



Refined prefrontal working memory network as a neuromarker for Alzheimer's disease

EUNHO KIM,^{1,7} JIN-WOO YU,^{1,7} BOMIN KIM,¹ SUNG-HO LIM,^{1,2} 
SANG-HO LEE,³ KWANGSU KIM,⁴ GOWOON SON,⁴ HYEON-AE
JEON,^{2,4} CHEIL MOON,^{2,3,4} JOON SAKONG,^{5,6} AND JI-WOONG
CHOI^{1,2,*} 

¹Department of Information and Communication Engineering, DGIST, Daegu 42988, Republic of Korea

²Brain Engineering Convergence Research Center, DGIST, Daegu 42988, Republic of Korea

³Convergence Research Advanced Centre for Olfaction, DGIST, Daegu 42988, Republic of Korea

⁴Department of Brain and Cognitive Sciences, DGIST, Daegu 42988, Republic of Korea

⁵Department of Occupational and Environmental Medicine, Yeungnam University Hospital, Daegu 42415, Republic of Korea

⁶Department of Preventive Medicine and Public Health, College of Medicine, Yeungnam University, Daegu 42415, Republic of Korea

⁷These authors equally contributed to this work

*jwchoi@dgist.ac.kr

Abstract: Detecting Alzheimer's disease (AD) is an important step in preventing pathological brain damage. Working memory (WM)-related network modulation can be a pathological feature of AD, but is usually modulated by untargeted cognitive processes and individual variance, resulting in the concealment of this key information. Therefore, in this study, we comprehensively investigated a new neuromarker, named "refined network," in a prefrontal cortex (PFC) that revealed the pathological features of AD. A refined network was acquired by removing unnecessary variance from the WM-related network. By using a functional near-infrared spectroscopy (fNIRS) device, we evaluated the reliability of the refined network, which was identified from the three groups classified by AD progression: healthy people (N=31), mild cognitive impairment (N=11), and patients with AD (N=18). As a result, we identified edges with significant correlations between cognitive functions and groups in the dorsolateral PFC. Moreover, the refined network achieved a significantly correlating metric with neuropsychological test scores, and a remarkable three-class classification accuracy (95.0%). These results implicate the refined PFC WM-related network as a powerful neuromarker for AD screening.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative brain disease that causes gradual cognitive decline [1]. Pathological progression of AD precedes symptomatic presentation [2]; thus the functional/structural damage caused by AD accumulates for several years before diagnosis [2,3]. As this damage is currently incurable, strategies have been designed to delay the progression of AD at an early stage, to prevent cognitive impairment [3–5]. Mild cognitive impairment (MCI) is defined as a mental state with minor cognitive decline, which is often considered preclinical AD [6–9]. Therefore, to provide a medical intervention before irreversible damage affects the patients, early detection of the AD pathology is important [3,10,11].

One of the initial characteristics of AD pathology is the dysfunction of working memory (WM) [12–13], the part of memory that provides temporary storage and manipulation of the information required for performing various cognitive tasks [14]. Cognitive degeneration, such as WM decline, is known to have a close relationship with the alteration of the brain network

[15–17]. Changes in brain networks caused by AD pathology have been frequently investigated to provide useful evidence for the pathological properties [18,19].

Functional connectivity (FC) is a temporal correlation in brain activity between different brain regions, which is an effective indicator of an individual's brain network [20–22]. In a recent study, FC gained interest as a neurological biomarker (neuromarker) because its variance was largely determined by stable individual properties, rather than trial variations [23–25]. Moreover, it was reported that task-induced FC provided more detailed information about individual cognitive abilities than resting-state FC [26,27]. Therefore, through the change in task-induced FC, recent functional imaging studies showed a better understanding of the decline in the WM ability of patients with AD pathology [28], and attempted to detect AD in the early stage [29].

However, task-induced FC includes trivial FC variance that covers the FC changes triggered by the task; individual characteristics can yield FC variances [30,31]. A variety of untargeted processes, other than the actual targeted cognitive process, may also influence the task-induced FC [19,32]. As unrelated cognitive processes or physiological noise inflate the FC estimation, a previous study tried to attenuate the unrelated cognitive process from the task-induced FC [33]. This process prevents possible distortion through cognitive processes, rather than from those that caused by AD. Therefore, for a reliable AD detection, we refined the FC to prevent data unrelated to AD pathology affecting the result of the screening.

In this study, we focused on the pathological alteration of the prefrontal cortex (PFC) of AD. The PFC contains the core network of the WM process, including the default mode network and fronto-parietal network, which are known to be damaged in AD pathology [34,35]. The alteration can be detected by functional near-infrared spectroscopy (fNIRS) [36–38]. As fNIRS has various advantages, such as low measurement cost and high portability [39,40], there have been many attempts to screen AD through fNIRS signal-based biomarkers [29,41–46]. Li et al. demonstrated that the concentration changes of oxy-hemoglobin (ΔHbO) signal was highly correlated to a neuropsychological test result, the mini-mental state examination (MMSE) score [43]. Another study reported that the ΔHbO concentration score in PFC could be a good indicator for classifying patients with MCI and healthy older adults [29,44–46].

Although the fNIRS-measured biomarker has been successful in classifying AD as well as in inferring simple cognitive indices with high accuracy, there are still problems to be solved. Early AD is well known for its marked decrease in memory and executive function, including WM. For a proposed biomarker to well characterize AD, the sensitivity affected by the inference of specific cognitive functions, such as memory or executive function, should be reviewed. Furthermore, due to the nature of neurodegenerative diseases, AD biomarkers need to be validated on a broad spectrum of AD pathologies for continuous therapeutic monitoring.

Therefore, we aimed to comprehensively investigate whether the unrelated variance removed brain network, named the refined network, was a good biomarker that inferred multiple cognitive functions from the pathological changes in AD, and further support the high classification performance. We hypothesized that a refined network estimated from PFC by using fNIRS could be a helpful AD neuromarker for the early detection of AD pathological change and multi-domain cognitive state.

To verify our hypothesis, we designed two WM tasks, and corresponding control tasks, to estimate the WM-unrelated FC variance of WM task. We eliminated FC variances from FC maps obtained during WM tasks by subtracting FC maps obtained during control tasks. From the variance removed FC map, the “refined” PFC WM-related network was identified. Next, we investigated whether the refined WM-related network improved the performance of the three-class classifications: healthy control (HC), patients with MCI, and patients with AD. However, the process of refining the network may cause the elimination of proper individual features. Therefore, to evaluate whether personal features related to cognition were adequately

preserved, we confirmed the effect of the refining method of the WM-related network on the prediction of cognitive scores.

In this study, we achieved a high accuracy for AD pathological state classification and prediction performance for cognitive scores. Our results show that a refined WM-related network is a powerful neuromarker for disease detection and assessing cognitive abilities. These results suggest that early detection of AD with a sufficient level of clinical accuracy is possible through the prefrontal fNIRS device. Thus, this study may contribute to improving the quality of lives of many potential patients with AD in the future.

2. Materials and methods

2.1. Participants

For the experiment, 18 patients with AD [77.8 ± 6.6 years, 4 males (M)/14 females (F), education 4.4 ± 3.8 years, clinical dementia rating (CDR) score 1.08 ± 0.65, Korean version of the mini-mental state examination (K-MMSE) score 19.4 ± 6.2] and 11 patients with MCI [78.2 ± 5.4 years, 2M/9F, education 3.2 ± 4.8 years, CDR score 0.95 ± 0.61, K-MMSE score 21.0 ± 4.7] participated in this study. All patients were recruited from a local adult day-care center in Daegu, South Korea. Diagnostic records were used to evaluate each participant's AD pathological condition. Participants comprising a control group [75.9 ± 4.6 years, 3M/28F, education 5.7 ± 2.6 years, CDR score 0.45 ± 0.33, K-MMSE score 25.3 ± 2.7] were recruited from the local community in Daegu, South Korea. For all recruited participants, medical examination, K-MMSE [47], and CDR [48] were provided by Yeongnam University Medical Center. The demographics of the three groups are described in Table 1. All participants took no psychiatric medications or had any medical history of neuropsychiatric disease except MCI and AD. They also did not have severe vascular disease, which can cause hemodynamic dysfunction. There were no significant differences ($p \geq 0.05$) in sex ratio ($p=0.48$, $\chi^2_2=1.49$, Kruskal-Wallis test), education ($p=0.11$, $F_{2,57}=2.30$, one-way ANOVA), or age ($p=0.33$, $F_{2,57}=1.14$ one-way ANOVA) between the groups. There were significant differences in the CDR score ($p < 0.001$, $\chi^2_2=15.95$, Kruskal-Wallis test) and K-MMSE score ($p < 0.001$, $F_{2,57}=11.29$, one-way ANOVA). This research process was approved by the DGIST Institutional Review Board (DGIST-190401-HR-007-02). Before the experiment, all participants and their guardians were informed of the purpose of the experiment, process, and procedure and signed a consent form for the experiment.

Table 1. The demographics of HC, MCI, and AD groups

Demographics	HC (N=31)	MCI (N=11)	AD (N=18)	<i>p</i> value
Age (years) ^b	75.9 ± 4.6	78.2 ± 5.4	77.8 ± 6.6	0.33
ex (Male/Female) ^c	3/28	2/9	4/14	0.48
Education (years) ^b	5.7 ± 2.6	3.2 ± 4.8	4.4 ± 3.8	0.11
CDR score ^c	0.45 ± 0.33	0.95 ± 0.61	1.08 ± 0.65	<0.001 ^a
K-MMSE score ^b	25.3 ± 2.7	21.0 ± 4.7	19.4 ± 6.2	<0.001 ^a

HC: healthy controls; MCI: mild cognitive impairment; AD: Alzheimer's disease

^aThe *p* values representing significant differences ($p < 0.05$).

^bThe *p* value was calculated by one-way ANOVA.

^cThe *p* value was calculated by Kruskal-Wallis test.

2.2. Experimental procedure

During the experiment, all participants underwent a cognitive examination and three fNIRS measurement sessions. All sessions were conducted in a quiet room. To avoid contamination of the resulting data with fatigue, each experimental session occurred at least one week apart.

In the cognitive examination, a neuropsychological test battery, called the Seoul Neuropsychological Screening Battery-Second Edition (SNSB-II) [49], was performed on all participants by qualified neuropsychologists, to evaluate each participant's WM ability. The SNSB-II is a test battery used for the clinical diagnosis of dementia in Korea. It includes several standard cognitive tests that are re-standardized according to the Korean age and education cohort (see Table S1). The SNSB-II aggregates the results of various standard cognitive tests and provides cognitive scores for five cognitive domains: attention, language, memory, visuospatial function, and frontal executive function. The WM ability noted in this study was primarily assessed via memory and frontal executive function domain scores.

