1 Metrics

1.1 Regression Metrics

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{3}$$

$$PCC = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{4}$$

$$FC_{PCC}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{5}$$

$$FN = \sqrt{\sum_{i,j}|A_{ij} - B_{ij}|^2} \tag{6}$$

1.2 Classification Metrics

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN} \tag{9}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{10}$$

2 Performance

Table 1: Comparative experimental results on UKB. The best result is **bolded**, and the second best is underlined. Metrics include Mean Absolute Error (MAE), Root Mean Square Error ($RMSE$), Mean Absolute Percentage Error ($MAPE$), Correlation Coefficient (PCC), Functional Connectivity PCC (FC_{PCC}), and Feature Norm (FN).

Category	Model	MAE	$RMSE$	$MAPE$	PCC	FC_{PCC}	FN
Transformer	BrainTransformer [35]	20.2087	30.0437	0.2419	0.0029	0.0705	120.99
	iTransformer [34]	19.5380	29.0149	0.2346	0.1045	0.2305	128.69
	PatchTST [36]	18.9038	27.9970	0.2298	0.2039	0.7732	34.96
Linear	DLinear [38]	15.0546	21.4078	0.1969	0.5116	0.9703	15.87
	TSMixer [37]	15.9711	22.9913	0.2037	0.4818	0.8163	36.21
	LightTS [39]	15.1007	21.5178	0.1968	0.5094	0.9688	17.17
Graph	MSGNet [31]	19.7222	29.2989	0.2363	0.0520	0.2977	94.97
	FourierGNN [33]	17.5481	25.5135	0.2193	0.3612	0.8776	25.30
	ForecastGrapher [32]	31.2625	46.1902	0.3742	0.0471	0.6004	46.83
LLM Model	One Fit All [41]	15.1127	21.4374	0.1985	0.5020	0.9766	13.25
	SimMTM [40]	15.9283	22.8618	0.2045	0.4537	0.9080	22.15
	BOLD-Cast	**5.9595**	**15.8012**	**0.0744**	**0.9594**	**0.9960**	**4.80**

Table 2: Experimental results on **HCP-YA**, **HCP-A**, **HCP-D**, and **ABIDE**

Method	MAE	$RMSE$	$MAPE$	PCC	FC_{PCC}	FN
HCP-YA	7.3909	10.2241	0.1045	0.9562	0.9929	5.0612
HCP-A	10.5163	14.2336	0.1263	0.9847	0.9864	5.6261
HCP-D	12.6548	16.3170	0.1549	0.9722	0.9830	6.5438
ABIDE	10.1812	18.4301	0.1868	0.9614	0.9459	7.4615

