognitive sprecturm, EasyCog provides a **proof-of-concept** for truly low-burden screening. The validation of consumer-grade, easy-to-wear sensors in real-world settings further lowers the barrier of deploying such assessments at scale, moving beyond the confines of specialized lab environments.

The dataset opens several key research domains. The synchronized EEG and gaze data enables the discovery of multimodal biomarkers, i.e., researchers can investigate how attentional deficits correlate with changes in neural connectivity in specific dementia subtypes. Futhermore, the rich stimulus set allows for fine-grained analysis of cognitive domain deficits, linking specific task performance to the sub-scores of MoCA. Most importantly, EasyCog can serve as a benchmark for developing new models that bridge modalities, which has implications for both brain-computer interfaces and understanding the mechanisms of visual attention.

Despite the promising potential, EasyCog has limitations that define clear directions for future work. First, the dataset's current composition has an imbalanced in dementia conditions and was collected in clinical settings. Further data collection should aim to balance the disease cohorts and also extend to in-home settings to improve the model generalizability and ecological validity. Second, the stimulus protocol and sensor configuration can be further optimized with the evolution of hardwares. Systematic ablation studies are needed to identify the most important tasks and EEG channels for each specific sub-tasks, which may help minimize the protocol and sensor set. This could pave the way for a more accessible self-management assessment tool.

8 Concluding Remarks

In this work, we propose EasyCog, a novel multi-modal dataset specifically designed for low-burden and low-cost cognitive assessments. By utilizing easy-to-wear, low-cost electrodes activated by passive visual stimuli, EasyCog enables participants to engage in cognitive evaluations with minimal discomfort and effort. This innovative approach addresses the limitations of traditional assessment methods, such as the MoCA and MMSE, which often impose significant user burdens. Furthermore, we propose a benchmark that includes a novel model tailored for low-burden cognitive assessment, alongside four advanced baseline methods. Through extensive experiments, we demonstrate the effectiveness of EasyCog and our benchmark methods, showcasing the potential of integrating passive visual stimuli with low-cost sensors in practical cognitive evaluations. We envision that this dataset and our benchmark will provide valuable insights into cognitive function and impairment, ultimately promoting the development of low-burden and accessible cognitive assessment tools that benefit the elderly population.

Acknowledgments

The authors would like to thank all the anonymous reviewers for their insightful comments and all volunteers for their invaluable contributions. We also thank the Turing AI Computing Cloud (TACC) [74] and HKUST iSING Lab for computation resources on their platform. This research is supported in part by RGC under Contract CERG 16206122, 16204523, SRFS2425-6S05, AoE/E-601/22-R, and Contract R8015.

References

[1] 2025. GazeTrack Github. https://github.com/AugustoEst/gazetrack.
[2] 2025. TMSi cEEGrids. https://www.tmsi.artinis.com/eeg-accessories.
[3] 2025. Tobii Accuracy. https://thume.ca/2016/03/24/eye-tracker-reviews-pupil-labs-tobii-eyex-eye-tribe-tobii-x2-30/.
[4] 2025. Tobii Eye Tracker 4C. https://help.tobii.com/hc/en-us/sections/360001811457-Tobii-Eye-Tracker-4C.

[5] 2025. Wiki of Airpod Max. https://en.wikipedia.org/wiki/AirPods_Max.

[6] Alzheimer's Disease International (ADI). [n. d.]. Numbers of people with dementia worldwide. https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/.

[7] Md Fahim Anjum, Arturo I Espinoza, Rachel C Cole, Arun Singh, Patrick May, Ergun Y Uc, Soura Dasgupta, and Nandakumar S Narayanan. 2024. Resting-state EEG measures cognitive impairment in Parkinson's disease. *npj Parkinson's Disease* 10, 1 (2024), 6.

[8] Ingrid Arevalo-Rodriguez, Nadja Smailagic, Marta Roqué i Figuls, Agustín Ciapponi, Erick Sanchez-Perez, Antri Giannakou, Olga L Pedraza, Xavier Bonfill Cosp, and Sarah Cullum. 2015. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane database of systematic reviews* 3 (2015).

[9] Mihai Bace, Vincent Becker, Chenyang Wang, and Andreas Bulling. 2020. Combining gaze estimation and optical flow for pursuits interaction. In *ACM Symposium on Eye Tracking Research and Applications*. 1–10.

