

Supplementary Material

Task-evoked multimodal wearable sensing for community-based
cognitive impairment assessment

Supplementary Contents

I. Extended Results

- **Supplementary Fig. 1:** Wearable multimodal data acquisition workflow 5
- **Supplementary Fig. 2:** Multimodal cognitive impairment assessment model 6
- **Supplementary Fig. 3:** Correlation analysis of multimodal features with cognitive and emotional assessments 7
- **Supplementary Fig. 4:** Comparing EEG and speech features across stages of cognitive decline 8
- **Supplementary Fig. 5:** Model performance and ablation analyses for cognitive impairment classification 9
- **Supplementary Fig. 6:** Participant recruitment, exclusion, and cohort assignment flowchart 10
- **Supplementary Table 1:** Demographic characteristics of the internal development and external validation cohorts 11
- **Supplementary Fig. 7:** Distribution of demographic characteristics and clinical assessment scores in the internal development cohort 12
- **Supplementary Fig. 8:** Distribution of demographic characteristics and clinical assessment scores in the external validation cohort 13
- **Supplementary Fig. 9:** Distribution of cognitive assessment scores in the internal development cohort 14
- **Supplementary Fig. 10:** Distribution of cognitive assessment scores in the external validation cohort 15
- **Supplementary Fig. 11:** Distribution of cognitive assessment scores in the overall cohort 16
- **Supplementary Fig. 12:** Distribution of cognitive assessment scores and age across cohorts 17
- **Supplementary Table 2:** Comparison of assessment performance across different modalities 18

•	Supplementary Fig. 13: ROC and PR curves of assessment performance across different modalities	19
•	Supplementary Table 3: Comparison of assessment performance across cognitive tasks (T0–T10)	20
•	Supplementary Fig. 14: ROC and PR curves of assessment performance across cognitive tasks	21
•	Supplementary Table 4: Comparison of assessment performance across model architectures under all modalities	22
•	Supplementary Fig. 15: ROC and PR curves of assessment performance across model architectures under all modalities	22
•	Supplementary Table 5: Comparison of assessment performance across model architectures using EEG signals	23
•	Supplementary Fig. 16: ROC and PR curves of assessment performance across model architectures using EEG signals	23
•	Supplementary Table 6: Comparison of assessment performance across model architectures using ECG signals	24
•	Supplementary Fig. 17: ROC and PR curves of assessment performance across model architectures using ECG signals	24
•	Supplementary Table 7: Comparison of assessment performance across model architectures using video signals	25
•	Supplementary Fig. 18: ROC and PR curves of assessment performance across model architectures using video signals	25
•	Supplementary Table 8: External validation performance for distinguishing NC from MCI/DE under different prevalence settings	26
•	Supplementary Fig. 19: External validation performance for distinguishing NC from MCI/DE under different prevalence settings	26
•	Supplementary Table 9: External validation performance for distinguishing NC from MCI under different prevalence settings	27
•	Supplementary Fig. 20: External validation performance for distinguishing NC from MCI under different prevalence settings	27
•	Supplementary Fig. 21: Distribution of multimodal features across MoCA sub-domains	28

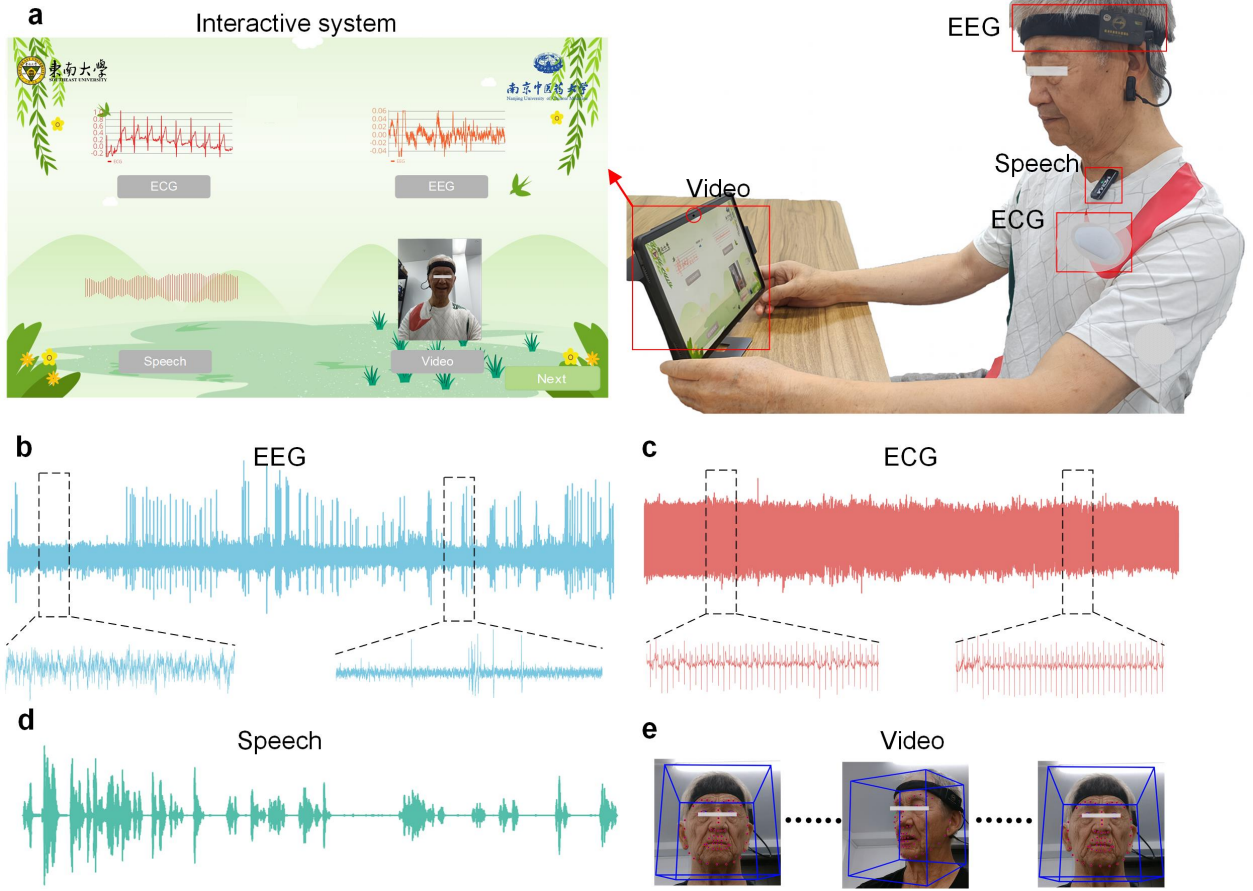
• Supplementary Fig. 22: Dynamic trajectories of multimodal signals across cognitive tasks	29
• Supplementary Fig. 23: Correlation between demographic characteristics and model probability scores	30
• Supplementary Fig. 24: Association between model probability scores and conventional cognitive assessment scores	30
• Supplementary Fig. 25: Model discrimination performance in MoCA-defined borderline cognitive ranges	31
• Supplementary Fig. 26: Model probability scores across dementia severity groups	31

II. Extended Methods

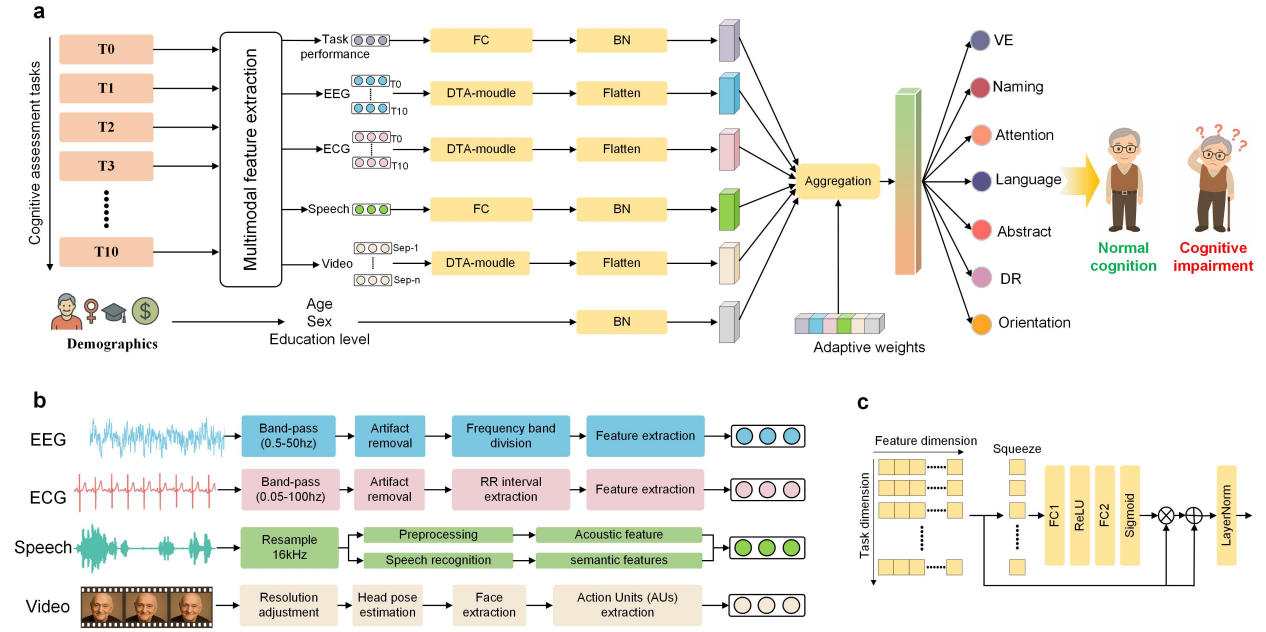
• Supplementary Fig. 27: Schematic of recruitment timeline, clinical assessment, devices, and experimental environment	32
• Supplementary Fig. 28: Education-adjusted cut-off criteria for clinical cognitive impairment assessment	34
• Supplementary Fig. 29: Task description of multidimensional cognitive assessment	37
• Supplementary Table 10: Overview of cognitive tasks (T0–T10)	38
• Supplementary Table 11: Mapping of tasks to MoCA cognitive domains	38
• Supplementary Table 12: Duration of each cognitive task in the assessment protocol	39
• Supplementary Table 13: EEG preprocessing and artifact mitigation procedure	40
• Supplementary Table 14: EEG feature categories, definitions, and their relevance to cognitive impairment	41
• Supplementary Table 15: Expanded ECG and HRV feature categories, definitions, and their relevance to cognitive impairment	42
• Supplementary Table 16: Speech feature categories, definitions, and their relevance to cognitive impairment	43
• Supplementary Table 17: Expanded Video feature categories, definitions, and their relevance to cognitive impairment	44
• Supplementary Table 18: Meta-features describing the temporal evolution of physiological and behavioral signals	45

- **Supplementary Table 19:** Robustness analysis under modality-missing settings 46
in the external validation cohort
- **Supplementary Fig. 30:** Baseline characteristics of included and excluded partic- 47
ipants

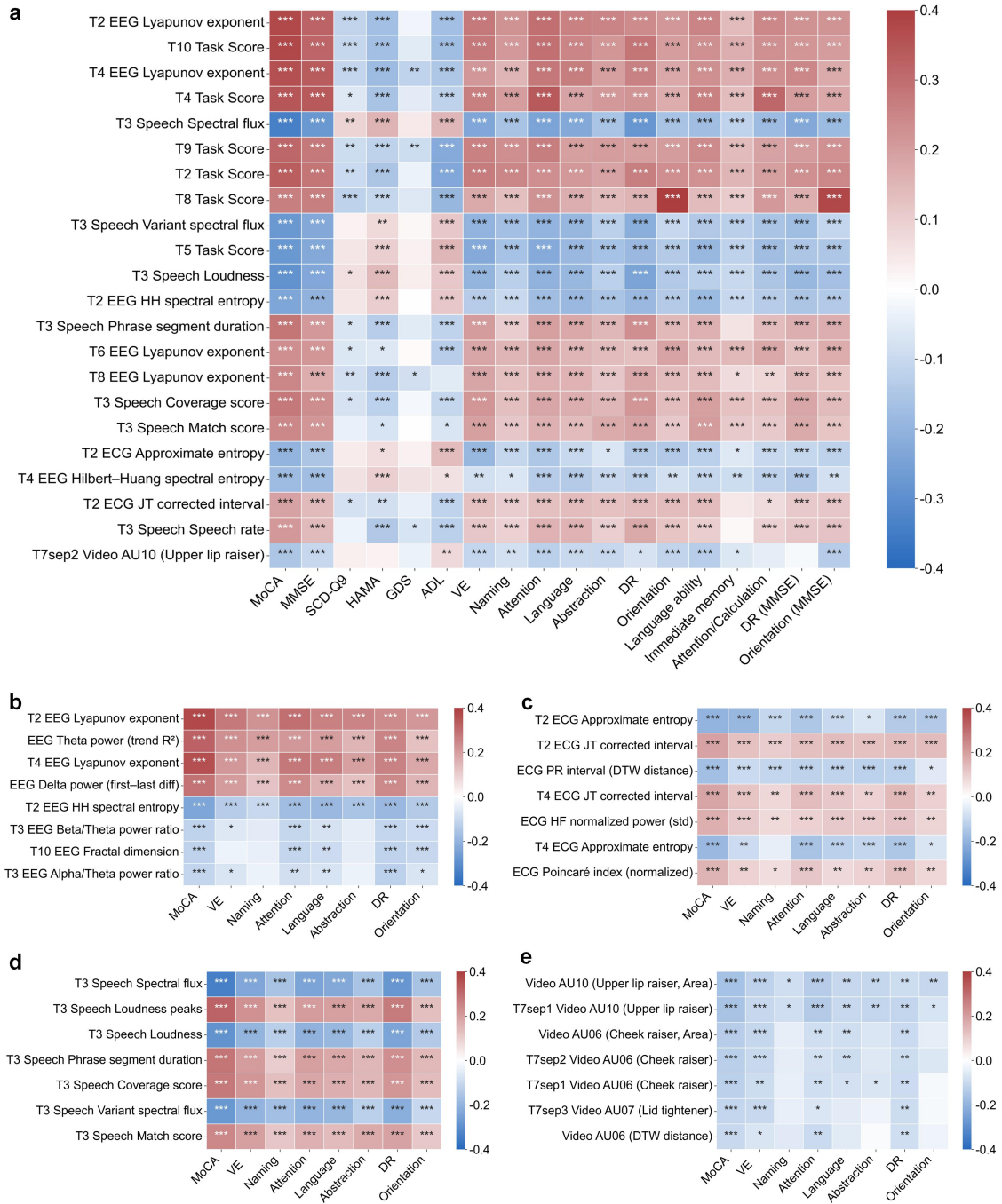
I. Extended Results



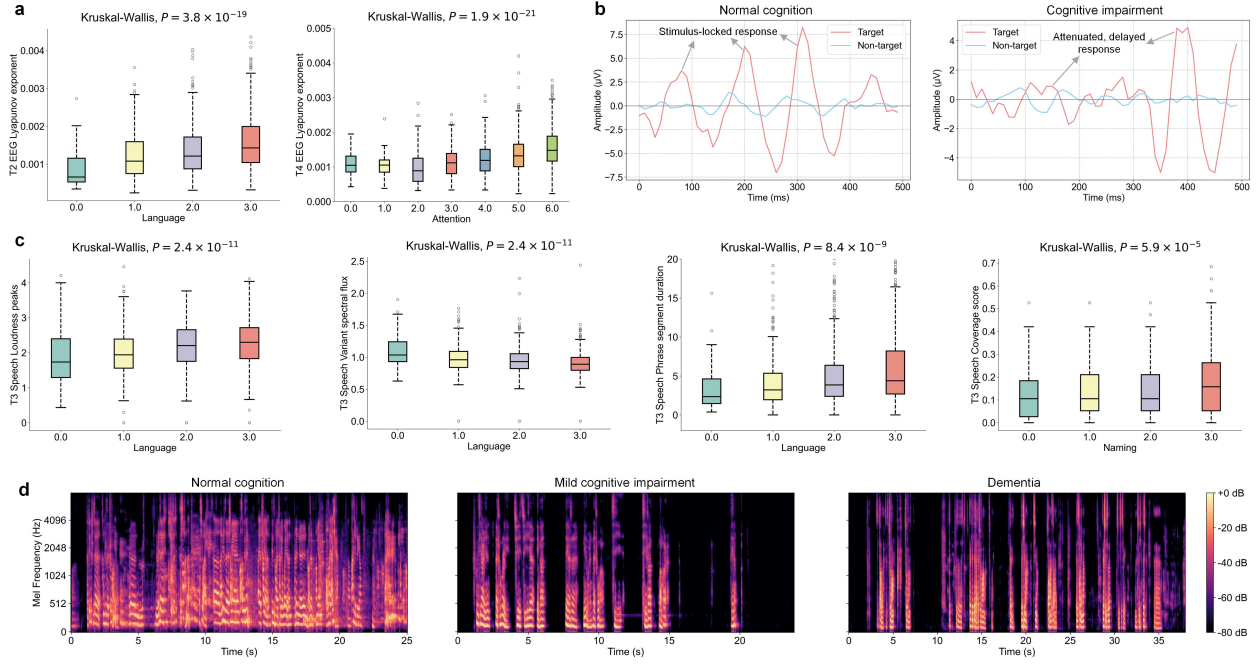
Supplementary Fig. 1: Wearable multimodal data acquisition workflow. **a**, System interface and real-world data acquisition setup. A participant is equipped with multiple non-invasive sensors, including an EEG headband, ECG electrodes, and a speech microphone, while interacting with a tablet-based cognitive task interface. The interactive screen displays real-time EEG and ECG waveforms, integrates live video, and records synchronized audio to enable comprehensive multimodal data collection and visual feedback. **b**, Raw EEG waveforms recorded during the task, with a zoomed-in region highlighting event-related potential changes corresponding to specific task stimuli. **c**, ECG waveform showing periodic QRS complexes, with a magnified section emphasizing cardiac rhythm characteristics, demonstrating stable acquisition of high-quality ECG signals. **d**, Speech waveform collected during the cognitive task, used for subsequent acoustic and semantic content analysis. **e**, Synchronized video stream, with the participant's face highlighted for facial expression analysis. Action units extracted from this region provide visual-behavioral features to support cognitive state evaluation.



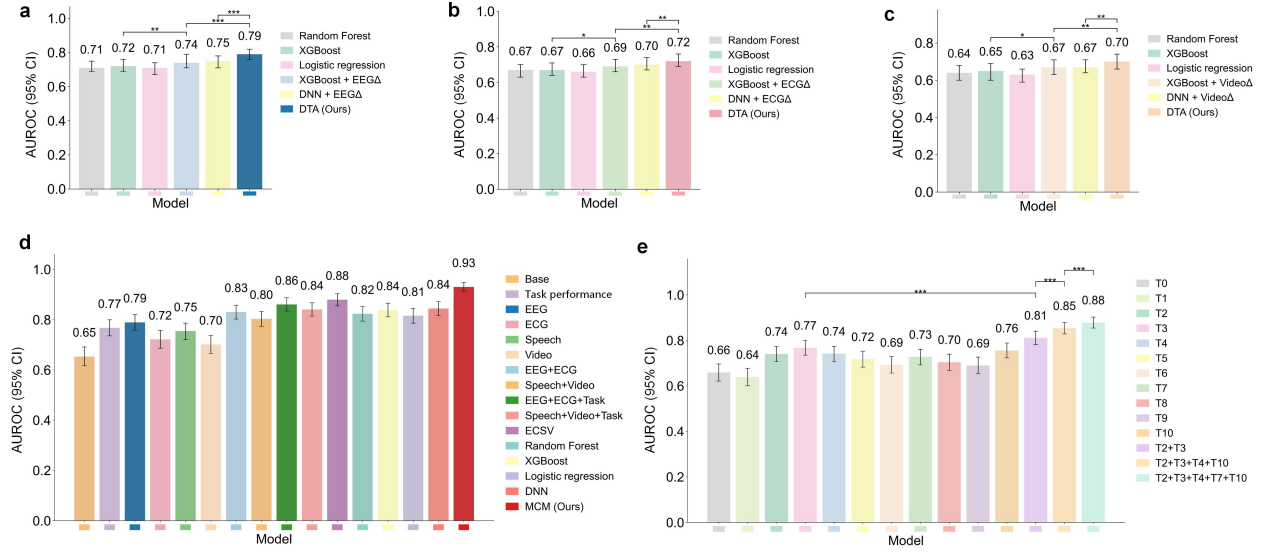
Supplementary Fig. 2: Multimodal cognitive impairment assessment model. **a**, Multimodal task-related features are processed through four stages: feature encoding, modality fusion, cognitive function supervision, and impairment classification. We employ weak supervision targeting seven specific cognitive function dimensions at intermediate nodes to guide domain-specific feature learning. The model outputs predicted cognitive function scores and impairment risk concurrently. **b**, Preprocessing and feature extraction pipeline for multimodal signals. **c**, Architecture of the dynamic task adapter (DTA) module.



Supplementary Fig. 3: Correlation analysis of multimodal features with cognitive and emotional assessments. **a**, Correlation heatmap between multimodal features and cognitive assessment scales, as well as emotional scales. **b**, Correlation analysis of EEG-derived features with the same scales. **c**, Correlation analysis of ECG-derived features. **d**, Correlation analysis of speech-derived features. **e**, Correlation analysis of video-derived features. Each cell represents the strength and direction of the correlation, with darker colors indicating stronger associations. Significance levels are indicated as * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$, with all p values adjusted for multiple comparisons using the false discovery rate (FDR).

