

## A DATA DETAILS

### A.1 EEG Acquisition

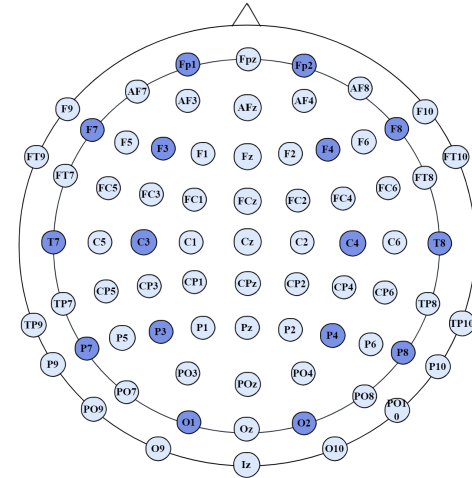
*Mild Cognitive Impairment (MCI) Dataset.* The MCI dataset was acquired from a local hospital using a 64-channel Brain Products EEG system. This system was configured as Figure A1 with the following electrodes: Fp1, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP9, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, TP10, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, and Iz. The electrodes at AFz and FCz were designated as ground and reference, respectively, and were excluded from the data analysis. A sampling rate of 5000 Hz was utilized during the EEG recording process. Data from the ground and reference channels were excluded from the dataset. A total of one hundred participants with cognitive decline were initially recruited. Following the inspection of medication history and other case-relevant factors (conducted by clinical physicians) as well as strict age-matching with the control group (carried out by engineering personnel), a final cohort comprising 46 subjects in the MCI group and 43 subjects in the HC group was deemed eligible for the study. Participants were instructed to sit comfortably and maintain a state of eyes-closed relaxation for an 8-minute recording session.

*Alzheimer’s Disease (AD) Dataset.* The AD dataset was procured from clinical monitoring at an external center, utilizing a 16-channel Syntop EEG amplifier. This system employed the International 10-20 system for electrode distribution, a configuration derived from the more comprehensive 10-10 system but with fewer channels, specifically: Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, and O2. The electrodes utilized for the AD data collection are marked in a darker color in Figure A1. Data capture for each participant lasted one minute, operating at a sampling frequency of 1,024 Hz. The linked earlobes, A1 and A2, served as the reference points. Under expert supervision, each participant contributed two samples—one with eyes closed and the other with eyes open. Each sample was ensured to have a minimum effective duration of 8 seconds, devoid of any significant artifacts or poor recording intervals.

*Parkinson’s Disease (PD) Datasets.* Two PD datasets were collected from the University of New Mexico (NMU Center) and the University of Iowa (IU Center), employing a 64-channel Brain Vision EEG system. This system utilized the International 10-10 system for electrode distribution, as depicted in Figure A1, encompassing the following channels: Fp1, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT9, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, FT10, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP9, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, TP10, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, and O2. During EEG data acquisition, the sampling rate was set to 500 Hz. Concurrently, the online reference was configured to channels CPz and Pz for the NMU and IU centers, respectively. Data from these reference channels were excluded from the publicly accessible repository. At the NMU Center, resting-state EEG recordings were obtained from

54 subjects under both eyes-open and eyes-closed conditions. At the IU Center, recordings were conducted with 28 subjects solely under the eyes-open condition. The PD and Normal Control (NC) groups are well-balanced according to the participant numbers across both datasets. To critically evaluate the adaptability of the proposed MACS framework under varying data collection conditions, it was initially trained using data from closed-eye sessions at the NMU center. Subsequently, the performance of the trained model was assessed by applying it directly to open-eye data from the IU center.

*Ethical Declaration.* The acquisition of EEG data was conducted in strict adherence to ethical guidelines, with approval from the relevant ethical committee, which remains confidential to comply with anonymity rules. Informed consent was duly obtained from all participants prior to their involvement in the study.



