Deep Counterfactual-Guided MRI Feature Representation and Quantitatively Interpretable Alzheimer's Disease Prediction

Kwanseok Oh[1], Da-Woon Heo[1], Ahmad Wisnu Mulyadi[2], Wonsik Jung[2], Eunsong Kang[2], Kun Ho Lee[3,4,5], and Heung-Il Suk[1,2,*]

[1]Department of Artificial Intelligence, Korea University, Seoul 02841, Korea.
[2]Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Korea.
[3]Gwangju Alzheimer's & Related Dementia Cohort Research Center, Chosun University, Gwangju 61452, Republic of Korea.
[4]Department of Biomedical Science, Chosun University, Gwangju 61452, Republic of Korea.
[5]Korea Brain Research Institute, Daegu 41062, Korea.
[*]Corresponding author: hisuk@korea.ac.kr

A Data Description

To demonstrate the validity of our study, we used the ADNI-1, ADNI-2, and GARD datasets in all the experiments. The ADNI dataset consisted of 1,540 subjects, which included cognitively normal (CN) subjects (n = 433), stable mild cognitive impairment (sMCI) subjects (n = 497), progressive mild cognitive impairment (pMCI) subjects (n = 251), and AD subjects (n = 359) in the ADNI-1 and ADNI-2 studies. The GARD dataset consisted of 745 subjects, which included CN subjects (n = 261), MCI subjects (n = 375), and AD subjects (n = 109) in the GARD studies. Because the GARD dataset was comprised of three class categories (*i.e.*, CN, MCI, and AD), we grouped the sMCI and pMCI subjects in the ADNI dataset as MCI subjects to maintain consistency with the GARD dataset.

For the preprocessing of raw brain images, we first performed neck removal (FSL [1] v6.0.1/v6.0.4 robustfov), followed by a brain extraction procedure using the HD-BET [2] to eliminate non-brain tissue from the image of the whole head. Note that we have used default parameters for these preprocessing steps, whereas the version of FSL was different in that ADNI and GARD datasets are preprocessed to FSL v6.0.1 and FSL v6.0.4, respectively. These skull-stripped images were aligned to the MNI152 template using the linear image registration tool (FLIRT) from the FMRIB software library (v6.0.1/v6.0.4). We then conducted down-scaling by 2× (to mitigate/anticipate the prevailing memory issue because of high dimensionality), quantile normalization at 10% and 90%, and zero-mean unit-variance normalization sequentially to use the images as input for the deep learning model training. As a result, we obtained the preprocessed 3D MRI scans, each with a dimensionality of $96 \times 114 \times 96$.

Table S1: Demographic details for the ADNI (*i.e.*, ADNI-1 and ADNI-2) and GARD datasets. Gender is presented as male/female. Age, education, and MMSE scores are indicated as mean ± standard deviation.

Dataset	Category	Gender	Age	Education	MMSE
ADNI-1	CN	119/112	76.0±5.0	16.1±2.8	29.1±1.0
	sMCI	148/75	74.8±7.7	15.5±3.2	27.3±1.8
	pMCI	103/65	74.7±7.0	15.7±2.8	26.6±1.7
	AD	103/97	75.6±7.7	14.7±3.2	23.3±2.0
ADNI-2	CN	95/107	73.5±6.6	16.4±2.6	29.0±1.3
	sMCI	151/123	71.3±7.5	16.3±2.6	28.2±1.6
	pMCI	45/38	72.9±7.2	16.2±2.3	27.1±1.8
	AD	91/68	74.9±8.1	15.7±2.7	23.1±2.1
GARD	CN	120/141	70.2±6.1	11.0±4.6	27.2±2.2
	MCI	176/199	71.6±7.1	10.5±4.8	25.4±3.2
	AD	56/53	74.9±7.1	8.30±4.8	19.0±5.4

B Predictive Model Performance

As a pre-trained predictive model for generating counterfactual maps, we adopted ResNet18 [3] as a convolutional neural network (CNN)-based predictive model owing to its widespread use and proven performance, even in the medical field. Our predictive model was trained for three scenarios (*i.e.*, CN *vs.* MCI, MCI *vs.* AD, and CN *vs.* AD) on ADNI and GARD datasets, respectively, and the classification results are reported in Table S2.

Table S2: Classification outcomes for ResNet18 in various binary scenarios using r-sMRI data on ADNI and GARD datasets.

Model	Dataset	Scenario	AUC	Accuracy	Sensitivity	Specificity
ResNet18	ADNI	CN *vs.* MCI	0.6209±0.056	0.6479±0.037	0.7487±0.146	0.4931±0.218
		MCI *vs.* AD	0.7696±0.087	0.7965±0.065	0.6943±0.091	0.8449±0.083
		CN *vs.* AD	0.8878±0.036	0.8898±0.036	0.8686±0.043	0.9070±0.055
	GARD	CN *vs.* MCI	0.6179±0.027	0.6254±0.026	0.8073±0.084	0.4385±0.103
		MCI *vs.* AD	0.7084±0.074	0.7959±0.018	0.6546±0.181	0.9122±0.036
		CN *vs.* AD	0.8343±0.044	0.8703±0.029	0.7455±0.099	0.9231±0.038

C Details of Hyperparameters and Implementations

We implemented our proposed LiCoL and all comparative models, including ML and DL models (*i.e.*, predictive model and the LEAR framework's CMG) using Python 3.9 and TensorFlow2 on an Intel Xeon Silver4216 CPU @ 2.10 GHz and a single NVIDIA RTX A6000 GPU on Ubuntu 18.04. Specifically, ML-based model implementation was conducted via the Scikit-learn library [4] in the Python module, which includes a wide variety of machine learning algorithms.