[10] Kyoungwon Baik, Jin Ho Jung, Seong Ho Jeong, Seok Jong Chung, Han Soo Yoo, Phil Hyu Lee, Young H Sohn, Seung Wan Kang, and Byoung Seok Ye. 2022. Implication of EEG theta/alpha and theta/beta ratio in Alzheimer's and Lewy body disease. *Scientific reports* 12, 1 (2022), 18706.

[11] Martin G Bleichner and Stefan Debener. 2017. Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG. *Frontiers in human neuroscience* 11 (2017), 163.

[12] Sarah Blum, Reiner Emkes, Falk Minow, Jan Anlauff, Andrea Finke, and Stefan Debener. 2020. Flex-printed forehead EEG sensors (fEEGrid) for long-term EEG acquisition. *Journal of neural engineering* 17, 3 (2020), 034003.

[13] Matous Cejnek, Oldrich Vysata, Martin Valis, and Ivo Bukovsky. 2021. Novelty detection-based approach for Alzheimer's disease and mild cognitive impairment diagnosis from EEG. *Medical & Biological Engineering & Computing* 59 (2021), 2287–2296.

[14] Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini, and Tzyy-Ping Jung. 2018. Evaluation of artifact subspace reconstruction for automatic EEG artifact removal. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 1242–1245.

[15] Xiaokang Chen, Zhiyu Wu, Xingchao Liu, Zizheng Pan, Wen Liu, Zhenda Xie, Xingkai Yu, and Chong Ruan. 2025. Janus-pro: Unified multimodal understanding and generation with data and model scaling. *arXiv preprint arXiv:2501.17811* (2025).

[16] Michael D Crutcher, Rose Calhoun-Haney, Cecelia M Manzanares, James J Lah, Allan I Levey, and Stuart M Zola. 2009. Eye tracking during a visual paired comparison task as a predictor of early dementia. *American Journal of Alzheimer's Disease & Other Dementias®* 24, 3 (2009), 258–266.

[17] Kirk R Daffner, LFM Scinto, S Weintraub, JE Guinessey, and MM Mesulam. 1992. Diminished curiosity in patients with probable Alzheimer's disease as measured by exploratory eye movements. *Neurology* 42, 2 (1992), 320–320.

[18] Rebecca Davis. 2021. The Feasibility of Using Virtual Reality and Eye Tracking in Research With Older Adults With and Without Alzheimer's Disease. *Frontiers Aging Neurosci.* 13 (2021).

[19] CJ De Gans, P Burger, ES Van den Ende, J Hermanides, PWB Nanayakkara, RJBJ Gemke, F Rutters, and DJ Stenvers. 2024. Sleep assessment using EEG-based wearables–A systematic review. *Sleep Medicine Reviews* (2024), 101951.

[20] Étienne de Villers-Sidani, Patrice Voss, Daniel Guitton, J Miguel Cisneros-Franco, Nils A Koch, and Simon Ducharme. 2023. A novel tablet-based software for the acquisition and analysis of gaze and eye movement parameters: a preliminary validation study in Parkinson's disease. *Frontiers in Neurology* 14 (2023), 1204733.

[21] Arnaud Delorme and Scott Makeig. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* 134, 1 (2004), 9–21.

[22] Antonella Gallo, Anna Maria Martone, Rosa Liperoti, Maria Camilla Ciprian, Sara Camilli Francesca Ibba, Fiammetta Maria Rognoni, Francesco Landi, and Massimo Montalto. 2024. Mild cognitive impairment and microbiota: what is known and future perspectives. *Frontiers in Medicine* 11 (2024), 1410246. doi:10.3389/fmed.2024.1410246

[23] James E Galvin and Carl H Sadowsky. 2012. Practical guidelines for the recognition and diagnosis of dementia. *The Journal of the American Board of Family Medicine* 25, 3 (2012), 367–382.

[24] Melissa C Gibbs, Jenna Huxley, Megan Rose Readman, Megan Polden, Oliver Bredemeyer, Trevor J Crawford, and Chrystalina A Antoniades. 2024. Naturalistic Eye Movement Tasks in Parkinson's Disease: A Systematic Review. *Journal of Parkinson's Disease* 14, 7 (2024), 1369–1386.