The hemodynamic signals in the PFC were measured during the WM task through an fNIRS device. For fNIRS measurements, all participants performed two different WM tasks: Delayed Match-To-Sample (DMTS) task [15] and Digit Span Test (DST) [38], via the touchscreen at eye-level height. All participants were asked not to move their heads as much as possible, not to speak, and to focus on the screen before the measurement started. Between the two tasks, participants removed the device from their heads and rested for at least 10 min. Each task consisted of 15 WM trials and 15 control trials. Between each trial, a 10-s inter-stimulus interval (ISI) was inserted, as in previous studies, for a clear separation of the brain signal from each trial, avoiding long sessions causing mental fatigue in patients with AD [38,43]. Before the first trial of each task, a description of the task and a sufficient number of practice trials were provided (Fig. 1(A)). The WM trial was designed to provide a WM load. The WM trial consisted of three steps: 5 s of an encoding phase, 5 s of a maintenance phase, and 10 s of a retrieval phase. The control trial consisted of a 10 s control phase mimicking the retrieval phase of WM trials, without a WM load, to estimate the baseline bias caused by non-WM processes (Fig. 1(B)).

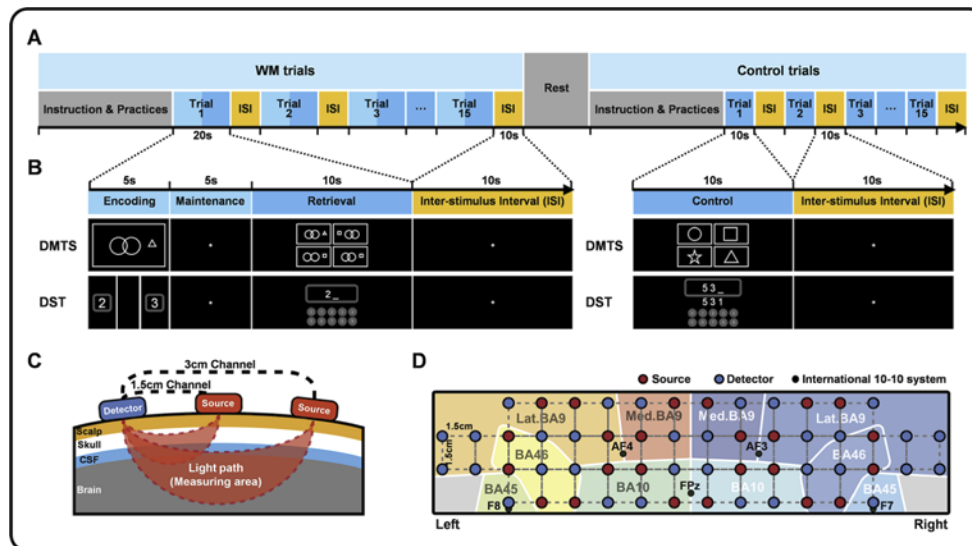


Fig. 1. Experimental design for the fNIRS measurements. (A) The timeline of the WM task. (B) The structure of two trial types, WM trial and control trial, and the stimulus example per two WM tasks: DMTS and DST. (C) A schematic diagram of fNIRS channels with different SDD. (D) The source-and-detector configuration of NIRSIT with BAs and 10-10 system reference position. BAs: Brodmann areas; DMTS: Delayed Match-To-Sample task; DST: Digit Span Test; fNIRS: functional near-infrared spectroscopy; NIRSIT: portable wearable continuous-wave fNIRS device

In the WM trial of the DMTS task, participants performed a visual-based memory recall task using a set of images inspired by the Benton Visual Retention Test [50] as visual stimuli. In the encoding phase of the WM trial, an image was randomly selected from the set as a stimulus and shown on the screen. Participants were asked to remember the stimulus. During the maintenance phase, they were asked to keep staring at the point located at the center of the screen. In the retrieval phase, participants were asked to identify which of the four images were shown during the encoding phase. The control trial of the DMTS task used four images, composed of a simple shape, as stimuli. During the control phase, participants were instructed to look for a star shape among the other shapes presented. In the WM trial of DST, participants performed a verbal-based memory retrieval task that used sequences of 2–4 digit numbers as a stimulus instead of images. In the control trial, they were instructed to type the number that appeared on the touchscreen. Regardless of the task, participants were asked to look at the screen during all ISI (Fig. 1(B)). The study was only executed on participants who were able to successfully follow the instruction and perform at least 10 WM trials and 10 control trials in both tasks, across all sessions.

2.3. fNIRS device

A portable wearable continuous-wave fNIRS device (NIRSIT, OBELAB, Seoul, Korea) was used in this study [51]. NIRSIT uses a channel consisting of one source and one detector for measurement. The source emits near-infrared light at wavelengths of 780 nm and 850 nm that can penetrate the skull. The detector records the light's attenuated intensity from the source, scattered by the inner structure of the human forehead, at a sampling rate of 8.138 Hz. Each channel can measure the brain signals coming from different depths according to source-detector distance (SDD) (Fig. 1(C)). NIRSIT consists of 24 sources and 32 detectors providing 204 channels with different SDDs over 1 cm, which are capable of measuring brain hemodynamic signals [52,53]. The configurations of all channels are presented in Figure S1. Although multiple channels share the same source or detector, the signal processing techniques used in the device enable each channel to acquire a signal at the same time.

NIRSIT was worn on the participant's forehead. FPz of the international 10-10 system was used as a reference point for putting on the device. The sources' and detectors' locations on the device and the Brodmann areas (BA), measurable with the device (BA 9, BA 10, BA 45, and BA 46), are shown in Fig. 1(D). Each location of BA was estimated by mapping the reference point and channel positions on the ICBM-152 template. In this study, we divided BA9 into lateral BA9 (Lat. BA9) and medial BA9 (Med. BA9) according to the location for the detailed description of the BA9 area.

2.4. Signal preprocessing

During the task execution, the fNIRS device measured optical density (OD) signals. For each channel, a signal-to-noise power ratio (SNR) was calculated using the power ratio of the OD signal to the noise. Signals from channels with low SNR (<10 dB) were rejected.

It should be acknowledged that the OD signal could be distorted by motion artifacts and baseline drift. Motion artifacts mainly occur when the sensor slips due to head movement. In our experiment, the task was designed to be performed with a touch screen to prevent the participants from moving their heads to see the keypad. Moreover, a portable fNIRS device, NIRSIT, which is stable against the effects of head movement was also used. Therefore, there were not many motion artifacts in the signal, and they were corrected for in the OD signal with the spline interpolation. However, baseline drift can be caused by the short ISI of 10 s. Due to the nature of the hemodynamic response, the short (<20 s) ISI may induce baseline bias due to undershoot [29,44], although a previous study showed that the undershoot did not occur significantly at short ISI of 10 s [43]. To prevent a false positive rate inflation of FC due to the distortion, systematic baseline drift was eliminated by 4th-order polynomial detrending from OD (Fig. S2).

The corrected OD signals were converted into ΔHbO and concentration changes of deoxy-hemoglobin (ΔHbR) via the modified Beer-Lambert law. The ΔHbO and ΔHbR signals were filtered by a Butterworth band-pass filter with a 4th-order infinite impulse response in the 0.009-0.2 Hz bandwidth to remove physiological noise. Since the concentration change is a relative value, the average of hemodynamic response signal for 20 s before the start of the task was selected as the baseline.

2.5. Hemodynamic response analysis

Prior to identifying WM-related networks, it was necessary to verify that the hemodynamic response elicited by the WM process were being measured. First, we extracted block-averaged ΔHbO and ΔHbR signals per participant. Channels that showed an average signal over the retrieval phase that differed significantly from the baseline in all participants were selected as channels of interest (COI) through one-tailed, one-sample *t*-test. The right tailed test was used for the ΔHbO signal and the left tailed was used for the ΔHbR signal.

To compare the hemodynamic response of each task group, the signal from each COI was averaged. The group effect was evaluated by one-way ANOVA of averaged COI signal over retrieval phase. Multiple comparisons were performed to evaluate the effect by group, corrected with Tukey-Kramer post hoc test. The *p* value threshold of statistical significance is 0.05.

2.6. WM-related network extraction and refinement

In previous studies, significant differences in the PFC between patients with MCI and HC were found in the retrieval phase [37,54]. In addition, the ΔHbO signal is considered a more reliable metric, with a stronger correlation with the fMRI signal than the ΔHbR signal [55]. Therefore, the ΔHbO signals generated during the retrieval phase and the control phase were used for WM-related network analysis. To calculate the FC for each trial, we extracted the ΔHbO signals generated during the retrieval phase and the control phase. When extracting the signal, an offset of 2 s after the end of the phase was additionally included in consideration of the hemodynamic delay. Thus, we extracted a 12 s signal of ΔHbO hemodynamics from each trial. Signals extracted from “failed” trials were discarded (Fig. 2(A)).

From each trial, the FC was calculated through Pearson’s correlation of time-series signals between each channel to characterize the brain network. For each trial, diagonal elements and duplicates were removed from a correlation matrix, and FC maps were obtained with a total of 20,706 connectivity values. The FC map of each trial was averaged according to the task (DMTS task and DST) and trial types (control and WM) to estimate a robust correlation coefficient of the full time-series in extracted phase (retrieval and control phased). In theory, this averaging process is equivalent to Pearson’s correlation of full time-series from the retrieval phase if the mean and standard deviation between trials were the same (see Supplement 1). Therefore, the averaging process strongly corrected for values that distort Pearson’s correlation coefficient, such as trial-wise artifacts that have not been removed. The averaged FC map was normalized using Fisher Z transformation for statistical testing. Through this process, the WM FC, which is a WM-related FC map with task-unrelated variance, and control FC, which is an estimation of task-unrelated variance included in WM FC, were obtained per task, following the trial type. To eliminate fluctuations not related to the task, the contrast FC, which was obtained by subtracting the control FC from WM FC, was calculated (Fig. 2(A)).

To identify WM-related networks, we obtained the significant connectivity involved in the WM process for each disease condition. The significant connectivity was identified by a one-sample *t*-test performed with the participants’ FC maps on each group. We extracted a Boolean matrix with edges exceeding a significant threshold ($p=0.05$) from the *t*-test results of each group. The WM-related network may be altered as the disease progresses. Therefore, the Boolean matrix of each group aggregated to the selected WM-related FC mask through Boolean OR operator.