The best-performing model hyperparameters are shown in Tables S3, S4, S5, and S6. The predictive model, LiCoL, and CMG used the same hyperparameters for all scenarios (*i.e.*, CN *vs.* MCI, MCI *vs.* AD, and CN *vs.* AD) and datasets (*i.e.*, ADNI and GARD) in each given task. Note that the arguments of ML-based methods that are not mentioned in the table are default values.

Table S3: ADNI and GARD models' hyperparameters for the predictive model training.

Epochs & Batch Size	150, 12
Optimizer	Adam ($\beta_1 = 0.9, \beta_2 = 0.999$)
Learning Rate	0.0001
Exponential Decay Rate	0.98

Table S4: ADNI and GARD models' hyperparameters for the LiCoL training.

Epochs & Batch Size	100, 6
Optimizer	Adam ($\beta_1 = 0.9, \beta_2 = 0.999$)
(Embedding) D	256
Learning Rate	0.0001
Exponential Decay Rate	0.98

Table S5: ADNI and GARD models' hyperparameters for the CMG training.

Epochs & Batch Size	100, 3
Optimizer	Adam ($\beta_1 = 0.9, \beta_2 = 0.999$)
Learning Rate	$\mathcal{G}_\phi, \mathcal{D}_\psi = 0.01$
Exponential Decay Rate	1.0
Weight Constants	$\lambda_1 = 10, \lambda_2 = 10, \lambda_3 = 10, \lambda_4 = 5, \lambda_5 = 1, \lambda_6 = 1, \lambda_7 = 5e^{-6}$

Table S6: Hyperparameters for comparative ML-based methods training.

Arguments	**Models**	**CN ↔ MCI**		**MCI ↔ AD**		**CN ↔ AD**	
		ADNI	**GARD**	**ADNI**	**GARD**	**ADNI**	**GARD**
Estimators	RF	15	15	10	10	5	5
C	SVM	5	5	5	5	2	2
	LR	2	2	1	1	1	2
Max depth	DTs	2	2	3	3	3	3
Alpha	Lasso	0.1	0.1	0.1	0.1	0.1	0.1
	Ridge	1	1	0.9	0.9	1	1

D Details of Longitudinal Samples

We produced ground-truth for the counterfactual maps using MRIs from longitudinal subjects. First, we gathered MRIs from subjects who converted from CN (baseline) to MCI and AD at any later point in time. There were 12 subjects and 9 subjects gathered, respectively, from ADNI and GARD. Note that in the GARD longitudinal dataset, because there was only one subject exhibiting all the clinical stages (*i.e.*, CN, MCI, and AD), longitudinal samples that corresponded to each scenario were selected and used.

Table S7: Description of ADNI longitudinal subjects for the qualitative and quantitative evaluations.

Subject ID	**Cognitively normal (CN)** Image ID	**Mild cognitive impairment (MCI)** Image ID	**Alzheimer's disease (AD)** Image ID
023_S_0061	9046	401795	473765
123_S_0106	10126	213947	865961
131_S_0123	10042	292389	475755
005_S_0223	11645	26115	143296
037_S_0467	14861	169074	372375
129_S_0778	20543	205611	388698
024_S_0985	27607	342890	396124
023_S_1190	35585	780686	1135165
002_S_4262	259653	397601	788894
029_S_4385	285589	421501	642470
098_S_4506	286987	714545	1073489
051_S_5285	396288	1055626	1182436

Table S8: Description of GARD longitudinal subjects for the qualitative and quantitative evaluations.

Object ID	**Cognitively normal (CN)** Image ID	**Mild cognitive impairment (MCI)** Image ID	**Alzheimer's disease (AD)** Image ID
271	-	14050409	15120104
61	14041307	15112501	17071905
319	-	14051704	18010803
408	14053104	-	18030704
6367	17040504	-	20030203
198	14072008	15072801	-
2901	15071403	17030803	-
121	17051601	18062103	-
269	14050312	16030905	-

E Criteria of Threshold Value Selection

Table S9: An ablation study of the validation performance (AUC) comparison according to each threshold (*i.e.*, percentile and p-value) in the AD-effect map and statistical map to derive the significant ROIs. We refer to the set of AD-effect ROIs and statistical ROIs as $\mathcal{R}_{\text{eff}}$ and $\mathcal{R}_{\text{stat}}$, respectively. To assess the validity of those ROIs, we exploited our LiCoL as a classifier for the prediction.