[25] Ying Gu, Evy Cleeren, Jonathan Dan, Kasper Claes, Wim Van Paesschen, Sabine Van Huffel, and Borbála Hunyadi. 2017. Comparison between scalp EEG and behind-the-ear EEG for development of a wearable seizure detection system for patients with focal epilepsy. *Sensors* 18, 1 (2017), 29.

[26] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9729–9738.

[27] David Hoffman, Ian Haislip, and Christina Cool. 2024. Estimated Time to Properly Apply Electroencephalogram (EEG) Electrodes: A Survey (P11-1.009). In *Neurology*, Vol. 102. Lippincott Williams & Wilkins Hagerstown, MD, 6212.

[28] Björn Holtze, Marc Rosenkranz, Manuela Jaeger, Stefan Debener, and Bojana Mirkovic. 2022. Ear-EEG measures of auditory attention to continuous speech. *Frontiers in Neuroscience* 16 (2022), 869426.

[29] Julie Hugo and Mary Ganguli. 2014. Dementia and cognitive impairment: epidemiology, diagnosis, and treatment. *Clinics in geriatric medicine* 30, 3 (2014), 421.

[30] Cosimo Ieracitano, Nadia Mammone, Amir Hussain, and Francesco C Morabito. 2020. A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks* 123 (2020), 176–190.

[31] Jittrapol Intarasirisawat, Chee Siang Ang, Christos Efstratiou, Luke William Feidhlim Dickens, and Rupert Page. 2019. Exploring the Touch and Motion Features in Game-Based Cognitive Assessments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3 (2019), 87:1–87:25.

[32] Belmir Jesus Jr, Raymundo Cassani, William J McGeown, Marco Cecchi, KC Fadem, and Tiago H Falk. 2021. Multimodal prediction of Alzheimer's disease severity level based on resting-state EEG and structural MRI. *Frontiers in human neuroscience* 15 (2021), 700627.

[33] Parunyou Julayanont, Natalie Phillips, Howard Chertkow, and Ziad S Nasreddine. 2013. Montreal Cognitive Assessment (MoCA): concept and clinical review. *Cognitive screening instruments: A practical approach* (2013), 111–151.

[34] Chris Kalafatis, Mohammad Hadi Modarres, Panos Apostolou, Haniye Marefat, Mahdiyeh Khanbagi, Hamed Karimi, Zahra Vahabi, Dag Aarsland, and Seyed-Mahdi Khaligh-Razavi. 2021. Validity and cultural generalisability of a 5-minute AI-based, computerised cognitive assessment in mild cognitive impairment and Alzheimer's dementia. *Frontiers in Psychiatry* 12 (2021), 706695.

[35] Netiwit Kaongoen, Jaehoon Choi, Jin Woo Choi, Haram Kwon, Chaeeun Hwang, Guebin Hwang, Byung Hyung Kim, and Sungho Jo. 2023. The future of wearable EEG: A review of ear-EEG technology and its applications. *Journal of neural engineering* 20, 5 (2023), 051002.

[36] Michael T. Knierim, Christian Zimny, Gabriel Ivucic, and Tobias Röddiger. 2025. Advancing Wearable BCI: Headphone EEG for Cognitive Load Detection in Lab and Field. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 1 (2025), 9:1–9:26.

[37] Nils A Koch, Patrice Voss, J Miguel Cisneros-Franco, Alexandre Drouin-Picaro, Fama Tounkara, Simon Ducharme, Daniel Guitton, and Étienne de Villers-Sidani. 2024. Eye movement function captured via an electronic tablet informs on cognition and disease severity in Parkinson's disease. *Scientific Reports* 14, 1 (2024), 9082.

[38] Emmanouil Ktistakis, Vasileios Skaramagkas, Dimitris Manousos, Nikolaos S Tachos, Evanthia Tripoliti, Dimitrios I Fotiadis, and Manolis Tsiknakis. 2022. COLET: A dataset for COgnitive workLoad estimation based on eye-tracking. *Computer Methods and Programs in Biomedicine* 224 (2022), 106989.

[39] Marta Kutas and Kara D Federmeier. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology* 62, 1 (2011), 621–647.

[40] Allan M Landes, Susan D Sperry, Milton E Strauss, and David S Geldmacher. 2001. Apathy in Alzheimer's disease. *Journal of the American Geriatrics Society* 49, 12 (2001), 1700–1707.