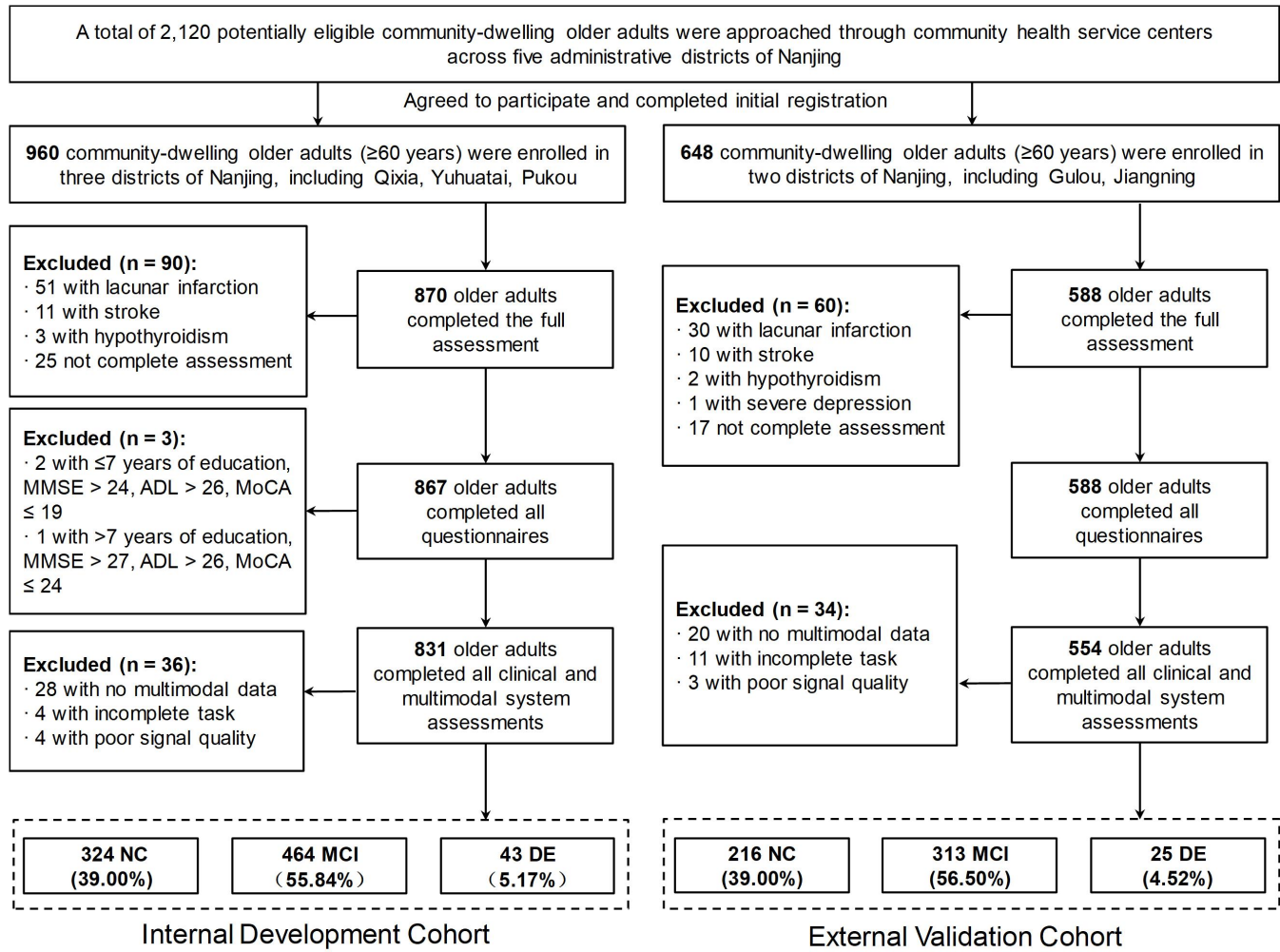


Supplementary Fig. 4: Comparing EEG and speech features across stages of cognitive decline. EEG and speech features and their differences across individuals with normal cognition (NC), mild cognitive impairment (MCI), and dementia (DE). **a**, Relationships between EEG nonlinear complexity and cognitive functions across tasks. In the T2 (object categorization) task, frontal EEG complexity showed significant group differences and positive associations with language and abstraction scores; in the T4 (calculation) task, complexity similarly distinguished groups in relation to attention performance. **b**, Event-related potential waveforms from the T5 (attention counting) task, illustrating target and non-target evoked responses and highlighting the reduced, delayed, and unstable components in cognitively impaired individual compared with the normal cognition counterpart. **c**, Speech-derived features, including spectral flux, loudness variation, pause-related measures, and semantic indicators-within language and naming cognitive functions. **d**, Mel-spectrograms from representative NC, MCI, and DE individuals, showing progressive degradation of spectral structure with increasing cognitive impairment.



Supplementary Fig. 5: Model performance and ablation analyses for cognitive impairment classification. **a**, EEG ablation analysis showing the contribution of model architecture and trajectory-based features. **b**, ECG ablation analysis. **c**, Video ablation analysis. **d**, Multimodal ablation analysis across different signal combinations. **e**, Task ablation analysis across multiple cognitive tasks. Among them, MCM denotes the proposed multimodal cognitive impairment assessment model, and DTA represents the dynamic task adapter module. EEGΔ, ECGΔ, and videoΔ indicate the statistical feature variations across tasks for each respective modality.

Cohort Characteristics



Supplementary Fig. 6: Participant recruitment, exclusion, and cohort assignment flowchart

Supplementary Table 1: Demographic characteristics of the internal development and external validation cohorts

	Total (n=1385)	NC (n=540)	Cognitive Impairment (n=845)	
			MCI (n=777)	DE (n=68)
Internal Development Cohort (n = 831)				
Participants, n	831	324	464	43
Sex (female, male, %)	(68.7, 31.3)	(71.6, 28.4)	(69.2, 30.8)	(41.9, 58.1)
Age (years), mean ± SD	71.53 ± 7.51	70.11 ± 7.22	71.50 ± 6.75	82.53 ± 8.44
Height (cm), mean ± SD	160.65 ± 7.39	161.02 ± 7.30	160.44 ± 7.43	159.80 ± 7.94
Weight (kg), mean ± SD	61.78 ± 10.14	62.71 ± 11.02	61.23 ± 9.52	59.73 ± 8.08
Years of education, mean ± SD	8.12 ± 3.67	8.63 ± 3.82	7.61 ± 3.42	9.90 ± 4.14
Education level (%) ^a	(33.6, 36.9, 21.8, 7.6, 0.1)	(37.3, 22.5, 28.7, 11.1, 0.3)	(32.3, 47.6, 16.4, 3.7, —)	(18.6, 30.2, 27.9, 23.3, —)
Marital status (%) ^b	(0.7, 78.7, 2.2, 18.4)	(—, 83.3, 0.9, 15.7)	(0.9, 77.8, 3.0, 18.3)	(4.7, 53.5, 2.3, 39.5)
Living status (%) ^c	(17.7, 57.0, 9.4, 11.1, 4.8)	(15.4, 61.4, 9.9, 12.7, 0.6)	(20.5, 57.5, 9.7, 11.0, 1.3)	(4.7, 18.6, 2.3, 74.4, —)
Disposable income (%) ^d	(10.0, 32.2, 19.0, 13.0, 14.2, 11.6)	(7.1, 26.5, 19.8, 15.1, 16.7, 14.8)	(10.8, 37.3, 19.2, 12.5, 11.9, 8.4)	(23.8, 19.0, 11.9, 2.4, 21.4, 21.4)
Family history of Alzheimer's (% , yes, no)	(7.7, 92.3)	(6.8, 93.2)	(8.6, 91.4)	(4.8, 95.2)
Number of children, mean ± SD	1.63 ± 0.85	1.52 ± 0.74	1.67 ± 0.88	2.07 ± 1.10
External Validation Cohort (n = 554)				
Participants, n	554	216	313	25
Sex (female, male, %)	(65.9, 34.1)	(63.9, 36.1)	(68.1, 31.9)	(56.0, 44.0)
Age (years), mean ± SD	71.43 ± 7.40	70.05 ± 6.59	71.52 ± 7.24	82.12 ± 7.28
Height (cm), mean ± SD	161.34 ± 7.70	162.06 ± 7.58	160.80 ± 7.71	162.24 ± 8.57
Weight (kg), mean ± SD	61.97 ± 9.67	61.75 ± 9.50	62.21 ± 9.65	60.31 ± 12.20
Years of education, mean ± SD	7.83 ± 3.65	8.38 ± 3.70	7.46 ± 3.42	7.60 ± 5.19
Education level (%) ^a	(37.2, 37.5, 19.7, 5.2, 0.4)	(38.0, 27.8, 24.5, 9.3, 0.5)	(36.1, 44.7, 17.3, 1.9, —)	(44.0, 32.0, 8.0, 12.0, 4.0)
Marital status (%) ^b	(0.9, 77.9, 2.0, 19.2)	(—, 83.8, 1.4, 14.8)	(1.0, 76.0, 2.2, 20.8)	(8.0, 52.0, 4.0, 36.0)
Living status (%) ^c	(16.3, 61.8, 10.5, 8.5, 2.9)	(16.3, 65.1, 7.4, 10.7, 0.5)	(17.3, 63.3, 11.8, 6.7, 1.0)	(4.0, 16.0, 20.0, 12.0, 48.0)
Disposable income (%) ^d	(8.3, 31.2, 22.6, 15.4, 12.8, 9.7)	(5.1, 27.3, 26.4, 15.8, 14.8, 10.6)	(8.9, 34.8, 19.8, 16.0, 11.8, 8.6)	(28.0, 20.0, 24.0, 4.0, 8.0, 16.0)
Family history of Alzheimer's (% , yes, no)	(6.3, 93.7)	(6.5, 93.5)	(6.7, 93.3)	(0.0, 100.0)
Number of children, mean ± SD	1.60 ± 0.83	1.50 ± 0.77	1.62 ± 0.82	2.08 ± 1.26

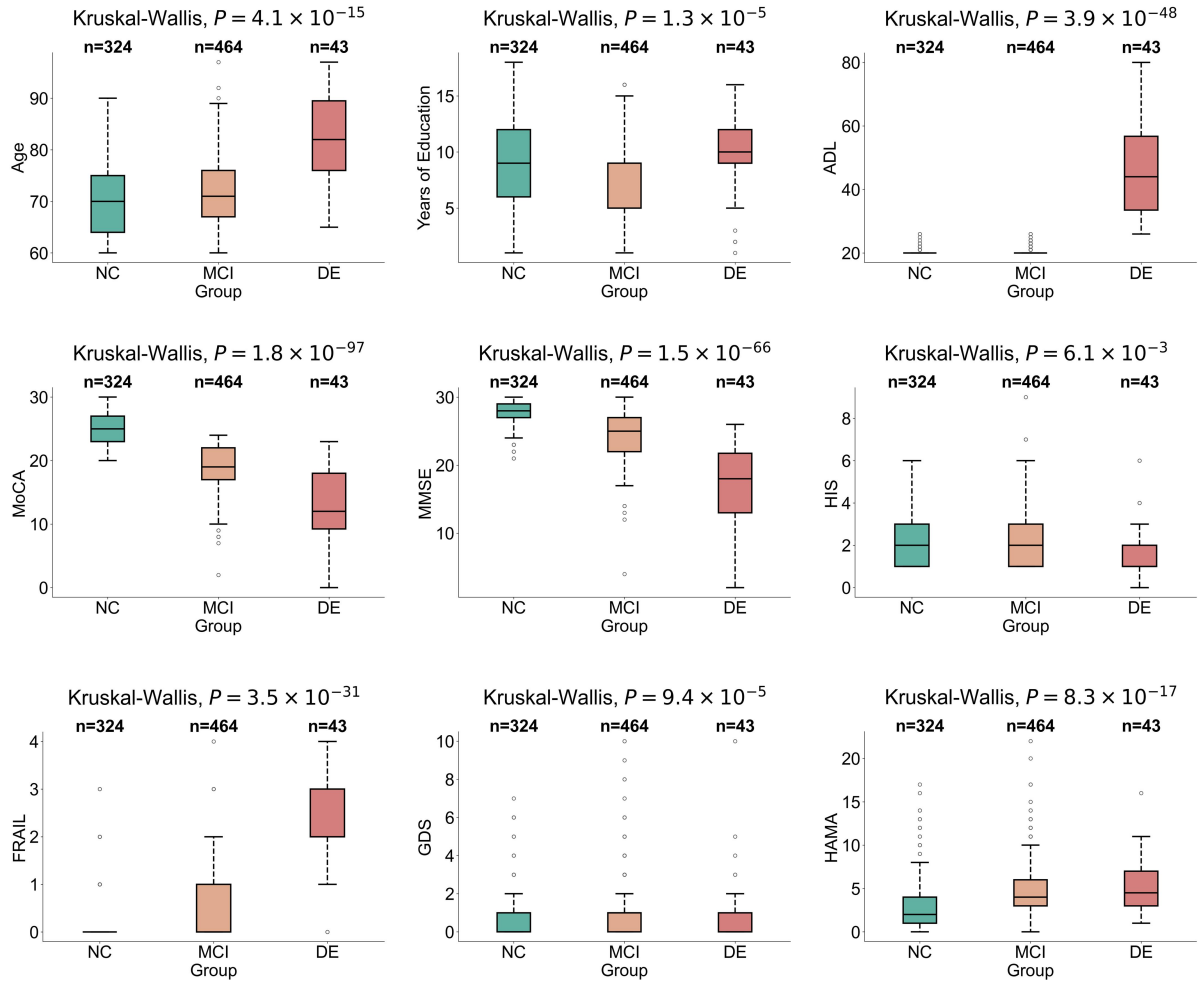
NC = *Normal Cognition*; MCI = *Mild Cognitive Impairment*; DE = *Dementia*.

^a Education level: (1) primary school; (2) middle school; (3) high school/technical secondary school; (4) college/university; (5) master's degree.

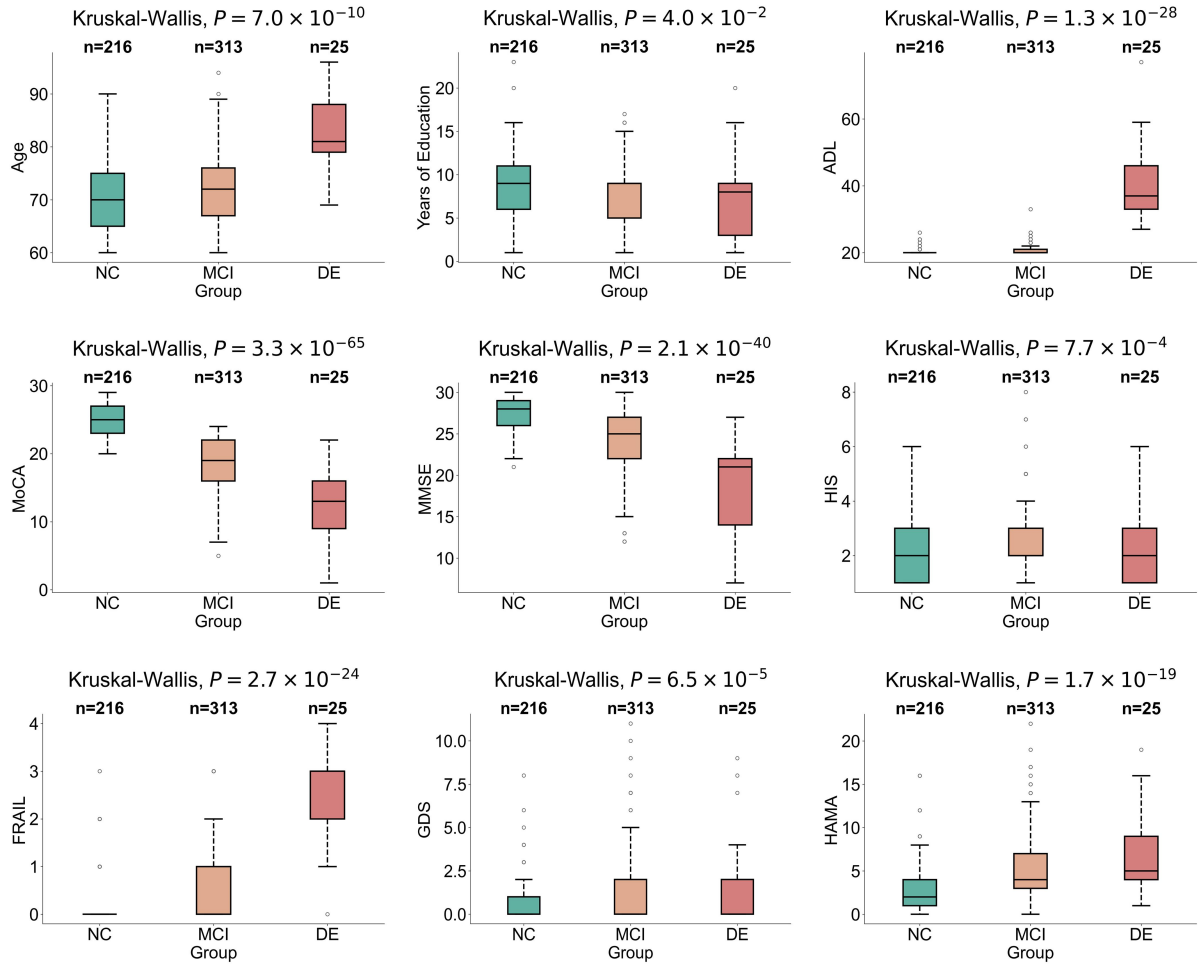
^b Marital status: (1) unmarried; (2) married; (3) divorced; (4) widowed.

^c Living status: (1) living alone; (2) with spouse; (3) with children; (4) with spouse and children; (5) other.

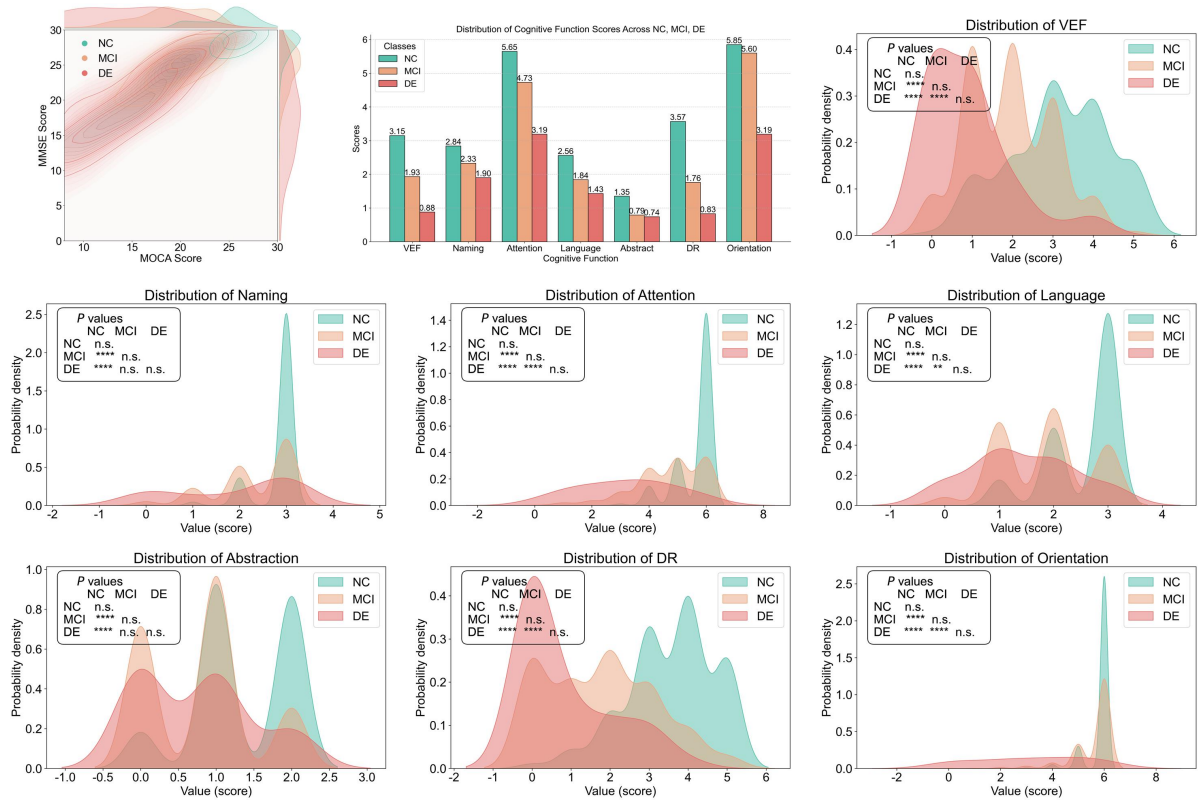
^d Disposable income (monthly, RMB): (1) <1000; (2) 1000–2000; (3) 2001–3000; (4) 3001–4000; (5) 4001–5000; (6) $\geq$ 5000.