Datasets	**Scenarios**	**Percentile ($\mathcal{R}_{\text{eff}}$)**			**P-value ($\mathcal{R}_{\text{stat}}$)**		
		0.99	**0.95**	**0.90**	**0.05**	**0.01**	**0.001**
ADNI	CN *vs.* MCI	**0.7794±0.023**	0.7549±0.015	0.7362±0.027	0.7084±0.036	**0.7168±0.033**	0.6817±0.041
	MCI *vs.* AD	**0.7965±0.025**	0.7594±0.014	0.7422±0.035	0.7346±0.021	**0.7623±0.021**	0.7232±0.019
	CN *vs.* AD	**0.9412±0.014**	0.9268±0.021	0.9121±0.009	0.8827±0.011	**0.9084±0.011**	0.8724±0.016
GARD	CN *vs.* MCI	**0.7906±0.036**	0.7694±0.017	0.7516±0.037	0.7192±0.031	**0.7314±0.027**	0.7008±0.053
	MCI *vs.* AD	**0.7997±0.022**	0.7639±0.016	0.7523±0.031	0.7302±0.023	**0.7586±0.028**	0.7186±0.024
	CN *vs.* AD	**0.9274±0.016**	0.9233±0.026	0.9171±0.019	0.8926±0.020	**0.9176±0.012**	0.8885±0.025

F ML-based Models' Classification Performance

Table S10: Classification outcomes in various scenarios using statistical ROIs with comparative ML-based models on ADNI and GARD datasets. For the ML-based models, we use six models, which are random forest (RF), support vector machine (SVM), logistic regression (LR), decision tree (DTs), Lasso regression, and Ridge regression.

Scenarios	Datasets	Model	AUC	Accuracy	Sensitivity	Specificity
CN *vs.* MCI	ADNI	RF	0.5721±0.036	0.5255±0.038	0.6676±0.071	0.5231±0.057
		SVM	0.5278±0.036	0.5481±0.031	0.6361±0.046	0.5313±0.037
		LR	0.5046±0.041	0.5065±0.044	0.5298±0.057	0.6287±0.051
		DTs	0.5281±0.032	0.5486±0.027	0.6075±0.031	0.6278±0.023
		Lasso	0.5187±0.038	0.5192±0.041	0.5639±0.052	0.6232±0.042
		Ridge	0.5225±0.044	0.5276±0.037	0.5774±0.041	0.6386±0.031
	GARD	RF	0.5808±0.044	0.5761±0.036	0.6673±0.063	0.5894±0.075
		SVM	0.5667±0.054	0.5819±0.042	0.7091±0.048	0.5276±0.069
		LR	0.5337±0.062	0.5438±0.038	0.5014±0.072	0.6781±0.046
		DTs	0.5583±0.045	0.5479±0.051	0.6998±0.037	0.5613±0.045
		Lasso	0.5471±0.052	0.5698±0.036	0.6228±0.051	0.5790±0.035
		Ridge	0.5534±0.036	0.5535±0.041	0.6416±0.034	0.5619±0.032
MCI *vs.* AD	ADNI	RF	0.6247±0.035	0.6342±0.036	0.4683±0.041	0.7013±0.044
		SVM	0.6317±0.043	0.6452±0.035	0.5186±0.030	0.6869±0.046
		LR	0.5969±0.043	0.6267±0.048	0.5317±0.062	0.6484±0.057
		DTs	0.6123±0.031	0.6238±0.037	0.5074±0.045	0.7227±0.031
		Lasso	0.6093±0.046	0.6311±0.039	0.4889±0.036	0.6849±0.042
		Ridge	0.6079±0.043	0.6214±0.036	0.5025±0.033	0.6798±0.038
	GARD	RF	0.6371±0.041	0.6821±0.037	0.7017±0.048	0.4818±0.056
		SVM	0.6418±0.032	0.6039±0.031	0.7984±0.089	0.5372±0.063
		LR	0.5901±0.062	0.5718±0.056	0.7852±0.047	0.4727±0.101
		DTs	0.6326±0.022	0.6578±0.027	0.5082±0.046	0.7894±0.055
		Lasso	0.6189±0.034	0.6084±0.026	0.5468±0.053	0.7172±0.038
		Ridge	0.6212±0.043	0.6173±0.027	0.5369±0.046	0.7222±0.053
CN *vs.* AD	ADNI	RF	0.7713±0.054	0.7841±0.052	0.6588±0.092	0.7781±0.074
		SVM	0.8198±0.043	0.8195±0.031	0.6991±0.067	0.8712±0.055
		LR	0.6823±0.033	0.7053±0.028	0.6393±0.061	0.7692±0.035
		DTs	0.7614±0.032	0.7916±0.036	0.6431±0.047	0.8257±0.038
		Lasso	0.7681±0.036	0.7803±0.024	0.6989±0.035	0.8008±0.042
		Ridge	0.7287±0.034	0.7558±0.032	0.7181±0.025	0.7993±0.041
	GARD	RF	0.7483±0.051	0.7217±0.041	0.5531±0.116	0.8019±0.046
		SVM	0.7619±0.035	0.7886±0.048	0.6417±0.056	0.6985±0.051
		LR	0.6801±0.021	0.6917±0.015	0.6219±0.048	0.7391±0.054
		DTs	0.7368±0.021	0.7192±0.031	0.5051±0.019	0.6792±0.021
		Lasso	0.7312±0.038	0.7381±0.037	0.6583±0.029	0.7338±0.023
		Ridge	0.7402±0.033	0.7309±0.022	0.6723±0.034	0.7513±0.027

Table S11: Classification outcomes in various scenarios using cGMs as augmented data with comparative ML-based models on ADNI and GARD datasets. Here, the symbols * and ** for AUC and accuracy denote the statistical significance with the Wilcoxon signed-rank test at $p < 0.05$ and $p < 0.01$, respectively, when compared with our model's performance (*i.e.*, Ours). Notably, despite learning by adding augmented data, our proposed method outperforms the ML-based models on all the evaluation criteria (*i.e.*, AUC, accuracy, sensitivity, and specificity) in all scenarios and datasets.