[41] Kenneth M Langa and Deborah A Levine. 2014. The diagnosis and management of mild cognitive impairment: a clinical review. *Jama* 312, 23 (2014), 2551–2561.

[42] Moritz Langner, Peyman Toreini, and Alexander Maedche. 2024. Cognitive state detection with eye tracking in the field: an experience sampling study and its lessons learned. *i-com* 23, 1 (2024), 109–129.

[43] Lauren KS Lei, Bess YH Lam, Daniel WL Lai, Xue Bai, Jessica Li, Zhi Zou, and Chetwyn CH Chan. 2022. Stability of montreal cognitive assessment in individuals with mild cognitive impairment: potential influence of practice effect. *Journal of Alzheimer's Disease* 87, 3 (2022), 1401–1412.

[44] Quentin Lenoble, Xavier Corveleyn, Sébastien Szaffarczyk, Florence Pasquier, and Muriel Boucart. 2018. Attentional capture by incongruent object/background scenes in patients with Alzheimer disease. *Cortex* 107 (2018), 4–12.

[45] Yi Jin Leow, JDJ Wang, Ashwati Vipin, Gurveen Kaur Sandhu, See Ann Soo, Dilip Kumar, Adnan Azam Mohammed, FZB Zailan, Faith Phemie Hui En Lee, Smriti Ghildiyal, et al. 2024. Biomarkers and cognition study, Singapore (BIOCIS): protocol, study design, and preliminary findings. *The Journal of Prevention of Alzheimer's Disease* 11, 4 (2024), 1093–1105.

[46] Jingyi Lin, Tianyan Xu, Xuan Yang, Qijie Yang, Yuan Zhu, Meidan Wan, Xuewen Xiao, Sizhe Zhang, Ziyu Ouyang, Xiangmin Fan, et al. 2024. A detection model of cognitive impairment via the integrated gait and eye movement analysis from a large Chinese community cohort. *Alzheimer's & Dementia* 20, 2 (2024), 1089–1101.

[47] Joseph R Manns, Craig EL Stark, and Larry R Squire. 2000. The visual paired-comparison task as a measure of declarative memory. *Proceedings of the National Academy of Sciences* 97, 22 (2000), 12375–12379.

[48] Silvia Marino, Edoardo Sessa, Giuseppe Di Lorenzo, Pietro Lanzafame, Gabriella Scullica, Alessia Bramanti, Francesco La Rosa, Giancarlo Iannizzotto, Placido Bramanti, and Paolo Di Bella. 2007. Quantitative analysis of pursuit ocular movements in Parkinson's disease by using a video-based eye tracking system. *European Neurology* 58, 4 (2007), 193–197.

[49] Roy C Martin, Shannon M Annis, Laurie Z Darling, Virginia Wadley, Lindy Harrell, and Daniel C Marson. 2003. Loss of calculation abilities in patients with mild and moderate Alzheimer disease. *Archives of neurology* 60, 11 (2003), 1585–1589.

[50] Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

[51] Arnd Meiser, Francois Tadel, Stefan Debener, and Martin G Bleichner. 2020. The sensitivity of ear-EEG: evaluating the source-sensor relationship using forward modeling. *Brain topography* 33 (2020), 665–676.

[52] Kyriaki Mengoudi, Daniele Ravì, Keir X. X. Yong, Silvia Primativo, Ivanna M. Pavisic, Emilie Brotherhood, Kirsty Lu, Jonathan M. Schott, Sebastian J. Crutch, and Daniel C. Alexander. 2020. Augmenting Dementia Cognitive Assessment With Instruction-Less Eye-Tracking Tests. *IEEE J. Biomed. Health Informatics* 24, 11 (2020), 3066–3075.

[53] Andreas Miltiadous, Katerina D Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G Tsalikakis, Pantelis Angelidis, Markos G Tsipouras, Euripidis Glavas, Nikolaos Giannakeas, et al. 2023. A dataset of scalp EEG recordings of Alzheimer's disease, frontotemporal dementia and healthy subjects from routine EEG. *Data* 8, 6 (2023), 95.

[54] Robert J Molitor, Philip C Ko, and Brandon A Ally. 2015. Eye movements in Alzheimer's disease. *Journal of Alzheimer's disease* 44, 1 (2015), 1–12.