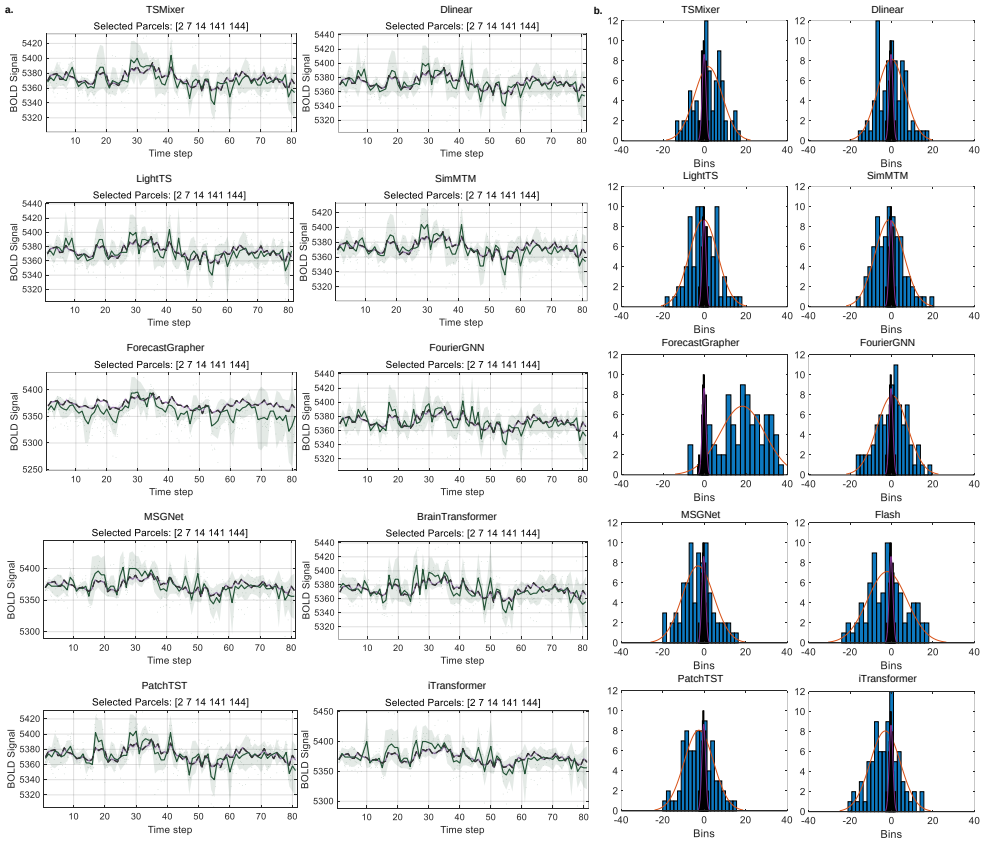

Fig. 1: Benchmarking BOLD-Cast on UK Biobank fMRI dataset. **a**, Representative BOLD signal trajectories from two resting-state networks (parietal and supramarginal), each visualized by averaging the signals across their constituent parcels. Individual parcels are shown as scattered points. Ground-truth (black), 10 SOTAs (green) are overlaid. Shaded bands indicate inter-parcel variability. **b**, Histogram of voxel-wise prediction errors, aggregated across all samples. BOLD-Cast (purple), 10 SOTAs (blue) are overlaid.

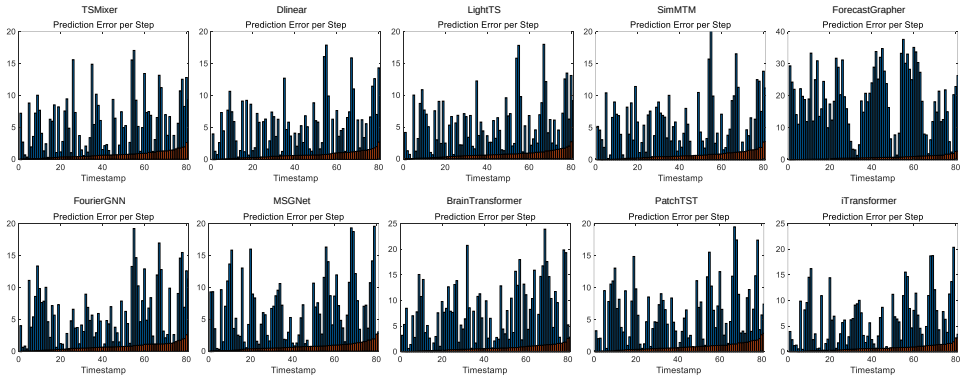

Fig. 2: Step-wise absolute prediction errors across time. Compared to 10 SOTAs, BOLD-Cast yields consistently smaller and more stable errors across the horizon. BOLD-Cast (red), 10 SOTAs (blue) are overlaid.

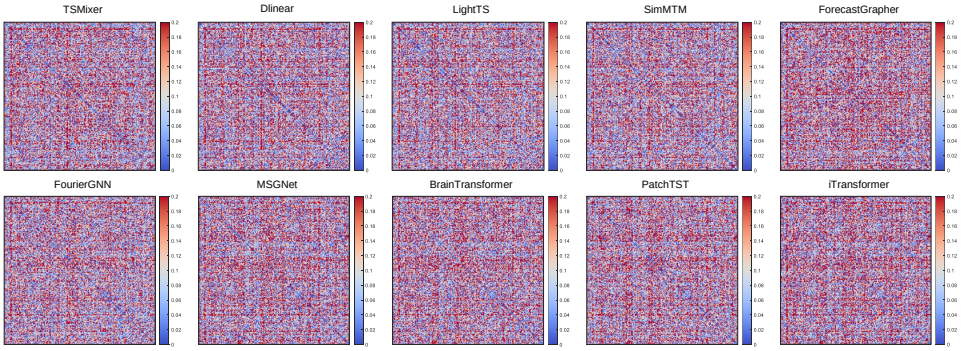

Fig. 3: Heatmaps of FC deviation from ground-truth FC.