Scenarios	Datasets	Model	AUC	Accuracy	Sensitivity	Specificity
CN *vs.* MCI	ADNI	RF	0.6992±0.028**	0.6669±0.032**	0.6837±0.022	0.7818±0.025
		SVM	0.6721±0.021**	0.6882±0.028**	0.6655±0.025	0.7385±0.011
		LR	0.6015±0.018**	0.6085±0.020**	0.6692±0.021	0.7127±0.035
		DTs	0.6381±0.014**	0.6288±0.025**	0.6595±0.034	0.7571±0.019
		Lasso	0.6583±0.024**	0.6376±0.023**	0.6717±0.034	0.7281±0.028
		Ridge	0.6678±0.027**	0.6413±0.022**	0.6778±0.026	0.7418±0.023
		Ours	**0.7678±0.013**	**0.7562±0.015**	**0.7143±0.021**	**0.8182±0.015**
	GARD	RF	0.7083±0.024**	0.6983±0.031**	0.6935±0.026	0.7689±0.021
		SVM	0.7145±0.022**	0.7059±0.031**	0.7281±0.044	0.6798±0.031
		LR	0.6691±0.038**	0.6583±0.033**	0.5891±0.041	0.7289±0.018
		DTs	0.6783±0.031**	0.6737±0.031**	0.7082±0.022	0.6081±0.034
		Lasso	0.6829±0.025**	0.6915±0.031**	0.6684±0.024	0.6286±0.027
		Ridge	0.6895±0.031**	0.6943±0.027**	0.6781±0.025	0.6319±0.037
		Ours	**0.7778±0.020**	**0.7662±0.019**	**0.7485±0.024**	**0.8246±0.017**
MCI *vs.* AD	ADNI	RF	0.7382±0.013*	0.7795±0.011	0.7714±0.022	0.8825±0.028
		SVM	0.7519±0.013	0.7329±0.021**	0.6657±0.030	0.7785±0.029
		LR	0.6904±0.021**	0.7081±0.023**	0.5879±0.034	0.7885±0.018
		DTs	0.7029±0.023**	0.7342±0.019**	0.6286±0.025	0.8085±0.031
		Lasso	0.7182±0.027**	0.7424±0.021*	0.6471±0.026	0.7912±0.025
		Ridge	0.7205±0.034*	0.7388±0.019**	0.6368±0.015	0.7867±0.021
		Ours	**0.7727±0.018**	**0.7921±0.016**	**0.8014±0.022**	**0.8913±0.019**
	GARD	RF	0.7197±0.028**	0.7451±0.021*	0.8596±0.038	0.7287±0.031
		SVM	0.7284±0.025*	0.7386±0.028*	0.8329±0.034	0.7391±0.036
		LR	0.6745±0.031**	0.7003±0.035**	0.7685±0.019	0.5579±0.023
		DTs	0.6974±0.024**	0.7228±0.014**	0.7984±0.032	0.6092±0.039
		Lasso	0.7183±0.031**	0.7352±0.024*	0.7641±0.038	0.6481±0.024
		Ridge	0.7196±0.019*	0.7315±0.023*	0.7787±0.024	0.6296±0.041
		Ours	**0.7753±0.016**	**0.7814±0.014**	**0.8917±0.011**	**0.7955±0.023**
CN *vs.* AD	ADNI	RF	0.9082±0.018*	0.9086±0.012*	0.8918±0.023	0.8571±0.017
		SVM	0.9167±0.009*	0.9167±0.011	0.8889±0.023	0.9444±0.015
		LR	0.7819±0.024*	0.7811±0.030**	0.7015±0.045	0.8492±0.011
		DTs	0.8668±0.013**	0.8345±0.018**	0.6993±0.021	0.9017±0.020
		Lasso	0.8791±0.011**	0.8591±0.013**	0.8097±0.026	0.8921±0.019
		Ridge	0.8788±0.017**	0.8681±0.022**	0.7841 ±0.028	0.8312±0.033
		Ours	**0.9371±0.005**	**0.9285±0.011**	**0.9271±0.017**	**0.9591±0.015**
	GARD	RF	0.8818±0.021*	0.8967±0.012*	0.8913±0.024	0.8791±0.032
		SVM	0.8774±0.021*	0.8533±0.024**	0.8912±0.014	0.8778±0.012
		LR	0.8123±0.014**	0.8174±0.021**	0.6787±0.012	0.8025±0.011
		DTs	0.8192±0.007**	0.8352±0.016**	0.6517±0.031	0.8291±0.020
		Lasso	0.8541±0.023**	0.8278±0.024**	0.8513±0.024	0.8291±0.023
		Ridge	0.8603±0.022**	0.8391±0.027**	0.8587±0.017	0.8336±0.014
		Ours	**0.9304±0.008**	**0.9193±0.007**	**0.9441±0.010**	**0.9283±0.012**