[55] Anu Muraja-Murro, Esa Mervaala, Susanna Westeren-Punnonen, Pasi Lepola, Juha Töyräs, Sami Myllymaa, Petro Julkunen, Anne-Mari Kantanen, Reetta Kälviäinen, and Katja Myllymaa. 2015. Forehead EEG electrode set versus full-head scalp EEG in 100 patients with altered mental state. *Epilepsy & behavior* 49 (2015), 245–249.

[56] Diederick C Niehorster, Richard Andersson, and Marcus Nyström. 2020. Titta: A toolbox for creating PsychToolbox and Psychopy experiments with Tobii eye trackers. *Behavior research methods* 52, 5 (2020), 1970–1979.

[57] OpenBCI. [n. d.]. OpenBCI Cyton+Daisy Board. https://shop.openbci.com/products/cyton-daisy-biosensing-boards-16-channel.

[58] Pongsatorn Paholpak, Liang Li-Jung, Drew R Carr, Elvira Jimenez, Robin J Barrows, Valeiry Sabodash, and Mario F Mendez. 2016. Prolonged visual facial grasp in frontotemporal dementia. *Journal of Alzheimer's Disease* 53, 1 (2016), 327–335.

[59] Angelina J Polsinelli and Liana G Apostolova. 2022. Atypical Alzheimer disease variants. *CONTINUUM: Lifelong Learning in Neurology* 28, 3 (2022), 676–701.

[60] Megan Rose Readman, Megan Polden, Melissa Chloe Gibbs, Lettie Wareing, and Trevor J Crawford. 2021. The potential of naturalistic eye movement tasks in the diagnosis of Alzheimer's disease: a review. *Brain sciences* 11, 11 (2021), 1503.

[61] Daniel Rojas-Líbano, Gabriel Wainstein, Ximena Carrasco, Francisco Aboitiz, Nicolás Crossley, and Tomás Ossandón. 2019. A pupil size, eye-tracking and neuropsychological dataset from ADHD children during a cognitive task. *Scientific data* 6, 1 (2019), 25.

[62] Mark Sanderson-Cimino, Jeremy A Elman, Xin M Tu, Alden L Gross, Matthew S Panizzon, Daniel E Gustavson, Mark W Bondi, Emily C Edmonds, Graham ML Eglit, Joel S Eppig, et al. 2022. Cognitive practice effects delay diagnosis of MCI: Implications for clinical trials. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 8, 1 (2022), e12228.

[63] Glória SA Siqueira, Paula de MS Hagemann, Daniela de S Coelho, Flávia Heloísa Dos Santos, and Paulo HF Bertolucci. 2019. Can MoCA and MMSE be interchangeable cognitive screening tools? A systematic review. *The Gerontologist* 59, 6 (2019), e743–e763.

[64] Vasileios Skaramagkas, Emmanouil Ktistakis, Dimitris Manousos, Eleni Kazantzaki, Nikolaos S Tachos, Evanthia Tripoliti, Dimitrios I Fotiadis, and Manolis Tsiknakis. 2023. eSEE-d: Emotional state estimation based on eye-tracking dataset. *Brain Sciences* 13, 4 (2023), 589.

[65] Jingnan Sun, Anruo Shen, Yike Sun, Xiaogang Chen, Yunxia Li, Xiaorong Gao, and Bai Lu. 2024. Adaptive spatiotemporal encoding network for cognitive assessment using resting state EEG. *npj Digital Medicine* 7, 1 (2024), 375.

[66] Jingnan Sun, Yike Sun, Anruo Shen, Yunxia Li, Xiaorong Gao, and Bai Lu. 2024. An ensemble learning model for continuous cognition assessment based on resting-state EEG. *npj Aging* 10, 1 (2024), 1.

[67] Dietmar Rudolf Thal, Lea Tenenholz Grinberg, and Johannes Attems. 2012. Vascular dementia: different forms of vessel disorders contribute to the development of dementia in the elderly brain. *Experimental gerontology* 47, 11 (2012), 816–824.