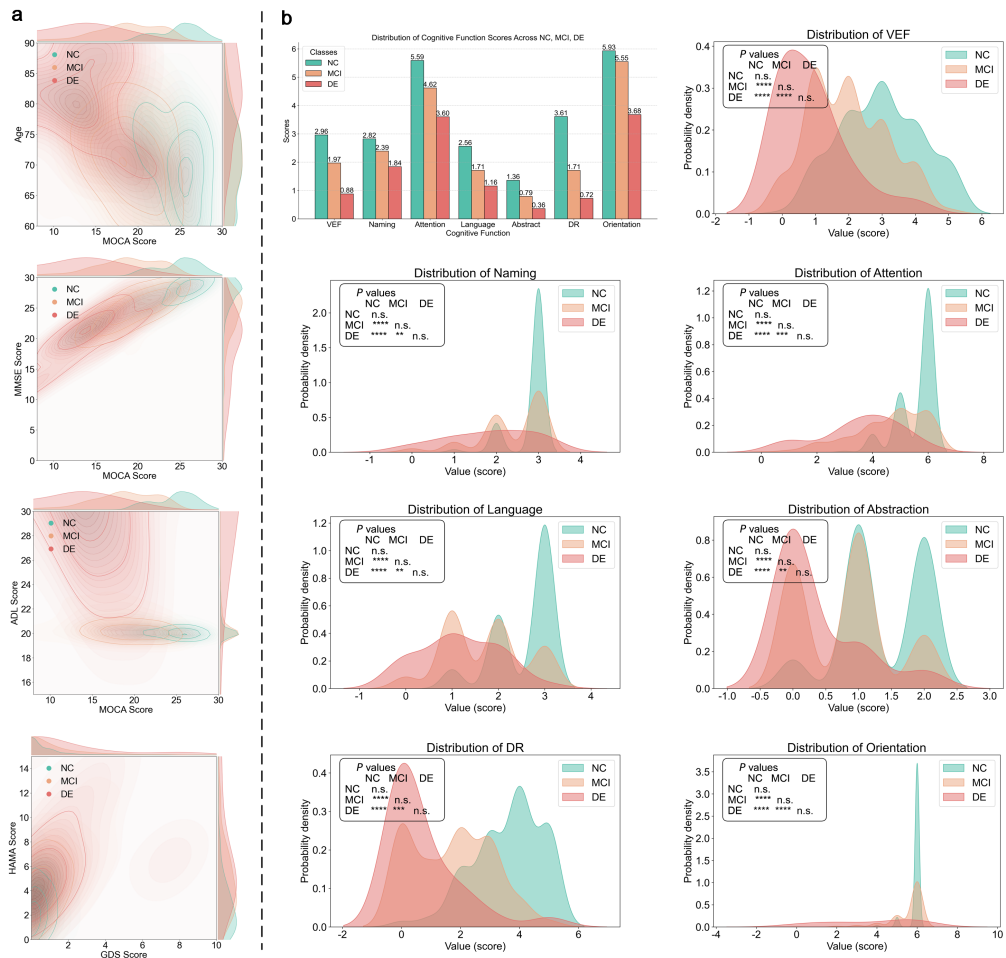
Supplementary Fig. 7: Distribution of demographic characteristics and clinical assessment scores in the internal development cohort. Boxplots show the distributions of age, years of education, ADL, MoCA, MMSE, HIS, FRAIL, GDS, and HAMA across NC, MCI, and DE groups. Kruskal–Wallis tests were used to compare group differences.



Supplementary Fig. 8: Distribution of demographic characteristics and clinical assessment scores in the external validation cohort. Boxplots show the distributions of age, years of education, ADL, MoCA, MMSE, HIS, FRAIL, GDS, and HAMA across NC, MCI, and DE groups. Kruskal-Wallis tests were used to compare group differences.

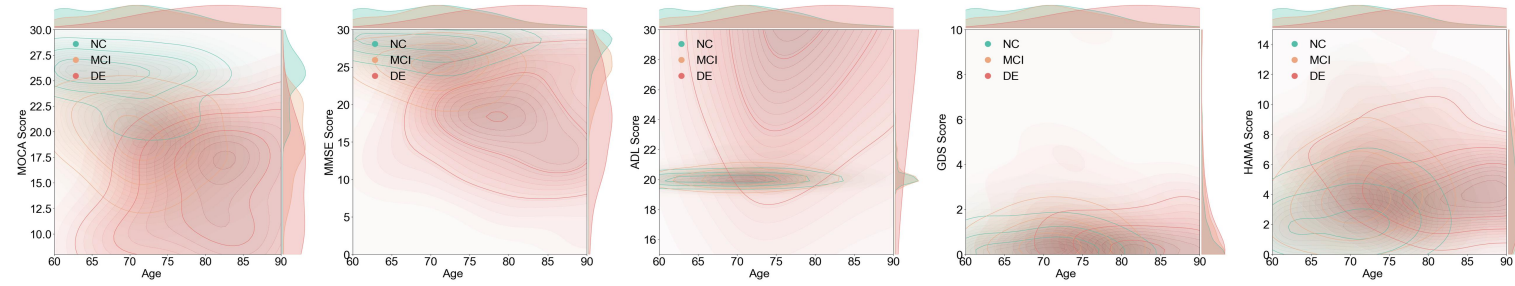


Supplementary Fig. 9: Distribution of cognitive assessment scores in the internal development cohort

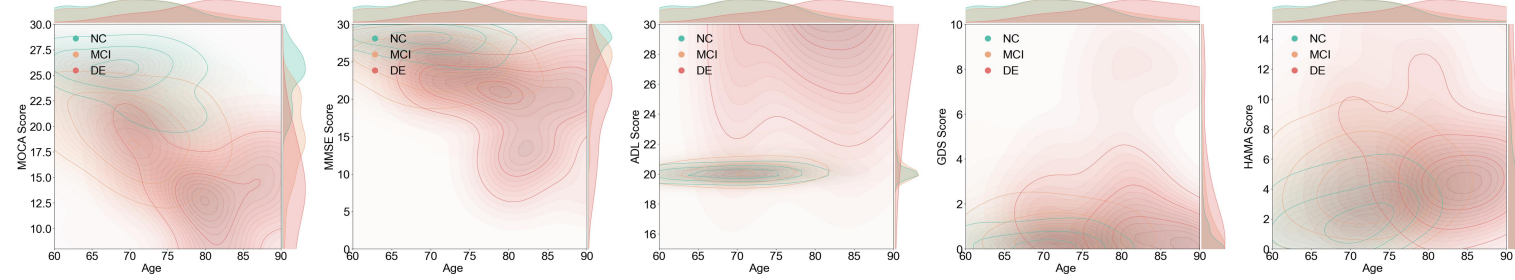


Supplementary Fig. 10: Distribution of cognitive assessment scores in the external validation cohort. **a**, Distribution of neuropsychological scores. **b**, Distribution of MoCA subdomains across cognitive statuses.

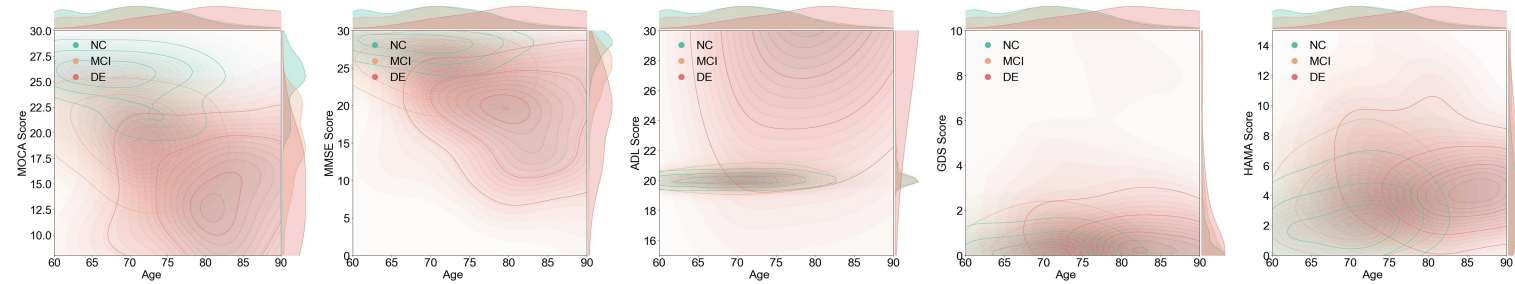
Internal Development Cohort



External Validation Cohort



Overall Cohort



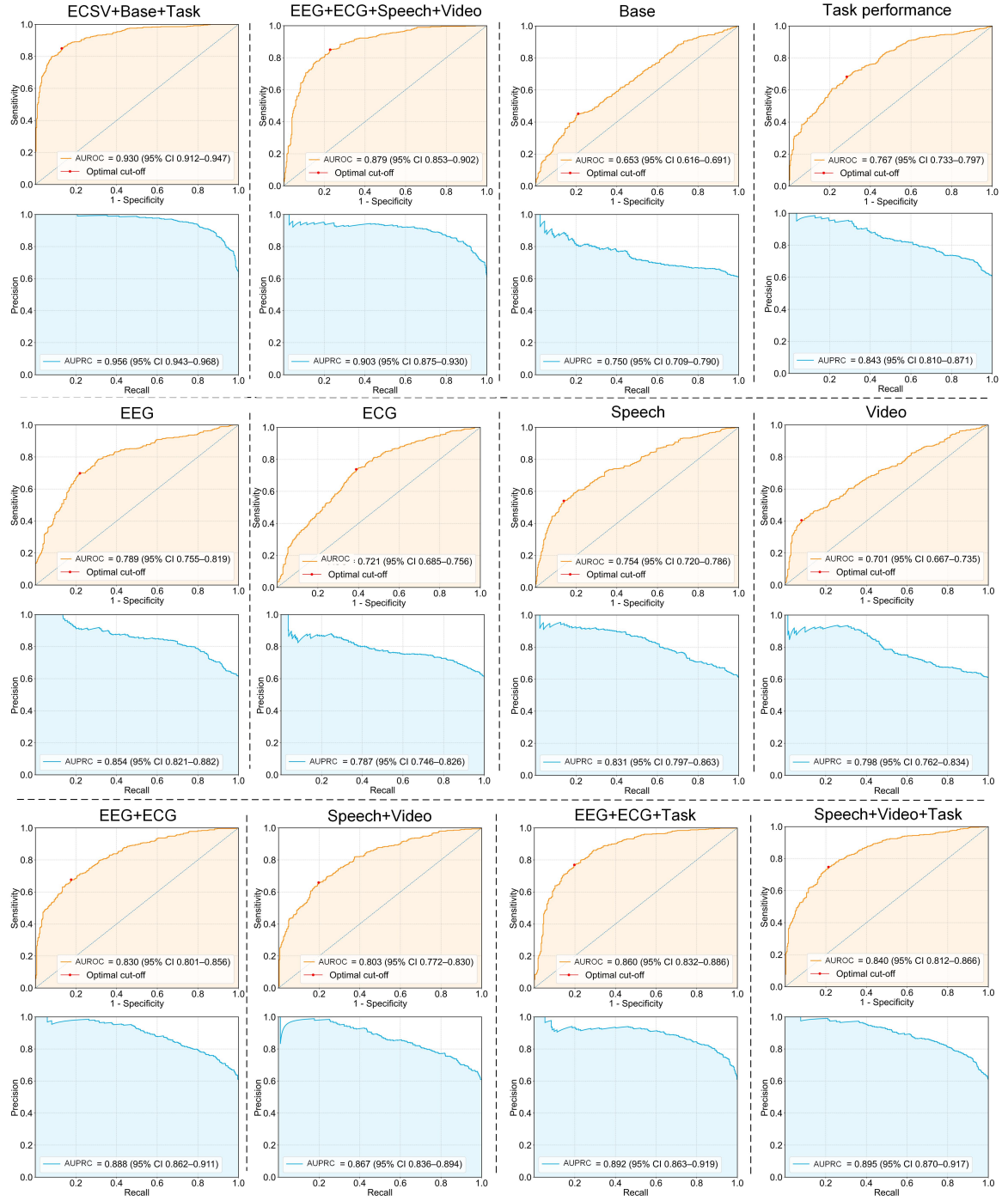
Supplementary Fig. 12: Distribution of cognitive assessment scores and age across cohorts. The figure illustrates the distribution of three groups, normal cognition (NC), mild cognitive impairment (MCI), and dementia (DE) in both overall scales and specific cognitive domains. It presents the two-dimensional joint probability density distributions between age and cognitive-related scales, including the MoCA, MMSE, ADL, HAMA, and GDS.

Multimodal Model Performance in the Internal Development Cohort

Supplementary Table 2: Comparison of assessment performance across different modalities

Modality	Accuracy % (95% CI)	AUROC % (95% CI)	AUPR % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
Demographic	58.4 (54.9–61.7)	65.3 (61.6–69.1)	75.0 (70.9–79.0)	45.2 (40.8–49.6)	79.0 (74.2–83.3)	77.1 (71.9–81.8)	47.9 (43.6–52.3)
Task performance	69.6 (66.3–72.7)	76.7 (73.5–79.9)	84.3 (81.0–87.1)	68.2 (64.0–72.3)	71.6 (66.4–76.5)	79.0 (74.9–82.7)	59.0 (54.0–63.9)
EEG	73.0 (69.9–76.0)	78.9 (75.8–82.0)	85.4 (82.1–88.2)	69.8 (65.6–73.8)	78.1 (73.2–82.5)	83.3 (79.4–86.7)	62.3 (57.4–67.0)
ECG	68.8 (65.6–72.0)	72.1 (68.6–75.7)	78.7 (74.6–82.6)	73.8 (69.7–77.5)	61.1 (55.6–66.5)	74.8 (70.8–78.5)	59.8 (54.3–65.1)
Speech	66.5 (63.2–69.8)	75.4 (72.1–78.6)	83.1 (79.7–86.3)	54.0 (49.6–58.4)	86.1 (81.9–89.7)	85.9 (81.6–89.5)	54.5 (50.1–58.9)
Video	60.5 (57.1–63.9)	70.1 (66.5–73.6)	79.8 (76.2–83.4)	40.4 (36.1–44.8)	92.0 (88.5–94.7)	88.7 (83.9–92.5)	49.7 (45.6–53.7)
EEG + ECG	73.4 (70.3–76.4)	83.0 (80.2–85.7)	88.8 (86.2–91.1)	67.7 (63.4–71.7)	82.4 (77.8–86.4)	85.8 (81.9–89.0)	61.9 (57.2–66.6)
Speech + Video	71.5 (68.3–74.5)	80.3 (77.3–83.2)	86.7 (83.6–89.4)	65.9 (61.6–70.0)	80.2 (75.5–84.4)	83.9 (79.9–87.4)	60.0 (55.3–64.7)
EEG + ECG + Task	78.2 (75.3–81.0)	86.0 (83.4–88.7)	89.2 (86.3–91.9)	76.9 (73.0–80.5)	80.2 (75.5–84.4)	85.9 (82.4–89.0)	69.0 (64.0–73.6)
Speech + Video + Task	76.3 (73.3–79.1)	84.0 (81.4–86.7)	89.5 (87.0–91.7)	74.8 (70.7–78.5)	78.7 (73.8–83.0)	84.6 (80.9–87.8)	66.6 (61.6–71.3)
ECSV	81.9 (79.2–84.5)	87.9 (85.4–90.3)	90.3 (87.5–93.0)	85.0 (81.6–88.0)	77.2 (72.2–81.6)	85.3 (82.0–88.3)	76.7 (71.7–81.2)
Base+Task+ECSV	85.8 (83.2–88.1)	93.0 (91.3–94.7)	95.6 (94.3–96.8)	85.0 (81.6–88.0)	87.0 (82.9–90.5)	91.1 (88.2–93.5)	78.8 (74.2–82.9)

Task performance refers to participants’ scores and time during cognitive task; EEG = electroencephalogram; ECG = electrocardiogram; Speech = acoustic features; Video = facial/behavioral features; ECSV = multimodal fusion of EEG, ECG, Speech, and Video; Base+Task+ECSV combines all available feature groups. All performance metrics were obtained via stratified 5-fold cross-validation on the internal development dataset and are reported as the micro-average across all folds. All metrics are reported as percentages (0–100%) with 95% confidence intervals (CI) [1]. The 95% CI of area under the receiver operating characteristic curve (AUROC) were computed using the non-parametric method of DeLong [2]. The 95% CI of area under the precision–recall curve (AUPRC) were computed using 2000 bootstrap resampling. The 95% CI of accuracy, sensitivity, specificity, PPV, and NPV were obtained using the Clopper–Pearson method. Bold values indicate the overall best-performing modality combination.

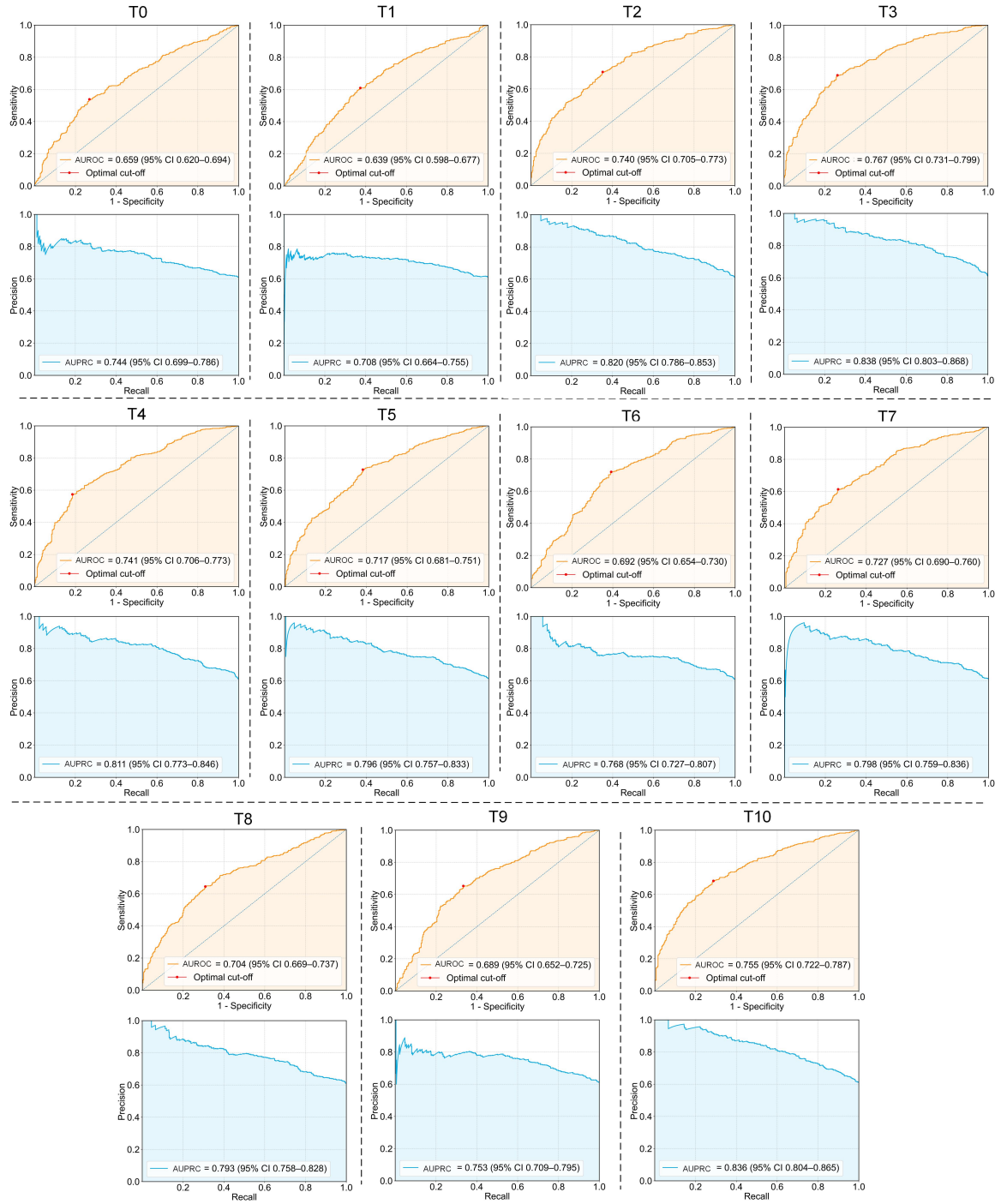


Supplementary Fig. 13: ROC and PR curves of assessment performance across different modalities

Supplementary Table 3: Comparison of assessment performance across cognitive tasks (T0–T10)