G Intra-Subject ADness Index Visualization

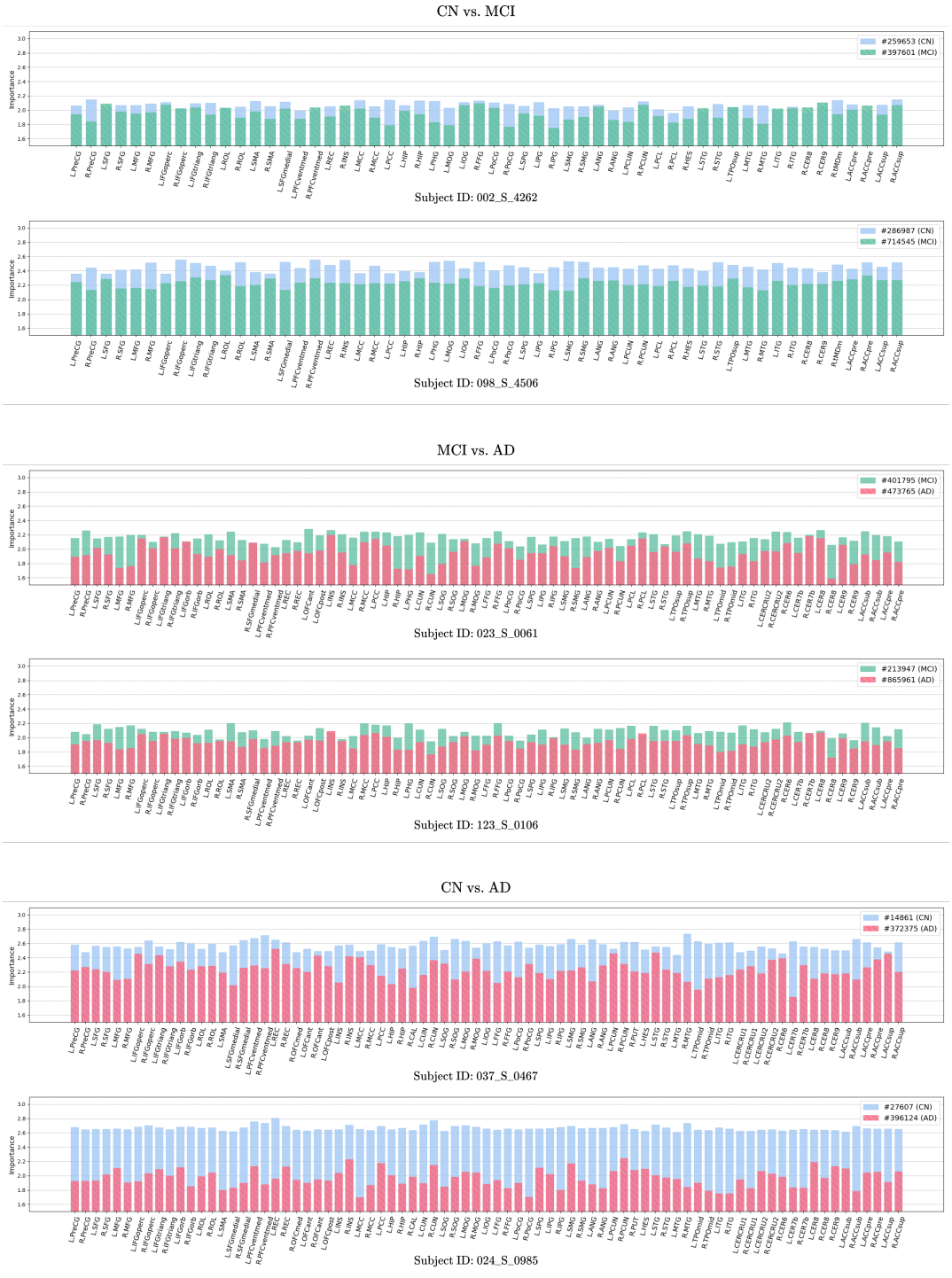

Figure S1: Illustration of importance scores (*i.e.*, $\mathbf{Ax}_s$) for each AD-effect ROI $\mathcal{R}_{\text{eff}}$ based on randomly selected longitudinal samples on ADNI. Here, "importance" refers to the quantitative figures of the ADness index that contributed to the prediction of each patient's disease. Note that the subject ID of the selected longitudinal sample is presented below the center of each bar plot, and the image ID corresponding to the clinical stage is exhibited in the legend in the upper right corner.