3 Scalebility

Table 3: Scalability evaluation of the proposed model on the UKB dataset. Here, "1" denotes prediction with 1× sequence length, "2" denotes 2× prediction length, and "3" denotes 3× prediction length.

UKB	MAE	$RMSE$	$MAPE$	PCC	FC_{PCC}	FN
1	5.960272	15.80382	0.074366	0.959444	0.99737	4.80696
2	14.16711	27.88933	0.169294	0.749358	0.87334	27.4937
3	18.17853	32.16586	0.215275	0.588536	0.82327	32.8177

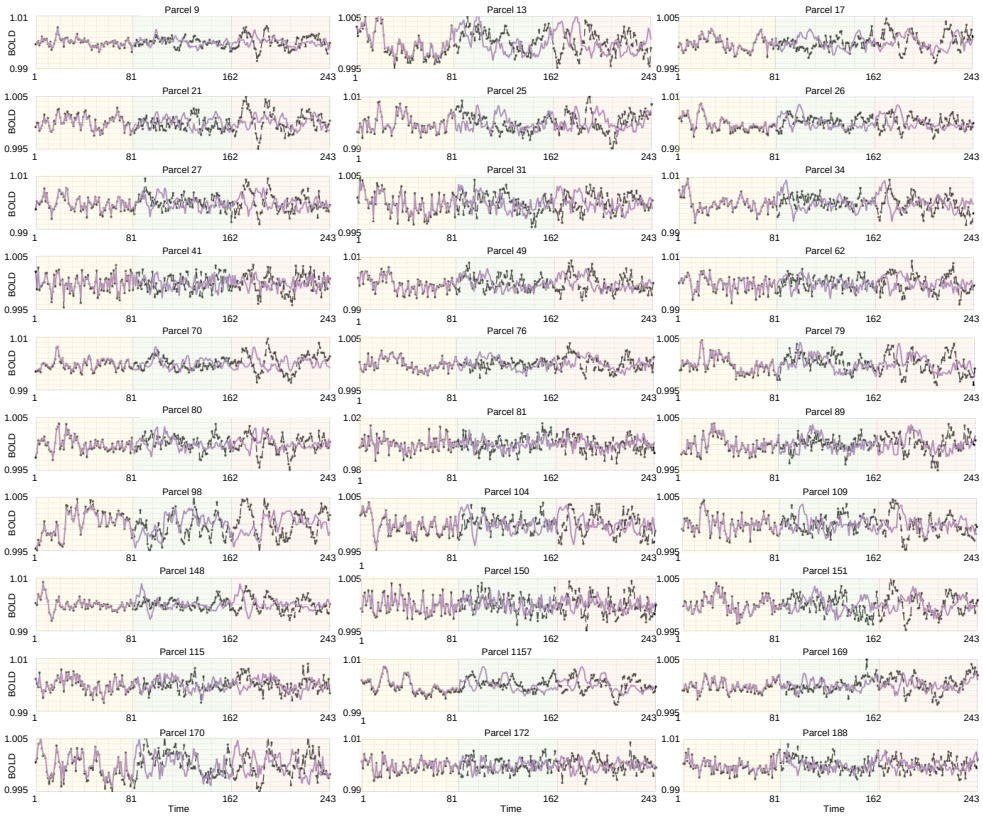

Fig. 4: Time-series prediction results of the proposed BOLD-Cast model. The plots show a comparison between the ground-truth BOLD signals (black lines) and the self-tested predictions (colored lines) for 30 randomly selected parcels from the CC200 atlas. Predictions corresponding to 1×, 2×, and 3× the input length are displayed in yellow, green, and orange, respectively. Each sub-plot represents a distinct parcel, arranged in a grid layout, with the x-axis indicating time and the y-axis representing the blood-oxygen-level dependent (BOLD) signal. The results demonstrate close alignment between the predicted and actual time series across various parcels and prediction horizons.

4 Generalization

Table 4: Results predicted directly on **HCP-YA** using the model trained on UKB.