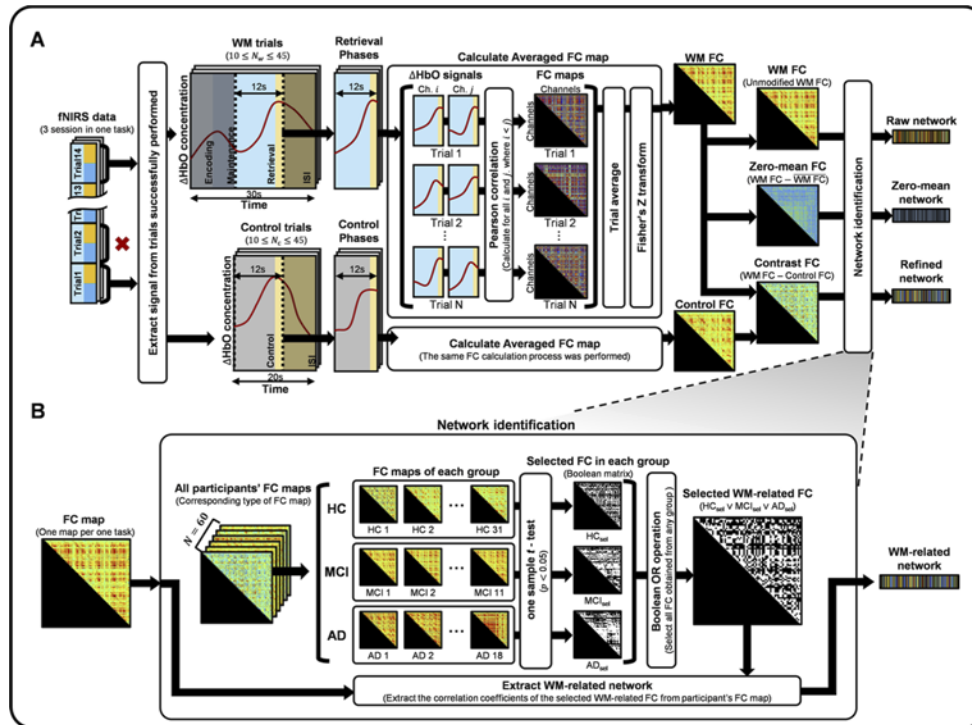


Fig. 2. Procedure for networks extraction from measured fNIRS signal. (A) Calculation of variance removed FC map and conventional FC map using for network extraction, which is used as a comparison. ‘Ch.’ means channel. Overlined WM FC indicates a mean value of WM FC. (B) Identification of WM-related networks from the FC maps which were calculated at A. FC: functional connectivity; fNIRS: functional near-infrared spectroscopy; WM FC: working memory functional connectivity; HC: healthy controls; MCI: mild cognitive impairment; AD: Alzheimer’s disease

The edge selection mask was thus the same in all participants (Fig. 2(B)). A refined WM-related network was obtained by masking the contrast FC, the removal of task-unrelated variance from the connectivity map, with aggregated Boolean matrix. To compare the network refining effect with the conventional metric, we identified and extracted the network edges with the same procedure for two conditions without the variance removal process. A raw network was extracted from the unmodified WM FC, conventional functional connectivity map. A zero-mean network was extracted from the zero-mean FC, calculated by subtracting the mean of the WM FC from itself, to adjust the average of functional connectivity value to zero (Fig. 2(A)).

2.7. Similarity test for validation of refinement

To ensure that the WM-related network is reliable through variance removal, it is necessary to confirm the variance estimation by using a control FC. A similarity test was conducted to show that there was a WM-task-unrelated FC that was common between WM FC and control FC. We examined whether the similarity calculated by Pearson’s correlation coefficient between two FC vectors in participant-wise task FC and control FC was more similar compared to other metrics. The tests were performed via one-way ANOVA, and multiple comparisons were Bonferroni-corrected.

2.8. WM network analysis

We investigated anatomical features of the refined networks for each task. First, we identified anatomical locations of the edges that were strengthened or weakened when performing the task from the extracted network. Significant edges of WM networks that can describe the memory cognitive domain score were also classified. Permutation analysis of linear regression coefficients was performed for the weight analysis of the linear support vector regression (SVR) model that showed a statistically significant prediction performance in the memory domain. A total of 10^5 permutations were performed. Next, using an empirical distribution of averaged-regression coefficient following the permutation, we extracted ($p < 0.05$) edges that significantly contributed to memory domain score estimation.

Furthermore, a one-way ANOVA was performed to identify the edges showing significant inter-group differences ($p < 0.05$), and differences between three different group pairs (HC vs. MCI, HC vs. AD, MCI vs. AD) in the extracted network. Multiple compared p values were corrected by the Bonferroni correction.

2.9. Classification of disease state

We envisaged that the refinement of the WM-related network would remove the irrelevant variance, better revealing the features of each pathological group. Therefore, we investigated whether the disease classification performance for each network could be improved by applying the refining method. An artificial neural network (ANN) model, using all of connectivity values in the network as a feature set, was used to evaluate the classification performance of each network. Because the number of selected edges differed, “zero-padding” was used to align the number of input features extracted from the network to compare the performance of each network using the same ANN model, without affecting the output of the neural network model. In addition, the classification performance according to the structure of the ANN model was also investigated for each network, as the structure of the ANN model that achieves the best performance in each network may be different [56,57]. The five ANN model structures used in this study are described in Table 2.

Table 2. The artificial neural network model structures

Numbers of	Structure 1	Structure 2	Structure 3	Structure 4	Structure 5
Layers	2	2	3	3	3
Nodes in layer 1	68	128	128	256	512
Nodes in layer 2	16	16	48	48	96
Nodes in layer 3			16	16	16

To evaluate model performance, leave-one-out cross-validation (LOOCV) was used. In each LOOCV loop, we formed 60 data sets with different test data. This was done by dividing the data vector of the participants into a single participant’s data (as the test data), and the remaining 59 participants’ data (as training data). For the evaluation, the classification models derived from an optimization procedure were used to evaluate the test data (Fig. 3(A)).

In the optimization procedure, the training data was subdivided into nine participants’ data (as validation data), consisting of randomly selected three participants’ data from each pathological group, and remaining training data of 50 participants for model optimization. The ANN model was developed on the training set and optimized to maximize the classification performance of the validation set to prevent overfitting of the model [58]. As the model performance may vary depending on initial weights and selected validation data, a total of 30 models were trained to statistically obtain the general classification results. Therefore, the test data was evaluated with a total of 30 models, and their classification results were averaged (Fig. 3(B)).

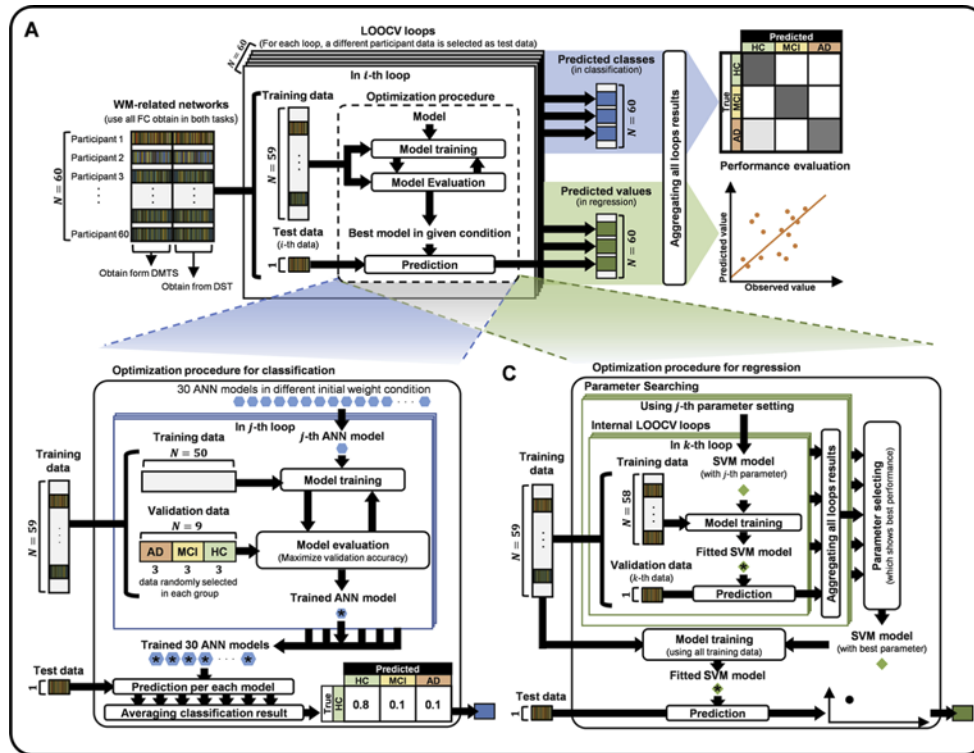


Fig. 3. Performance evaluation of cognitive score prediction performance and disease stage classification. (A) LOOCV loop for performance evaluation. (B) The optimization procedure of the classification model. (C) The optimization procedure of the regression model. AD: Alzheimer's disease; ANN: artificial neural network; LOOCV: leave-one-out cross-validation; HC: healthy controls; MCI: mild cognitive impairment; SVM: support vector machine

As an outcome of 60 loops of LOOCV, 60 test data evaluation results were obtained as a model's disease screening performance. Evaluation metrics including confusion matrix and receiver operating characteristics (ROC) curve were used for the detailed assessment of the model's classification performance. The one-versus-rest method was used to obtain the ROC curve and Macro-F1 score (see Supplement 1).

2.10. Prediction of cognitive scores

Even if the refinement helped to achieve a good performance in group classifications, there was a possibility that the refinement of the WM-related network even removed useful individual features. Therefore, we confirmed that refined FC also preserved or emphasized the information on individual cognitive characteristics. The SVR algorithm was used to predict individual performance scores. As it is known that nonlinear analysis techniques have advantages in characterizing psychiatric disorders, such as Alzheimer's disease [59], a non-linear regression model was generated, and its performance was investigated.

Following the classification framework described above, LOOCV was used for performance evaluation. In each loop, the 60 participant's data vector was divided into one test data and 59 training data. The test data were evaluated by the SVR model, which was trained and optimized through the optimization procedure. A total of 60 loops were generated, and the performance of

the final model was quantified as the Pearson's correlation coefficient between the true value of the test data and the model prediction result (Fig. 3(A)).

To choose an optimal set of parameters for the SVR model, an optimization procedure was executed. The training data were subdivided into one validation data and 58 training data. The SVR model was optimized with the subdivided 58 training data for one set of parameters. This procedure was repeated 59 times. In each loop, hyperparameters C and γ of the nonlinear SVR were determined by a grid-search method to optimize the regression model. Optimal C and γ were searched for each of the 9 conditions (3^{-4} , 3^{-3} , ..., 3^3 , 3^4). The performance for each parameter set was obtained and the highest performance parameter set was chosen (Fig. 3(C)).

We performed 10^3 permutations to evaluate the significance of the observed performance of the model. The participants' SNSB-II cognitive domain scores were randomly shuffled and regressed according to the previous framework. This allowed the creation of the correlation coefficient distribution of the null model. The significance of the model's performance was statistically verified by a comparison with the coefficient distribution of the null model. The significance of the model's performance was statistically verified as follows; $p = (1 + \text{Rank of the observed correlation coefficient}) / (1 + \text{Permutation number})$.

3. Results

3.1. Behavior analysis

The cognitive ability score of the SNSB-II was collected to confirm the decrease in cognitive ability with AD pathology. Group characteristics of the performance results for each domain of SNSB-II are described in Table S2. The behavioral data, and accuracy and response time for the task showed that the reduction in WM capacity had affected the task performance. Table 3 shows the behavioral characteristics and SNSB-II cognitive ability score of each group.