**Figure A1: Topographic mapping of electrode positions utilized for EEG dataset collection. The electrodes F9, F10, FT9, FT10, P9, P10, PO9, PO10, O9, and O10 were not included in the 64-channel Brain Products EEG system used for the Mild Cognitive Impairment (MCI) dataset collection. Additionally, the electrodes Fpz, F9, F10, P9, P10, PO9, PO10, O9, O10, and Iz were excluded from the 64-channel Brain Vision EEG system used for the collection of two Parkinson’s Disease (PD) datasets. The electrodes marked in darker color indicate inclusion in the 16-channel Syntop EEG system for the Alzheimer’s Disease (AD) data collection.**

### A.2 EEG Preprocessing

To ensure efficiency, the raw EEG datasets undergo preprocessing through an automated standard pipeline, as described by Ávila et al. (2023) [1]. This pipeline includes filtering, outlier detection, re-referencing, and independent component analysis, designed to eliminate artifacts attributable to low-frequency drifts, high-frequency noise, head movements, cardiac activity, and eye movements. Due to the extensive manual processing and expert monitoring conducted on the AD clinical dataset, we bypassed the standard preprocessing

pipeline for this preprocessed data. A total of 63 effective electrodes were employed for the two PD datasets, each utilizing a different reference electrode. For the MCI dataset, 62 effective electrodes were used, while the AD dataset employed 16 electrodes. To ensure temporal resolution consistency, all EEG recordings were uniformly downsampled to 250 Hz. For spatial alignment in the transferability test, linear interpolation was employed before inputting the data into the MACS framework to maintain consistency across the datasets.

## B IMPLEMENTATION DETAILS

### B.1 MACS Encoder Architecture

Here, we provide additional details on the MACS *Encoder's* two fundamental structures - the convolution starter  $g_{str}$  and the temporal clipper  $g_{clip}$  - that are omitted from the main body of the text for brevity.

*Convolution Starter.* The convolution starter  $g_{str}$  constitutes a spatial one-dimensional convolution layer and a temporal convolution layer. Each layer is equipped with  $d$  kernels, with the spatial layer utilizing kernels of size  $(d, 1)$  and the temporal layer employing kernels of size  $(1, 25)$ , where  $d$  denotes the number of EEG electrodes. We selected a kernel size of 25 sampling points for the EEG's temporal layer, corresponding to the preprocessed sampling rate of 250 Hz. This kernel size enables the convolutional filter to extract features from a receptive field representing 100 ms of neural activities. The stride for both convolution operations is set to 1, ensuring a comprehensive analysis of the input signals. This architecture is designed to effectively capture generalized spatial and temporal patterns within the EEG data.

*Temporal Clipper.* The output generated by the convolution starter  $g_{str}$  is a feature embedding denoted by  $X_{d,T_s}$ . This embedding is partitioned into  $i$  non-overlapping feature clips  $\{X_{d,t_i}\}_{i=1}^I$  by the temporal clipper  $g_{clip}$ . Consequently, the chosen value of  $I$  plays a critical role in determining the temporal scale at which the dynamic functional network is analyzed in the subsequent manifold-based dynamic attention block  $g_{datt}$ . For instance, to observe brain synchronization conditions at a one-second temporal scale,  $I$  should be configured as  $T_s/fs$  for the embedding  $X_{d,T_s}$ . Experiments have been conducted to evaluate the effects of the selected hyperparameter, as illustrated in Table 3 of the main text. These experiments ensure that the temporal resolution is appropriately aligned with the desired analytical scale.

### B.2 Evaluation Metrics

The metrics used to assess the model's performance include accuracy, precision, recall, and the F-1 score. Building on the model trained at the fragment level, an ensemble strategy is employed to generate subject-level predictions. This approach involves averaging the probabilities at the fragment level and then determining the subject-level label using the optimal threshold, which is identified by employing Youden's J statistic [2]. *It is crucial to note that this strategy is consistently applied across all comparative methods to ensure fairness in the evaluation process.*

## C EXTENDED EXPERIMENTAL RESULTS

### C.1 Comparison Study

To guarantee fairness in the comparative analysis, all state-of-the-art (SOTA) methods implemented identical fold divisions, fragment partitioning, cross-validation approaches, and evaluation strategies.

Table A1 delineates the comprehensive results of this comparative study, showcasing full evaluation metrics. While MACS may not achieve the highest scores in every single metric, it demonstrates superior overall performance in the recognition of both MCI and PD, surpassing SOTA methods.