Task	Accuracy % (95% CI)	AUROC % (95% CI)	AUPRC % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
T0	61.4 (58.0–64.7)	65.9 (62.1–69.6)	74.4 (69.9–78.6)	53.8 (49.4–58.3)	73.1 (68.0–77.9)	75.8 (71.1–80.2)	50.3 (45.7–54.9)
T1	61.6 (58.2–64.9)	63.9 (60.0–67.7)	70.8 (66.4–75.5)	60.9 (56.5–65.2)	62.7 (57.1–67.9)	71.9 (67.4–76.1)	50.6 (45.6–55.6)
T2	68.4 (65.1–71.5)	74.0 (70.6–77.3)	82.0 (78.6–85.3)	70.6 (66.4–74.5)	64.8 (59.3–70.0)	75.8 (71.7–79.6)	58.5 (53.2–63.6)
T3	70.8 (67.5–73.8)	76.7 (73.5–80.0)	83.8 (80.3–86.8)	68.8 (64.6–72.8)	73.8 (68.6–78.5)	80.4 (76.4–84.0)	60.2 (55.2–65.1)
T4	66.8 (63.5–70.0)	74.1 (70.7–77.4)	81.1 (77.3–84.6)	57.4 (53.0–61.7)	81.5 (76.8–85.6)	82.9 (78.6–86.7)	55.0 (50.4–59.5)
T5	68.5 (65.2–71.6)	71.7 (68.2–75.2)	79.6 (75.7–83.3)	72.8 (68.7–76.6)	61.7 (56.2–67.0)	74.8 (70.8–78.6)	59.2 (53.7–64.5)
T6	67.6 (64.3–70.8)	69.2 (65.6–72.9)	76.8 (72.7–80.7)	72.0 (67.9–75.9)	60.8 (55.3–66.2)	74.2 (70.1–78.0)	58.1 (52.7–63.4)
T7	66.2 (62.9–69.4)	72.7 (69.2–76.1)	79.8 (75.9–83.6)	61.3 (56.9–65.6)	73.8 (68.6–78.5)	78.5 (74.2–82.5)	54.9 (50.1–59.7)
T8	66.3 (63.0–69.5)	70.4 (66.8–73.9)	79.3 (75.8–82.8)	64.5 (60.2–68.7)	69.1 (63.8–74.1)	76.6 (72.3–80.5)	55.4 (50.5–60.4)
T9	65.8 (62.5–69.0)	68.9 (65.3–72.6)	75.3 (70.9–79.5)	65.3 (61.0–69.4)	66.7 (61.2–71.8)	75.4 (71.1–79.4)	55.1 (50.0–60.1)
T10	69.6 (66.3–72.7)	75.5 (72.3–78.8)	83.6 (80.4–86.5)	68.4 (64.2–72.5)	71.3 (66.0–76.2)	78.9 (74.7–82.6)	59.1 (54.0–64.0)
T2+T3	74.4 (71.3–77.3)	81.1 (78.1–84.0)	87.4 (84.6–89.8)	75.0 (70.9–78.7)	73.5 (68.3–78.2)	81.5 (77.7–85.0)	65.2 (60.1–70.1)
T2+T3+T4+T10	78.3 (75.4–81.1)	85.3 (82.8–87.9)	89.8 (87.4–92.0)	81.9 (78.2–85.1)	72.8 (67.6–77.6)	82.5 (78.9–85.7)	72.0 (66.8–76.7)
T2+T3+T4+T7+T10	80.5 (77.6–83.1)	87.8 (85.4–90.2)	91.4 (89.2–93.4)	79.1 (75.3–82.6)	82.7 (78.2–86.7)	87.7 (84.4–90.6)	71.7 (66.8–76.2)

T0 represents the resting-state acquisition. T1–T10 correspond to ten specific tasks designed based on multidimensional cognitive functions. Each row (T0–T10) reports model performance using all available modalities collected under the corresponding task condition. All performance metrics were obtained via stratified 5-fold cross-validation on the internal development dataset and are reported as the micro-average across all folds. All metrics are reported as percentages (0–100%) with 95% confidence intervals (CI) [1]. The 95% CI of area under the receiver operating characteristic curve (AUROC) were computed using the non-parametric method of DeLong [2]. The 95% CI of area under the precision–recall curve (AUPRC) were computed using 2000 bootstrap resampling. The 95% CI of accuracy, sensitivity, specificity, PPV, and NPV were obtained using the Clopper–Pearson method.

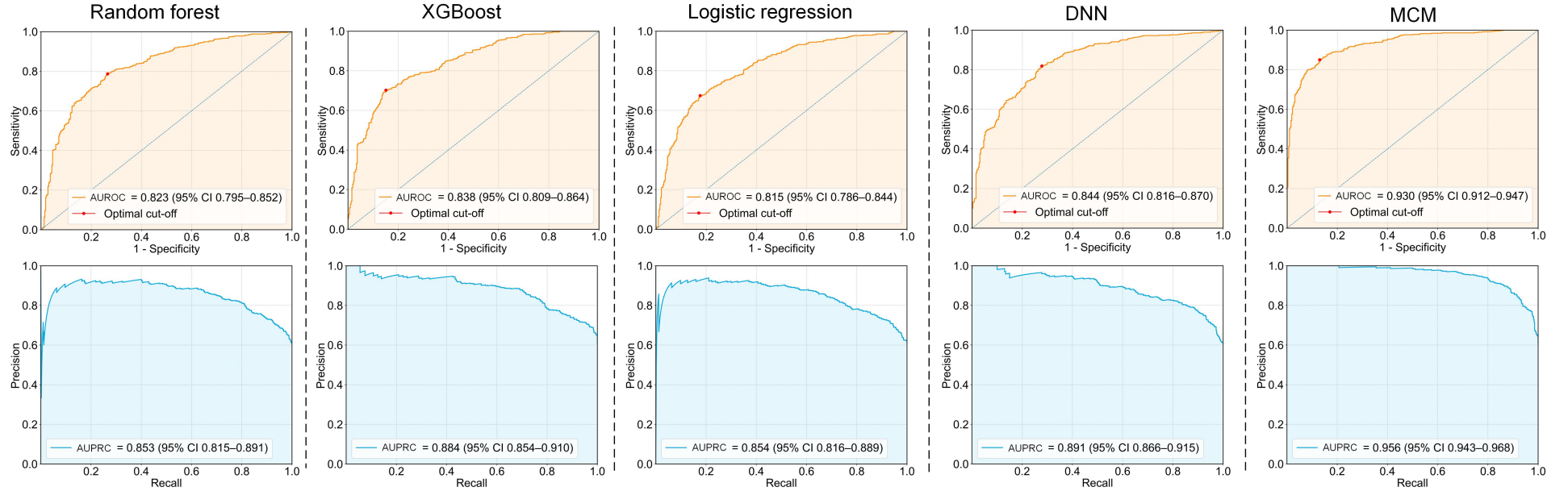


Supplementary Fig. 14: ROC and PR curves of assessment performance across cognitive tasks

Supplementary Table 4: Comparison of assessment performance across model architectures under all modalities

Model architecture	Accuracy % (95% CI)	AUROC % (95% CI)	AUPRC % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
Random Forest	76.7 (73.6–79.5)	82.3 (79.4–85.2)	85.3 (81.5–89.1)	78.7 (74.9–82.2)	73.5 (68.3–78.2)	82.3 (78.6–85.6)	68.8 (63.6–73.6)
XGBoost	75.9 (72.9–78.8)	83.8 (81.1–86.5)	88.4 (85.4–91.0)	70.2 (66.0–74.2)	84.9 (80.5–88.6)	87.9 (84.3–90.9)	64.6 (59.8–69.1)
Logistic Regression	73.3 (70.1–76.3)	81.5 (78.6–84.5)	85.4 (81.6–88.9)	67.5 (63.2–71.5)	82.4 (77.8–86.4)	85.7 (81.9–89.0)	61.8 (57.0–66.4)
DNN	78.1 (75.1–80.9)	84.4 (81.7–87.1)	89.1 (86.6–91.5)	81.9 (78.2–85.1)	72.2 (67.0–77.0)	82.2 (78.6–85.4)	71.8 (66.6–76.6)
MCM (Ours)	85.8 (83.2–88.1)	93.0 (91.3–94.7)	95.6 (94.3–96.8)	85.0 (81.6–88.0)	87.0 (82.9–90.5)	91.1 (88.2–93.5)	78.8 (74.2–82.9)

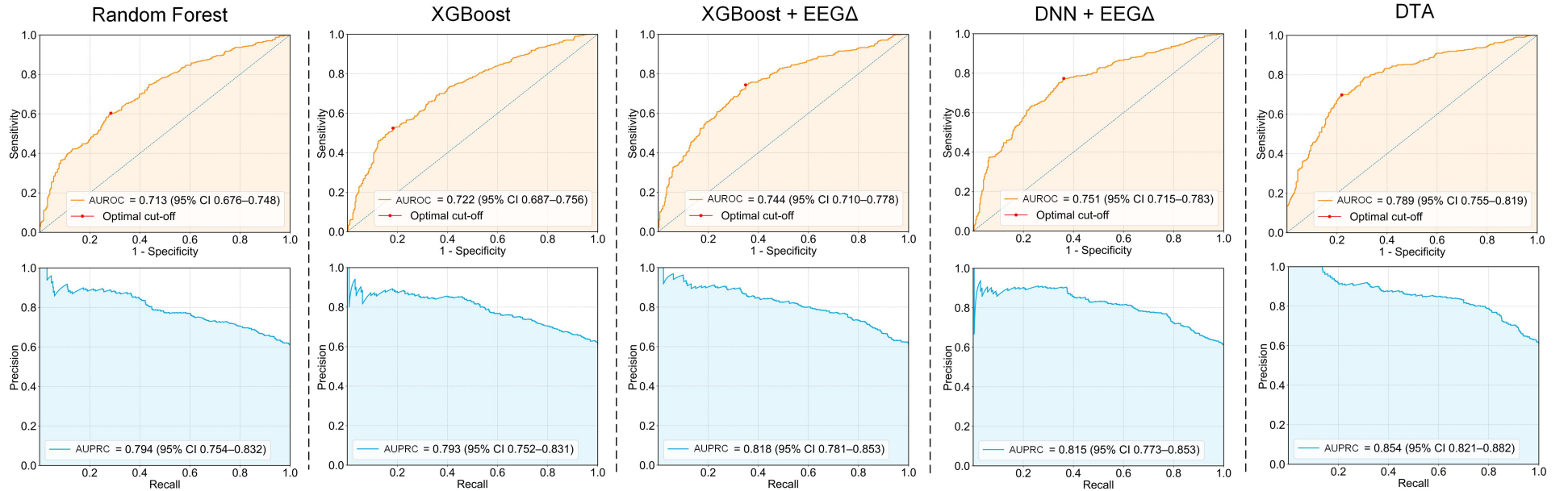
Random Forest (ensemble of decision trees with bootstrap aggregation), XGBoost (gradient boosting decision trees with regularization), Logistic Regression (generalized linear model with a logit link), DNN (deep neural network with fully connected layers), and MCM (the proposed multimodal cognitive assessment model) were compared. All performance metrics were obtained via stratified 5-fold cross-validation on the internal development dataset and are reported as the micro-average across all folds. All metrics are reported as percentages (0–100%) with 95% confidence intervals (CI) [1]. The 95% CI of area under the receiver operating characteristic curve (AUROC) were computed using the non-parametric method of DeLong [2]. The 95% CI of area under the precision–recall curve (AUPRC) were computed using 2000 bootstrap resampling. The 95% CI of accuracy, sensitivity, specificity, PPV, and NPV were obtained using the Clopper–Pearson method.

**Supplementary Fig. 15:** ROC and PR curves of assessment performance across model architectures under all modalities

Supplementary Table 5: Comparison of assessment performance across model architectures using EEG signals

Model	Accuracy % (95% CI)	AUROC % (95% CI)	AUPRC % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
Random Forest	64.7 (61.4–68.0)	71.3 (67.8–74.8)	79.4 (75.4–83.2)	60.4 (55.9–64.6)	71.6 (66.4–76.5)	76.9 (72.4–80.9)	53.6 (48.8–58.4)
XGBoost	63.9 (60.5–67.2)	72.2 (68.7–75.7)	79.3 (75.2–83.1)	52.5 (48.0–56.9)	81.8 (77.1–85.8)	81.8 (77.2–85.9)	52.4 (47.9–56.8)
Logistic Regression	65.6 (62.2–68.8)	70.6 (67.0–74.2)	78.4 (74.4–82.1)	65.1 (60.8–69.2)	66.4 (60.9–71.5)	75.2 (70.9–79.1)	54.8 (49.8–59.8)
XGBoost + EEG Δ	70.8 (67.5–73.8)	74.4 (71.0–77.8)	81.8 (78.1–85.3)	74.4 (70.3–78.1)	65.1 (59.7–70.3)	76.9 (73.0–80.6)	61.9 (56.5–67.1)
DNN + EEG Δ	72.1 (68.9–75.1)	75.1 (71.7–78.4)	81.5 (77.3–85.3)	77.3 (73.4–80.9)	63.9 (58.4–69.1)	77.0 (73.1–80.6)	64.3 (58.8–69.5)
DTA (Ours)	73.0 (69.9–76.0)	78.9 (75.8–82.0)	85.4 (82.1–88.2)	69.8 (65.6–73.8)	78.1 (73.2–82.5)	83.3 (79.4–86.7)	62.3 (57.4–67.0)

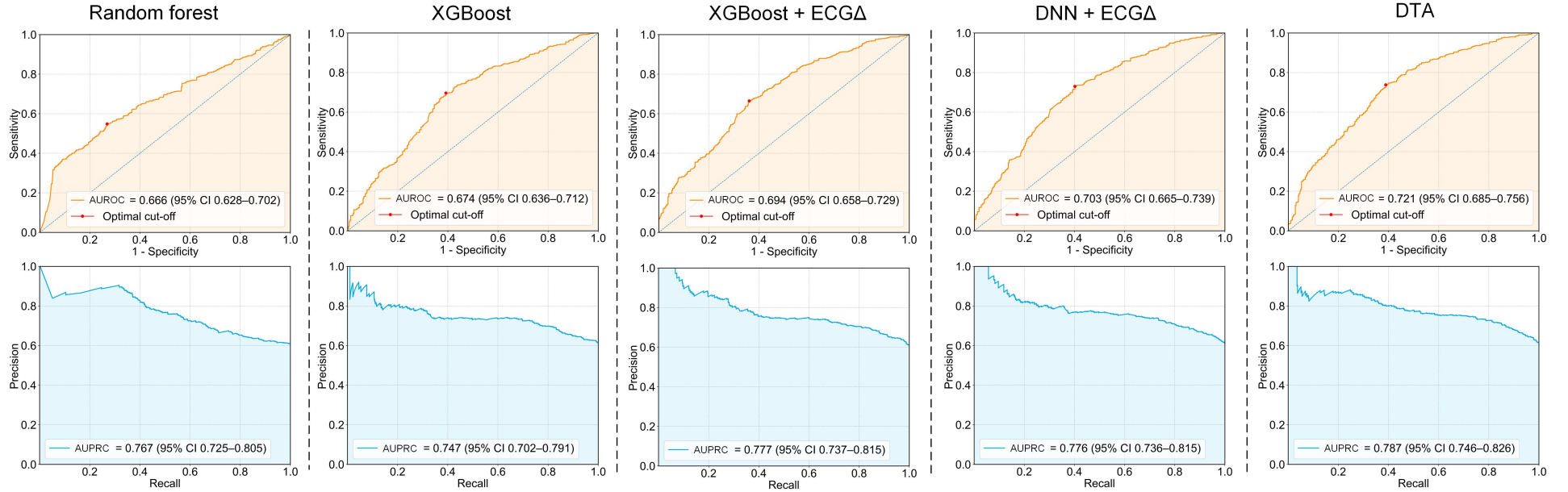
Comparison of EEG-based ablation models, including classical machine learning models (Random Forest, XGBoost, Logistic Regression), enhanced variants with EEG Δ features, a deep neural network with fully connected layers (DNN), and the proposed dynamic task-adaptive model (DTA). EEG Δ incorporate trajectory-level EEG features, which capture task-related signal dynamics. The proposed dynamic task-adaptive model integrates multimodal representations and adaptive weighting, achieving the best overall performance. All metrics are reported in the same format as Table 4.

**Supplementary Fig. 16:** ROC and PR curves of assessment performance across model architectures using EEG signals

Supplementary Table 6: Comparison of assessment performance across model architectures using ECG signals

Model	Accuracy % (95% CI)	AUROC % (95% CI)	AUPRC % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
Random Forest	62.0 (58.6–65.3)	66.6 (62.9–70.2)	76.7 (72.5–80.5)	54.8 (50.4–59.2)	73.1 (68.0–77.9)	76.2 (71.5–80.4)	50.9 (46.2–55.5)
XGBoost	66.2 (62.9–69.4)	67.4 (63.6–71.1)	74.7 (70.2–79.1)	69.6 (65.4–73.6)	60.8 (55.3–66.2)	73.5 (69.4–77.4)	56.1 (50.8–61.4)
Logistic Regression	57.6 (54.2–61.0)	66.4 (62.8–70.1)	76.4 (72.3–80.3)	35.9 (31.7–40.2)	91.7 (88.1–94.4)	87.1 (81.8–91.3)	47.7 (43.8–51.8)
XGBoost + ECG Δ	65.5 (62.1–68.7)	69.4 (65.8–73.1)	77.7 (73.7–81.5)	66.5 (62.2–70.6)	63.9 (58.4–69.1)	74.2 (69.9–78.2)	54.9 (49.7–60.0)
DNN + ECG Δ	67.9 (64.6–71.0)	70.3 (66.6–73.9)	77.6 (73.6–81.5)	73.0 (68.9–76.8)	59.9 (54.3–65.3)	74.0 (69.9–77.8)	58.6 (53.1–64.0)
DTA (Ours)	68.8 (65.6–72.0)	72.1 (68.6–75.7)	78.7 (74.6–82.6)	73.8 (69.7–77.5)	61.1 (55.6–66.5)	74.8 (70.8–78.5)	59.8 (54.3–65.1)

Comparison of ECG-based ablation models, including classical machine learning models (Random Forest, XGBoost, Logistic Regression), enhanced variants with ECG Δ features, a deep neural network with fully connected layers (DNN), and the proposed dynamic task-adaptive model (DTA). ECG Δ incorporate trajectory-level ECG features, which capture task-related signal dynamics. The proposed dynamic task-adaptive model integrates multimodal representations and adaptive weighting, achieving the best overall performance. All metrics are reported in the same format as Table 4.

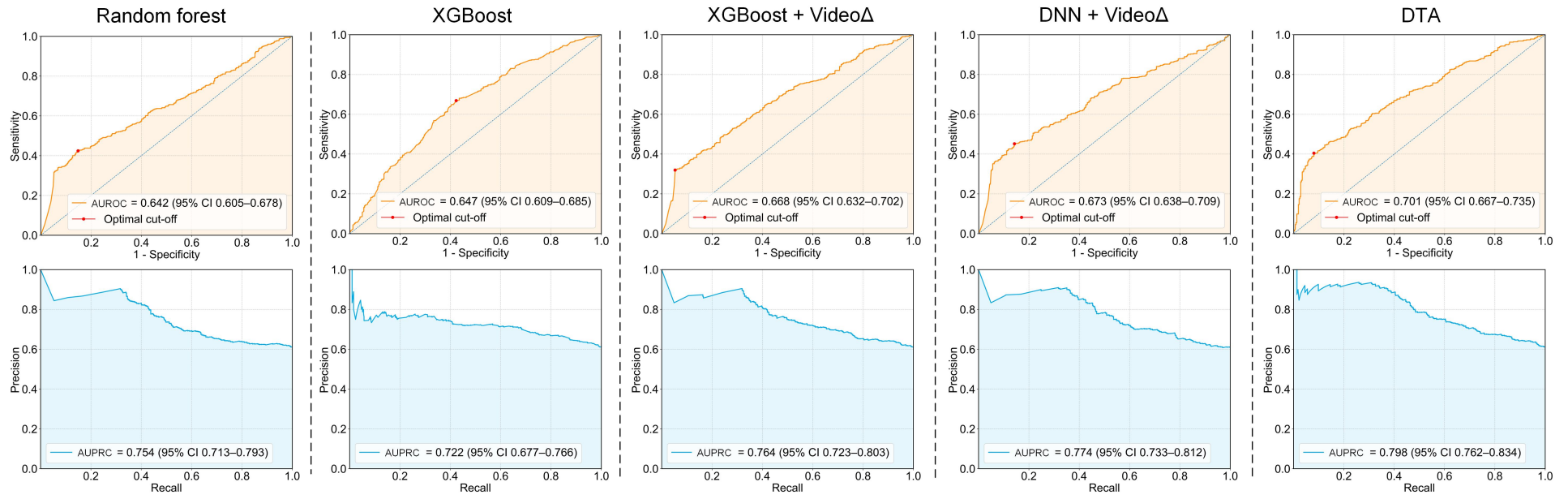


Supplementary Fig. 17: ROC and PR curves of assessment performance across model architectures using ECG signals

Supplementary Table 7: Comparison of assessment performance across model architectures using video signals

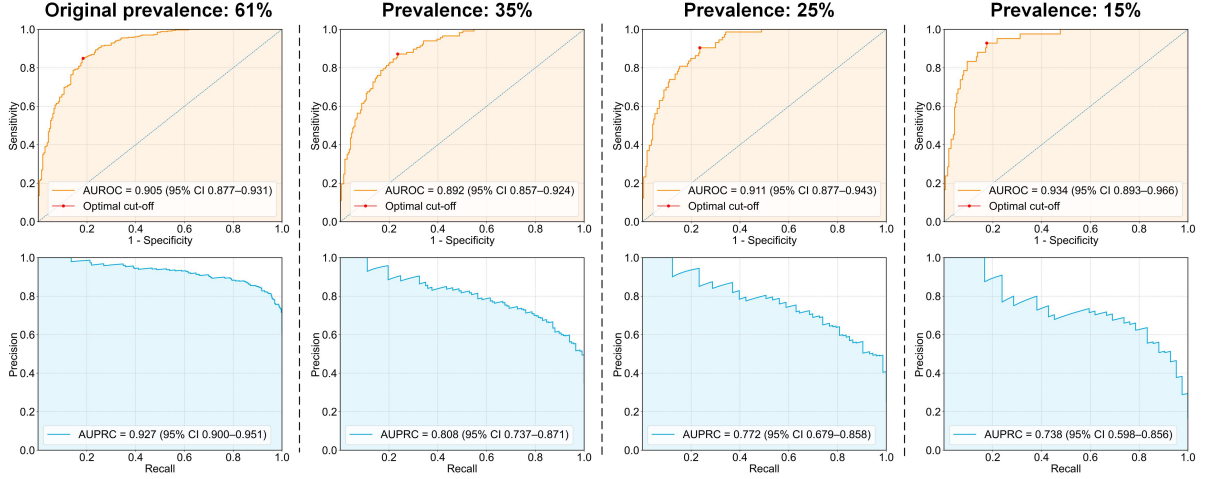
Model	Accuracy % (95% CI)	AUROC % (95% CI)	AUPRC % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
Random Forest	59.1 (55.7–62.5)	64.2 (60.5–67.9)	75.4 (71.3–79.3)	42.4 (38.1–46.8)	85.2 (80.8–88.9)	81.7 (76.5–86.2)	48.6 (44.4–52.8)
XGBoost	63.3 (59.9–66.6)	64.7 (60.9–68.6)	72.2 (67.7–76.6)	66.9 (62.6–71.0)	57.7 (52.1–63.2)	71.2 (66.9–75.2)	52.7 (47.3–58.0)
Logistic Regression	64.0 (60.7–67.3)	62.5 (58.6–66.3)	71.4 (67.2–75.7)	77.9 (74.0–81.4)	42.3 (36.8–47.9)	67.9 (63.9–71.7)	55.0 (48.6–61.3)
XGBoost + Video Δ	56.4 (53.0–59.8)	66.8 (63.1–70.5)	76.4 (72.3–80.3)	32.0 (27.9–36.2)	94.8 (91.7–96.9)	90.5 (85.2–94.4)	47.1 (43.2–51.0)
DNN + Video Δ	61.0 (57.6–64.3)	67.3 (63.7–70.9)	77.4 (73.3–81.2)	45.2 (40.8–49.6)	85.8 (81.5–89.4)	83.3 (78.3–87.5)	50.0 (45.8–54.2)
DTA (Ours)	60.5 (57.1–63.9)	70.1 (66.5–73.6)	79.8 (76.2–83.4)	40.4 (36.1–44.8)	92.0 (88.5–94.7)	88.7 (83.9–92.5)	49.7 (45.6–53.7)

Comparison of video-based ablation models, including classical machine learning models (Random Forest, XGBoost, Logistic Regression), enhanced variants with video Δ features, a deep neural network with fully connected layers (DNN), and the proposed dynamic task-adaptive model (DTA). Video Δ incorporate trajectory-level video features, which capture task-related signal dynamics. The proposed dynamic task-adaptive model integrates multimodal representations and adaptive weighting, achieving the best overall performance. All metrics are reported in the same format as Table 4.