[68] Pedro A Valdes-Sosa, Lidice Galan-Garcia, Jorge Bosch-Bayard, Maria L Bringas-Vega, Eduardo Aubert-Vazquez, Iris Rodriguez-Gil, Samir Das, Cecile Madjar, Trinidad Virues-Alba, Zia Mohades, et al. 2021. The Cuban Human Brain Mapping Project, a young and middle age population-based EEG, MRI, and cognition dataset. *Scientific Data* 8, 1 (2021), 45.

[69] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[70] Barbara Vicini Chilovi, Marta Conti, Marina Zanetti, Ilenia Mazzù, Luca Rozzini, and Alessandro Padovani. 2009. Differential impact of apathy and depression in the development of dementia in mild cognitive impairment patients. *Dementia and geriatric cognitive disorders* 27, 4 (2009), 390–398.

[71] Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. 2025. CBraMod: A Criss-Cross Brain Foundation Model for EEG Decoding. In *The Thirteenth International Conference on Learning Representations*.

[72] Dominik Welke and Edward A Vessel. 2022. Naturalistic viewing conditions can increase task engagement and aesthetic preference but have only minimal impact on EEG quality. *NeuroImage* 256 (2022), 119218.

[73] Juan Xu, Ming Jiang, Shuo Wang, Mohan S Kankanhalli, and Qi Zhao. 2014. Predicting human gaze beyond pixels. *Journal of vision* 14, 1 (2014).

[74] Kaiqiang Xu, Decang Sun, Hao Wang, Zhenghang Ren, Xinchen Wan, Xudong Liao, Zilong Wang, Junxue Zhang, and Kai Chen. 2025. Design and Operation of Shared Machine Learning Clusters on Campus. In *ASPLOS (1)*. ACM, 295–310.

[75] Ying Xu, Chi Zhang, Baobao Pan, Qing Yuan, and Xu Zhang. 2024. A portable and efficient dementia screening tool using eye tracking machine learning and virtual reality. *NPJ Digital Medicine* 7, 1 (2024), 219.

[76] Yasunori Yamada, Kaoru Shinkawa, Masatomo Kobayashi, Miyuki Nemoto, Miho Ota, Kiyotaka Nemoto, and Tetsuaki Arai. 2024. Distinct eye movement patterns to complex scenes in Alzheimer's disease and Lewy body disease. *Frontiers in Neuroscience* 18 (2024).

[77] Cindy Yeo, Wee Shiong Lim, Mark Chan, Xin Qin Ho, Philomena Vasantha Anthony, Huey Charn Han, and Mei Sian Chong. 2016. Severe impairment rating scale: A useful and brief cognitive assessment tool for advanced dementia for nursing home residents. *American Journal of Alzheimer's Disease & Other Dementias®* 31, 1 (2016), 87–96.

[78] Baiqiao Zhang, Xiangxiang Li, Yunfan Zhou, Juan Liu, Chao Zhou, Weiying Liu, and Yulong Bian. 2024. Are We in the Zone? Exploring the Features and Method of Detecting Simultaneous Flow Experiences Based on EEG Signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4 (2024), 152:1–152:42.

[79] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. 2017. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412* (2017).

A Appendix: Information of Passive Visual Stimuli

This section describes the details of stimuli design and selection. Based on the stimuli rationale in Section 3.2, we need to collect pictures to serve as visual stimuli for the video. The duration for each picture is set to between 3 and 5 seconds, as determined by previous studies [52, 76]. To minimize participant fatigue and maintain attention, we limited the overall test duration to 10 minutes, including a 3-minute closed-eye resting period—a duration considered appropriate by experienced clinicians. Therefore, each task is allocated 10 different pictures, except for the moving target task, which contains only 6 videos of 5 seconds each to avoid fatigue from prolonged pursuit.

Wherever possible, we selected stimuli from established sources used in cognitive research. All pictures in Tasks 1 and 2, as well as the pictures featuring people in Tasks 3 and 4, are randomly selected from the image dataset proposed by Xu et al.[73], which has also been utilized for visual analysis in dementia research[76]. For the non-social pictures in Tasks 3 and 4, which need to resemble the corresponding social pictures, we search for images on Google using keywords that illustrate the background of the social pictures.