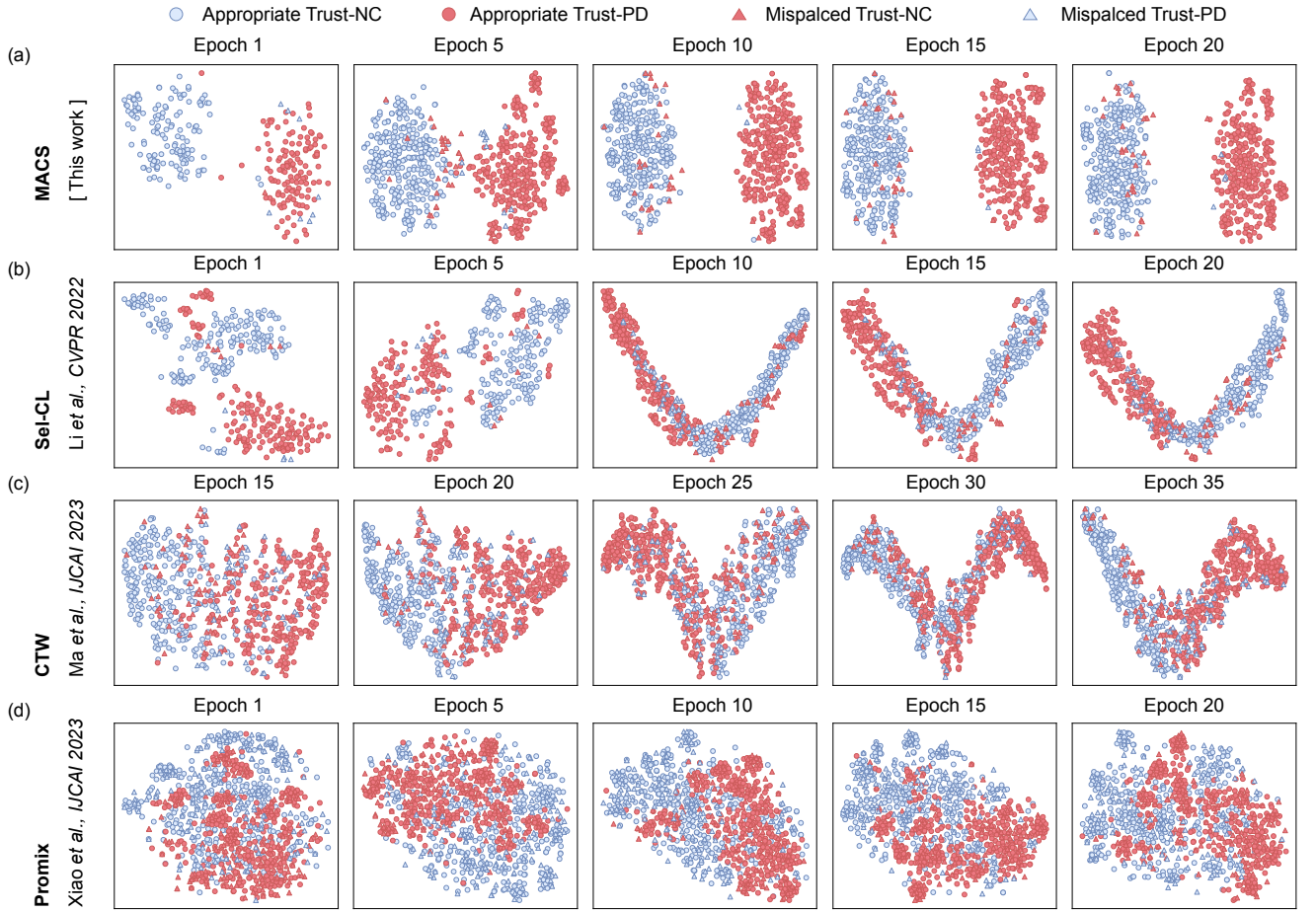
Qualitatively, the MACS framework was compared with three SOTA methods by visualizing the learning progress. This was achieved using t-SNE mappings to depict the embedding features of trusted samples identified by the model across various epochs. Based on their original ground truth labels, before they were reclassified as part of the unreliable annotation sets through label interchanging, the samples identified by the model as trusted were categorized as either appropriately trusted or misplaced. The main text details the model training process on the MCI dataset, showcasing the efficiency of the MACS framework. Complementing this, Figure A2 in the Appendix displays the results obtained from the PD dataset.

For Sel-CL [3], Promix [4], and our proposed MACS framework, training began at the initial epoch, and thus we visualize the progress starting from epoch 1 with intervals of 5 epochs. For the CTW method [5], which includes a warm-up phase before training, t-SNE visualization commenced after the warm-up phases, beginning from epoch 15 with the same 5-epoch intervals.