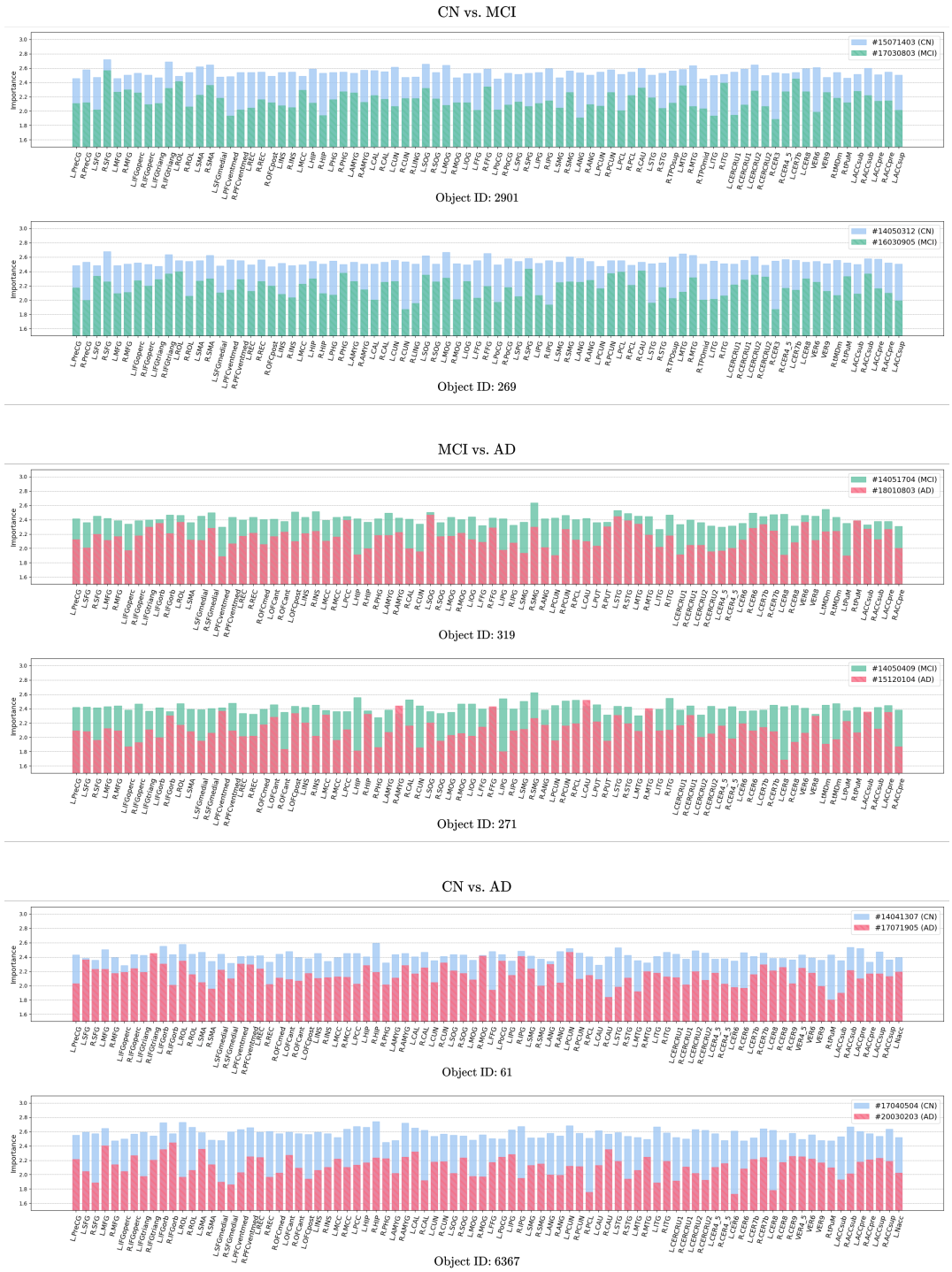

Figure S2: Illustration of importance scores (*i.e.*, $\mathbf{Ax}_s$) for each AD-effect ROI $\mathcal{R}_{\text{eff}}$ based on randomly selected longitudinal samples on GARD. Here, "importance" refers to the quantitative figures of the ADness index that contributed to the prediction of each patient's disease. Note that the object ID of the selected longitudinal sample is presented below the center of each bar plot, and the image ID corresponding to the clinical stage is exhibited in the legend in the upper right corner.

H Proof of LiCoL's Transparency

The neural network of our proposed LiCoL possesses inherent interpretability, as it was developed without any non-linearity, except for the counterfactual-guided attention map calculation (*i.e.*, usage of the softmax function for the AD-effect ROIs scoring). That is, solely by examining the respectively produced counterfactual-guided attention map, LiCoL readily allows for a numerical interpretation, based on the persuasive rationale via the ADness index regardless of additional module training or computation, to infer the classification decision. Therefore, we have shown the mathematical development of how LiCoL smoothly transforms into a linear function, as shown below.