**Supplementary Fig. 18:** ROC and PR curves of assessment performance across model architectures using video signals

Supplementary Table 8: External validation performance for distinguishing NC from MCI/DE under different prevalence settings. The table reports model performance in the external validation cohort for binary classification between cognitively normal individuals (NC) and individuals with cognitive impairment (MCI/DE). Metrics are reported as percentages with 95% confidence intervals.

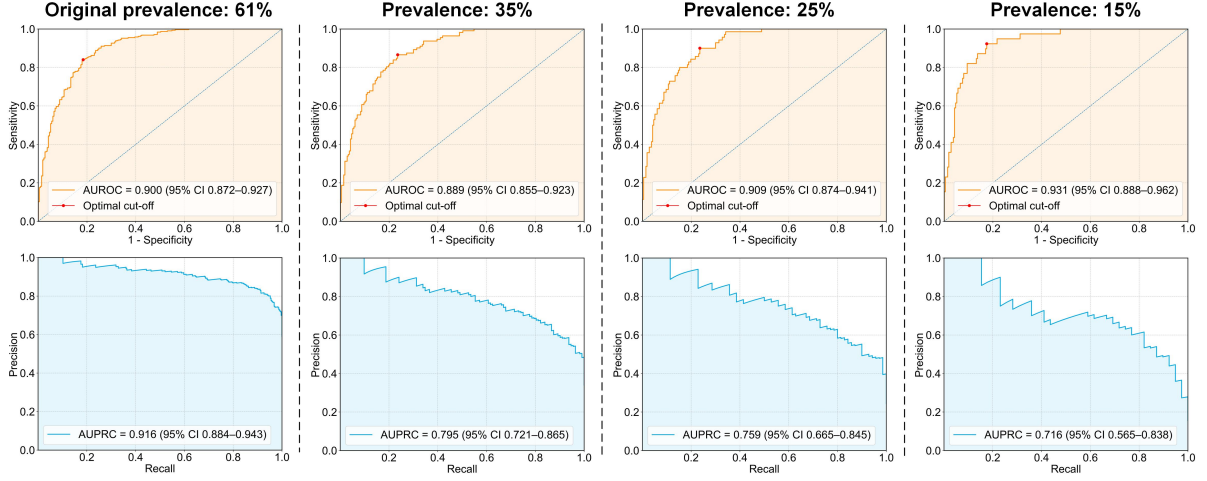
Prevalence	Accuracy (%)	AUROC (%)	AUPRC (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
61% (original)	83.6 (80.3–86.6)	90.5 (87.9–93.1)	92.7 (90.0–95.1)	84.9 (80.6–88.6)	81.6 (75.8–86.5)	87.8 (83.7–91.1)	77.6 (71.7–82.9)
35%	80.2 (75.6–84.4)	89.2 (85.9–92.6)	80.8 (73.7–87.1)	87.2 (79.7–92.6)	76.5 (70.3–82.0)	66.7 (58.6–74.1)	91.7 (86.7–95.3)
25%	80.0 (74.9–84.4)	91.1 (87.8–94.4)	77.2 (67.9–85.8)	90.4 (81.2–96.1)	76.5 (70.3–82.0)	56.4 (46.9–65.6)	96.0 (91.8–98.4)
15%	84.3 (79.2–88.5)	93.4 (90.0–96.8)	73.8 (59.8–85.6)	92.9 (80.5–98.5)	82.5 (76.8–87.4)	51.3 (39.6–63.0)	98.3 (95.2–99.7)



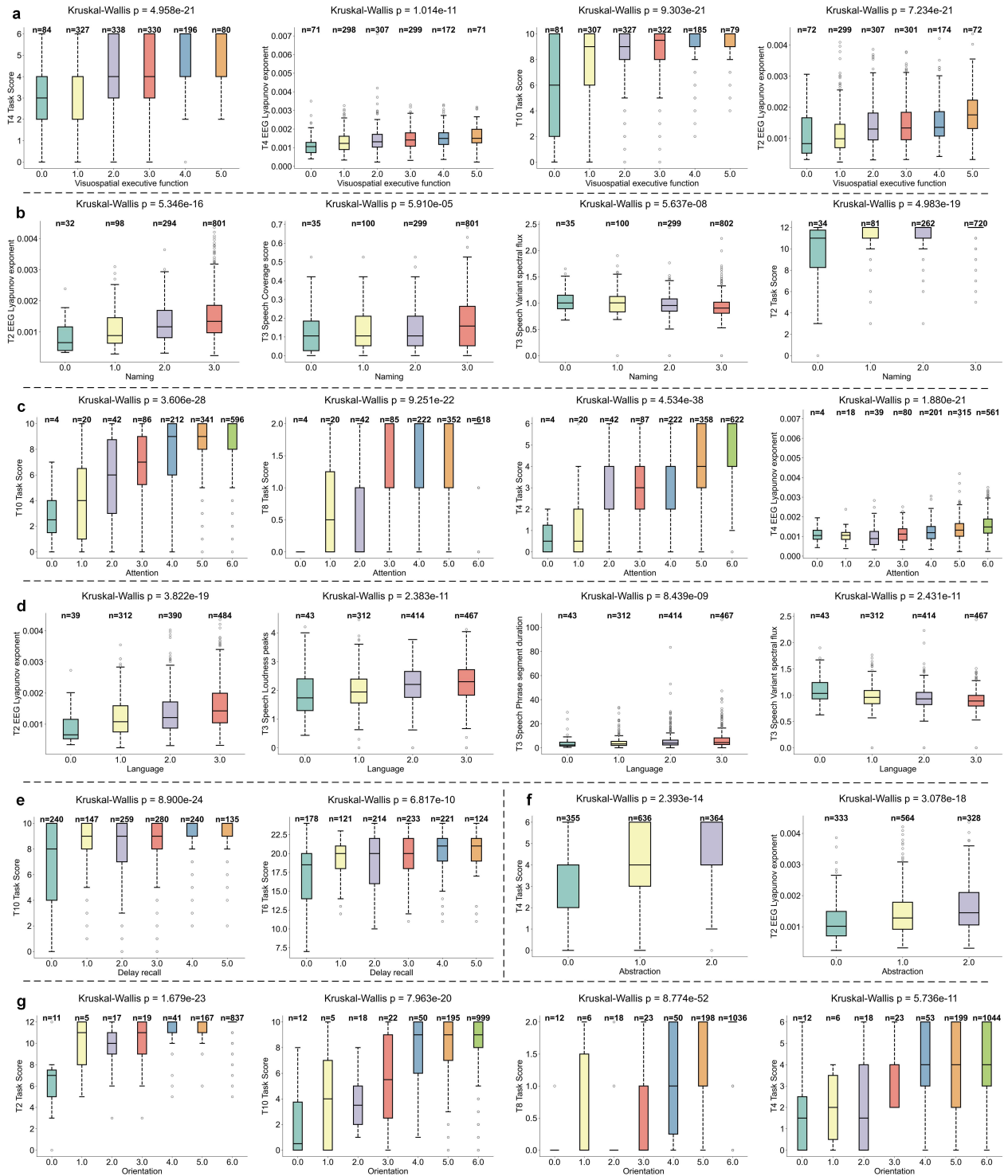
Supplementary Fig. 19: External validation performance for distinguishing NC from MCI/DE under different prevalence settings. Model performance was evaluated under the original prevalence distribution and adjusted prevalence settings of 35%, 25%, and 15%.

Supplementary Table 9: External validation performance for distinguishing NC from MCI under different prevalence settings. The table reports model performance in the external validation cohort for binary classification between cognitively normal individuals (NC) and individuals with mild cognitive impairment (MCI). Metrics are reported as percentages with 95% confidence intervals.

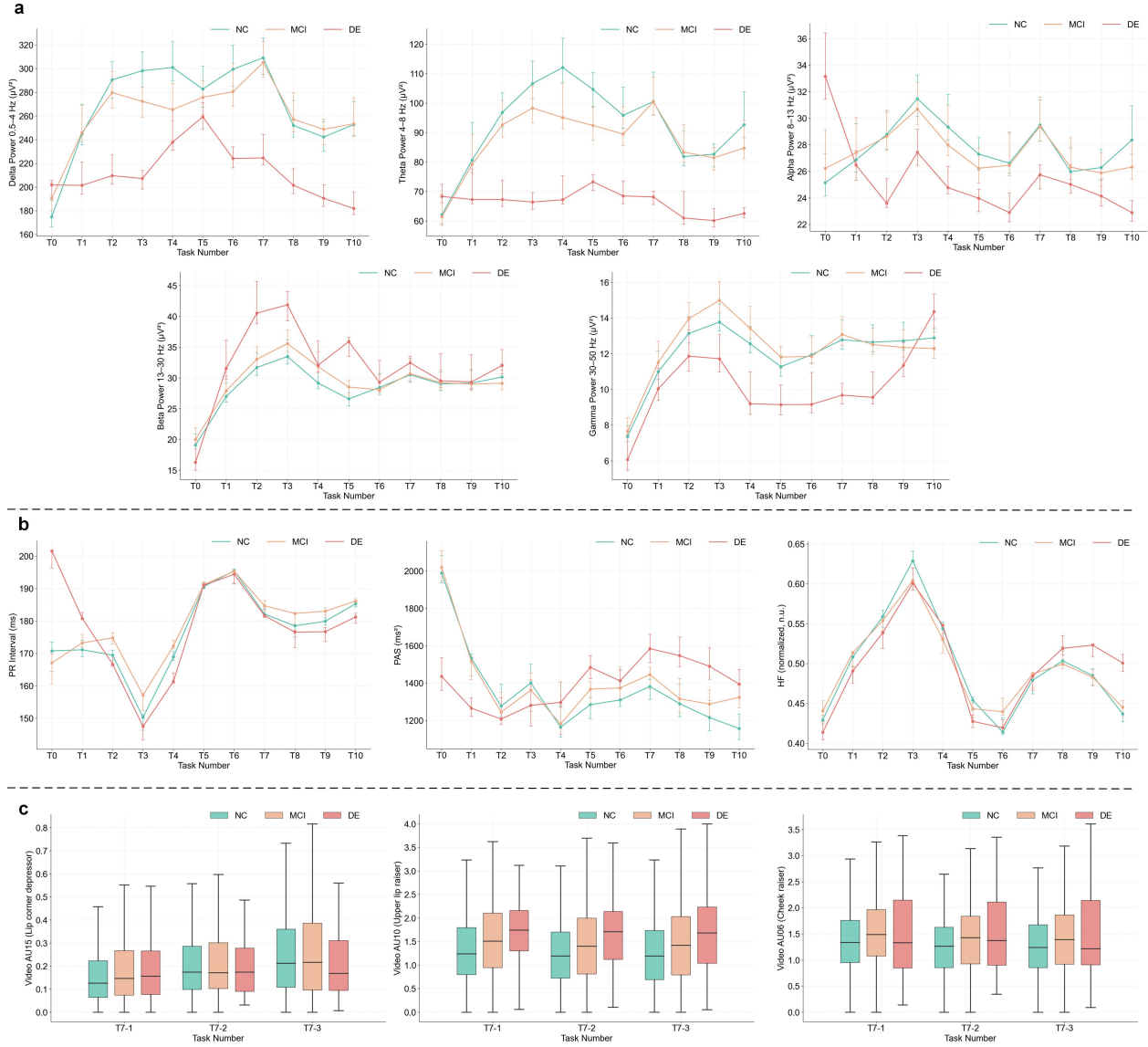
Prevalence	Accuracy (%)	AUROC (%)	AUPRC (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
61% (original)	83.0 (79.5–86.1)	90.0 (87.3–92.7)	91.6 (88.4–94.3)	84.0 (79.5–87.9)	81.6 (75.8–86.5)	86.8 (82.5–90.4)	78.0 (72.0–83.2)
35%	79.9 (75.2–84.1)	88.9 (85.5–92.4)	79.5 (72.1–86.5)	86.6 (78.9–92.3)	76.5 (70.3–82.0)	65.5 (57.3–73.2)	91.7 (86.7–95.3)
25%	79.8 (74.7–84.3)	90.9 (87.5–94.3)	75.9 (66.5–84.5)	90.0 (80.5–95.9)	76.5 (70.3–82.0)	55.3 (45.7–64.6)	96.0 (91.8–98.4)
15%	84.1 (78.9–88.4)	93.1 (89.5–96.7)	71.6 (56.5–83.8)	92.3 (79.1–98.4)	82.5 (76.8–87.4)	49.3 (37.4–61.3)	98.3 (95.2–99.7)



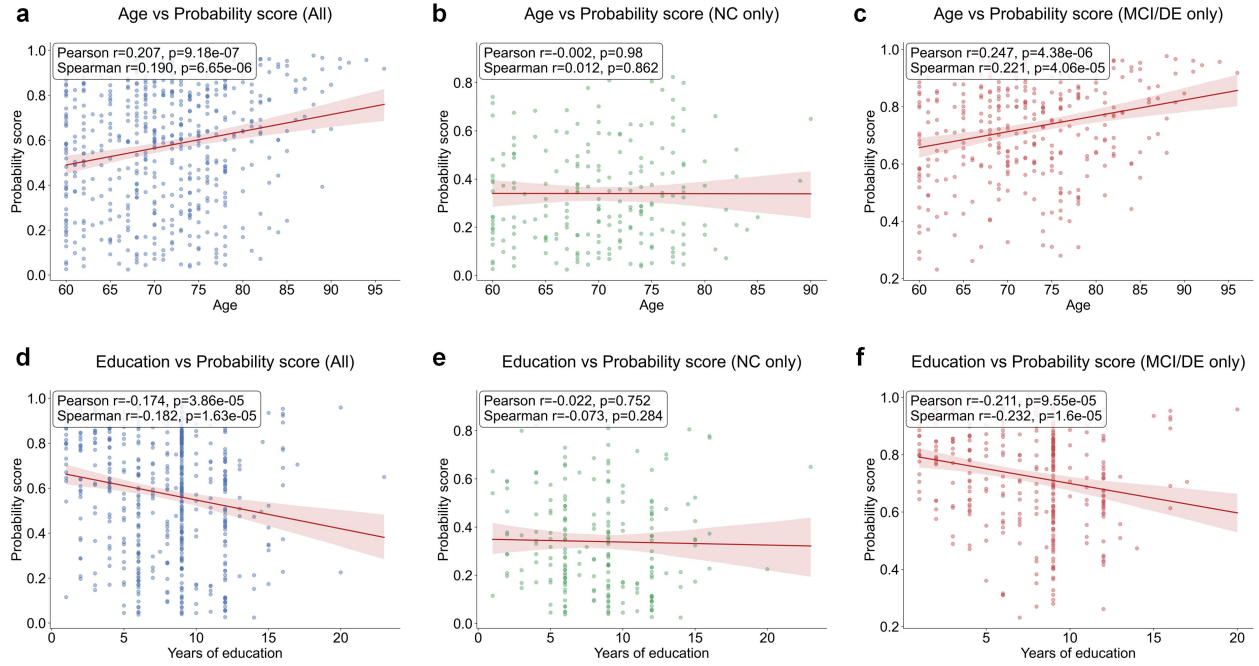
Supplementary Fig. 20: External validation performance for distinguishing NC from MCI under different prevalence settings. Model performance was evaluated under the original prevalence distribution and adjusted prevalence settings of 35%, 25%, and 15%.



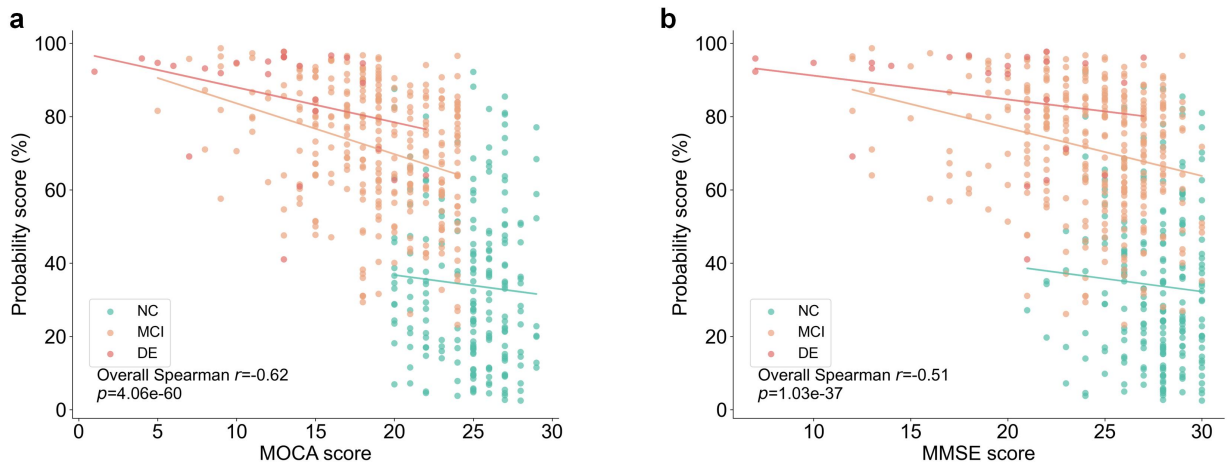
Supplementary Fig. 21: Distribution of multimodal features across MoCA subdomains. **a–g**, Box plots showing the distribution of key multimodal physiological and behavioral features across score levels in the seven MoCA subdomains: visuospatial and executive (a), naming (b), attention (c), language (d), abstract (e), delayed recall (f), and orientation (g). The figure highlights the effectiveness and consistency of multimodal features in representing cognitive function across domains.