Method	MAE	$RMSE$	$MAPE$	PCC	FC_{PCC}	FN
BOLD-Cast	2.5344	3.5230	0.0334	0.9862	0.9963	5.0043
ForecastGrapher [32]	20.4581	27.2546	0.2594	0.4152	0.8258	35.0849
FourierGNN [33]	22.1698	29.6735	0.2790	0.3157	0.8306	25.3962
MSGNet [31]	22.4069	29.8570	0.2792	0.0478	0.2188	99.4921
BrainTransformer [35]	24.6038	32.7973	0.3065	0.0010	0.0496	82.6730
iTransformer [34]	22.3124	29.7395	0.2782	0.0841	0.2036	130.2870
PatchTST [36]	20.7234	27.6625	0.2611	0.4106	0.9183	18.0014
SimMTM [40]	20.0715	26.7488	0.2553	0.4273	0.8767	22.7491
DLinear [38]	19.4890	26.0447	0.2495	0.4790	0.9583	16.6756
LightTS [39]	19.4474	25.9918	0.2487	0.4801	0.9554	18.3423
TSMixer [37]	20.7952	27.8729	0.2635	0.4215	0.7786	34.5488
One Fit All [41]	19.8574	26.6801	0.2541	0.4500	0.9478	14.7505

Table 5: Results predicted directly on **HCP-A** using the model trained on UKB.

Method	MAE	$RMSE$	$MAPE$	PCC	FC_{PCC}	FN
BOLD-Cast	5.9359	13.8357	0.0765	0.9555	0.9902	6.3423
ForecastGrapher [32]	18.2882	32.1008	0.2174	0.3678	0.7046	39.7370
FourierGNN [33]	39.0462	46.2622	0.2310	0.3044	0.8385	26.0966
MSGNet [31]	19.7092	26.5311	0.2468	0.0334	0.2214	102.3460
BrainTransformer [35]	58.1294	67.0210	0.6858	0.0015	0.0520	84.9479
iTransformer [34]	19.7286	26.5496	0.2472	0.0627	0.2155	128.1340
PatchTST [36]	18.4321	32.7639	0.2116	0.3683	0.8987	21.2430
SimMTM [40]	17.9924	32.7336	0.2120	0.3900	0.8560	27.9462
DLinear [38]	29.7842	35.8975	0.2063	0.4556	0.9600	16.9521
LightTS [39]	30.0381	36.2487	0.2057	0.4549	0.9565	18.7103
TSMixer [37]	33.4622	41.3429	0.2206	0.3975	0.7694	37.9118
One Fit All [41]	16.0127	21.1898	0.2110	0.4130	0.9492	15.4822

Table 6: Results predicted directly on **HCP-D** using the model trained on UKB.

Method	MAE	$RMSE$	$MAPE$	PCC	FC_{PCC}	FN
BOLD-Cast	7.2284	10.6253	0.0928	0.9867	0.9964	5.0443
ForecastGrapher [32]	38.7098	52.2318	0.5005	0.5022	0.7571	43.1136
FourierGNN [33]	45.1674	60.4220	0.5826	0.2646	0.7581	36.9197
MSGNet [31]	51.3598	68.4068	0.6508	0.0454	0.2477	95.2288
BrainTransformer [35]	65.4545	84.1394	0.8115	0.0140	0.0501	82.1174
iTransformer [34]	51.8758	68.9667	0.6575	0.0625	0.1950	121.1810
PatchTST [36]	37.6136	51.0902	0.4802	0.5144	0.7881	36.8174
SimMTM [40]	35.3273	48.2508	0.4562	0.5394	0.8346	31.3364
DLinear [38]	34.6864	46.6826	0.4521	0.5538	0.9280	21.1483
LightTS [39]	34.6591	46.6462	0.4504	0.5547	0.9254	22.0792
TSMixer [37]	42.4833	57.5424	0.5532	0.3926	0.6270	48.8311
One Fit All [41]	33.4636	45.9985	0.4342	0.5729	0.9192	21.1432

Table 7: Results predicted directly on **ABIDE** using the model trained on UKB.