As the model trains, the desired outcome is that more samples are correctly selected, which enhances representation learning and facilitates the formation of class-specific clusters. The observations on both the MCI dataset (Figure 5 in the main text) and the PD dataset (Figure A2) have demonstrated the efficiency of the MACS framework.

### C.2 Hyperparameter Tuning Study

*Memory Length.* The multi-view contrastive loss is a pivotal component in our MACS framework, as it significantly depends on the adequacy of positive and negative pairs for effective contrastive learning. However, the prevalent scarcity of clinical data poses a substantial challenge. To circumvent the constraints imposed by batch size and to capitalize on the more extensively available data during the learning process, we adopt a storage strategy as outlined by [6]. This strategy involves retaining the preceding  $M$  features in memory, with  $M$  denoting the memory length. As illustrated in Table A2, the influence of memory length on the model's performance was assessed for both MCI and PD datasets via an N-fold cross-validation experiment. This hyperparameter tuning study determined the optimal memory lengths to be 300 for the MCI dataset and 100 for the PD dataset. The results reveal a bell-shaped relationship between memory length and model performance, suggesting an optimal point appears to correlate with the dataset size, as evidenced by the MCI dataset containing approximately three times more EEG fragments than the PD dataset.



**Figure A2: Qualitative comparison of the MACS framework with state-of-the-art methods: t-SNE visualization of latent distribution in Parkinson's disease (PD) data throughout training progress.**

*k*-Nearest Neighbors. The parameter  $K$  determines the number of nearest neighbors selected to infer potential labels, using the  $k$ -Nearest Neighbors algorithm in the *Stratifier*  $fs_t$  module of the MACS framework. A thorough exploration of  $K$ , employing  $N$ -fold cross-validation, is summarized in Table A2. In the experiments, the minimum value of  $K$  and the interval are contingent upon the batch size of MCI and PD datasets. Similar to memory length,  $K$  exhibits an 'inverted-U relationship'. Thus,  $K$ 's optimal configuration should be well-coordinated with the batch sizes.

*Temporal Scale.* As delineated in Appendix B.1, the variable  $I$  dictates the temporal scale employed for analyzing characteristics within the dynamic functional networks. To this end, a series of experiments were conducted, exploring the effects of temporal scales, extending from the millisecond level to the second level. The findings presented in Table A2 indicate that adopting a second-level temporal scale is necessary for achieving robust and improved outcomes. Specifically, the temporal scale was configured to 2 seconds for the MCI dataset and 1 second for the PD dataset. The decision to configure the maximum temporal scale for PD data at

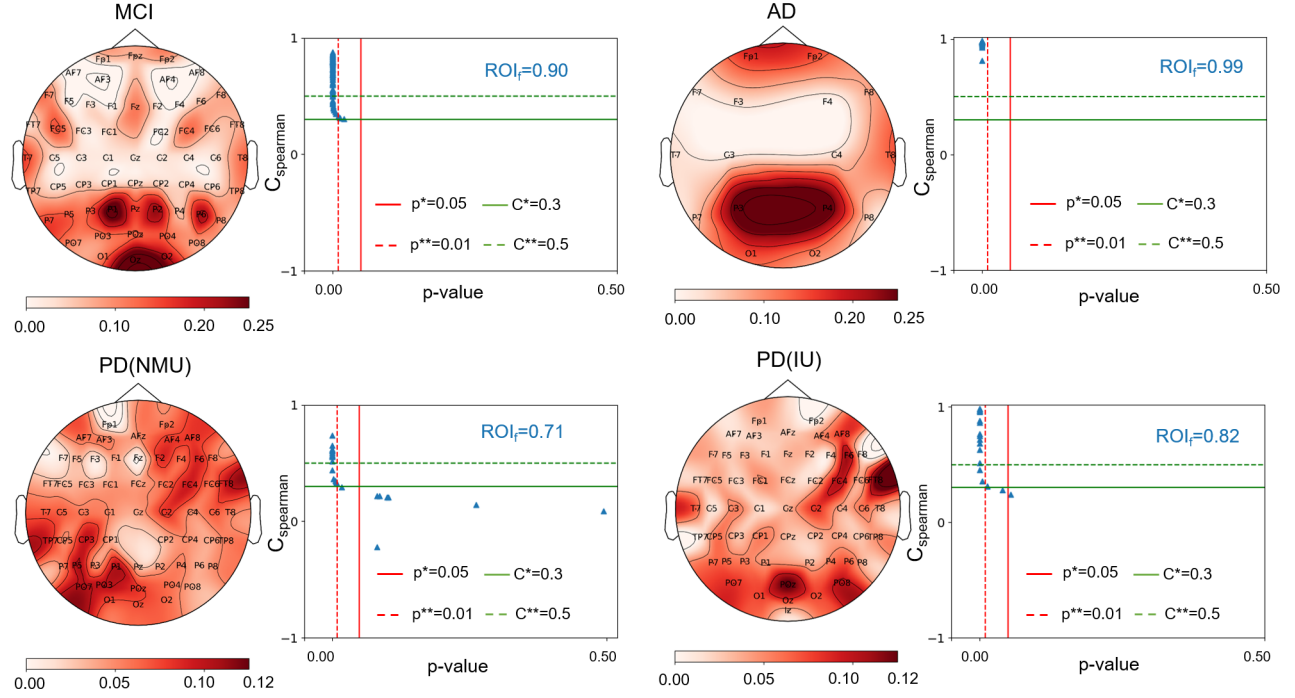
1 second was informed by the minimal length of PD data, which stands at 2 seconds. This configuration was strategically chosen to ensure that a minimum of one sample is capable of yielding two distinct data segments, thus meeting the structural prerequisites of our model. The emergence of more effective EEG markers at the larger temporal scales, as presented in Table A2, may be attributable to the 'slowing' phenomenon in brain activity, which is more prominently manifested at broader temporal scales in patients with neurodegenerative diseases [10, 11]