In the behavioral data, three groups showed significant ($p < 0.05$) differences in accuracy ($p < 0.001$, $F_{2,57} = 7.83$, one-way ANOVA) and response time ($p = 0.003$, $F_{2,57} = 6.46$, one-way ANOVA) in the DMTS task. On the post-hoc test, the accuracy of the HC group was significantly higher than that of the MCI and AD groups (HC vs. MCI, $p = 0.002$; HC vs. AD, $p = 0.006$, Tukey-Kramer post hoc test), but there was no significant difference between the MCI and AD groups ($p = 0.987$, Tukey-Kramer post hoc test). However, in terms of response time, the AD group had a slower response time than the other two groups (HC vs. AD, $p = 0.003$; MCI vs. AD, $p = 0.039$, Tukey-Kramer post hoc test), but there was no difference between the HC and MCI groups ($p = 0.980$, Tukey-Kramer post hoc test). In the DST, the three groups also showed a significant difference in accuracy ($p < 0.001$, $F_{2,57} = 10.72$, one-way ANOVA) and response time ($p < 0.001$, $F_{2,57} = 19.39$, one-way ANOVA). In the post-hoc test, the HC group showed a significantly higher accuracy than the MCI and AD groups (HC vs. MCI, $p < 0.001$; HC vs. AD, $p = 0.003$, Tukey-Kramer post hoc test), but no significant differences were observed between the two groups, MCI and AD ($p = 0.608$, Tukey-Kramer post hoc test). The response times showing significant differences between each group had a similar trend as the accuracies (HC vs. MCI, $p < 0.001$; HC vs. AD, $p < 0.001$; MCI vs. AD, $p = 0.953$, Tukey-Kramer post hoc test).

Additionally, we identified group characteristics across the cognitive domains. All three groups showed significant group effects in attention ($F_{2,57} = 8.20$, $p < 0.001$), visuospatial function ($F_{2,57} = 3.44$, $p = 0.039$), memory ($F_{2,57} = 20.78$, $p < 0.001$), and frontal executive function ($F_{2,57} = 11.63$, $p < 0.001$) domain scores in one-way ANOVA. In the post-hoc comparison, the scores in the HC group were significantly higher than the scores in the MCI and AD groups in attention (HC vs. MCI, $p = 0.002$; HC vs. AD, $p = 0.018$, Tukey-Kramer post hoc test), memory (HC vs. MCI, $p < 0.001$; HC vs. AD, $p < 0.001$, Tukey-Kramer post hoc test), and frontal executive function (HC vs. MCI, $p = 0.002$; HC vs. AD, $p < 0.001$, Tukey-Kramer post hoc test) domains. The visuospatial function domain scores did not have any significant difference in multiple comparison correction (HC vs. MCI, $p = 0.114$; HC vs. AD, $p = 0.078$; MCI vs. AD, $p = 0.987$,

Table 3. Behavioral data and SNSB-II cognitive ability score

Behavioral Characteristics	HC (N=31) ^a	MCI (N=11) ^a	AD (N=18) ^a	$F_{2,57}$ ^b	p value ^b	HC vs. MCI ^c	HC vs. AD ^c	MCI vs. AD ^c
DMTS								
Accuracy	0.66±0.17	0.49±0.15	0.48±0.19	7.83	<0.001 ^d	0.002 ^d	0.006 ^d	0.987
Response time (s)	5.27±1.08	5.36±1.03	6.53±1.52	6.46	0.003	0.980	0.003 ^d	0.039 ^d
DST								
Accuracy	0.82±0.18	0.53±0.21	0.61±0.25	10.72	<0.001 ^d	<0.001 ^d	0.003 ^d	0.608
Response time (s)	4.59±1.32	7.19±1.66	7.01±1.83	19.39	<0.001 ^d	<0.001 ^d	<0.001 ^d	0.953
SNSB-II								
Attention score	8.0±1.8	5.6±1.4	6.4±2.1	8.20	<0.001 ^d	0.002 ^d	0.018 ^d	0.484
Language score	-0.6±0.6	-1.0±0.6	-1.0±0.8	2.65	0.079	0.237	0.111	0.997
Visuospatial function score	-0.6±0.9	-1.4±1.0	-1.3±1.2	3.44	0.039	0.114	0.078	0.987
Memory score	-0.5±0.6	-1.5±0.5	-1.7±0.8	20.78	<0.001 ^d	<0.001 ^d	<0.001 ^d	0.808
Frontal executive function score	-0.9±0.8	-1.9±0.7	-1.9±0.8	11.63	<0.001 ^d	0.002 ^d	<0.001 ^d	0.994

DMTS: Delayed Match-To-Sample task; DST: Digit Span Test; SNSB-II: Seoul Neuropsychological Screening Battery-Second Edition;

^aMeans and standard deviations are shown.

^bThe F values and the p values were calculated through one-way ANOVA.

^cThe p values of binary comparisons were derived by the Tukey-Kramer post hoc test.

^dThe p values representing significant differences ($p<0.05$).

Tukey-Kramer post hoc test). No significant differences were found in comparisons between the other groups.

These results indicate a significant decrease in the WM capacity in the early stages of AD pathology. This was similarly reflected in the cognitive domain scores related to WM, such as attention, memory, and frontal executive function ability. As the severity progressed from MCI to AD, most of the cognitive and behavioral scores related to WM, excluding the response time of the DMTS task, decreased slightly. However, this difference was insignificant.

3.2. Hemodynamic response during the WM task

We investigated the hemodynamic response magnitude of each task condition. Figure 4 is a summary of the COI for hemodynamic responses in two WM tasks. In all cases, a task-related hemodynamic response was observed. Group comparisons were conducted for the mean COI hemodynamic response magnitude, which were obtained during the retrieval phase. In the DMTS task, the ΔHbO concentration showed significant differences between groups ($p=0.016$, $F_{2,57}=4.44$, one-way ANOVA). In post hoc analysis, the ΔHbO concentration was significantly higher in the MCI group during the DMTS task than in the HC group ($p=0.016$, Tukey-Kramer post hoc test). However, in the DST task, the ΔHbO concentration showed no significant differences between groups ($p=0.553$, $F_{2,57}=0.60$, one-way ANOVA). Additionally, there was no significant group effect in the ΔHbR concentration on the hemodynamic response magnitude during the DMTS task ($p=0.575$, $F_{2,57}=0.56$, one-way ANOVA) and DST task ($p=0.831$, $F_{2,57}=0.19$, one-way ANOVA).

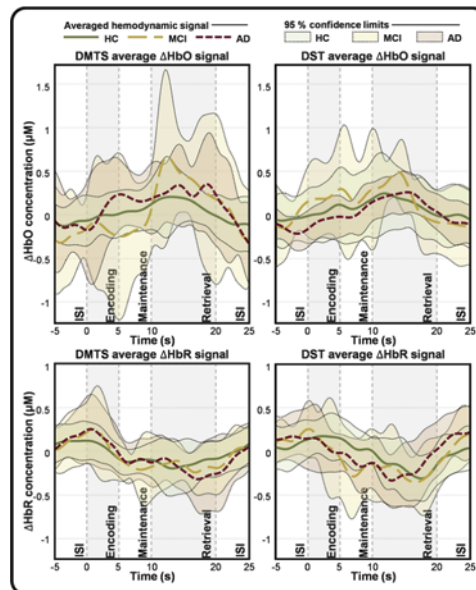


Fig. 4. Averaged COI hemodynamic response signal for each task (DMTS and DST) and trial type (WM and control). The thick line represents the group-by-group mean of hemodynamic response signal, and the shaded area represents the 95% confidence area. DMTS: Delayed Match-To-Sample task; DST: Digit Span Test; WM: working memory; AD: Alzheimer's disease; HC: healthy controls; MCI: mild cognitive impairment; ΔHbO : change in oxy-hemoglobin; ΔHbR : change in deoxy-hemoglobin.

3.3. Similarity test between WM FC and control FC

We examined whether the similarity calculated in participant-wise WM FC and control FC was more similar than other metrics. First, the calculated similarity values were grouped into two types based on whether the participants were the same (intra-participant: intra) or different (inter-participant: inter). The grouped pairs were divided again into four subtypes of pairs according to the same/difference in the task (DMTS task and DST) and trial type (WM and control). We selected the closest pair in intra-participant (same task) and inter-participant (same task and same trial type) for detailed analyses. Additional information of the pairs is described in Table S3.

As shown in Fig. 5(A), the similarity of intra-participant was significantly higher than that of inter-participant ($p < 0.0001$). This trend was also seen in the comparison of subtypes between intra-participant and inter-participant described in Fig. 5(B) ($p < 0.0001$, between the closest pair of intra and inter: $p < 0.0001$, between other pairs of intra and inter: $p < 0.0001$, between the closest pair of intra and other pairs of inter: $p < 0.0001$, between other pairs of intra and the closest pair of inter). These results showed that there was a common FC trend within participants. When comparing subtypes in intra-participant, the closest pairs had significantly higher similarity than the other pairs ($p < 0.001$). However, these features were not seen in inter-participant ($p > 0.05$) (Fig. 5(B)). This indicated that there was an intra-individual FC trend common between the WM and Control FC maps.

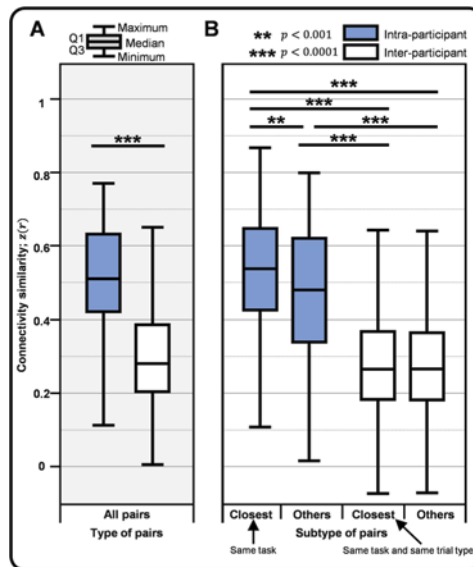


Fig. 5. Comparison of similarity test results for each type of pairs. (A) Result for all pairs per intra-participant and inter-participant. (B) Results for the closest pair and the other pairs in each case. A closest pair is a pair of the same task in the intra-participant, and a pair of the same task and the same trial type in the inter-participant.