In Task 5, which requires incongruent elements, we employed a large generative model DALL-E 3 to create two types of pictures: objects with abnormal properties (e.g., a horse with an abnormal number of legs) and objects placed in incongruent backgrounds (e.g., a lion sitting in a classroom) [17, 44]. For the video featuring an object moving at a constant velocity within naturalistic scenes, we filmed moving ships and aircraft from different directions and adjusted their velocities accordingly. In Task 7, we follow the sentence structure of the reading task from the Aphasia Battery of Chinese (ABC) scale, which was generated by a large language model. In Task 8, a presented picture will occur alongside a similar picture, where the similar picture is also sourced from Google using keywords extracted by a visual language model. For Task 9, the pictures directly depict simple one-digit arithmetic.

After collecting all the pictures, we concatenate them into a video, with a 2-second interval between each task. The total video duration is 430 seconds, and examples for each task are shown in Figure 4. To ensure clinical validity, three experienced clinicians reviewed the full video and confirmed that the stimuli appropriately covered all core cognitive domains assessed by the MoCA and MMSE scales.

B Appendix: Details of CogAssess

B.1 CogAssess Implementation

This section presents the implementation details of the proposed benchmark method, CogAssess. It is a three-layer hierarchical deep learning model that extracts cognitive features from the EEG signals.

Picture Layer: Visual Stimuli Alignment with Sample Contrastive Learning. The picture-layer part extracts EEG features corresponding to the single picture that serves as the basic unit to form the entire video stimuli. Since each picture lasts from 3-5 seconds, we use the synchronized 3-second EEG signals as input with specific regularization to guide the model towards cognitive features from the short period of inputs.

(1) **Visual Stimuli Alignment.** We first need to align the varying EEG signals with the cognitive visual stimuli for shared patterns at the picture level. To bridge the modality gap between visual pictures and brain electric signals, we adopt an EEG-Visual Alignment module. We firstly acquire a robust and comprehensive visual picture embedding from a leading large visual language model (VLM) Janus-Pro-7B [15]. Then we align the EEG features from the EEG foundation model with the visual embeddings from the VLM model with cross-attention. This cross-attention is similar to the self-attention module [69] but it learns the query module from visual pictures to extract more correlated EEG features. Here we leverage the visual task classification as the proxy optimization target to extract features discriminative to different visual tasks inside the video [52].

(2) **Sample-wise Contrastive Learning.** Note that our target is to discriminate picture-layer EEG features across different tasks and different subjects. If the proxy visual task classification is well optimized, the features can be overfitted with the visual stimuli contents with no cognitive features preserved. To preserve the deep cognitive

differences, we apply contrastive learning [26] to make EEG features across different pictures discriminative to other samples. Specifically, we adopt the pink noise data augmentation [65] to generate two views from the same picture-layer EEG input as the positive pairs and regard other samples as the negative ones. The contrast learning aims to pull the positive pairs closer and to push the negative pairs farther. Therefore, it serves as a form of regularization that helps maintain distinctions between different samples while also making the model robust to augmented noise.

Task Layer: Intra-task Aggregation and Inter-task Augmentation. With the picture-layer features, we need to extract the task-representative features from multiple pictures.

(1) Intra-task Aggregation. Since each task contains around 10 pictures, different subjects may behave differently. We adopt a dynamic aggregation module based on attention [69] on the extracted features from the pictures within the same task. Considering the nine visual tasks with different focuses on cognitive triggers, we classify the intra-task aggregation into two conditions: (1) Intra-isomorphic task: the pictures inside the same task are designed for the same functions, *i.e.* Tasks 1, 2, 5, 6, 8, 9. (2) Intra-contrast task: the pictures inside the same task contrast with adjacent pictures to amplify the response difference, *i.e.* Tasks 3, 4, 7. For the intra-isomorphic tasks, we regard their features as similar across the intra-task pictures and directly feed them to the attention module for feature aggregation. For the intra-contrast task, we form the contrastive pairs and extract the corresponding feature differences, which are further processed as the task features with another attention module.

(2) Inter-task Augmentation. With the above intra-task aggregation, we can acquire task-layer features. We observe the tasks in the clinical scales share the order invariance that the overall cognitive evaluation can be conducted without a fixed order. Based on this insight, we design an order-randomizing data augmentation method to force our task features invariant to different sequence orders. It creates multiple task feature sequences with different orders and forces the cognitive regression to be consistent across different orders. Let the augmented ratio be β and the number of tasks in the visual stimulus is $N_{task} = 10$. During each augmentation, we first select $\beta * N_{task}$ tasks and then shuffle these selected tasks to form a new sample. With the ratio β, we can control the similarity between the original and the new distribution. In our experiments, we augment the data 100 times with $\beta = 0.6$ by empirical experience, resulting in a dataset of reasonable size for training.