### C.3 Region of Interest Localized by MACS

We applied a class-discriminative activation algorithm inspired by the Grad-CAM method [12] to localize the region of interest (ROI) identified by the MACS framework while analyzing abnormalities in MCI, AD, and PD based on EEG signals.

To enhance our understanding, we computed the average activation across all samples within each dataset, resulting in the generation of four ROI heatmaps corresponding to MCI, AD, PD (MNU center), and PD (IU center). Concurrently, we aimed to assess





**Figure A3: The class-discriminative activation heatmaps generated by the MACS framework highlight localized Regions of Interest (ROIs) for detecting abnormalities in EEG signals under MCI, AD, and PD conditions. These heatmaps were averaged across all samples within each dataset, including the MCI, AD, and PD (from both the NMU and IU centers) datasets. The scatter plots illustrate the consistency between each sample and the averaged results, statistically quantified by the Spearman coefficient ( $C^*/C^{**}$ , indicating meaningful correlation) and the corresponding p-values ( $p^*/p^{**}$ , signifying significance).**

the uniformity of the sample-specific responses relative to their respective averaged ROI activation maps. To achieve this, we utilized the Spearman rank correlation method, as detailed in the referenced pipeline [13], calculating the correlation coefficients between the averaged activation maps and each sample, along with an overall quantification metric termed  $ROI_f$ . The higher  $ROI_f$  indicates greater consistency between all samples and their average result.

Here, a Spearman correlation coefficient ( $C_{Spearman}$ ) of 1 (or -1) indicates a perfect positive (or negative) correlation, while a coefficient of 0 signifies the absence of correlation. In each heatmap's right panel, the visualization includes blue triangle markers that identify individual samples. The correlation significance within these heatmaps is demarcated by two lines: a dotted green line at  $C_{Spearman} = 0.3$  differentiates between negligible and meaningful correlations, and a solid green line at  $C_{Spearman} = 0.5$  highlights correlations ranging from moderate to perfect. To establish the statistical relevance of these correlations, p-values were calculated. Correlations with p-values below 0.05 are marked by a solid red line, indicating statistical significance, while a dotted red line at  $p = 0.01$  delineates a more rigorous threshold for significance.

Similar patterns are observed in conditions of MCI and AD, notably with significant ROIs in the parietal-occipital and frontal-temporal lobes. This may be indicative of a pathological continuum from MCI to AD, with MCI being an earlier stage in the progression

towards the more pronounced dementia typical of AD. This evidence supports that the MACS framework effectively captures the core patterns in AD, potentially providing EEG evidence of compensatory mechanisms in the early stage [14]. Analogous ROIs, such as the right frontal-temporal lobes and left temporal-parietal lobes, are observed in PD conditions, despite the data originating from two different centers. This observation suggests that the model is capable of capturing shared features across diverse datasets, thereby enhancing its transferability.