3.4. Anatomy of the WM-related network

We identified the anatomical distribution of the WM-related network. Overall, regardless of the task, the dorsolateral PFC (DLPFC), across BA 46 and Lat. BA9, and the dorsomedial PFC (DMPFC), across Med. BA9, seemed to form multiple significant edges with each other. During the DMTS task, intra-hemispheric connectivity in the left hemisphere was enhanced significantly. In contrast, the connectivity within the right hemisphere, right DMPFC and left DMPFC, and

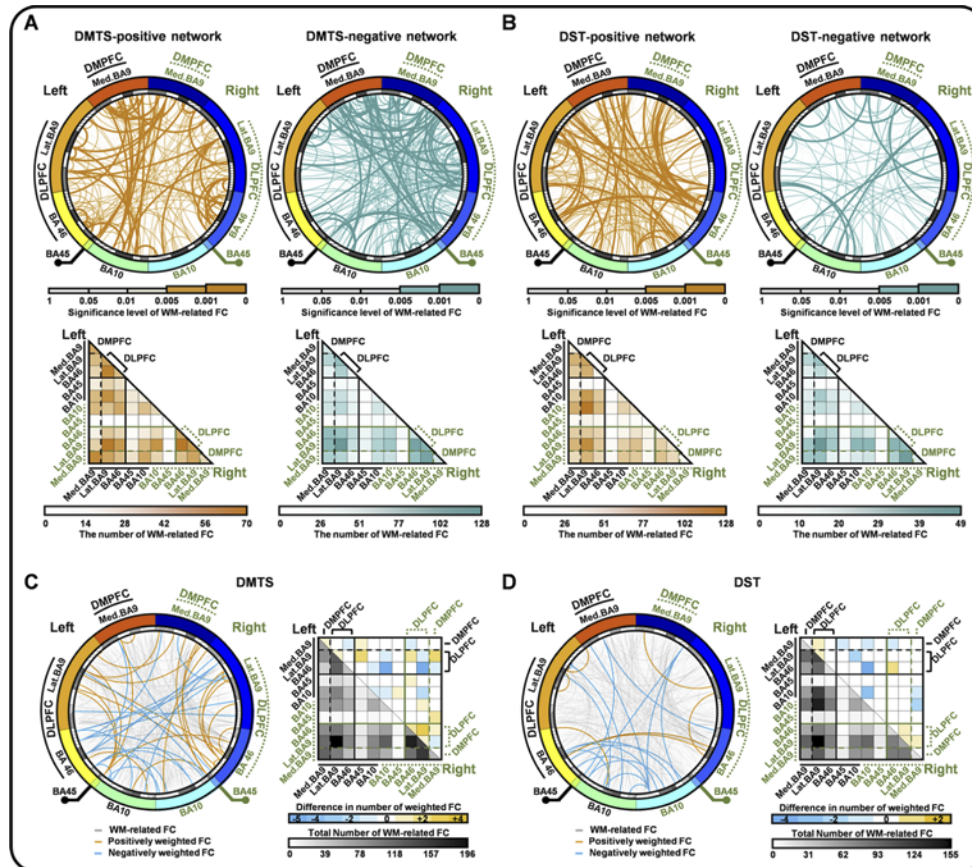


Fig. 6. Anatomical distribution of FC altered during the WM process. (A-B) Statistically ($p < 0.05$) enhanced (task-positive) and weakened (task-negative) FC in WM-related networks when performing tasks. (C-D) Statistically ($p < 0.05$) weighted FC in WM ability estimation model among the WM-related FC. Lower triangular matrix shows the number of WM-related FC. Upper triangular matrix shows the difference in number of positively (orange) / negatively (light blue) weighted FC from the pair of the BAs. BAs: Brodmann areas; DLPFC: dorsolateral prefrontal cortex; DMPFC: dorsomedial prefrontal cortex; DMTS: Delayed Match-To-Sample task; DST: Digit Span Test; FC: functional connectivity; WM: working memory

between right DLPFC and left DMPFC were significantly decreased when performing the DMTS task (Fig. 6(A)). Likewise, the DST task significantly strengthened the connectivity between the left and right DLPFC. Additionally, the connectivity between the right DLPFC and both DMPFCs was significantly enhanced. Conversely, the left frontopolar PFC, including BA 10, showed a statistically significant decline in connectivity between both DLPFCs (Fig. 6(B)). The anatomical distribution of the network edges, which are significantly related to the memory ability, was also analyzed (Fig. 6(C), 6(D)). In both tasks, the edges contributing negative weights to the memory score prediction were more concentratedly found in the inter-hemispheric regions than in the intra-hemispheric regions. The FC formed between the left BA46 and right BA10 predicted a poor memory score. Conversely, FCs with positive weights were evenly distributed over the PFC network.

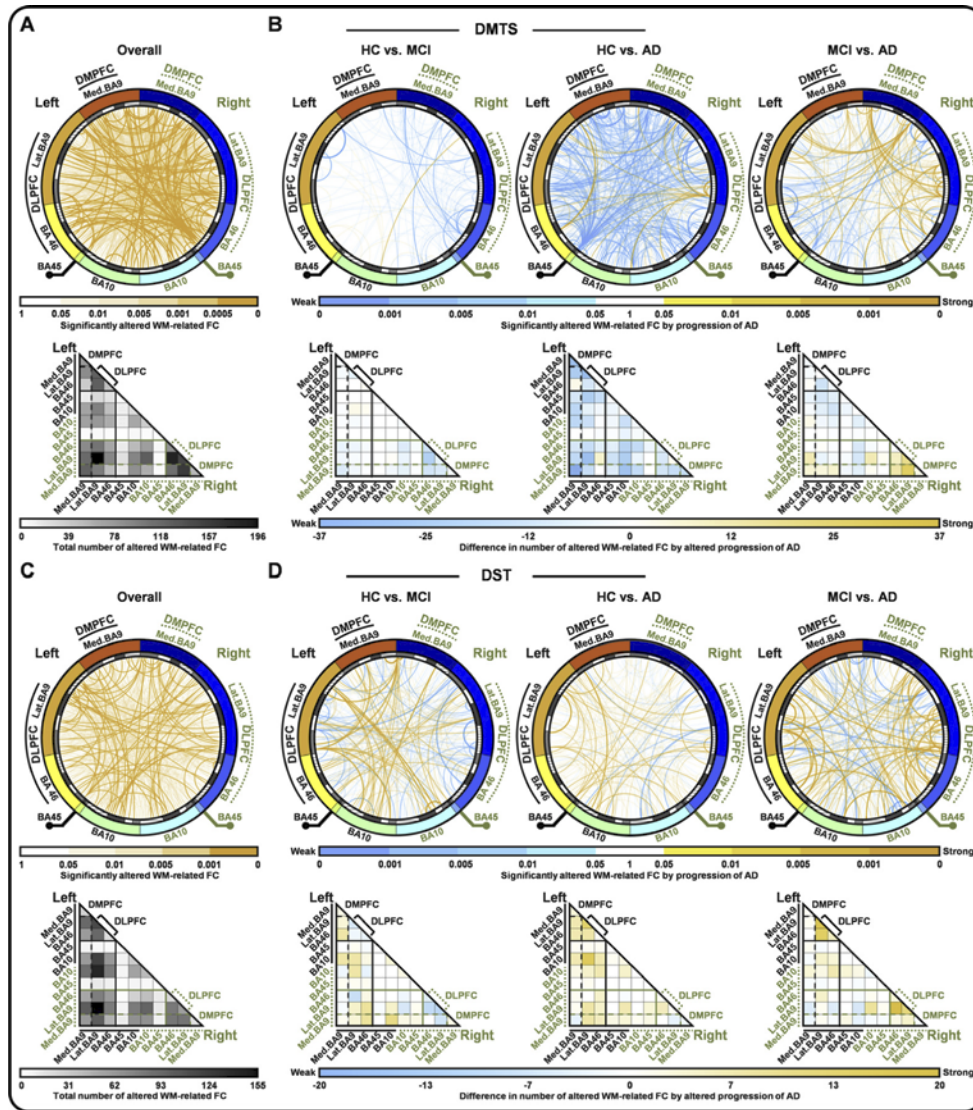


Fig. 7. Anatomical distribution of the FC differed during the AD pathological condition. Edges showing statistically significant differences ($p < 0.05$) among the WM-related FC displayed. (A) Edges showing significant differences in DMTS. (B) Edges that become stronger (orange) / weaker (light blue) as the AD pathological condition becomes more serious in DMTS. (C) Edges showing significant differences in DST. (D) Edges that altered as the AD pathological condition becomes more serious in DST. AD: Alzheimer's disease; BAS: Brodmann areas; DLPFC; dorsolateral prefrontal cortex; DMPFC: dorsomedial prefrontal cortex; DMTS: Delayed Match-To-Sample task; DST: Digit Span Test; FC: functional connectivity; WM: working memory; HC: healthy controls; MCI: mild cognitive impairment.

We also investigated how group characteristics contributed to the distribution of network edges. In the DMTS task, the DLPFC and DMPFC showed significant inter-group differences in intra-regional and inter-regional connectivity (Fig. 7(A)). The DMPFC and DLPFC in the right hemisphere recruited stronger connectivity in the HC group than in the other groups. In the

MCI group, the left hemisphere connectivity was strongly recruited, but the right hemisphere connectivity was relatively weak. (Figure 7(B)). Conversely, during the DST, the total number of edges with significant group differences was less than that during the DMTS task. However, even in DST, the DLPFC and DMPFC there were still areas with significant inter-group differences. Interestingly, the differences between groups of features extracted from DST were particularly strong in the left hemisphere (Fig. 7(C)). Compared to the HC and MCI groups, the AD group showed strong connectivity in the DLPFC. Conversely, compared to the HC group, the MCI group showed stronger left hemisphere connectivity and relatively weak right hemisphere connectivity (Fig. 7(D)).

3.5. Disease state classification performance

To investigate the effect of the refining method on disease state classification, we evaluated the classification performances of the ANN models trained by three kinds of networks that were extracted from the FC with different individual unnecessary variance removal methods. Averaged classification accuracies collected from 30 tests of five ANN structures, from each network, are shown in Table 4. Each network showed a significant difference in the number of features depending on whether individual variance was removed or not. The number of features of each network was 35896 in the raw network, 39264 in the zero-mean network, and 5860 in the refined network. Similarly, the classification performance depended on the network. The refined network achieved the best value (95.0% accuracy), whereas the raw network and zero-mean network achieved up to 58.5% and 57.7%, respectively. In all cases, the models using the refined network as a feature showed higher classification accuracy than the models using the zero-mean network and the raw network. However, the number of features was not the main cause of the difference in accuracy. When the number of features was adjusted for equally, by adjusting the threshold for edge selection, the overall accuracy trends were the same (See Table S4). Conversely, the difference in accuracy evaluated by models using different structures was relatively insignificant. In each network, the difference in the accuracy value between the models using different structures showing the best accuracy and the lowest accuracy was only 4.8%p in the raw network, 4.2%p in the zero-mean network, and 2.7%p in the refined network.

Table 4. Disease state classification accuracy

Network	Structure 1	Structure 2	Structure 3	Structure 4	Structure 5
Raw	58.5% ^a	54.0%	56.7%	55.8%	53.7%
Zero-mean	57.7% ^a	57.2%	57.2%	57.5%	53.5%
Refined	93.7%	92.3%	95.0% ^a	93.7%	92.3%

^aThe highest classification score of each network.