Subject Layer: Cognitive Regression. Finally, we aggregate all task embeddings from the same subject by concatenation the task features with the shape of $[10, n_{dim}]$, where n_{dim} is the target feature dimension and 10 is the number of tasks with random orders. We then apply mix-up augmentation [79] to further extend the dataset. The regression model is a 2-layer transformer with two linear layers to regress both the overall scores and each detailed question score in the clinical scales as a consistency regularization.

Implementation. We implement the benchmark model with PyTorch 2.0.1 on an NVIDIA RTX 4090 GPU. The EEG encoder is a pre-trained CBraMod model [71] with an extra linear layer for feature transformer and then conducts mask reconstruction on our data collection with a mask ratio of 75% for 100 epochs. The EEG-Visual alignment part firstly extracts the last token from the high-dimension visual token embeddings from Janus-Pro [15] and maps the token from 4096-dim to 200-dim. Then a cross-attention module aligns both modalities, whose outputs are transformed with a linear layer for MoCo-based contrastive learning [26]. It is optimized with the cross entropy loss for visual task classification and contrastive loss with the same weight. The task layer has two attention modules with 200-dim for intra-isomorphic tasks and intra-contrast tasks separately. The subject layer has a 2-layer transformer model for cognitive score regression. The final output layers are two linear layers, where the first one outputs 2 scores for MoCA/MMSE estimations, and the latter one outputs the 13 subscores of each clinical scale for detailed regularization. It is optimized with a multi-task loss, which contains the L1 loss for both MoCA and MMSE scores with a weight of 1 and the L1 loss for the subscores with a weight of 0.1. The benchmark model has an optimizer of Adam with a cosine-scheduled learning rate that increases from 0 to $1e^{-3}$

Design Modules				Validation Set				Test Set			
				MoCA		MMSE		MoCA		MMSE	
VA	MoCo	TA	Rand. Ord.	MAE	PCC	MAE	PCC	MAE	PCC	MAE	PCC
✓	✓	✓	✓	<u>5.339</u>	<u>0.453</u>	<u>4.291</u>	<u>0.526</u>	**6.975**	**0.213**	**5.659**	**0.310**
✓	✓	✓	✗	5.495	0.402	4.554	0.430	7.483	<u>0.115</u>	6.101	0.148
✗	✓	✓	✓	5.967	0.402	5.070	0.297	<u>6.972</u>	0.011	<u>5.839</u>	<u>0.200</u>
✗	✗	✓	✓	**5.156**	**0.502**	**4.663**	**0.518**	7.019	0.110	6.125	0.097
✗	✗	✗	✓	5.412	0.474	4.629	0.519	8.013	0.062	6.681	0.066
✗	✗	✗	✗	5.133	0.456	4.497	0.480	9.186	-0.024	8.035	-0.030

Table 10. Ablation study of CogAssess, 'VA': 'visual alignment', 'TA': 'task cross attention'.

in the first 20 epochs and then decreases to $5e^{-5}$ from 20 to 100 epochs. We adopt an early stop strategy that stops the training if the error on the validation set no more decreases for 5 epochs.

B.2 Impact on the CogAssess Modules.

We then conduct an ablation study on our CogAssess model to demonstrate the effectiveness of each design, and the results are shown in Table 10. By adding the task random ordering augmentation, the performance on the test set slightly increases, presenting better data generalizability. When we further add the task cross attention, the model achieves the highest performance on the validation set, while losing the ability to generalize to other data. Although the involvement of MoCA brings significant performance degradation on the validation set, it dramatically improves the generalized performance on the test set. Finally, when all components are activated, the model achieves the best balance of the ability of cognitive information understanding and generalizability, demonstrating the effectiveness of our designs on CogAssess.

Key Insight: This result shows the promising results from the EasyCog dataset and CogAssess that it can learn a high correlation on the validation set which indicates better cognitive features. The EEG signals are varying and we need to align them with the stimuli settings to make the model focus more on the visual-triggered features for better generalizability.