## D DATA AVAILABILITY

The PD datasets are publicly accessible and can be downloaded from <http://predict.cs.unm.edu/>. Access to the MCI and AD datasets will be granted upon request to the corresponding author and subject to the approval of the collaborating hospitals following the publication of this work.

## E CODE AVAILABILITY

To facilitate the reproduction of our experiment, we have made the core code available anonymously at <https://anonymous.4open.science/r/EEG-Disease-MACS-0B4A>. Upon official publication, the complete project will be released publicly through our research group's GitHub repository, ensuring full access for the community.

Table A1: Comparative study of state-of-the-art methods for learning with well-annotated and unreliable annotated data.

Scenarios	Methods	Ratio	[MCI] 4-Fold				[PD] 3-fold				Overall Accuracy	
			Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1		
Clean Annotation	[7] [Ju et al., TNNLS, 2022]	Tensor-CSPNet	/	80.78(7.93)	81.57(10.34)	74.57(14.03)	77.70(11.63)	75.92(6.41)	81.94(6.36)	66.67(11.11)	73.20(8.16)	78.35(3.44)
	[8] [Pan et al., NeurIPS, 2022]	MAtt	/	81.97(3.96)	85.59(11.00)	75.80(13.60)	79.56(8.29)	79.63(11.56)	73.94(12.11)	<b>96.30(6.41)</b>	83.07(7.56)	80.80(1.65)
	[9] [Wang et al., NeurIPS, 2023]	COMET	/	73.25(26.02)	71.24(21.39)	69.70(26.69)	70.09(24.82)	75.47(9.75)	77.24(0.71)	75.54(11.06)	74.92(14.17)	74.36(1.57)
	<b>Encoder in MACS</b>	Final Epoch	/	<b>87.65(4.31)</b>	<b>88.89(10.39)</b>	<b>84.48(7.43)</b>	<b>86.47(7.73)</b>	<b>85.18(8.48)</b>	<b>90.00(10.00)</b>	81.48(23.13)	<b>83.46(11.67)</b>	<b>86.42(1.75)</b>
Unreliable Annotation	[5][Ma et al., IJCAI, 2023]	CTW	0.3	75.40(10.28)	74.37(15.59)	83.80(17.57)	76.75(9.80)	75.93(3.21)	70.71(3.50)	<b>88.89(0.00)</b>	78.73(2.20)	75.67(0.37)
			0.5	65.17(9.34)	68.01(19.02)	<b>75.82(17.90)</b>	<b>68.33(5.39)</b>	29.63(6.41)	23.59(20.94)	37.01(33.95)	28.79(25.85)	47.40(25.13)
	[4][Xiao et al., IJCAI, 2023]	Promix	0.3	55.14(18.94)	68.20(27.52)	60.13(22.21)	56.17(11.49)	79.63(8.49)	85.61(12.92)	74.08(23.13)	77.34(10.97)	67.39(17.32)
			0.5	42.94(24.19)	50.10(35.05)	33.07(25.86)	35.19(23.16)	42.59(25.66)	40.56(30.56)	37.04(27.96)	38.50(29.25)	42.77(0.25)
	[5] [Ma et al., IJCAI, 2023]	CTW Encoder + Promix	0.3	79.75(4.69)	74.80(11.99)	90.28(11.45)	80.83(7.21)	81.48(6.41)	89.74(17.77)	77.78(22.22)	80.25(8.15)	80.62(1.22)
	[4] [Xiao et al., IJCAI, 2023]		0.5	39.13(12.36)	48.31(35.83)	44.67(15.01)	42.23(15.87)	57.41(21.03)	66.40(19.93)	50.37(8.98)	54.17(18.17)	48.27(12.93)
	[3] [Li et al., CVPR, 2022]	Sel-CL	0.3	63.29(22.32)	73.49(11.27)	45.47(22.52)	54.82(19.42)	66.67(5.56)	68.81(7.84)	62.97(6.41)	65.42(4.49)	64.98(2.39)
			0.5	41.65(20.95)	47.56(15.55)	52.31(9.01)	48.42(10.22)	<b>66.67(9.62)</b>	<b>87.50(21.65)</b>	44.45(19.25)	55.54(17.76)	54.16(17.69)
	[3] [Li et al., CVPR, 2022]	Sel-CL+	0.3	74.16(7.73)	73.56(15.84)	76.15(15.33)	73.41(11.98)	57.41(21.03)	51.59(45.07)	37.04(32.08)	43.06(37.35)	65.79(11.84)
			0.5	50.74(20.52)	53.81(9.67)	53.78(6.80)	53.53(7.55)	64.82(3.21)	77.78(19.24)	51.85(25.66)	<b>56.57(17.50)</b>	57.78(9.96)
	[4][Xiao et al., IJCAI, 2023]	Encoder for Promix	0.3	85.38(4.41)	84.08(11.74)	88.89(10.39)	85.56(4.90)	83.33(5.56)	<b>93.33(11.55)</b>	74.08(16.97)	81.05(8.49)	84.36(1.45)
			0.5	59.39(16.10)	57.49(32.70)	46.28(29.29)	50.10(28.77)	33.33(5.56)	20.63(18.03)	22.22(22.22)	21.03(19.36)	46.36(18.43)
	[3] [Li et al., CVPR, 2022]	Encoder for Sel-CL	0.3	72.08(12.76)	85.52(17.61)	65.50(24.78)	70.05(9.37)	83.33(5.56)	89.63(10.02)	77.78(19.24)	81.51(9.04)	77.71(7.95)
			0.5	61.56(27.09)	63.28(26.34)	75.35(20.23)	66.15(21.75)	38.89(24.22)	32.22(42.21)	18.52(23.13)	23.49(29.89)	50.23(16.03)
	[3] [Li et al., CVPR, 2022]	Encoder for Sel-CL+	0.3	85.43(4.16)	<b>87.94(11.70)</b>	83.44(15.74)	84.18(7.01)	83.33(5.56)	<b>93.33(11.55)</b>	74.08(16.97)	81.05(8.49)	84.38(1.48)
			0.5	68.38(21.49)	81.18(26.52)	60.07(27.99)	63.89(17.70)	40.74(17.86)	40.36(19.19)	37.04(16.98)	38.47(17.93)	54.56(19.54)
	<b>MACS</b>	Final Epoch	0.3	<b>88.74(4.61)</b>	86.15(12.39)	<b>91.16(5.92)</b>	<b>88.18(7.23)</b>	<b>87.04(3.21)</b>	<b>93.33(11.55)</b>	81.48(6.41)	<b>86.40(1.90)</b>	<b>87.89(1.20)</b>
			0.5	<b>68.68(10.34)</b>	<b>83.47(13.61)</b>	52.39(24.08)	61.07(13.00)	57.41(21.03)	56.67(24.34)	<b>51.85(27.96)</b>	53.73(25.50)	<b>63.05(7.97)</b>