Method	MAE	$RMSE$	$MAPE$	PCC	FC_{PCC}	FN
BOLD-Cast	7.3172	10.1247	6.1070	0.8748	0.9846	9.5438
ForecastGrapher [32]	NaN	NaN	NaN	NaN	NaN	NaN
FourierGNN [33]	NaN	NaN	NaN	NaN	NaN	NaN
MSGNet [31]	NaN	NaN	NaN	NaN	NaN	NaN
BrainTransformer [35]	23.8716	33.7675	3.8878	0.0048	-0.0020	29.6531
iTransformer [34]	21.9021	31.2413	1.9725	0.0375	0.1203	24.9755
PatchTST [36]	22.8526	32.4712	4.1760	0.0649	0.1077	4.0259
SimMTM [40]	23.0017	32.7778	4.2146	0.0752	0.1102	4.1215
DLinear [38]	23.4458	33.3469	5.1087	0.1084	0.3826	11.2021
LightTS [39]	21.9021	31.2413	1.9725	0.0375	0.1203	24.9755
TSMixer [37]	10.8325	15.4289	1.9768	0.0533	0.1027	4.5713
One Fit All [41]	22.8819	32.6108	4.1938	0.0737	0.1096	4.0998

5 Explainability

Fig. 5: Individual FC matrices generated from BOLD-Cast predictions using the CC200 partitioning method and visualized with Yeo-17 network boundaries. Intra- and inter-network similarity scores are reported (e.g., intra-network: 0.51, inter-network: 0.19 in ABIDE; $\Delta = 0.32$). Typical resting-state modularity was preserved across all datasets, confirming that the shared latent representations learned by BOLD-Cast generalize across groups while maintaining neurobiologically meaningful structure.

6 Ablation Study

	Our	A_1	A_2	A_3	A_4	A_5	A_6
$Loss_{match}$	-	×	-	-	-	×	-
$Loss_{pear}$	-	-	×	-	-	×	-
$Loss_{recon}$	-	-	-	×	-	×	-
$Loss_{con}$	-	-	-	-	×	×	-
TE	-	-	-	-	-	-	×

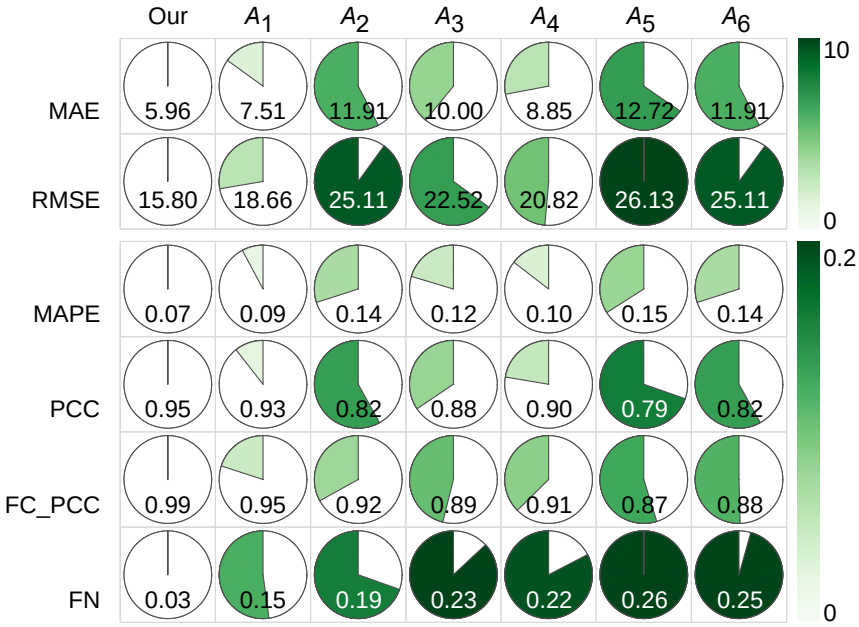

Fig. 6: Ablation experiment settings and results. The table above shows the ablation experiment settings. Modules marked with a cross are ablated, and modules marked with a check mark are retained. The figure below shows the degradation of model performance under different ablation settings. Greater degradation indicates larger sector areas.

7 fMRI Data Preprocessing Pipeline

Table 8: Summary of fMRI preprocessing pipelines for the datasets used in this study. All datasets were preprocessed following their official pipelines to ensure consistency and comparability.