We exploit the formulated LiCoL below to perform the mathematical development:

$$\bar{y} = \sigma\left(\mathrm{MP}_{\downarrow}\left(\mathrm{MP}_{\rightarrow}\left(g\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d}}\right)\mathbf{V}\right) + \mathbf{x}_q\right)\right) \tag{1}$$

For brevity, we substitute the expression $\mathbf{A}$ for the counterfactual-guided attention map $g\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d}}\right)$ in the first equation:

$$\bar{y} = \sigma\left(\mathrm{MP}_{\downarrow}\left(\mathrm{MP}_{\rightarrow}\left(\mathbf{A}\mathbf{x}_s\mathbf{w}_v\right) + \mathbf{x}_q\right)\right), \;\; \text{for} \;\; \mathbf{V} = \mathbf{x}_s\mathbf{w}_v \tag{2}$$

Thereafter, factorization is carried out for ROI-wise mean pooling $\mathrm{MP}_{\downarrow}$ as:

$$\bar{y} = \sigma\left(\mathrm{MP}_{\downarrow}\,\mathrm{MP}_{\rightarrow}\left(\mathbf{A}\mathbf{x}_s\mathbf{w}_v\right) + \mathrm{MP}_{\downarrow}(\mathbf{x}_q)\right) \tag{3}$$

Because $\mathrm{MP}_{\downarrow}(\mathbf{x}_q)$ is a constant term, substituting it with c_1 and performing the factorization on the first term proceeds as follows:

$$\bar{y} = \sigma\left(\mathrm{MP}_{\downarrow}\,\mathrm{MP}_{\rightarrow}\left(\mathbf{A}\mathbf{x}_s\right)\mathrm{MP}_{\downarrow}\,\mathrm{MP}_{\rightarrow}\left(\mathbf{w}_v\right) + c_1\right) \tag{4}$$

Similarly, $\mathrm{MP}_{\downarrow}\,\mathrm{MP}_{\rightarrow}\left(\mathbf{w}_v\right)$ can also be defined as a constant term. If we substitute c_2, we eventually derive an equation in the form of a linear function:

$$\bar{y} = \sigma\left(\mathrm{MP}_{\downarrow}\,\mathrm{MP}_{\rightarrow}\left(\mathbf{A}\mathbf{x}_s\right)c_2 + c_1\right) \tag{5}$$

Consequently, within Eq. (5), we are able to prove that the LiCoL decision relied on the counterfactual-guided attention map $\mathbf{A}$ multiplied by each patient $\mathbf{x}_s$ based on scaling (*i.e.*, c_2) and bias (*i.e.*, c_1) to perform the prediction. Moreover, by utilizing this equation for the analysis, we account for where LiCoL attended individually for each patient among the AD-effect ROIs that primarily contributed to the prediction.

I Anatomical Description on AAL Atlas

Table S12: The anatomical regions defined in the automated anatomical labelling atlas 3 [5].

No.	Anatomical Description	Label AAL3	Possible Abbreviation
1,2	Precentral gyrus	Precentral	PreCG
3, 4	Superior frontal gyrus, dorsolateral	Frontal_Sup	SFG
5, 6	Middle frontal gyrus	Frontal_Mid	MFG
7, 8	Inferior frontal gyrus, opercular part	Frontal_Inf_Oper	IFGoperc
9, 10	Inferior frontal gyrus, triangular part	Frontal_Inf_Tri	IFGtriang
11, 12	IFG pars orbitalis,	Frontal_Inf_Orb	IFGorb
13, 14	Rolandic operculum	Rolandic_Oper	ROL
15, 16	Supplementary motor area	Supp_Motor_Area	SMA
17, 18	Olfactory cortex	Olfactory	OLF
19, 20	Superior frontal gyrus, medial	Frontal_Sup_Med	SFGmedial
21, 22	Superior frontal gyrus, medial orbital	Frontal_Med_Orb	PFCventmed
23, 24	Gyrus rectus	Rectus	REC
25, 26	Medial orbital gyrus	OFCmed	OFCmed
27, 28	Anterior orbital gyrus	OFCant	OFCant
29, 30	Posterior orbital gyrus	OFCpost	OFCpost
31, 32	Lateral orbital gyrus	OFClat	OFClat
33, 34	Insula	Insula	INS
35, 36	Anterior cingulate & paracingulate gyri	Cingulate_Ant	ACC
37, 38	Middle cingulate & paracingulate gyri	Cingulate_Mid	MCC
39, 40	Posterior cingulate gyrus	Cingulate_Post	PCC
41, 42	Hippocampus	Hippocampus	HIP
43, 44	Parahippocampal gyrus	ParaHippocampal	PHG
45, 46	Amygdala	Amygdala	AMYG
47, 48	Calcarine fissure and surrounding cortex	Calcarine	CAL
49, 50	Cuneus	Cuneus	CUN
51, 52	Lingual gyrus	Lingual	LING
53, 54	Superior occipital gyrus	Occipital_Sup	SOG
55, 56	Middle occipital gyrus	Occipital_Mid	MOG
57, 58	Inferior occipital gyrus	Occipital_Inf	IOG
59, 60	Fusiform gyrus	Fusiform	FFG
61, 62	Postcentral gyrus	Postcentral	PoCG
63, 64	Superior parietal gyrus	Parietal_Sup	SPG
65, 66	Inferior parietal gyrus, excluding supramarginal and angular gyri	Parietal_Inf	IPG
67, 68	SupraMarginal gyrus	SupraMarginal	SMG
69, 70	Angular gyrus	Angular	ANG
71, 72	Precuneus	Precuneus	PCUN
73, 74	Paracentral lobule	Paracentral_Lobule	PCL
75, 76	Caudate nucleus	Caudate	CAU
77, 78	Lenticular nucleus, Putamen	Putamen	PUT
79, 80	Lenticular nucleus, Pallidum	Pallidum	PAL
81, 82	Thalamus	Thalamus	THA
83, 84	Heschl's gyrus	Heschl	HES
85, 86	Superior temporal gyrus	Temporal_Sup	STG
87, 88	Temporal pole: superior temporal gyrus	Temporal_Pole_Sup	TPOsup
89, 90	Middle temporal gyrus	Temporal_Mid	MTG
91, 92	Temporal pole: middle temporal gyrus	Temporal_Pole_Mid	TPOmid
93, 94	Inferior temporal gyrus	Temporal_Inf	ITG
95, 96	Crus I of cerebellar hemisphere	Cerebellum_Crus1	CERCRU1
97, 98	Crus II of cerebellar hemisphere	Cerebellum_Crus2	CERCRU2
99, 100	Lobule III of cerebellar hemisphere	Cerebellum_3	CER3
101, 102	Lobule IV, V of cerebellar hemisphere	Cerebellum_4_5	CER4_5
103, 104	Lobule VI of cerebellar hemisphere	Cerebellum_6	CER6
105, 106	Lobule VIIB of cerebellar hemisphere	Cerebellum_7b	CER7b
107, 108	Lobule VIII of cerebellar hemisphere	Cerebellum_8	CER8
109, 110	Lobule IX of cerebellar hemisphere	Cerebellum_9	CER9
111, 112	Lobule X of cerebellar hemisphere	Cerebellum_10	CER10