Table A2: The investigation of hyper-parameters in MACS includes: 'Memory Length', which represents the number of previous mini-batches used for storing pairs; 'Temporal Scale', which denotes the scale of time considered for constructing dynamic functional networks in manifold mapping; and 'K Value', which indicates the number of nearest neighbors considered in the k-Nearest Neighbors approach.

[MCI] 4-Fold Cross-Validation						[PD] 3-Fold Cross-Validation				
Parameters	Configuration	Accuracy	Precision	Recall	F1	Configuration	Accuracy	Precision	Recall	F1
Memory Length	0	83.10(6.93)	78.41(17.67)	<b>92.46(8.95)</b>	83.64(9.09)	0	85.18(3.21)	85.46(4.78)	<b>85.19(6.41)</b>	85.15(3.37)
	200	87.60(4.46)	<b>87.29(12.65)</b>	86.30(5.96)	86.39(7.58)	50	83.33(0.00)	81.76(5.09)	83.01(5.58)	83.30(0.93)
	300	<b>88.74(4.61)</b>	86.15(12.39)	91.16(5.92)	<b>88.18(7.23)</b>	100	<b>87.04(3.21)</b>	<b>93.33(11.55)</b>	81.48(6.41)	<b>86.40(1.90)</b>
	400	86.31(5.55)	85.36(12.25)	85.60(13.86)	84.71(9.69)	200	85.18(3.21)	89.17(10.10)	81.48(6.41)	84.69(2.61)
	500	85.33(4.52)	84.67(13.37)	84.21(4.80)	84.07(7.77)	300	83.33(0.00)	89.17(10.10)	77.78(11.11)	82.19(2.11)
Temporal Scale	2s	<b>88.74(4.61)</b>	86.15(12.39)	91.16(5.92)	<b>88.18(7.23)</b>	1s	<b>87.04(3.21)</b>	<b>93.33(11.55)</b>	<b>81.48(6.41)</b>	<b>86.40(1.90)</b>
	1s	87.65(4.31)	<b>87.12(10.98)</b>	87.26(10.90)	86.75(7.87)	500ms	81.48(3.20)	89.17(10.10)	74.08(16.97)	79.33(6.90)
	500ms	74.06(6.09)	67.24(13.56)	<b>94.36(7.86)</b>	77.60(8.15)	250ms	83.33(0.00)	89.17(10.10)	77.78(11.11)	82.19(2.11)
K Value	15	86.46(3.87)	<b>93.03(8.13)</b>	76.92(18.90)	82.69(11.65)	8	83.33(0.00)	85.00(4.33)	<b>81.48(6.41)</b>	82.97(1.07)
	20	87.60(4.46)	82.68(13.28)	<b>93.94(7.01)</b>	87.39(7.30)	12	83.33(0.00)	89.17(10.10)	77.78(11.11)	82.19(2.11)
	25	<b>88.74(4.61)</b>	86.15(12.39)	91.16(5.92)	<b>88.18(7.23)</b>	16	<b>87.04(3.21)</b>	<b>93.33(11.55)</b>	<b>81.48(6.41)</b>	<b>86.40(1.90)</b>
	30	87.60(4.46)	82.96(12.50)	<b>93.94(7.01)</b>	87.61(6.87)	20	85.18(3.21)	<b>93.33(11.55)</b>	77.78(11.11)	83.90(3.76)