Preprocessing Step	**UKB**	**HCP-YA**	**HCP-A**	**HCP-D**	**ABIDE**
Processing Pipeline	FSL	HCP pipeline	HCP pipeline	HCP pipeline	AFNI
Slice Time Correction	O	O	O	O	O
Head Motion Correction	O	O	O	O	O
Motion Co-registration	FSL applywarp	O	O	O	O
Spatial Smoothing	X	X	X	X	6 mm
High-pass Filtering	O	Gaussian (200s)	Gaussian (200s)	Gaussian (200s)	0.009
Low-pass Filtering	X	X	X	X	0.08
Denoising	ICA-FIX	ICA-FIX	ICA-FIX	ICA-FIX	3dDeconvolve (nuisance regression)
Discarding First Timepoints	X	X	X	X	X

8 Model training

Table 9: Summary of key hyperparameters and settings for **BOLD-Cast**.

Parameter	**Default Value**	**Description**
Data Loader Configuration		
task_name	'long_term_forecast'	Task name
drop_last	True	Drop last incomplete batch
val_set_shuffle	False	Shuffle validation set
drop_short	True	Drop short sequences
his_len	81	Historical sequence length
pre_len	81	Prediction length
Model Configuration		
epoch	100	Training epoch
lr	0.001	Learning rate
bs	32	Batch size
patience	5	Early stop patience
loss	MSE	Loss function
is_training	1	Training status (1=train, 0=test)
llm_ckp_dir	gpt2	Pre-trained LLM checkpoint
dropout	0.1	Dropout rate
enc_in	7	Encoder input size
dec_in	7	Decoder input size

Table 10: Summary of key hyperparameters and settings for **Brain-Transformer/iTransformer/Dlinear/TSMixer**.

Parameter	Default Value	Description
	Data Loader Configuration	
task_name	'long_term_forecast'	Task name
drop_last	True	Drop last incomplete batch
val_set_shuffle	False	Shuffle validation set
drop_short	True	Drop short sequences
seq_len	162	Input sequence length
his_len	81	Start token length
pre_len	81	Prediction sequence length
moving_avg	25	Window size of moving average
d_ff	2048	Dimension of fcn
features	'M'	Forecasting task
freq	's'	Freq for time features encoding
d_model	512	Dimension of model
c_out	7	Output size
	Model Configuration	
epoch	100	Training epoch
lr	0.001	Learning rate
bs	32	Batch size
patience	5	Early stop patience
loss	MSE	Loss function
is_training	1	Training status (1=train, 0=test)
llm_ckp_dir	gpt2	Pre-trained LLM checkpoint
dropout	0.1	Dropout rate
enc_in	7	Encoder input size
dec_in	7	Decoder input size

Table 11: Summary of key hyperparameters and settings for **LightTS/PatchTST**.

Parameter	**Default Value**	**Description**
	Data Loader Configuration	
task_name	'long_term_forecast'	Task name
drop_last	True	Drop last incomplete batch
val_set_shuffle	False	Shuffle validation set
drop_short	True	Drop short sequences
seq_len	162	Input sequence length
his_len	81	Start token length
pre_len	81	Prediction sequence length
moving_avg	25	Window size of moving average
d_ff	2048	Dimension of fcn
features	'M'	Forecasting task
freq	's'	Freq for time features encoding
d_model	64	Dimension of model
c_out	7	Output size
	Model Configuration	
epoch	100	Training epoch
lr	0.001	Learning rate
bs	32	Batch size
patience	5	Early stop patience
loss	MSE	Loss function
is_training	1	Training status (1=train, 0=test)
llm_ckp_dir	gpt2	Pre-trained LLM checkpoint
dropout	0.1	Dropout rate
enc_in	190	Encoder input size
dec_in	190	Decoder input size
stride	8	Stride
patch_len	16	Patch length
head_dropout	0.0	Head dropout

Table 12: Summary of key hyperparameters and settings for **One Fit All**.