We achieved the best classification performance from a model using structure 3 and the refined network. Figure 8(A) shows the detailed results of the best performance classification. The overall classification accuracy was 95.0%. The macro-averaged F1 score was 95.2%. The binary classification performance between the two groups was also obtained. To calculate the performances, we considered whether the data in the selected two groups were correctly classified. The classification accuracy was 99.3% in HC vs. MCI classification (F1 score: 99.1%), 94.4% in HC vs. AD classification (F1 score: 93.7%), and 89.9% in MCI vs. AD classification (F1 score: 89.6%). During the classification, some patients with AD were misclassified as HC.

Since the misclassification is unacceptable in disease screening, a detailed analysis was conducted to determine whether these patients with AD could be classified correctly by adjusting the classification threshold values. Figure 8(B) shows the ROC curve and area under the curve (AUC) for each group. Although most data show outstanding classification results (AUC>0.9), a few AD data, which were inaccurately classified as HC, were rarely classified accurately. For

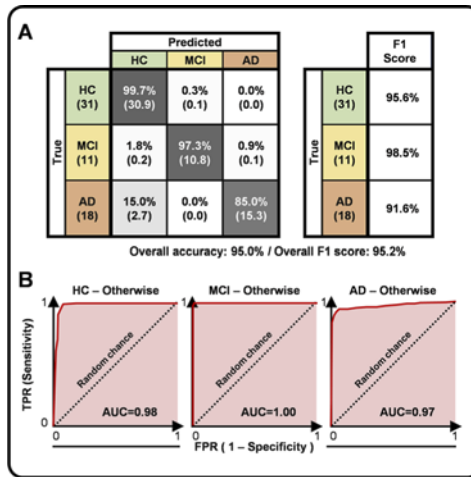


Fig. 8. Detailed results of the best classification performance model. (A) Confusion matrix and F1 score. (B) ROC curves of each class. AD: Alzheimer’s disease; HC: healthy controls; MCI: mild cognitive impairment.

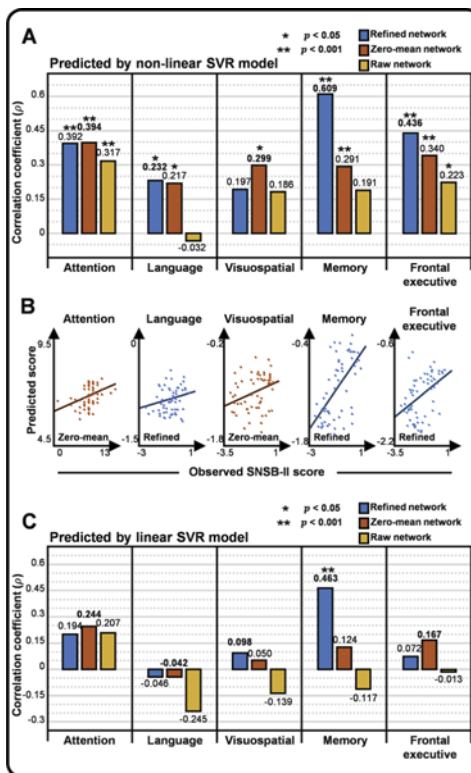


Fig. 9. SNSB-II cognitive domains score prediction result. (A) Predictive performances of five cognitive domains per networks. (B) Prediction result of best predictive network for each domain. (C) Linear model predictive performance for each cognitive domain. SNSB-II: Seoul Neuropsychological Screening Battery-Second Edition; SVR: support vector regression.

MCI, it showed an ideal classification performance ($AUC \approx 1$). This result implies that MCI can be completely distinguished from other groups.

3.6. Prediction performance of cognitive characteristics

To investigate the effect of the refined network on the prediction performance of an individual's cognitive ability, the prediction performances derived from three different network extraction conditions were compared using the nonlinear SVR model. Figure 9(A) shows the Pearson's correlation coefficient (r) between the predicted five cognitive domain scores and the true scores. The refined network achieved the highest performance in the prediction of language ($r=0.232$, $p=0.025$, permutation test), memory ($r=0.609$, $p<0.001$, permutation test), and frontal executive function domains ($r=0.436$, $p<0.001$, permutation test). Meanwhile, the zero-mean network achieved the highest prediction performance in the prediction of the attention ($r=0.394$, $p<0.001$, permutation test) and visuospatial function domains ($r=0.299$, $p=0.020$, permutation test). In the case of memory and frontal executive function, the refined network achieved dramatically higher performance than the zero-mean network. Additionally, the raw network without any refinement was inferior to the predicted performance of the refined and zero-mean networks in all cases. The prediction results of each cognitive domain score are presented in Fig. 9(B), using the best network that achieved the highest performance in each cognitive domain. The prediction performances of the linear SVR model are shown in Fig. 9(C). In all cases, linear regression was less predictive than nonlinear regression. In contrast to nonlinear regression, the linear regression model did not achieve high performance excluding the memory domain score.

4. Discussion

This study showed that the prefrontal fNIRS can identify WM-related networks, which characterize changes in AD pathological cognitive abilities and group effects. It also confirmed that eliminating task-independent variances can enhance the performance of group classification by emphasizing individual cognitive features related to the task. Importantly, local PFC cortical connectivity strongly revealed individual and group clinical information. This fact suggests that the refined PFC WM-related network is a feature that not only distinguishes AD pathological groups, but also provides a good clue for personalized care.

4.1. WM-related network contains the information of the WM process and AD pathological alteration

In line with the converging evidence of a decline in WM capacity in MCI and AD, we confirmed that the WM was exacerbated by AD through clinical and behavioral indicators [13,60]. We also confirmed that the decrease in clinical memory indicators was significantly associated with intra-hemispheric FC. This result is similar to the findings of a recent research that used a connectome-based predictive model (CPM) to predict WM models of a group of older adults comprising patients with AD and MCI [28]. The memory-inferior network edge, located in the intra-hemispheric PFC, supports the compensation mechanism for functional compensation of patients with AD pathology. As mentioned by Avery et al., patients in the AD stage appear to be recruiting more hemisphere FC to compensate for their lowered functional abilities [28]. On the other hand, the significant difference between the groups of the WM-related network, predicted through refined FC, was found in a broader area than the important edge that infers cognitive ability. As previously reported, significant differences occur primarily in the connectivity with the DLPFC or DMPFC [61]. We also investigated the group-specific effects of the FC. The AD group recruited higher right hemispheric connectivity than the MCI group in the DMTS task, which is considered to compensate the injured brain network with recruiting additional resources [62]. In the DST, hyperconnectivity due to compensatory action tended to appear in milder AD. This may be because the decrease in the relative accuracy and response time of the AD and MCI

groups, compared to the HC group, was larger in the DST than in the DMTS task, and thus the demand for cognitive resources increased faster according to the severity of AD. We confirmed that the hyperconnectivity in AD pathology reported in several previous studies [63,64] was also reproduced in the fNIRS measurement.

4.2. Refining task-induced FC improves clinical characteristics

Traditionally, studies of AD pathology are usually inferred at the population level [65]. However, as individual characteristics are becoming increasingly scrutinized, recent studies have emphasized the need to consider the impact of individual differences on the group level [66]. The FC can be modulated by individual characteristics [27,67,68], or untargeted cognitive processes that occur with the task [32,33]. A recent study reported that individual variance contributed highly to the FC variance [23]. Indeed, in our results, participant-wise WM FC also showed the highest similarity to control FC. The variance makes the contribution of targeted cognitive processes and group effects to FC variance relatively low. We were able to emphasize the inter-group and personal traits by removing the unnecessary variance, which was not related to the task at the participant level. Variance elimination in group classification showed a positive effect on classification performance. This seems to be because irrelevant FC variance was reduced during the FC refinement. In the case of the predictive model for participants' cognitive scores, these positive effects only appeared in a task-related cognitive domain prediction. This implies that the cognitive processes caused by the task are reflected in the FC, which emphasizes objective evidence of the cognitive abilities involved. It may provide some explanation for the predictive performance improvement of the task-induced cognitive ability, which has been consistently reported in recent studies [26,27]. Therefore, our results suggest that the refinement of the WM-related network of PFC works as a better neuromarker by clarifying both cognitive process information and group information related to the task.

4.3. fNIRS device provides clinical information for AD screening

The portable fNIRS device is intuitive and inexpensive [39,40], which can make it convenient to be used as a disease screening device in local medical institutions. Therefore, many recent studies have investigated whether the fNIRS could screen psychiatric diseases, such as major depressive disorder [69] and attention deficit hyperactivity disorder [70]. As in this context, there is an ongoing effort to screen for AD pathology [71]. However, no study has tried to predict the performance of various cognitive domains and classify AD pathological groups simultaneously with fNIRS. Our results suggest that the portable fNIRS device can provide a high classification accuracy suitable for clinical AD screening. In this study, the classification model achieved a three-class classification accuracy of 95.0%. Particularly, the pre-AD screening, which is a classification between HC and MCI reached an accuracy of 99.3%. Our model's performance reached a performance level that is better than/similar to that of the three-class classification accuracy reported by conventional fMRI research [72]. Additionally, we used the fNIRS data to confirm the high prediction level of cognitive scores for individuals. These facts confirm that fNIRS can provide continuous information on group classification and cognitive abilities, increasing the possibility of fNIRS as a clinical screening tool in the future.

4.4. Alteration of FC by AD pathology is nonlinear

It is known that changes in the brain network caused by AD are nonlinear [73]. In particular, the initial changes in MCI and AD do not change linearly with changes in cognitive ability [74]. We found that the MCI group showed higher activation in the DMTS task than the other two groups. Although not statistically significant, this trend was also maintained in the DST. In the case of FC, the MCI and AD groups recruited some hyperconnectivity compared to milder AD pathology. These results followed the consistently reported findings in activation [44,42,75], and functional

connectivity [76,77], known as reward mechanisms. Compensation causes hyperactivation and hyperconnectivity, while recruiting more brain regions to meet cognitive demand [62]. Combined with degenerative brain disease, the compensatory mechanisms and decreased activity due to structural disruption of the brain compete, leading to a non-linear relationship between cognitive state and brain state [78]. In our suggested model, we confirmed that the prediction of cognitive ability according to AD was more accurately predicted in nonlinear predictive models. This suggests that in population-level predictions where nonlinear network changes occur, nonlinear prediction contains more insight than linear prediction. The effect of nonlinear changes due to the compensation mechanism is also seen in the classifier performance. MCI and AD could be separated almost completely through ANN, which means that the WM network patterns of MCI and AD are very different. Given the fact that the behavioral performances of the two groups are not well-differentiated ($p > 0.05$), these results may suggest that cognitive abilities could be sustained by the compensatory mechanism produced by FC alterations.

4.5. Limitations

Although this study showed that the refined WM-related network is a useful AD neuromarker, detailed analysis of important features in classification and nonlinear prediction of individual characteristics have not yet been conducted. Analyzing these features would allow us to gather more clinical evidence of the generalizable possibility of our models. Therefore, in future studies, it is necessary to consider the weights of classification models and nonlinear regression models to clarify changes in networks related to AD pathological conditions.

