



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

MedTS-TTT: Test-Time Training for Medical Time Series Classification

Mingzhi Chen, Yiyu Gui, and Guibo Luo*

Guangdong Provincial Key Laboratory of Ultra High Definition Immersive Media Technology, Shenzhen Graduate School, Peking University, China
luogb@pku.edu.cn

Abstract. Medical time series (MedTS) signals such as electroencephalography (EEG) and electrocardiography (ECG) support many clinical applications. However, substantial subject-level heterogeneity often induces subject-level distribution shift, causing a fixed parameter set to generalize poorly to unseen individuals. Compared with domain adaptation methods that often depend on extra adaptation components or target-batch statistics, Test-Time Training (TTT) provides a more practical solution for sequential clinical data by enabling online adaptation from unlabeled test samples. However, many representative TTT methods require iterative inner-loop optimization, increasing test-time overhead. In this paper, we propose MedTS-TTT, a test-time training framework for medical time series modeling. MedTS-TTT is built upon Closed-Loop Self-Alignment Test-Time Training (CLSA-TTT) and a Gated Convolutional Backbone (GCB). CLSA-TTT constructs a token-level self-supervised target and performs a single-step fast-weight update for intra-layer closed-loop alignment, enabling rapid sample-wise adaptation without iterative inner-loop optimization. GCB combines CLSA-TTT-based fast adaptation and token-level fusion with a gated convolutional branch to balance local dynamic modeling and information-flow control. On 4 public datasets (2 EEG and 2 ECG) with subject-independent splits, MedTS-TTT achieves 11 top-1 rankings out of 12 evaluations across 9 baselines and 3 metrics. The code is publicly available at <https://github.com/mingzhi-c/MedTS-TTT>.

Keywords: Medical time series · Test-Time Training · Gated Convolutional Backbone.

1 Introduction

Medical time series (MedTS) data, such as electroencephalography (EEG) and electrocardiography (ECG), are widely used in clinical decision-making and continuous health monitoring. Recent deep-learning methods have shown strong performance in end-to-end classification on these signals. For example, end-to-end EEG models have been explored for neurodegenerative disorders such as

* Corresponding author.

Alzheimer’s disease [9, 27] and Parkinson’s disease [2, 11]. In cardiac care, ECG models have achieved strong results for arrhythmia detection [8, 10] and myocardial infarction detection [16, 22]. These advances highlight the broad potential of end-to-end MedTS modeling across clinical applications.

Recent studies have explored general-purpose backbones for unified temporal and inter-channel modeling. EEGMamba [6] combines bidirectional state-space modeling with task-aware MoE for multi-task EEG classification. MedFormer [19], MedGNN [4], and MedSpaformer [21] respectively employ multi-granularity patching, multi-resolution spatiotemporal graph learning, and dynamic token sparsification. Despite these architectural advances, a fundamental challenge remains: MedTS exhibit substantial subject-level heterogeneity, often leading to subject-level distribution shift, as shown in Fig. 1(a). Consequently, a fixed parameter set may generalize poorly to unseen individuals, even under the same dataset and acquisition protocol.

One line of work addresses cross-subject generalization through domain adaptation (DA), including source-free unsupervised adaptation for personalized medical time series analysis [26]. While effective in controlled settings, DA-based methods are often difficult to deploy in real clinical workflows because they usually introduce additional adaptation modules and frequently rely on target batches or relatively stable target statistics. In practice, clinical signals often arrive sequentially, and subject-level distribution shift may evolve over time. These constraints motivate a lightweight adaptation mechanism that can operate directly on unlabeled test samples.

Test-Time Training (TTT) provides a promising alternative by updating model parameters at inference time using self-supervised objectives on test inputs [15, 7, 5]. TTT is particularly suitable here because it enables online adaptation from each incoming test sample without target labels or explicit target-domain training. This makes TTT a natural fit for real-world MedTS applications, where data are often observed sequentially and adaptation must be performed under limited computational and annotation resources. However, many existing TTT methods rely on iterative inner-loop optimization to perform adaptation, which increases test-time computational overhead and complicates use in latency-sensitive scenarios. For MedTS applications, where efficient per-sample inference is often required, a more efficient TTT design is needed.