Parameter	**Default Value**	**Description**
	Data Loader Configuration	
task_name	'long_term_forecast'	Task name
drop_last	True	Drop last incomplete batch
val_set_shuffle	False	Shuffle validation set
drop_short	True	Drop short sequences
seq_len	162	Input sequence length
his_len	81	Start token length
pre_len	81	Prediction sequence length
moving_avg	25	Window size of moving average
d_ff	2048	Dimension of fcn
features	'M'	Forecasting task
freq	's'	freq for time features encoding
d_model	768	Dimension of model
is_gpt	1	Use GPT
pretrain	1	Use pretained model
patch_size	16	Use GPT
gpt_layer	3	layer number of GPT
	Model Configuration	
epoch	20	Training epoch
lr	0.0001	Learning rate
bs	8	Batch size
patience	5	Early stop patience
loss	MSE	Loss function
is_training	1	Training status (1=train, 0=test)
llm_ckp_dir	gpt2	Pre-trained LLM checkpoint
dropout	0.1	Dropout rate
enc_in	7	Encoder input size
dec_in	7	Decoder input size

Table 13: Summary of key hyperparameters and settings for **SimMTM**.

Parameter	**Default Value**	**Description**
		Data Loader Configuration
drop_last	True	Drop last incomplete batch
val_set_shuffle	False	Shuffle validation set
drop_short	True	Drop short sequences
seq_len	162	Input sequence length
his_len	81	Start token length
pre_len	81	Prediction sequence length
moving_avg	25	Window size of moving average
d_ff	2048	Dimension of fcn
features	'M'	Forecasting task
freq	's'	Freq for time features encoding
d_model	512	Dimension of model
c_out	7	Output size
		Model Configuration
task_name	'finetune'	Task name
epoch	100	Training epoch
lr	0.001	Learning rate
bs	32	Batch size
patience	5	Early stop patience
loss	MSE	Loss function
is_training	1	Training status (1=train, 0=test)
llm_ckp_dir	gpt2	Pre-trained LLM checkpoint
dropout	0.1	Dropout rate
enc_in	7	Encoder input size
dec_in	7	Decoder input size
temperature	0.2	temperature
lm	3	Average masking length
positive_nums	3	Masking series numbers
rbtp	1	(0: rebuild the embedding of oral series; 1: rebuild oral series)
masked_rule	'geometric'	(geometric, random, masked tail, masked head)
mask_rate	7	Mask ratio
stride	12	Stride
patch_len	12	Patch length
head_dropout	0.1	Head dropout

Table 14: Summary of key hyperparameters and settings for **ForecastGrapher**.

Parameter	Default Value	Description
Data Loader Configuration		
task_name	'long_term_forecast'	Task name
drop_last	True	Drop last incomplete batch
val_set_shuffle	False	Shuffle validation set
drop_short	True	Drop short sequences
seq_len	162	Input sequence length
his_len	81	Start token length
pre_len	81	Prediction sequence length
moving_avg	25	Window size of moving average
d_ff	2048	Dimension of fcn
features	'M'	Forecasting task
freq	's'	Freq for time features encoding
d_model	512	Dimension of model
c_out	190	Output size
num_nodes	190	To create graph
subgrpah_size	3	Neighbors number
k	3	Number of GNN block
num_dim	10	Node embbed to dim dimentions
Model Configuration		
epoch	100	Training epoch
lr	0.001	Learning rate
bs	32	Batch size
patience	5	Early stop patience
loss	MSE	Loss function
is_training	1	Training status (1=train, 0=test)
llm_ckp_dir	gpt2	Pre-trained LLM checkpoint
dropout	0.1	Dropout rate
enc_in	7	encoder input size
dec_in	7	decoder input size

Table 15: Summary of BrainGB and PTGB implementation details in our downstream tasks.

Aspect	BrainGB (Classification)	PTGB (Regression)
Task Type	ASD diagnosis, Sex prediction	fMRI-based phenotype regression
Output Layer	Softmax classifier (original)	Linear projection layer (no softmax)
Loss Function	Cross-entropy loss	Mean squared error
Evaluation Metrics	Accuracy, Precision, F1-score, Specificity	MAE, RMSE, MPAE, PCC
Graph Input	Static FC matrix per subject	Static FC matrix per subject
Graph Backbone	GIN / GAT (unchanged)	GIN / GAT (unchanged)
Pooling Method	Global mean or attention pooling	Global mean or attention pooling
Optimizer	Adam	Adam
Learning Rate	0.001	0.001
Epochs	100 (early stop)	100 (early stop)
Batch Size	32	32
Model Structure	3 GNN layers + 2-layer MLP	3 GNN layers + 2-layer MLP