Additionally, our data were gathered from a small number of people per group compared to the predictor variable. This can weaken the statistical power of network estimation. Thus, the network was extracted without p correction. Therefore, in future studies, it is necessary to acquire additional data to identify an accurate anatomical model of WM-related networks. Also, the data used to train the model in this study were small and a bit biased towards specific groups. Therefore, it is necessary to additionally recruit AD and MCI groups to prevent such bias.

5. Conclusions

We have shown that the WM-related network of the PFC is an appropriate neuromarker for AD classification and cognitive score prediction. In identifying WM-related networks, removing variances was important in classification and prediction. Furthermore, we confirmed the credibility of a portable fNIRS device as a clinical instrument by achieving high classification accuracy and personal characteristic prediction performance. This research could benefit many potential Alzheimer's patients by providing opportunities to improve their quality of future life.

Funding. National Research Foundation of Korea (2017M3A9G8084463, 2017R1E1A1A01077393); Ministry of Science and ICT, South Korea (21-IJRP-01).

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2017R1E1A1A01077393). This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No.2017M3A9G8084463). This work was supported by the DGIST R&D Program of the Ministry of Science and ICT (21-IJRP-01).

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data Availability. fNIRS measurement data and behavioral data collected during the experiment will be made available from the corresponding author for the purpose of reproducing research results. The code that supports the findings of this study is available from the corresponding author, upon a reasonable request or a formal data-sharing agreement.

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. A. Kumar and A. Singh, "A review on Alzheimer's disease pathophysiology and its management: an update," *Pharmacol. Rep.* **67**(2), 195–203 (2015).
2. G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack Jr, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, and C. H. Phelps, "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," *Alzheimers Dement.* **7**(3), 263–269 (2011).
3. R. C. Petersen, "Early diagnosis of Alzheimer's disease: is MCI too late?" *Curr. Alzheimer Res.* **6**(4), 324–330 (2009).
4. J. L. Cummings, R. Doody, and C. Clark, "Disease-modifying therapies for Alzheimer disease: challenges to early intervention," *Neurology* **69**(16), 1622–1634 (2007).
5. J. Folch, D. Petrov, M. Etcheto, S. Abad, E. Sánchez-López, M. L. García, J. Olloquequi, C. Beas-Zarate, C. Auladell, and A. Camins, "Current research therapeutic strategies for Alzheimer's disease treatment," *Neural Plast.* **2016**, 1–15 (2016).
6. J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, and L. Berg, "Mild cognitive impairment represents early-stage Alzheimer disease," *Arch. Neurol.* **58**(3), 397–405 (2001).
7. R. C. Petersen, "Mild cognitive impairment as a diagnostic entity," *J. Intern. Med.* **256**(3), 183–194 (2004).
8. R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, "Mild cognitive impairment: clinical characterization and outcome," *Arch. Neurol.* **56**(3), 303–308 (1999).
9. R. C. Petersen, R. Doody, A. Kurz, R. C. Mohs, J. C. Morris, P. V. Rabins, K. Ritchie, M. Rossor, L. Thal, and B. Winblad, "Current concepts in mild cognitive impairment," *Arch. Neurol.* **58**(12), 1985–1992 (2001).
10. M. Ewers, C. Walsh, J. Q. Trojanowski, L. M. Shaw, R. C. Petersen, C. R. Jack Jr, H. H. Feldman, A. L. W. Bokde, G. E. Alexander, P. Scheltens, B. Vellas, B. Dubois, M. Weiner, and H. Hampel, and North American Alzheimer's Disease Neuroimaging Initiative, "Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance," *Neurobiol. Aging* **33**(7), 1203–1214 (2012).
11. R. Green, V. Clarke, N. Thompson, J. Woodard, and R. Letz, "Early detection of Alzheimer disease: methods, markers, and misgivings," *Alzheimer. Dis. Assoc. Disord.* **11**(Suppl. 5), S84 (1996).
12. A. D. Baddeley, S. Bressi, S. Della Sala, R. Logie, and H. Spinnler, "The decline of working memory in Alzheimer's disease: A longitudinal study," *Brain* **114**(6), 2521–2542 (1991).
13. A.-M. Kirova, R. B. Bays, and S. Lagalwar, "Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease," *Biomed. Res. Int.* **2015**, 1–9 (2015).
14. A. D. Baddeley, "Working memory," *Science* **255**(5044), 556–559 (1992).
15. C. L. Grady, M. L. Furey, P. Pietrini, B. Horwitz, and S. I. J. B. Rapoport, "Altered brain functional connectivity and impaired short-term memory in Alzheimer's disease," *Brain* **124**(4), 739–756 (2001).
16. F. Z. Yetkin, R. N. Rosenberg, M. F. Weiner, P. D. Purdy, and C. M. Cullum, "fMRI of working memory in patients with mild cognitive impairment and probable Alzheimer's disease," *Eur. Radiol.* **16**(1), 193–206 (2006).
17. M. Filippi, S. Basaia, E. Canu, F. Imperiale, A. Meani, F. Caso, G. Magnani, M. Falautano, G. Comi, A. Falini, and F. Agosta, "Brain network connectivity differs in early-onset neurodegenerative dementia," *Neurology* **89**(17), 1764–1772 (2017).
18. M. R. Brier, J. B. Thomas, A. Z. Snyder, T. L. Benzinger, D. Zhang, M. E. Raichle, D. M. Holtzman, J. C. Morris, and B. M. Ances, "Loss of intranetwork and internetwork resting state functional connections with Alzheimer's disease progression," *J. Neurosci.* **32**(26), 8890–8899 (2012).
19. Z. Dai, C. Yan, K. Li, Z. Wang, J. Wang, M. Cao, Q. Lin, N. Shu, M. Xia, Y. Bi, and Y. He, "Identifying and mapping connectivity patterns of brain network hubs in Alzheimer's disease," *Cereb. Cortex* **25**(10), 3723–3742 (2015).
20. J. D. Medaglia, M.-E. Lynall, and D. S. Bassett, "Cognitive network neuroscience," *J. Cogn. Neurosci.* **27**(8), 1471–1491 (2015).
21. K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak, "Functional connectivity: the principal component analysis of large (pet) data sets," *J. Cereb. Blood Flow Metab.* **13**(1), 5–14 (1993).
22. S. Tak and J. C. Ye, "Statistical analysis of fNIRS data: a comprehensive review," *NeuroImage* **85**, 72–91 (2014).
23. C. Gratton, T. O. Laumann, A. N. Nielsen, D. J. Greene, E. M. Gordon, A. W. Gilmore, S. M. Nelson, R. S. Coalson, A. Z. Snyder, B. L. Schlaggar, N. U. F. Dosenbach, and S. E. Petersen, "Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation," *Neuron* **98**(2), 439–452.e5 (2018).
24. M. D. Rosenberg, E. S. Finn, D. Scheinost, X. Papademetris, X. Shen, R. T. Constable, and M. M. Chun, "A neuromarker of sustained attention from whole-brain functional connectivity," *Nat. Neurosci.* **19**(1), 165–171 (2016).
25. C.-W. Woo, L. J. Chang, M. A. Lindquist, and T. D. Wager, "Building better biomarkers: brain models in translational neuroimaging," *Nat. Neurosci.* **20**(3), 365–377 (2017).
26. A. S. Greene, S. Gao, D. Scheinost, and R. T. Constable, "Task-induced brain state manipulation improves prediction of individual traits," *Nat. Commun.* **9**(1), 2807 (2018).
27. R. Jiang, N. Zuo, J. M. Ford, S. Qi, D. Zhi, C. Zhuo, Y. Xu, Z. Fu, J. Bustillo, J. A. Turner, V. D. Calhoun, and J. Sui, "Task-induced brain connectivity promotes the detection of individual differences in brain-behavior relationships," *NeuroImage* **207**, 116370 (2020).

28. E. W. Avery, K. Yoo, M. D. Rosenberg, A. S. Greene, S. Gao, D. L. Na, D. Scheinost, T. R. Constable, and M. M. Chun, "Distributed patterns of functional connectivity predict working memory performance in novel healthy and memory-impaired individuals," *J. Cogn. Neurosci.* **32**(2), 241–255 (2020).
29. D. Yang, K.-S. Hong, S.-H. Yoo, and C.-S. Kim, "Evaluation of neural degeneration biomarkers in the prefrontal cortex for early identification of patients with mild cognitive impairment: an fNIRS study," *Front. Hum. Neurosci.* **13**, 317 (2019).
30. B. Cai, G. Zhang, W. Hu, A. Zhang, P. Zille, Y. Zhang, J. M. Stephen, T. W. Wilson, V. D. Calhoun, and Y.-P. Wang, "Refined measure of functional connectomes for improved identifiability and prediction," *Hum. Brain Mapp.* **40**(16), 4843–4858 (2019).
31. W. Jin, H. Zhu, P. Shu, S. Tong, and J. Sun, "Extracting individual neural fingerprint encoded in functional connectivity by silencing indirect effects," *IEEE Trans. Biomed. Eng.* **67**(8), 2253–2265 (2020).
32. M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nat. Rev. Neurosci.* **8**(9), 700–711 (2007).
33. N. Al-Aidroos, C. P. Said, and N. B. Turk-Browne, "Top-down attention switches coupling between low-level and high-level areas of human visual cortex," *Proc. Natl. Acad. Sci. U. S. A.* **109**(36), 14675–14680 (2012).
34. F. Agosta, M. Pievani, C. Geroldi, M. Copetti, G. B. Frisoni, and M. Filippi, "Resting state fMRI in Alzheimer's disease: beyond the default mode network," *Neurobiol. Aging* **33**(8), 1564–1578 (2012).
35. X. He, W. Qin, Y. Liu, X. Zhang, Y. Duan, J. Song, K. Li, T. Jiang, and C. Yu, "Abnormal salience network in normal aging and in amnesic mild cognitive impairment and Alzheimer's disease," *Hum. Brain Mapp.* **35**(7), 3446–3464 (2014).
36. F. Scholkmann, S. Kleiser, A. J. Metz, R. Zimmermann, J. M. Pavia, U. Wolf, and M. Wolf, "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *NeuroImage* **85**, 6–27 (2014).
37. J.-W. Yu, S.-H. Lim, B. Kim, E. Kim, K. Kim, S. K. Park, Y. S. Byun, J. Sakong, and J.-W. Choi, "Prefrontal functional connectivity analysis of cognitive decline for early diagnosis of mild cognitive impairment: a functional near-infrared spectroscopy study," *Biomed. Opt. Express* **11**(4), 1725–1741 (2020).
38. R. Li, T. Nguyen, T. Potter, and Y. Zhang, "Dynamic cortical connectivity alterations associated with Alzheimer's disease: An EEG and fNIRS integration study," *Neuroimage Clin.* **21**, 101622 (2019).
39. F. Irani, S. M. Platek, S. Bunce, A. C. Ruocco, and D. Chute, "Functional near infrared spectroscopy (fNIRS): an emerging neuroimaging technology with important applications for the study of brain disorders," *Clin. Neuropsychol.* **21**(1), 9–37 (2007).
40. M. Soltanlou, M. A. Sitnikova, H. C. Nuerk, and T. Dresler, "Applications of functional near-infrared spectroscopy (fNIRS) in studying cognitive development: The case of mathematics and language," *Front. Psychol.* **9**, 277 (2018).
41. K. Uemura, H. Shimada, T. Doi, H. Makizako, K. Tsutsumimoto, H. Park, and T. Suzuki, "Reduced prefrontal oxygenation in mild cognitive impairment during memory retrieval," *Int. J. Geriatr. Psychiatry.* **31**(6), 583–591 (2016).
42. K. H. Yap, W. C. Ung, E. G. M. Ebenezer, N. Nordin, P. S. Chin, S. Sugathan, S. C. Chan, H. L. Yip, M. Kiguchi, and T. B. Tang, "Visualizing hyperactivation in neurodegeneration based on prefrontal oxygenation: a comparative study of mild Alzheimer's disease, mild cognitive impairment, and healthy controls," *Front. Aging Neurosci.* **9**, 287 (2017).
43. R. H. Li, G. X. Rui, W. Chen, S. Li, P. E. Schulz, and Y. C. Zhang, "Early detection of Alzheimer's disease using non-invasive near-infrared spectroscopy," *Front. Aging Neurosci.* **10**, 366 (2018).
44. D. Yang, R. Huang, S. H. Yoo, M. J. Shin, J. A. Yoon, Y. I. Shin, and K.-S. Hong, "Detection of mild cognitive impairment using convolutional neural network: temporal-feature maps of functional near-infrared spectroscopy," *Front. Aging Neurosci.* **12**, 141 (2020).
45. S.-H. Yoo, S.-W. Woo, M.-J. Shin, J. A. Yoon, Y.-I. Shin, and K.-S. Hong, "Diagnosis of mild cognitive impairment using cognitive tasks: A functional near-infrared spectroscopy study," *Curr. Alzheimer Res.* **17**(13), 1145–1160 (2021).
46. D. Yang and K.-S. Hong, "Quantitative assessment of resting-state for mild cognitive impairment detection: A functional near-infrared spectroscopy and deep learning approach," *J. Alzheimers Dis.* **80**(2), 647–663 (2021).
47. Y. Kang, D. L. Na, and S. Hahn, "A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients," *J. Korean Neurol. Assoc.* **15**(2), 300–308 (1997).
48. J. C. Morris, "The clinical dementia rating (CDR): Current version and scoring rules," *Neurology* **43**(11), 2412.2 (1993).
49. Y. Kang, D. L. Na, and S. Hahn, *Seoul Neuropsychological Screening Battery*, 2nd ed. (Human Brain Research & Consulting Co., 2012).
50. A. L. Benton, *Revised Visual Retention Test: Clinical and Experimental Applications*, 4th ed. (Psychological Corporation, 1974).
51. J. Shin, J. Kwon, J. Choi, and C.-H. Im, "Performance enhancement of a brain-computer interface using high-density multi-distance NIRS," *Sci. Rep.* **7**(1), 1–10 (2017).
52. S. Brigadoi and R. J. Cooper, "How short is short? Optimum source–detector distance for short-separation channels in functional near-infrared spectroscopy," *Neurophotonics* **2**(2), 025005 (2015).