REFERENCES

[1] Cristina Gil Ávila, Felix S Bott, Laura Tiemann, Vanessa D Hohn, Elisabeth S May, Moritz M Nickel, Paul Theo Zebhauser, Joachim Gross, and Markus Ploner. Discover-eeg: an open, fully automated eeg pipeline for biomarker discovery in clinical neuroscience. *Scientific Data*, 10:613, 2023.

[2] William J Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950.

[3] Shikun Li, Xiaobo Xia, Shiming Ge, and Tongliang Liu. Selective-supervised contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[4] Ruixuan Xiao, Yiwen Dong, Haobo Wang, Lei Feng, Runze Wu, Gang Chen, and Junbo Zhao. Promix: Combating label noise via maximizing clean sample utility. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, 2023.

[5] Peitian Ma, Zhen Liu, Junhao Zheng, Linghao Wang, and Qianli Ma. Ctw: confident time-warping for time-series label-noise learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, 2023.

[6] Diego Ortego, Eric Arazo, Paul Albert, Noel E O'Connor, and Kevin McGuinness. Multi-objective interpolation training for robustness to label noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[7] Ce Ju and Cuntai Guan. Tensor-cspnet: A novel geometric deep learning framework for motor imagery classification. *IEEE Transactions on Neural Networks and Learning Systems*, 34(12):10955–10969, 2023.

[8] Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. Matt: a manifold attention network for eeg decoding. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[9] Yihe Wang, Yu Han, Haishuai Wang, and Xiang Zhang. Contrast everything: A hierarchical contrastive framework for medical time-series. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

[10] Aurora D'Atri, Serena Scarpelli, Maurizio Gorgoni, Ilaria Truglia, Giulia Lauri, Susanna Cordone, Michele Ferrara, Camillo Marra, Paolo Maria Rossini, and Luigi De Gennaro. Eeg alterations during wake and sleep in mild cognitive impairment and alzheimer's disease. *Isience*, 24(4), 2021.

[11] D Jones, V Lowe, J Graff-Radford, H Botha, L Barnard, D Wiepert, MC Murphy, M Murray, M Senjem, J Gunter, et al. A computational model of neurodegeneration in alzheimer's disease. *Nature communications*, 13(1):1643, 2022.

[12] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 2017.

[13] Zhenxi Song, Bin Deng, Jiang Wang, and Guosheng Yi. An eeg-based systematic explainable detection framework for probing and localizing abnormal patterns in alzheimer's disease. *Journal of Neural Engineering*, 19(3):036007, 2022.

[14] Sinead Gaubert, Federico Raimondo, Marion Houot, Marie-Constance Corsi, Lionel Naccache, Jacobo Diego Sitt, Bertrand Hermann, Delphine Oudiette, Geoffroy Gagliardi, Marie-Odile Habert, et al. Eeg evidence of compensatory mechanisms in preclinical alzheimer's disease. *Brain*, 142(7):2096–2112, 2019.