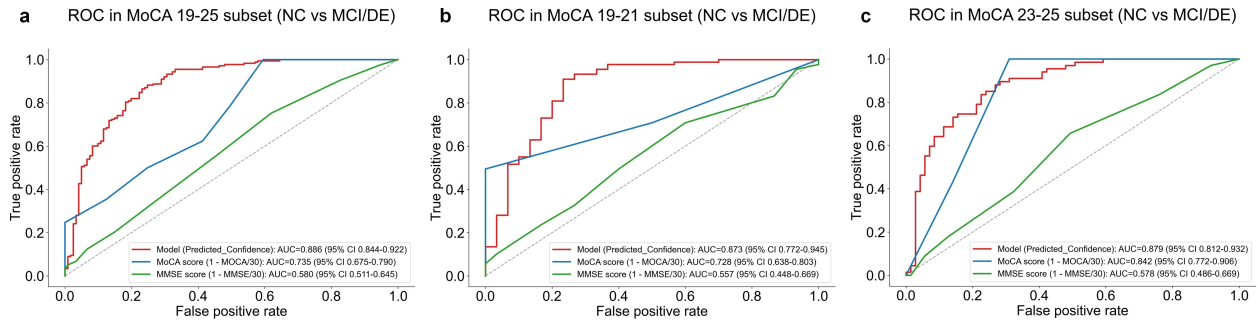
Supplementary Fig. 22: Dynamic trajectories of multimodal signals across cognitive tasks. **a**, EEG signals showing prefrontal band power modulation: Individuals with normal cognition (NC) and mild cognitive impairment (MCI) showed task-related increases in δ/θ and β/γ activity with preserved α reactivity, whereas those with dementia (DE) lacked low-frequency modulation, showed reduced γ , and exhibited abnormal β elevations. **b**, Heart rate variability (HRV) indices reflected autonomic differences across tasks, with normalized high-frequency power (HF_n) and phase space area (PAS) showing the most pronounced group effects. DE individuals exhibited abnormally elevated HF_n during recovery and a marked reduction in overall variability. **c**, Video-based facial signals with boxplots illustrating task-related changes in emotional activation features, highlighting attenuated modulation in MCI and abnormal or inconsistent responses in DE. Together, these task-related trajectories demonstrate alterations in neural oscillatory regulation, autonomic dynamics, and facial expression processing across cognitive states, providing dynamic multimodal markers for cognitive impairment assessment.



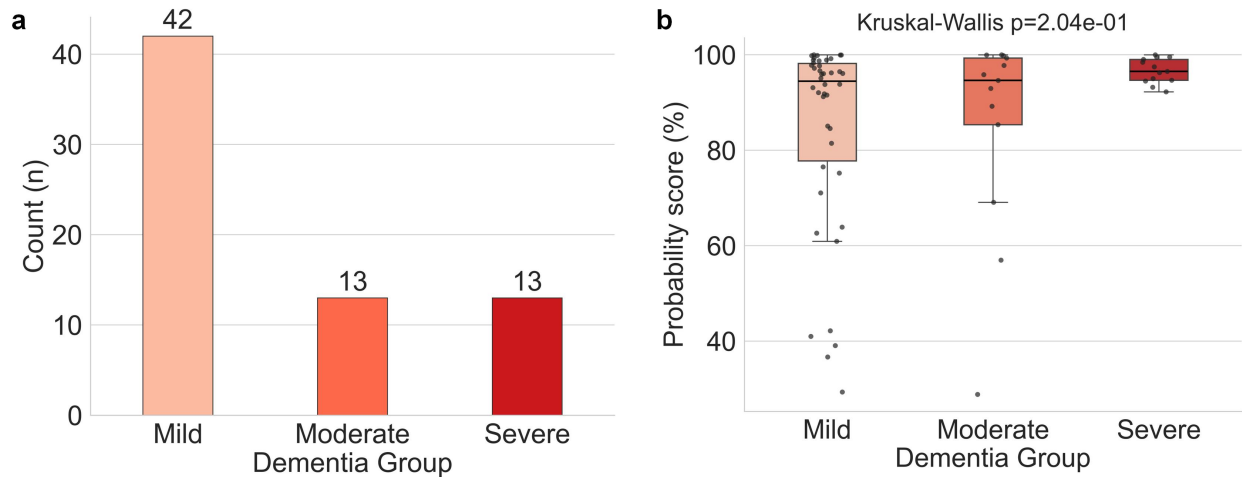
Supplementary Fig. 23: Correlation between demographic characteristics and model probability scores. **a–c**, Scatter plots showing the association between age and model-predicted probability scores in all participants (**a**), cognitively normal (NC) participants only (**b**), and participants with mild cognitive impairment or dementia (MCI/DE) only (**c**). **d–f**, Scatter plots showing the association between years of education and model-predicted probability scores in all participants (**d**), NC participants only (**e**), and MCI/DE participants only (**f**). Pearson and Spearman correlation coefficients with corresponding P values are reported in each panel. The regression lines and shaded areas indicate the fitted linear trends and 95% confidence intervals, respectively.



Supplementary Fig. 24: Association between model probability scores and conventional cognitive assessment scores. **a**, Scatter plot showing the relationship between model-predicted probability scores and MoCA scores across cognitively normal (NC), mild cognitive impairment (MCI), and dementia (DE) participants. **b**, Scatter plot showing the relationship between model-predicted probability scores and MMSE scores across the three assessment groups. Fitted trend lines are shown for each group, and the overall Spearman correlation coefficients with corresponding P values are reported in each panel.



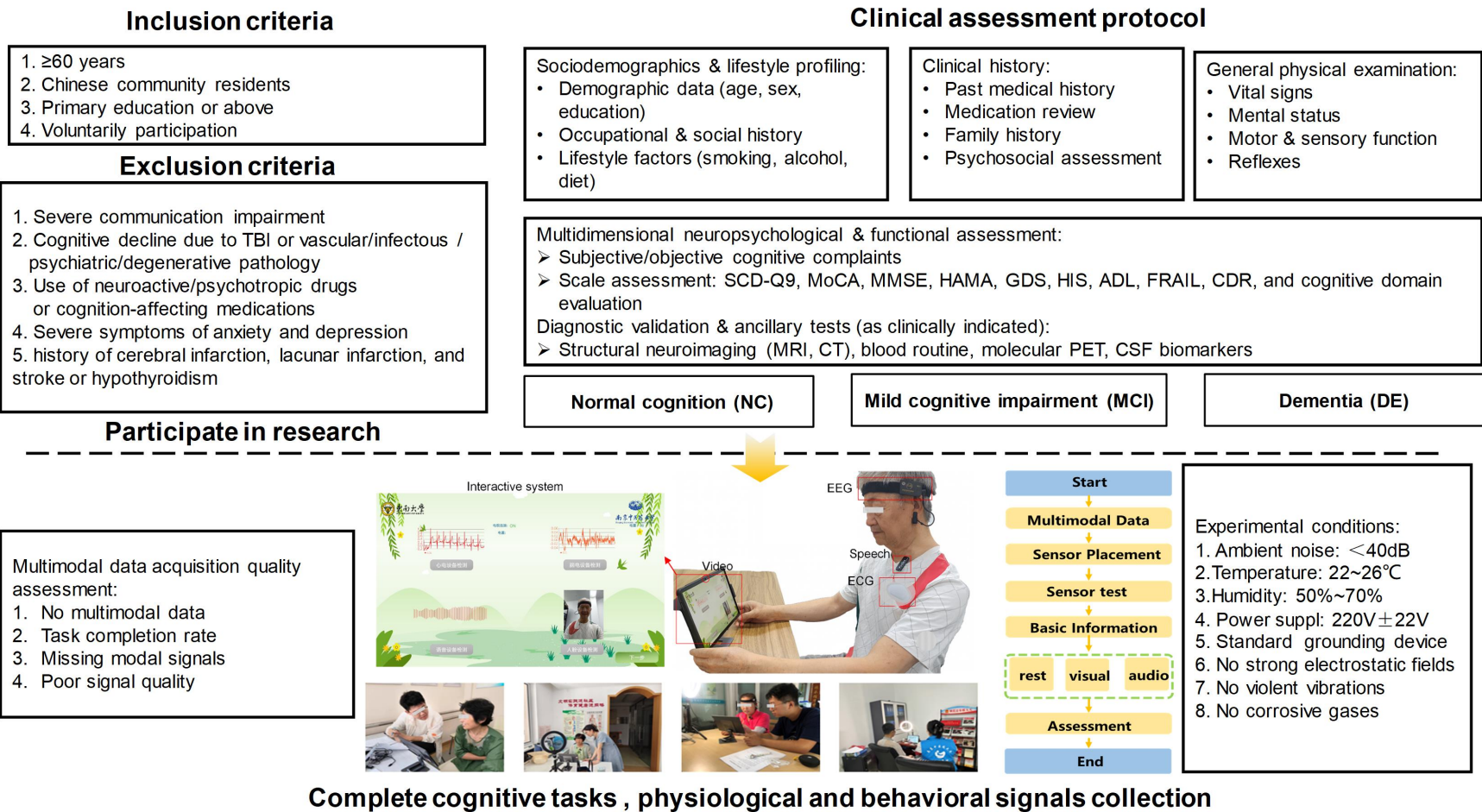
Supplementary Fig. 25: Model discrimination performance in MoCA-defined borderline cognitive ranges. **a**, Receiver operating characteristic (ROC) curves comparing the proposed model with MoCA- and MMSE-derived scores for distinguishing NC from MCI/DE participants in the MoCA 19–25 subset. **b**, ROC curves in the lower borderline MoCA 19–21 subset. **c**, ROC curves in the higher borderline MoCA 23–25 subset. Area under the ROC curve (AUC) values and 95% confidence intervals are reported in each panel.



Supplementary Fig. 26: Model probability scores across dementia severity groups. **a**, Distribution of participants across mild, moderate, and severe dementia groups. **b**, Boxplots showing model-predicted probability scores among different dementia severity groups. Individual dots represent participants, and the Kruskal–Wallis test was used to assess differences across severity groups.

II. Extended Methods

Study Population and Experimental Protocol



Supplementary Fig. 27: Schematic of recruitment timeline, clinical assessment, devices, and experimental environment.

Recruitment Process and Assessment Criteria

This study recruited eligible older adults from the community using a systematic screening procedure to ensure representativeness, safety, and compliance. Participants were recruited through community announcements and referrals from medical institutions. After an initial interview, they underwent detailed screening.

Inclusion criteria: (i) age ≥ 60 years; (ii) basic literacy and communication ability; and (iii) voluntary participation with signed informed consent.

Exclusion criteria: Participants were excluded if they met any of the following criteria: (1) cerebrovascular diseases, including cerebral embolism, cerebral thrombosis, cerebral infarction, moyamoya disease, or stroke; (2) neurological disorders, including Parkinson’s disease, brain tumors, epilepsy, or traumatic brain injury; (3) psychiatric disorders, including depression, schizophrenia, or chronic alcoholic psychosis (HAMA ≥ 29 , GDS-15 ≥ 12); (4) systemic diseases that may cause cognitive impairment, including hypothyroidism, vitamin B12 or folate deficiency, neurosyphilis, HIV infection, severe anemia, hepatic encephalopathy, or carbon monoxide poisoning; (5) use of any medications that may affect cognitive function, including antipsychotics, benzodiazepines, anticholinergics, antihistamines, theophyllines, digitalis, hypoglycemic agents, or chemotherapeutic drugs.

All participants provided written informed consent under a protocol approved by the Ethics Committee of Nanjing Hospital of Traditional Chinese Medicine. During screening, a brief neuropsychological scale was used for preliminary cognitive assessment to confirm participants’ ability to complete the study.

Clinical assessment workflow for cognitive impairment

1. Clinical assessment Criteria and Comprehensive Assessment System

Participants’ assessment classification (NC, MCI, DE) was based on a comprehensive clinical assessment, consisting of clinical interviews, informant reports, and standardized cognitive evaluations. The assessment criteria strictly followed the authoritative guidelines and clinical protocols listed below:

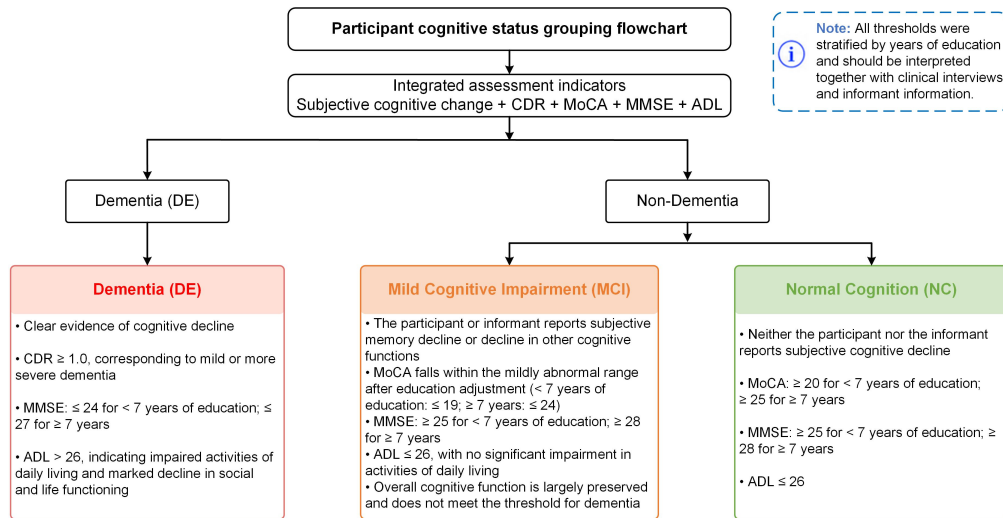
- Clinical diagnostic guidelines from the National Institute on Aging–Alzheimer’s Association (NIA-AA), including criteria for Alzheimer’s disease dementia and mild cognitive impairment due to Alzheimer’s disease [3, 4];
- The 2018 Chinese Guidelines for the Diagnosis and Treatment of Dementia and Cognitive Impairment, including diagnostic criteria for dementia and mild cognitive impairment [5, 6];
- The institutional clinical protocol of the Cognitive Impairment Diagnosis and Treatment Center at Xuanwu Hospital, Capital Medical University, which was implemented with reference to national dementia guidelines and established memory-clinic practices [7, 8].

2. Exact Cut-off Scores for Neurocognitive Tests

All participants completed a unified clinical assessment on the day of enrollment, including CDR, MoCA (Beijing version), MMSE, ADL, HAMA, and GDS-15. Considering the systematic impact of education level on cognitive scale scores, all cut-off values were stratified and adjusted based on years of education (< 7 years and ≥ 7 years). The specific cut-off criteria for each group are as follows [9]:

- **Normal Cognition (NC):** Neither the participant nor the informant reported subjective cognitive decline; MoCA (≥ 20 for < 7 years of education; ≥ 25 for ≥ 7 years); MMSE (≥ 25 for < 7 years; ≥ 28 for ≥ 7 years); ADL ≤ 26 .
- **Mild Cognitive Impairment (MCI):** The participant or informant reported subjective memory or other cognitive decline; MoCA scores were within the education-adjusted mildly abnormal range (≤ 19 for < 7 years; ≤ 24 for ≥ 7 years); overall cognitive function was largely preserved, not meeting the threshold for dementia (MMSE: ≥ 25 for < 7 years of education; ≥ 28 for ≥ 7 years); ADL ≤ 26 , indicating that activities of daily living were not significantly affected.
- **Dementia (DE):** Clear evidence of cognitive decline; CDR ≥ 1.0 , corresponding to mild or more severe dementia symptoms; MMSE scores were within the education-adjusted significantly abnormal range (MMSE: ≤ 24 for < 7 years; ≤ 27 for ≥ 7 years); ADL > 26 , indicating impaired activities of daily living and a significant decline in social and life functions.

The specific adjudication workflow is illustrated in Figure 28.



Supplementary Fig. 28: Education-adjusted cut-off criteria for clinical cognitive impairment assessment

3. Consensus Adjudication Process

To eliminate measurement bias potentially caused by single-scale cut-offs, this study employed a consensus assessment framework for participant grouping. This framework

integrated quantitative cut-offs with clinical consensus, and the overall decision-making process was divided into the following two tiers:

- **Tier 1: Evidence determination based on cut-off scores.** Education-adjusted MoCA and MMSE scores, ADL and CDR scores, participant- or informant-reported cognitive changes, and dementia-staging information were converted into structured clinical evidence according to predefined criteria (detailed in Figure 28). All cognitive scale cut-offs were stratified by years of education (< 7 years and ≥ 7 years) to mitigate the systematic effect of educational attainment. On this basis, the evidence was categorized into three roles: participant- or informant-reported cognitive change served as *subjective clinical evidence*; MoCA and MMSE served as *objective cognitive evidence*, providing a quantitative depiction of cognitive performance; and ADL served as *functional evidence* to distinguish cognitive impairment with preserved daily functioning from dementia-level functional impairment. CDR was additionally considered when dementia-level impairment was suspected.
- **Tier 2: Consensus determination based on evidence combinations.** Participants were first assigned to a preliminary cognitive status group according to the predefined evidence patterns for DE, MCI, and NC. Rather than relying on a single scale or a simple majority-voting rule, the preliminary classification was based on the consistency among subjective cognitive changes, education-adjusted cognitive test performance, functional status, and dementia-staging information when applicable.

All preliminary grouping results underwent independent blinded review by two senior neurologists with extensive clinical experience. During this process, the neurologists comprehensively considered clinical history, informant reports, education-adjusted MoCA and MMSE scores, ADL scores, and CDR information when dementia-level impairment was suspected. The final label was established when both physicians reached agreement.

When a participant showed overlapping evidence across groups, when subjective complaints, objective cognitive scores, and functional status were inconsistent, or when the two physicians disagreed, additional clinical evidence was further reviewed. This included rechecking the original scale records, conducting supplementary clinical interviews with the participant or informant when necessary, and referring to available medical imaging, routine laboratory findings, or other relevant clinical records. In such cases, a third senior neurologist at the director level was introduced, and a consensus meeting was convened to comprehensively review the participant's multidimensional assessment data until a final adjudication was reached. Boundary cases for which no consensus classification could be reached were excluded from the main analysis cohort.

Multimodal experimental protocol

The system consisted of a dual-channel EEG acquisition device, a single-lead ECG sensor, a wireless lapel microphone (MOMA S1), and a Lenovo Xiaoxin Plus tablet (Lenovo TB-J607F) for synchronized multimodal data collection and presentation. The tablet’s built-in camera was used to capture facial expressions. Detailed information can be found in the accompanying test report (Test Report of Physiological Signal Acquisition Device).

The experiments were conducted in a quiet, temperature-controlled indoor environment with standard desks and chairs. The participant’s face was kept at a distance of approximately 30 cm from the screen to ensure signal quality. Environmental control conditions were:

- Noise: < 40 dB
- Temperature: $25\text{--}26^{\circ}\text{C}$
- Relative humidity: $\leq 70\%$
- Atmospheric pressure: $86\text{--}106$ kPa
- Power supply: $220\text{ V} \pm 22\text{ V}$ with proper grounding

All devices were calibrated prior to the study and complied with electrophysiological data acquisition standards to ensure safety and data quality.

The research protocol was approved by the Ethics Committee of Nanjing Hospital of Chinese Medicine (approval number: KY2022004). Data collection followed a standardized procedure and began after written informed consent was obtained.

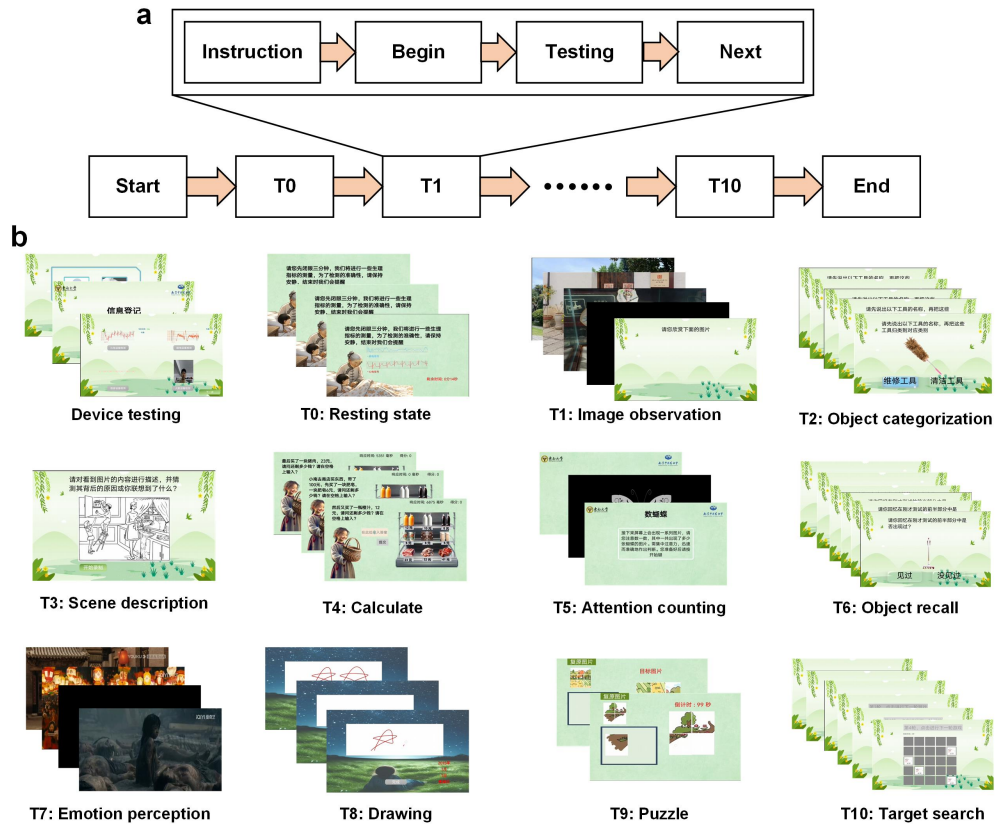
The entire experiment lasted approximately 60–90 minutes and included the following stages:

1. **Information collection and preparation:** Demographic data, medical history, and lifestyle information were collected. Participants completed standardized scales and were introduced to the procedure and equipment to reduce anxiety.
2. **Experiment initiation and signal acquisition:** Participants were guided to correctly wear EEG, ECG, and speech devices. Using a tablet-based application, they completed tasks including resting-state recording, cognitive tasks, image/video tasks, and speech tasks, with multimodal data automatically recorded.
3. **Completion and report generation:** Devices were removed, post-experiment heart rate was measured, and a preliminary assessment report was automatically generated.