No.	Anatomical Description	Label AAL3	Possible Abbreviation
113	Lobule I, II of vermis	Vermis_1_2	VER1_2
114	Lobule III of vermis	Vermis_3	VER3
115	Lobule IV, V of vermis	Vermis_4_5	VER4_5
116	Lobule VI of vermis	Vermis_6	VER6
117	Lobule VII of vermis	Vermis_7	VER7
118	Lobule VIII of vermis	Vermis_8	VER8
119	Lobule IX of vermis	Vermis_9	VER9
120	Lobule X of vermis	Vermis_10	VER10
121, 122	Thalamus, Anteroventral Nucleus	Thal_AV	tAV
123, 124	Lateral posterior	Thal_LP	tLP
125, 126	Ventral anterior	Thal_VA	tVA
127, 128	Ventral lateral	Thal_VL	tVL
129, 130	Ventral posterolateral	Thal_VPL	tVPL
131, 132	Intralaminar	Thal_IL	tIL
133, 134	Reuniens	Thal_Re	tRe
135, 136	Mediodorsal medial magnocellular	Thal_MDm	tMDm
137, 138	Mediodorsal lateral parvocellular	Thal_MDl	tMDl
139, 140	Lateral geniculate	Thal_LGN	tLGN
141, 142	Medial Geniculate	Thal_MGN	tMGN
143, 144	Pulvinar anterior	Thal_PuA	tPuA
145, 146	Pulvinar medial	Thal_PuM	tPuM
147, 148	Pulvinar lateral	Thal_PuL	tPuL
149, 150	Pulvinar inferior	Thal_PuI	tPuI
151, 152	Anterior cingulate cortex, subgenual	ACC_sub	ACCsub
153, 154	Anterior cingulate cortex, pregenual	ACC_pre	ACCpre
155, 156	Anterior cingulate cortex, supracallosal	ACC_sup	ACCsup
157, 158	Nucleus accumbens	N_Acc	Nacc
159, 160	Ventral tegmental area	VTA	VTA
161, 162	Substantia nigra, pars compacta	SN_pc	SNpc
163, 164	Substantia nigra, pars reticulata	SN_pr	SNpr
165, 166	Red nucleus	Red_N	RedN
167, 168	Locus coeruleus	LC	LC
169	Raphe nucleus, dorsal	Raphe_D	RapheD
170	Raphe nucleus, median	Raphe_M	RapheM

References

[1] Mark Jenkinson and Christian F. Beckmann and Timothy E.J. Behrens and Mark W. Woolrich and Stephen M. Smith: FSL. NeuroImage, **62**(2), 782–790 (2012)

[2] Isensee, Fabian and Schell, Marianne and Pflueger, Irada and Brugnara, Gianluca and Bonekamp, David and Neuberger, Ulf and Wick, Antje and Schlemmer, Heinz-Peter and Heiland, Sabine and Wick, Wolfgang and Bendszus, Martin and Maier-Hein, Klaus H. and Kickingereder, Philipp: Automated brain extraction of multisequence MRI using artificial neural networks. Human Brain Mapping, **40**(17), 4952–4964 (2019)

[3] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian: Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[4] Pedregosa, Fabian and Varoquaux, Gaël and Gramfort, Alexandre and Michel, Vincent and Thirion, Bertrand and Grisel, Olivier and Blondel, Mathieu and Prettenhofer, Peter and Weiss, Ron and Dubourg, Vincent and others: Scikit-learn: Machine learning in Python. the Journal of Machine Learning Research, **12**, 2825–2830 (2011)

[5] Rolls, E.T., Huang, C.C., Lin, C.P., Feng, J., and Joliot, M.: Automated anatomical labelling atlas 3. NeuroImage, **206**, 116189 (2020)