53. T. Funane, H. Atsumori, T. Katura, A. N. Obata, H. Sato, Y. Tanikawa, O. Eiji, and M. Kiguchi, "Quantitative evaluation of deep and shallow tissue layers' contribution to fNIRS signal using multi-distance optodes and independent component analysis," *NeuroImage* **85**, 150–165 (2014).
54. R. Heun, K. Freymann, M. Erb, D. T. Leube, F. Jessen, T. T. Kircher, and W. Grodd, "Mild cognitive impairment (MCI) and actual retrieval performance affect cerebral activation in the elderly," *Neurobiol. Aging* **28**(3), 404–413 (2007).
55. X. Cui, S. Bray, D. M. Bryant, G. H. Glover, and A. L. Reiss, "A quantitative comparison of NIRS and fMRI across multiple cognitive tasks," *NeuroImage* **54**(4), 2808–2821 (2011).
56. F. Dernoncourt and J. Y. Lee, "Optimizing neural network hyperparameters with gaussian processes for dialog act classification," in *2016 IEEE Spoken Language Technology Workshop (SLT)* (IEEE, 2016), pp. 406–413.
57. Y. Bengio, "Practical Recommendations for Gradient-Based Training of Deep Architectures," in *Neural networks: Tricks of the trade*, 2nd ed., G. Montavon, G. B. Orr, and K.-R. Müller, eds. (Springer, 2012).
58. L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," *Neural Netw.* **11**(4), 761–767 (1998).
59. R. Hornero, D. Abásolo, J. Escudero, and C. Gómez, "Nonlinear analysis of electroencephalogram and magnetoencephalogram recordings in patients with Alzheimer's disease," *Philos. Trans. A Math. Phys. Eng. Sci.* **367**(1887), 317–336 (2009).
60. D. Perpetuini, A. M. Chiarelli, C. Filippini, D. Cardone, P. Croce, L. Rotunno, N. Anzoletti, M. Zito, F. Zappasodi, and A. Merla, "Working memory decline in Alzheimer's disease is detected by complexity analysis of multimodal EEG-fNIRS," *Entropy* **22**(12), 1380 (2020).
61. A. Krause-Utz, B. M. Elzinga, N. Y. L. Oei, C. Paret, I. Niedtfeld, P. Spinhoven, M. Bohus, and C. Schmahl, "Amygdala and dorsal anterior cingulate connectivity during an emotional working memory task in borderline personality disorder patients with interpersonal trauma history," *Front. Hum. Neurosci.* **8**, 848 (2014).
62. F. G. Hillary and J. H. Grafman, "Injured brains and adaptive networks: the benefits and costs of hyperconnectivity," *Trends Cogn. Sci.* **21**(5), 385–401 (2017).
63. L. Bonanni, D. Moretti, A. Benussi, L. Ferri, M. Russo, C. Carrarini, F. Barbone, D. Arnaldi, N. W. Falasca, G. Koch, A. Cagnin, F. Nobili, C. Babiloni, B. Borroni, A. Padovani, M. Onofrij, and R. Franciotti, and the FTD Italian study group-SINDEM, "Hyperconnectivity in dementia is early and focal and wanes with progression," *Cereb. Cortex* **31**(1), 97–105 (2021).
64. A. P. Schultz, J. P. Chhatwal, T. Hedden, E. C. Mormino, B. J. Hanseeuw, J. Sepulcre, W. Huijbers, M. LaPoint, R. F. Buckley, K. A. Johnson, and R. A. Sperling, "Phases of hyperconnectivity and hypoconnectivity in the default mode and salience networks track with amyloid and tau in clinically normal individuals," *J. Neurosci.* **37**(16), 4323–4331 (2017).
65. J. D. Van Horn, S. T. Grafton, and M. B. Miller, "Individual variability in brain activity: a nuisance or an opportunity?" *Brain imaging Behav.* **2**(4), 327–334 (2008).
66. E. S. Finn, D. Scheinost, D. M. Finn, X. Shen, X. Papademetris, and R. T. Constable, "Can brain state be manipulated to emphasize individual differences in functional connectivity?" *NeuroImage* **160**, 140–151 (2017).
67. H. Xie, V. D. Calhoun, J. Gonzalez-Castillo, E. Damaraju, R. Miller, P. A. Bandettini, and S. Mitra, "Whole-brain connectivity dynamics reflect both task-specific and individual-specific modulation: A multitask study," *NeuroImage* **180**, 495–504 (2018).
68. E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable, "Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity," *Nat. Neurosci.* **18**(11), 1664–1671 (2015).
69. Y. Zhu, J. K. Jayagopal, R. K. Mehta, M. Erraguntla, J. Nuamah, A. D. McDonald, H. Taylor, and S.-H. Chang, "Classifying major depressive disorder using fNIRS during motor rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**(4), 961–969 (2020).
70. Y. Monden, I. Dan, M. Nagashima, H. Dan, M. Uga, T. Ikeda, D. Tsuzuki, Y. Kyutoku, Y. Gunji, D. Hirano, T. Taniguchi, H. Shimoizumi, E. Watanabe, and T. Yamagata, "Individual classification of ADHD children by right prefrontal hemodynamic responses during a go/no-go task as assessed by fNIRS," *Neuroimage Clin.* **9**, 1–12 (2015).
71. P. A. Cicalese, R. Li, M. B. Ahmadi, C. Wang, J. T. Francis, S. Selvaraj, P. E. Schulz, and Y. Zhang, "An EEG-fNIRS hybridization technique in the four-class classification of Alzheimer's disease," *J. Neurosci. Methods* **336**, 108618 (2020).
72. A. Gupta, M. Ayhan, and A. Maida, "Natural image bases to represent neuroimaging data," in *Proceedings of the 30th International Conference on Machine Learning*, S. Dasgupta and D. McAllester, eds. (JMLR.org, 2013), pp. 987–994.
73. H. Kim, K. Yoo, D. L. Na, S. W. Seo, J. Jeong, and Y. Jeong, "Non-monotonic reorganization of brain networks with Alzheimer's disease progression," *Front. Aging Neurosci.* **7**, 111 (2015).
74. P. A. Wilkosz, H. J. Seltman, B. Devlin, E. A. Weamer, O. L. Lopez, S. T. DeKosky, and R. A. Sweet, "Trajectories of cognitive decline in Alzheimer's disease," *Int. Psychogeriatr.* **22**(2), 281–290 (2010).
75. S. D. Han, W. S. Houston, A. J. Jak, L. T. Eyler, B. J. Nagel, A. S. Fleisher, G. G. Brown, J. Corey-Bloom, D. P. Salmon, L. J. Thal, and M. W. Bondi, "Verbal paired-associate learning by APOE genotype in non-demented older adults: fMRI evidence of a right hemispheric compensatory response," *Neurobiol. Aging* **28**(2), 238–247 (2007).

76. L. Koelewijn, T. M. Lancaster, D. Linden, D. C. Dima, B. C. Routley, L. Magazzini, K. Barawi, L. Brindley, R. Adams, K. E. Tansey, A. Bompas, A. Tales, A. Bayer, and K. Singh, "Oscillatory hyperactivity and hyperconnectivity in young APOE- ϵ 4 carriers and hypoconnectivity in Alzheimer's disease," *eLife* **8**, e36011 (2019).
77. Q. Ye, F. Su, H. Shu, L. Gong, C. Xie, Z. Zhang, and F. Bai, "The apolipoprotein E gene affects the three-year trajectories of compensatory neural processes in the left-lateralized hippocampal network," *Brain imaging Behav.* **11**(5), 1446–1458 (2017).
78. F. G. Hillary, C. A. Roman, U. Venkatesan, S. M. Rajtmajer, R. Bajo, and N. D. Castellanos, "Hyperconnectivity is a fundamental response to neurological disruption," *Neuropsychology* **29**(1), 59–75 (2015).