Throughout the session, participants remained seated and followed task instructions. The system was designed to ensure reproducibility of the procedure and completeness of the collected data.

Task Design and Mapping to Cognitive Functions

This study developed a structured multi-task cognitive assessment paradigm that encompasses key multidimensional indicators commonly used in the evaluation of cognitive impairment. The tasks were designed based on standardized procedures and implemented through interactive tablet-based interfaces, enabling instruction-driven cognitive testing. All tasks were conducted in a controlled environment, during which synchronized multimodal physiological and behavioral signals (including EEG, ECG, speech, and video) were recorded to continuously monitor cognitive responses. This design allowed for the analysis of multimodal features elicited under different cognitive task stimuli. Each task was standardized in both content and operation to ensure reproducibility and comparability across participants. The full assessment lasted approximately 15–20 minutes with a fixed task order. Details of the paradigm are provided in Table 10.



Supplementary Fig. 29: Task description of multidimensional cognitive assessment. **a**, Workflow of the multidimensional cognitive assessment. **b**, Interactive interfaces for specific cognitive tasks.

The cognitive tasks were designed based on the seven core domains assessed in the Montreal Cognitive Assessment (MoCA): visuospatial and executive function, naming, attention, language, abstract reasoning, delayed recall, and orientation. A total of 11 tasks (T0–T10) were included to comprehensively cover these domains through a multidimensional task structure. Detailed information can be found in the accompanying test report (Software Test Report of the Cognitive Assessment System Based on Wearable Multimodal Signals).

During the assessment, multimodal physiological and behavioral signals (EEG, ECG, speech, and video) were synchronously recorded, providing a basis for modeling multidimensional cognitive states. The mapping between the task paradigm and the MoCA cognitive domains is summarized in Table 11.

Supplementary Table 10: Overview of cognitive tasks (T0–T10)

Task Number	Task	Description
T0	Resting state	Record baseline physiological signals (EEG, ECG) during a 180-second eyes-closed resting state [10, 11].
T1	Image observation	Observe static images containing everyday scenes or objects to engage basic visuospatial processing [12, 13].
T2	Object category	Categorize visual objects based on shared features such as color, shape, or function, assessing abstract reasoning and naming ability [14, 15].
T3	Scene description	Verbally describe the content and elements of a displayed scene, targeting expressive language and semantic fluency [16, 17].
T4	Calculate	Perform mental arithmetic or estimation tasks based on object prices, evaluating numerical reasoning and calculation [18, 19].
T5	Attention counting	Identify and count target stimuli (e.g., butterflies) among distractors, measuring sustained and selective attention [20, 21].
T6	Object recall	Recall specific details about previously viewed items or scenes to assess short-term episodic memory [18, 22].
T7	Emotion perception	Watch emotional video clips and infer the underlying emotional content, probing affective comprehension and social cognition [12, 23].
T8	Drawing	Draw geometric patterns (e.g., a pentagram) and mark the current date, evaluating executive control and temporal orientation [24, 25].
T9	Puzzle	Solve visual puzzles by arranging fragmented images or completing partially presented figures, focusing on visuospatial integration [12, 26].
T10	Target search	Search for specific targets (e.g., sheep) within a complex visual scene to assess spatial memory and search strategy [27, 28].

Supplementary Table 11: Mapping of tasks to MoCA cognitive domains

Cognitive Domain	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
VEF (0–5)			✓		✓			✓	✓	✓	
Naming (0–4)			✓	✓				✓			
Attention (0–6)		✓			✓	✓		✓			✓
Language (0–3)			✓	✓							
Abstraction (0–2)			✓		✓					✓	
DR (0–5)			✓			✓			✓		
Orientation (0–6)			✓		✓			✓		✓	

Supplementary Table 12: Duration of each cognitive task in the assessment protocol

Task	Mean duration (min)	Standard deviation (min)
T0, Resting state	3.00	0.00
T1, Image observation	2.46	0.34
T2, Object categorization	1.30	0.64
T3, Scene description	1.35	0.62
T4, Calculation	1.15	0.53
T5, Attention counting	3.08	0.25
T6, Object recall	1.50	0.59
T7, Emotion perception	2.49	0.57
T8, Drawing	0.90	0.44
T9, Puzzle	1.29	0.50
T10, Target search	1.57	0.48
Total duration of all tasks	20.53	2.51

Preprocessing of Multimodal Signals and Feature Engineering

EEG Feature Categories and Relevance to Cognitive Impairment

EEG: Raw EEG signals were acquired at 250 Hz and band-pass filtered (0.5–50 Hz) to isolate relevant neural oscillations while removing low-frequency drift and high-frequency noise [29]. We employed a sliding-window statistical approach to dynamically assess signal quality. Artifact-contaminated segments (e.g., ocular or muscle artifacts) were identified and corrected using wavelet decomposition combined with wavelet quantile normalization (WQN) [30](<https://github.com/holcman-lab/wavelet-quantile-normalization>). The cleaned signals were reconstructed and downsampled to 100 Hz. The detailed preprocessing pipeline is summarized in Supplementary Table 13, and all preprocessing scripts have been released at <https://github.com/zmh56/M3-CIA-Feature-Engineering> to facilitate reproducibility. These preprocessed time-series served as the input for extracting nonlinear, spectral, and temporal features, as detailed in Supplementary Table 14.

Supplementary Table 13: EEG preprocessing and artifact mitigation procedure

Stage	Method	Description
Pre-filtering	Band-pass and notch filtering	EEG signals were filtered at 0.5–50 Hz with an additional 50 Hz notch filter to remove baseline drift, high-frequency noise, and power-line interference.
Artifact detection [31]	CNN–LSTM with sliding-window voting	Continuous wavelet transform was applied to each standardized window to obtain time–frequency energy maps. A CNN–LSTM model detected artifact-contaminated windows, and temporal voting was used to improve segmentation continuity.
Artifact reconstruction [30]	Wavelet Quantile Normalization	Detected artifact segments were reconstructed in the wavelet domain using the statistical distribution of adjacent clean segments from the same participant, thereby suppressing eye-blink, muscle, and motion-related artifacts while preserving physiological EEG rhythms.
Downsampling	Resampling to 100 Hz	Reconstructed signals were downsampled to 100 Hz to reduce computational cost while retaining the main temporal and spectral information.
Quality control	Rejection and missing-data marking	Trials with excessive artifact burden or persistent non-physiological amplitudes after reconstruction were marked as missing and processed using the missing-modality handling strategy.

Supplementary Table 14: EEG feature categories, definitions, and their relevance to cognitive impairment

Feature Category	Feature Name	Definition / Computation	Relevance to Cognitive Impairment	Refs.
Nonlinear complexity	Lyapunov Exponent	Measures system chaos and sensitivity to initial conditions	Reflects neural system flexibility and adaptability in cognitive processing	[32, 33]
	Higuchi Fractal Dim. (HFD)	Quantifies signal self-similarity and "jaggedness"	Indicates the complexity of brain functional geometry and damage severity	[34, 35]
	Lempel-Ziv Complexity (LZC)	Counts number of distinct substrings in binary sequence	Measures the richness and diversity of underlying neural patterns	[35]
Information entropy	Approximate Entropy (ApEn)	Quantifies time-series regularity (lower = more regular)	Captures the degree of neural signal irregularity associated with information processing	[36, 37]
	Sample Entropy (SampEn)	Robust ApEn variant, less dependent on data length	Indicates complexity of neural dynamics and robustness of information encoding	[37]
	Permutation Entropy (PE)	Complexity measure based on ordinal patterns	Characterizes temporal pattern diversity relevant to cognitive state transitions	[37]
	Fuzzy Entropy (FE)	Uses fuzzy membership functions, suited for nonstationary signals	Provides insight into tolerance of neural activity to uncertainty and noise	[33]
	Differential Entropy (DE)	Based on signal amplitude distribution (Gaussian assumption)	Represents variability in neural information content across cognitive states	[38]
Spectral power & ratios	δ Power (0.5–4 Hz)	Slow-wave activity	Reflects baseline neural regulation and large-scale cortical dynamics	[39]
	θ Power (4–8 Hz)	Related to relaxation/light sleep	Associated with attention modulation and working memory processes	[39]
	α Power (8–12 Hz)	Normal adult eyes-closed rhythm	Linked to cortical inhibition and functional network coordination	[39]
	β Power (12–30 Hz)	Active thinking and attention	Reflects engagement of attentional resources and task-related processing	[39]
	γ Power (>30 Hz)	Information processing and neural synchrony	Indicates fine-scale integration and higher-order cognitive functions	[39]
	Ratio1 (α/θ)	Ratio of α to θ power	Serves as an index of balance between inhibitory and attentional control systems	[39, 40]
	Ratio2 (β/θ)	Ratio of β to θ power	Highlights shifts in excitation–inhibition balance relevant to cognition	[40, 41]
Time-domain statistics	RMS Amplitude	Root Mean Square of signal amplitude	Characterizes overall signal power and amplitude variability related to neural activity	[29]
	Diff1/2/3 (Mean/Std)	First/second/third-order difference statistics	Capture higher-order signal dynamics linked to temporal processing capacity	[29]
	Skewness / Kurtosis	Measures asymmetry and tailedness of amplitude distribution	Detects deviations from Gaussianity often linked to pathological brain states	[35]
	Hjorth Parameters	Activity (var), Mobility (freq), Complexity (bandwidth)	Describes spectral properties via time-domain variances, sensitive to slowing	[35]
Waveform morphology	Peak Amplitude	Extreme voltage values in a window	Associated with transient neural synchrony magnitude	[35]
	Zero-crossing Rate (ZCR)	Rate of signal sign-changes per second	Proxy for dominant frequency shifts without full spectral analysis	[42]
	Line Length	Sum of absolute differences between consecutive points	Sensitive to amplitude and frequency changes, reflecting signal "busyness"	[43]

ECG Preprocessing and Feature Extraction

Supplementary Table 15: Expanded ECG and HRV feature categories, definitions, and their relevance to cognitive impairment

Feature Category	Feature Name	Definition / Computation	Relevance to Cognitive Impairment	Refs.
Morphological features	PR Interval (Mean/Std)	Interval from end of P wave to start of QRS	Reflects atrioventricular conduction linked to neurocardiac regulation	[41, 44]
	QRS Duration	Duration of QRS complex	Indicates ventricular conduction integrity with autonomic influence	[41, 44]
	QT/JT Interval	Ventricular repolarization duration	Provides markers of cardiac repolarization relevant to cognition	[41, 45]
	P/T Wave Amplitude	Amplitudes of atrial and ventricular repolarization	Represents cardiac electrical activity strength with systemic links	[46, 47]
	Signal Skewness	Index of waveform asymmetry	Reflects morphology changes related to neurocardiac coupling	[41]
	ST Segment Level	Displacement of ST segment from baseline	Sensitive to autonomic tone changes and physiological stress	[41]
Time-domain HRV	Mean RR	Average interval between normal heartbeats	Baseline metric of cardiac chronotropic control	[44, 48]
	SDNN	Std dev of NN intervals	Gold standard for overall HRV and autonomic adaptability	[41, 44]
	RMSSD	Root mean square of successive differences	Primary marker of vagally-mediated (parasympathetic) changes	[41, 44]
	pNN50	Percentage of successive RR differences > 50ms	Closely correlated with RMSSD, indicating rapid vagal changes	[41, 44]
	HRV TINN	Triangular interpolation of NN interval histogram	Captures geometric distribution of heart rate variability	[41]
	Triangular Index	Total NN intervals divided by histogram height	Robust measure of overall variability, less sensitive to artifacts	[45]
Frequency-domain HRV	VLF Power	Very Low Frequency power (0.0033–0.04 Hz)	Linked to thermoregulation, hormonal cycles, and slow mechanisms	[41]
	LF Power	Low Frequency power (0.04–0.15 Hz)	Reflects a mix of sympathetic and parasympathetic modulation	[41, 49]
	HF Power	High Frequency power (0.15–0.4 Hz)	Pure index of vagal (parasympathetic) modulation and respiration	[41, 49]
	LF/HF Ratio	Ratio of LF to HF power	Represents sympatho-vagal balance linked to stress and cognition	[41, 49]
	Total Power	Sum of energy in all frequency bands	Indicates the total variance and overall autonomic capacity	[41]
Nonlinear HRV	Poincaré SD1	Width of the Poincaré plot ellipse	Correlates with short-term HRV (parasympathetic activity)	[41]
	Poincaré SD2	Length of the Poincaré plot ellipse	Correlates with long-term HRV (sympathetic/global activity)	[41]
	HRV SampEn	Sample Entropy of RR intervals	Reflects heartbeat irregularity and physiological adaptability	[41]
	DFA α_1/α_2	Detrended Fluctuation Analysis scaling exponents	Describes fractal correlation properties of heart rate dynamics	[46, 47]
	Approximate Entropy	Measure of regularity similar to SampEn	Indicates the predictability of heart rate fluctuations	[41, 46]
Cardiorespiratory coupling	RSA Amplitude	Respiratory Sinus Arrhythmia magnitude	Direct measure of vagal tone and breathing-heart rate synch	[50]
	EDR	ECG-derived Respiration signal characteristics	Surrogate for respiratory effort when direct breath signal is absent	[50]
	Phase Synchronization	Phase locking value between ECG R-peaks and respiration	Quantifies the temporal coordination between cardiac and respiratory systems	[50]

ECG: Raw ECG signals were band-pass filtered (0.05–100 Hz) to suppress baseline wander and muscle artifacts [51]. A notch filter at 50 Hz was subsequently applied to eliminate power-line interference. R-peaks were detected using the Pan-Tompkins algorithm, followed by an adaptive refractory period check to minimize false positives. We applied a logical filter to identify ectopic beats or outliers (defined as RR intervals < 300 ms or > 1300 ms), which were corrected via cubic spline interpolation. This procedure yielded a clean, continuous RR interval time series (tachogram) for the computation of heart rate variability (HRV) metrics. Detailed definitions of the extracted ECG/HRV morphological and nonlinear features are provided in Supplementary Table 15.

Speech Signal Preprocessing and Feature Extraction

Supplementary Table 16: Speech feature categories, definitions, and their relevance to cognitive impairment

Category	Feature	Definition / Computation	Relevance to Cognitive Impairment	Refs.
Spectral features	MFCC (1-13)	Mel-frequency cepstral coefficients computed via Fourier transform and mel-scale filtering	Capture vocal tract spectral characteristics, providing markers of altered phonation patterns	[52]
	Formants (F1-F3)	Spectral peaks of vocal tract resonance calculated through LPC analysis	Indicate articulatory precision and resonance patterns associated with language production	[53]
	Spectral Centroid	The "center of mass" of the spectrum (brightness)	Reflects high-frequency energy loss often linked to articulatory slurring	[52]
	Spectral Flux	Rate of change of the power spectrum between frames	Indicates the dynamic changes in spectral shape, related to speech motor planning	[52]
Prosodic features	F0 (Mean/Std)	Fundamental frequency statistics derived from vocal fold vibration	Reflects vocal control, intonation flatness, and affective modulation	[53]
	Intensity	Root Mean Square (RMS) amplitude and its dynamic range	Indicates ability to modulate loudness, often reduced in apathy or depression	[54]
	Pitch Contour	Trajectory of F0 over time (e.g., rising/falling slopes)	Captures expressive prosody, which diminishes in cognitively impaired speech	[53]
Fluency & Temporal	Speech Rate	Syllables or words per second calculated excluding pause intervals	Represents efficiency of verbal output and psychomotor speed	[55]
	Pause Duration	Average length and frequency of silent intervals (>200ms) during speech	Reflects language planning delays, lexical retrieval deficits, and cognitive load	[55]
	Phonation Ratio	Ratio of active speech time to total recording duration	Indicates the density of speech content and sustained attention capacity	[56]
	Articulation Rate	Phoneme production rate measured during active speech periods	Reflects pure motor execution capacity independent of cognitive pausing	[54]
	Hesitation Ratio	Frequency of filled pauses (uh, um) relative to total words	Directly linked to word-finding difficulties (anomia) and executive dysfunction	[56]
Voice Quality	Jitter	Cycle-to-cycle variation in fundamental frequency (pitch perturbation)	Sensitive to lack of control over vocal fold vibration (neuromotor instability)	[54]
	Shimmer	Cycle-to-cycle variation in amplitude (loudness perturbation)	Reflects reduced glottal closure and breathiness potentially linked to vagal tone	[54]
	HNR	Harmonics-to-Noise Ratio (periodic vs. aperiodic energy)	Quantifies hoarseness or breathiness, indicating degradation in voice purity	[54, 55]
Semantic features	BERT Emb.	Contextual semantic vectors from transformer-based language models	Capture deep coherence and contextual relevance of spoken narratives	[57]
	Sem. Similarity	Cosine distance between speech content and target references	Indicates alignment of expression with target meaning, reflecting comprehension	[53]
	Sem. Coverage	Proportion of key narrative elements (Information Units) expressed	Reflects completeness of information encoding and memory retrieval	[58]
	Type-Token Ratio	Ratio of unique words (types) to total words (tokens)	Classic measure of lexical diversity; low TTR indicates repetitive speech	[58]
	Syntax Depth	Depth of dependency parse trees or clause length	Indicates the ability to construct complex grammatical structures	[58]

Speech: Audio recordings were first denoised using spectral subtraction and resampled to a standardized 16 kHz mono format. A pre-emphasis filter (coefficient $\alpha = 0.97$) was applied to flatten the spectral tilt and enhance high-frequency resolution for acoustic analysis. The signals were then segmented into 25 ms frames with a 10 ms step size for short-time feature computation using the openSMILE toolkit [59]. For semantic analysis, verbatim transcripts were generated via an automatic speech recognition system and tokenized for embedding extraction. A comprehensive list of acoustic, prosodic, and semantic features derived from this pipeline is presented in Supplementary Table 16.

Video Signal Preprocessing and Feature Extraction

Supplementary Table 17: Expanded Video feature categories, definitions, and their relevance to cognitive impairment

Category	Feature	Definition / Computation	Relevance to Cognitive Impairment	Refs.
Facial muscle actions	AU01 (Inner Brow)	Raising of the inner eyebrow (Frontalis muscle)	Linked to sadness, anxiety, and compensatory expressions during recall	[60, 61]
	AU04 (Brow Lowerer)	Lowering and contraction of eyebrows (Corrugator)	Primary indicator of cognitive load, confusion, and concentration effort	[60, 62]
	AU06 (Cheek Raiser)	Raising of cheeks (Orbicularis oculi)	Marker of genuine (Duchenne) smiles, often reduced in apathy	[63]
	AU12 (Lip Corner)	Pulling of lip corners upwards (Zygomaticus major)	Represents social responsiveness and positive affect signaling	[62, 64]
	AU15 (Lip Depressor)	Pulling of lip corners downwards	Associated with sadness and negative affect regulation	[65]
	AU10/17 (Mouth)	Upper lip raiser / Chin raiser	Associated with expressive modulation and tension in the mouth region	[66]
Head motor dynamics	Head Pose (Rx,Ry,Rz)	3D orientation angles (Pitch, Yaw, Roll) of the head	Reflects attention orientation and engagement with the interlocutor	[62, 67]
	Head Velocity	Rate of change in head translation and rotation	Indicates psychomotor agitation (high) or rigidity/hypomimia (low)	[67]
	Head Jerk	Rate of change of head acceleration	Captures smooth vs. spasmodic motor control anomalies	[67]
	Nodding Frequency	Frequency of vertical head oscillations	Reflects non-verbal communication and back-channeling behavior	[67]
Eye movement patterns	Gaze Angle (X, Y)	Horizontal and vertical direction of gaze vector	Reflects visual attentional control, distractibility, and eye contact	[67, 68]
	Blink Rate	Frequency of eye blinks per minute (AU45)	Associated with dopaminergic activity and cognitive fatigue/attention	[69]
	Fixation Duration	Average time gaze remains stable on a target	Indicator of information processing speed and focus stability	[69]
	AU07 (Lid Tightener)	Tightening of the eyelids (squinting)	Provides an indicator of tension, skepticism, or visual focusing	[69]
Facial kinematics & geometry	AUP (Frequency)	Activation probability of AUs over time	Temporal density of facial muscle engagement during tasks	[60, 62]
	AUI (Intensity)	Average regression value (0-5) of active AUs	Magnitude of facial expressiveness; reduced in "flat affect"	[60, 62]
	Expressive Entropy	Shannon entropy of AU intensity distributions	Quantifies the diversity and richness of facial expressions	[60]
	Landmark Velocity	Speed of facial landmark displacement (e.g., mouth)	Detects facial bradykinesia characteristic of neurodegeneration	[60]
	Asymmetry Index	Difference in movement between left/right face	Reflects lateralized motor control deficits or hemifacial weakness	[60]

Video: Raw video footage was processed using the OpenFace toolkit [67] for frame-by-frame behavioral analysis. The pipeline included face detection (filtering for the largest face to exclude bystanders), 68-point facial landmark localization, and affine transformation to align faces to a canonical frontal view. This normalization ensured that subsequent measurements were invariant to scale and in-plane rotation. Lighting variations were mitigated using histogram equalization on the aligned face patches. From these standardized sequences, we extracted facial action units (AUs), gaze vectors, and head pose dynamics. The full set of extracted video features and their relevance to cognitive impairment are listed in Supplementary Table 17.

Supplementary Table 18: Meta-features describing the temporal evolution of physiological and behavioral signals

Category	Metric Source	Definition / Computation
Directional Trend	Slope / Gradient	Linear regression slope of the feature curve over task stages
	Trend R^2	Goodness-of-fit of the linear trend
	Kendall's τ	Rank correlation coefficient with time
	First-Last Diff	Difference between the final and initial task segments
	Delta (Δ)	Step-wise change between adjacent task blocks
Trajectory Magnitude	Path Length	Sum of Euclidean distances between consecutive points
	Area	Area Under the feature Curve
	Mean Level	Average value across the entire trajectory
	Norm. Path	Path length normalized by duration or baseline
	Range	Difference between maximum and minimum values
Volatility & Scale	Std (Volatility)	Standard deviation of the feature sequence
	Curvature	Second derivative or deviation from linearity
	Skewness	Asymmetry of the feature distribution over time
Temporal Synchronization	DTW Distance	Dynamic Time Warping distance to a "healthy" template
	Cross-Corr	Correlation with task difficulty curve

Implementation details for missing-modality inference

In practical community-based deployment, incomplete multimodal inputs may occur because of sensor detachment, poor signal quality, or failure of a specific acquisition device. To improve the robustness of the proposed framework under such conditions, missing-modality inference was handled at three levels: feature preprocessing, model training, and multimodal fusion.

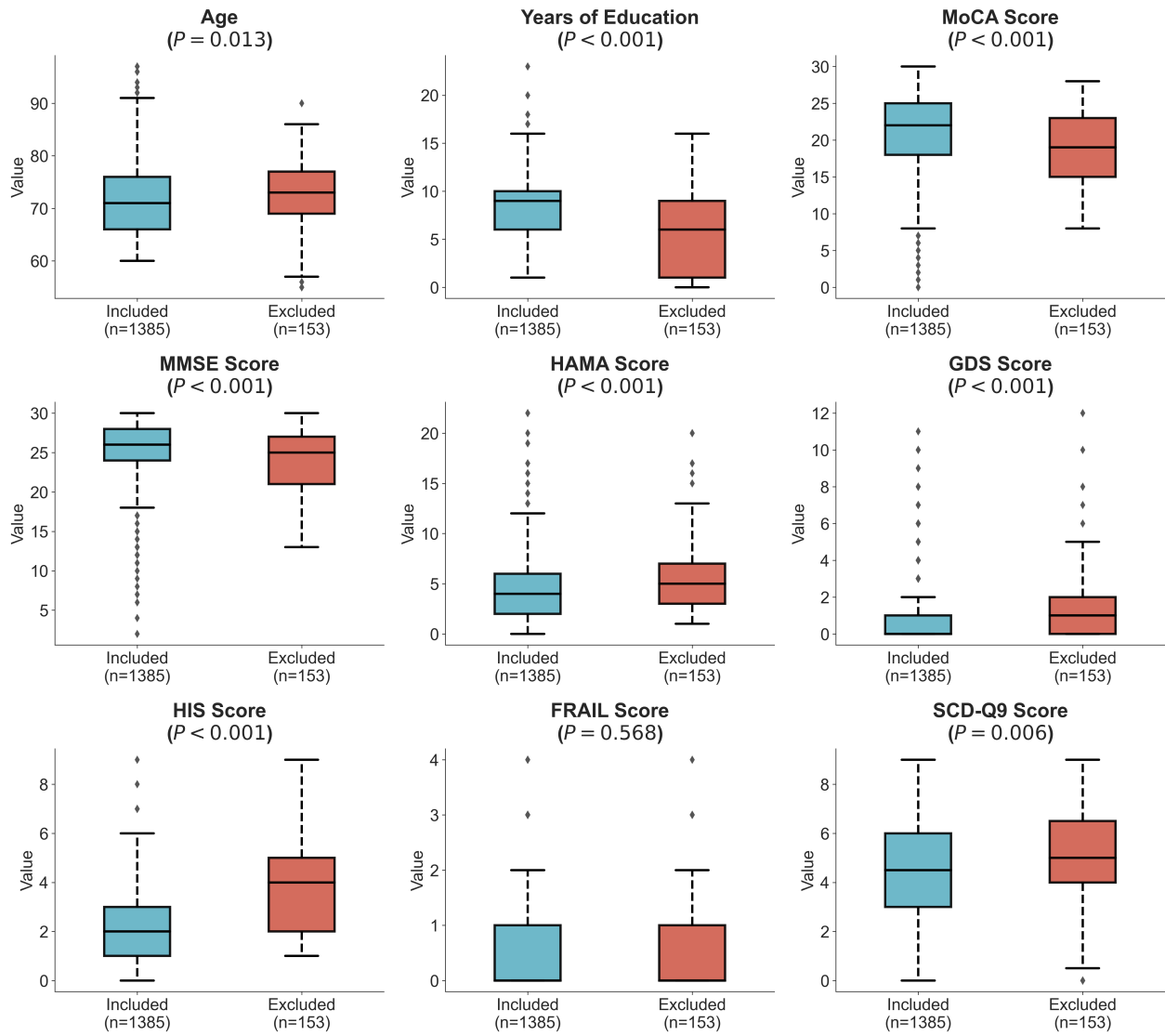
Feature-level structural placeholders. During inference, when task-specific features from a given modality were unavailable, median imputation was applied at the preprocessing stage. This step was used as a structural placeholder to preserve the fixed input dimensionality required by the network. Because the extracted features retained their original physical-value distributions before entering the model, median imputation reduced the risk of introducing extreme outliers. After entering the network, these imputed values were further processed by internal normalization layers, such as Layer Normalization or Batch Normalization, and were therefore mapped close to a neutral baseline in the learned feature space.

Training-level robustness with dropout. To reduce over-reliance on any single modality, dropout was introduced during model training. In the multimodal fusion network, dropout can be regarded as a regularization strategy that partially simulates the random loss of features, nodes, or modality-specific information. This encourages the model to use complementary information from the remaining modalities and improves its tolerance to partially missing inputs during inference.

Architecture-level dynamic routing. The dynamic task adapter further supported missing-modality inference by adaptively adjusting the contribution of different modalities. Inputs replaced by median values usually contain limited task-related variation and therefore provide less informative evidence for classification. Through the dynamic routing mechanism, the model can reduce the activation weight assigned to these uninformative inputs and rely more on available modalities with valid signal patterns. This design enables gradual performance degradation rather than abrupt failure when one or more modalities are missing.

Supplementary Table 19: Robustness analysis under modality-missing settings in the external validation cohort

EEG	ECG	Speech	Video	0%	20%	60%	100%
×	✓	✓	✓	0.905	0.889	0.879	0.864
✓	✓	×	✓	0.905	0.891	0.880	0.858
✓	✓	✓	×	0.905	0.890	0.869	0.868
✓	×	✓	✓	0.905	0.894	0.887	0.878
×	×	✓	✓	0.905	0.876	0.842	0.812
✓	✓	×	×	0.905	0.882	0.855	0.829



Supplementary Fig. 30: Baseline characteristics of included and excluded participants. Differences in demographic factors, cognitive scores, emotional symptoms, cerebrovascular burden, frailty, and subjective cognitive decline were compared between the final analytic cohort and excluded individuals with available clinical information.

References

- [1] Liew, T. M. *et al.* PENSIEVE-AI a brief cognitive test to detect cognitive impairment across diverse literacy. *Nature Communications* **16**, 2847 (2025).
- [2] DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated roc curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
- [3] McKhann, G. M. *et al.* The diagnosis of dementia due to alzheimer’s disease: Recommendations from the national institute on aging–alzheimer’s association workgroups on diagnostic guidelines for alzheimer’s disease. *Alzheimer’s & Dementia* **7**, 263–269 (2011).
- [4] Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to alzheimer’s disease: Recommendations from the national institute on aging–alzheimer’s association workgroups on diagnostic guidelines for alzheimer’s disease. *Alzheimer’s & Dementia* **7**, 270–279 (2011).
- [5] Writing Group of Chinese Guidelines for the Diagnosis and Treatment of Dementia and Cognitive Impairment. 2018 chinese guidelines for the diagnosis and treatment of dementia and cognitive impairment (i): Diagnostic criteria for dementia and its subtypes. *National Medical Journal of China* **98**, 965–970 (2018).
- [6] Writing Group of Chinese Guidelines for the Diagnosis and Treatment of Dementia and Cognitive Impairment, Du, Y. & Lyu, P. 2018 chinese guidelines for the diagnosis and treatment of dementia and cognitive impairment (v): Diagnosis and treatment of mild cognitive impairment. *National Medical Journal of China* **98**, 1294–1301 (2018).
- [7] Jia, J. *et al.* Diagnosis and treatment of dementia in neurology outpatient departments of general hospitals in china. *Alzheimer’s & Dementia* **12**, 446–453 (2016).
- [8] Jia, L. *et al.* Dementia in china: Epidemiology, clinical management, and research advances. *The Lancet Neurology* **19**, 81–92 (2020).
- [9] Jia, L. *et al.* Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. *The Lancet Public Health* **5**, e661–e671 (2020).
- [10] Jandackova, V. K., Scholes, S., Britton, A. & Steptoe, A. Midlife heart rate variability and cognitive decline: A large longitudinal cohort study. *International Journal of Clinical and Health Psychology* **24**, 100518 (2024).
- [11] Katayama, O. *et al.* Neurophysiological markers in community-dwelling older adults with mild cognitive impairment: an eeg study. *Alzheimer’s Research & Therapy* **15**, 217 (2023).

- [12] Harris, C. *et al.* Digital neuropsychology beyond computerized cognitive assessment: Applications of novel digital technologies. *Archives of Clinical Neuropsychology* **39**, 290–304 (2024).
- [13] Wang, Y. *et al.* Features of event-related potentials during retrieval of episodic memory in patients with mild cognitive impairment due to alzheimer’s disease. *Frontiers in Neuroscience* **17**, 1185228 (2023).
- [14] Mudar, R. A. *et al.* The effects of amnesic mild cognitive impairment on go/nogo semantic categorization task performance and event-related potentials. *Journal of Alzheimer’s Disease* **50**, 577–590 (2016).
- [15] Joubert, S., Gardy, L., Didic, M., Rouleau, I. & Barbeau, E. J. A meta-analysis of semantic memory in mild cognitive impairment. *Neuropsychology Review* **31**, 221–232 (2021).
- [16] Chou, C.-J. *et al.* Screening for early alzheimer’s disease: enhancing diagnosis with linguistic features and biomarkers. *Frontiers in Aging Neuroscience* **16**, 1451326 (2024).
- [17] Jafari, Z., Andrew, M. K. & Rockwood, K. J. Diagnostic utility of speech-based biomarkers in mild cognitive impairment: a systematic review and meta-analysis. *Age and Ageing* **54**, afaf316 (2025).
- [18] Huang, S. *et al.* Exploring the n-back task: insights, applications, and future directions. *Frontiers in Human Neuroscience* **19**, 1721330 (2025).
- [19] Benavides-Varela, S. *et al.* Anatomical substrates and neurocognitive predictors of daily numerical abilities in mild cognitive impairment. *Cortex* **71**, 58–67 (2015).
- [20] Nasiri, E. *et al.* A comprehensive review of attention tests: can we assess what we exactly do not understand? *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery* **59**, 26 (2023).
- [21] Mohamed, M., Mohamed, N. & Kim, J. G. P300 latency with memory performance: A promising biomarker for preclinical stages of alzheimer’s disease. *Biosensors* **14**, 616 (2024).
- [22] Manipa, V. *et al.* Resting-state eeg gamma power predicts immediate and delayed recall in healthy adults. *Cognitive Neurodynamics* **19**, 138 (2025).
- [23] Wang, C. *et al.* Visual event-related potentials under external emotional stimuli as early signs for mild cognitive impairment. *The Journal of Prevention of Alzheimer’s Disease* **11**, 1325–1338 (2024).

- [24] Yuan, J. *et al.* Association between the digital clock drawing test and neuropsychological test performance: Large community-based prospective cohort study. *Journal of Medical Internet Research* **23**, e27407 (2021).
- [25] Linari, I., Juantorena, G. E., Ibáñez, A., Petroni, A. & Kamienkowski, J. E. Unveiling trail making test: visual and manual trajectories indexing multiple executive processes. *Scientific Reports* **12**, 14265 (2022).
- [26] Che, J. *et al.* Executive function measures of participants with mild cognitive impairment: Systematic review and meta-analysis of event-related potential studies. *International Journal of Psychophysiology* **197**, 112295 (2024).
- [27] Eraslan Boz, H. *et al.* Visual search in alzheimer’s disease and amnestic mild cognitive impairment: An eye-tracking study. *Alzheimer’s & Dementia* **20**, 759–768 (2024).
- [28] Poos, J. M. *et al.* Short digital spatial memory test detects impairment in alzheimer’s disease and mild cognitive impairment. *Brain Sciences* **11**, 1350 (2021).
- [29] Song, T., Zheng, W., Song, P. & Cui, Z. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing* **11**, 532–541 (2018).
- [30] Dora, M. & Holcman, D. Adaptive single-channel EEG artifact removal with applications to clinical monitoring. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **30**, 286–295 (2022).
- [31] Zhao, M. *et al.* Multi-query cross-modal attention fusion for cognitive impairment recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **33**, 2520–2530 (2025).
- [32] Lee, K., Choi, K.-M., Park, S., Lee, S.-H. & Im, C.-H. Selection of the optimal channel configuration for implementing wearable EEG devices for the diagnosis of mild cognitive impairment. *Alzheimer’s Research & Therapy* **14**, 170 (2022).
- [33] Parbat, D. & Chakraborty, M. A novel methodology to study the cognitive load induced EEG complexity changes: Chaos, fractal and entropy based approach. *Biomedical Signal Processing and Control* **64**, 102277 (2021).
- [34] Higuchi, T. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena* **31**, 277–283 (1988).
- [35] Babiloni, C. *et al.* Abnormalities of cortical neural synchronization mechanisms in subjects with mild cognitive impairment due to Alzheimer’s and Parkinson’s diseases: An eeg study. *Journal of Alzheimer’s Disease* **59**, 339–358 (2017).

- [36] Pappalètera, C. *et al.* Approximate entropy analysis across electroencephalographic rhythmic frequency bands during physiological aging of human brain. *Geroscience* **45**, 1131–1145 (2023).
- [37] Cacciotti, A., Pappalètera, C., Miraglia, F., Rossini, P. M. & Vecchio, F. EEG entropy insights in the context of physiological aging and Alzheimer’s and Parkinson’s diseases: A comprehensive review. *GeroScience* **46**, 5537–5557 (2024).
- [38] Wang, L., Du, T., Zhao, L., Shi, Y. & Zeng, W. Research on the lateralization of brain functional complexity in mild cognitive impairment-Alzheimer’s disease progression based on multiscale lateralized brain entropy. *Biomedical Signal Processing and Control* **86**, 105216 (2023).
- [39] Güntekin, B., Emek-Savaş, D. D., Kurt, P., Yener, G. G. & Başar, E. Beta oscillatory responses in healthy subjects and subjects with mild cognitive impairment. *NeuroImage: Clinical* **3**, 39–46 (2013).
- [40] Azami, H. *et al.* Beta to theta power ratio in EEG periodic components as a potential biomarker in mild cognitive impairment and Alzheimer’s dementia. *Alzheimer’s Research & Therapy* **15**, 133 (2023).
- [41] Arco, J. E., Gallego-Molina, N. J., Ortiz, A., Arroyo-Alvis, K. & López-Pérez, P. J. Identifying HRV patterns in ECG signals as early markers of dementia. *Expert Systems with Applications* **243**, 122934 (2024).
- [42] Shenoy, R. R. & Seelamantula, C. S. Spectral zero-crossings: Localization properties and applications. *IEEE Transactions on Signal Processing* **63**, 3177–3190 (2015).
- [43] Koolen, N. *et al.* Line length as a robust method to detect high-activity events: Automated burst detection in premature EEG recordings. *Clinical Neurophysiology* **125**, 1985–1994 (2014).
- [44] Rykov, Y. G. *et al.* Predicting cognitive scores from wearable-based digital physiological features using machine learning: Data from a clinical trial in mild cognitive impairment. *BMC Medicine* **22**, 36 (2024).
- [45] Geng, D. *et al.* Effects of Alzheimer’s disease of varying severity on cardiac and autonomic function. *Brazilian Journal of Medical and Biological Research* **55**, e11504 (2022).
- [46] Mao, M. *et al.* Ventricular electrocardiographic signatures associated with dementia and plasma Alzheimer’s biomarkers in older adults: The mind-china study. *Alzheimer’s & Dementia* **19**, e060327 (2023).
- [47] Ahmad, M., Hughes, T. M., Kazibwe, R., Chen, L. Y. & Soliman, E. Z. Malignant left ventricular hypertrophy and risk of cognitive impairment or probable dementia: A post hoc analysis of the SPRINT trial. *Circulation* **148**, A12741–A12741 (2023).

- [48] Voss, A., Schroeder, R., Heitmann, A., Peters, A. & Perz, S. Short-term heart rate variability—influence of gender and age in healthy subjects. *PloS One* **10**, e0118308 (2015).
- [49] Beason-Held, L. L. *et al.* Health conditions associated with Alzheimer’s disease and vascular dementia. *Annals of Neurology* **93**, 805–818 (2023).
- [50] Wade, R. C. *et al.* ECG-based risk factors for adverse cardiopulmonary events and treatment outcomes in COPD. *European Respiratory Journal* **65** (2025).
- [51] Sarkar, P. & Etemad, A. Self-supervised ecg representation learning for emotion recognition. *IEEE Transactions on Affective Computing* **13**, 1541–1554 (2020).
- [52] Eyben, F. *et al.* The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing* **7**, 190–202 (2015).
- [53] Yang, Q., Li, X., Ding, X., Xu, F. & Ling, Z. Deep learning-based speech analysis for Alzheimer’s disease detection: A literature review. *Alzheimer’s Research & Therapy* **14**, 186 (2022).
- [54] Cummins, N. *et al.* A review of depression and suicide risk assessment using speech analysis. *Speech Communication* **71**, 10–49 (2015).
- [55] Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K. & Geraltz, D. S. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics* **20**, 50–64 (2007).
- [56] Azadmaleki, H. *et al.* SpeechCARE: dynamic multimodal modeling for cognitive screening in diverse linguistic and speech task contexts. *npj Digital Medicine* **8**, 677 (2025).
- [57] Ablimit, A., Botelho, C., Abad, A., Schultz, T. & Trancoso, I. Romberg, J. (ed.) *Exploring dementia detection from speech: Cross corpus analysis*. (ed. Romberg, J.) 2022-2022 *IEEE International Conference on Acoustics, Speech and Signal Processing*, 6472–6476 (2022).
- [58] Fard, A. P., Mahoor, M. H., Alsuhaibani, M. & Dodge, H. H. Linguistic-based mild cognitive impairment detection using informative loss. *Computers in Biology and Medicine* **176**, 108606 (2024).
- [59] Eyben, F., Wöllmer, M. & Schuller, B. Chen, C. W., Jr., R. S. & Lederman, A. (eds) *openSMILE: The munich versatile and fast open-source audio feature extractor*. (eds Chen, C. W., Jr., R. S. & Lederman, A.) *Proceedings of the 18th ACM International Conference on Multimedia*, 1459–1462 (2010).

- [60] Tian, Y.-I., Kanade, T. & Cohn, J. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, 97–115 (2001).
- [61] Zhou, Y. *et al.* Distinguishing apathy and depression in older adults with mild cognitive impairment using text, audio, and video based on multiclass classification and shapely additive explanations. *International Journal of Geriatric Psychiatry* **37** (2022).
- [62] Xu, X. *et al.* Facial aging, cognitive impairment, and dementia risk. *Alzheimer's Research & Therapy* **16**, 245 (2024).
- [63] Desai, R. *et al.* Affective symptoms and risk of progression to mild cognitive impairment or dementia in subjective cognitive decline: A systematic review and meta-analysis. *Ageing Research Reviews* **71**, 101419 (2021).
- [64] Sun, J., Dodge, H. H. & Mahoor, M. H. MC-ViViT: Multi-branch classifier-ViViT to detect mild cognitive impairment in older adults using facial videos. *Expert Systems with Applications* **238**, 121929 (2024).
- [65] Alsuhaibani, M., Dodge, H. H. & Mahoor, M. H. Mild cognitive impairment detection from facial video interviews by applying spatial-to-temporal attention module. *Expert Systems with Applications* **252**, 124185 (2024).
- [66] Zhou, Y. *et al.* Developing a machine learning model for detecting depression, anxiety, and apathy in older adults with mild cognitive impairment using speech and facial expressions: A cross-sectional observational study. *International journal of nursing studies* **146**, 104562 (2023).
- [67] Baltrusaitis, T., Zadeh, A., Lim, Y. C. & Morency, L.-P. Yin, L. & Busso, C. (eds) *OpenFace 2.0: Facial behavior analysis toolkit*. (eds Yin, L. & Busso, C.) *Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition*, 59–66 (2018).
- [68] Fei, Z. *et al.* A novel deep neural network-based emotion analysis system for automatic detection of mild cognitive impairment in the elderly. *Neurocomputing* **468**, 306–316 (2022).
- [69] Opwonya, J. *et al.* Saccadic eye movement in mild cognitive impairment and Alzheimer's disease: A systematic review and meta-analysis. *Neuropsychology Review* **32**, 193–227 (2022).