In this work, we propose MedTS-TTT, a test-time training framework for medical time series modeling. First, we introduce Closed-Loop Self-Alignment Test-Time Training (CLSA-TTT), which mitigates subject-level distribution shift during cross-subject inference by constructing a token-level self-supervised target and performing a single-step fast-weight update for intra-layer closed-loop alignment, enabling rapid sample-wise adaptation without iterative inner-loop optimization. Second, we propose a Gated Convolutional Backbone (GCB) for sequence modeling, which combines CLSA-TTT-based fast adaptation and token-level fusion with a gated convolutional branch to balance local dynamic modeling and information-flow control. We evaluate MedTS-TTT on 4 public datasets (2 EEG and 2 ECG) under subject-independent splits. MedTS-TTT

achieves 11 top-1 rankings out of 12 evaluations across 9 baselines and 3 metrics, demonstrating improved robustness under subject-level distribution shift and stronger practicality in real-world settings.

2 Method

We propose MedTS-TTT, an end-to-end test-time training framework for medical time series classification. It is designed to address subject-level distribution shift in real-world settings without requiring extra labels or architectural changes. As shown in Fig. 1, MedTS-TTT is built upon Closed-Loop Self-Alignment Test-Time Training (CLSA-TTT) and a Gated Convolutional Backbone (GCB). Specifically, a spatiotemporal tokenizer converts multi-channel signals into a token sequence, which is then processed by the GCB, where CLSA-TTT performs a single-step self-supervised fast-weight update at test time. The final prediction is obtained by aggregating token representations.

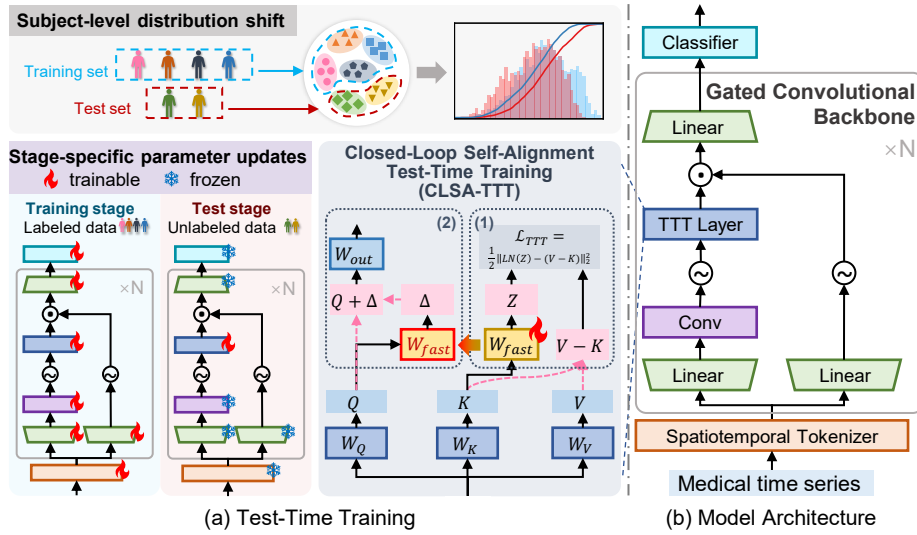


Fig. 1. The comprehensive architecture of MedTS-TTT. (a) Subject-level distribution shift setting and training/test-time adaptation protocol. (b) Overall architecture of the proposed MedTS-TTT framework.

2.1 Spatiotemporal Tokenizer for Medical Time Series

Given a medical time series segment $X \in \mathbb{R}^{B \times C \times T}$, we design a spatiotemporal tokenizer to map it into a token sequence $H \in \mathbb{R}^{B \times L \times D}$. Specifically, we apply channel-wise temporal encoding, cross-channel spatial fusion, and patch-wise

projection (with positional embeddings) to obtain $\tilde{H}$, which serves as the input for subsequent token-level adaptation.

Temporal encoding. To extract shared features of short-term dynamics and spectral structure within each channel, we apply channel-wise temporal convolutions to obtain sequential representations $U = \phi(\mathcal{F}_t(X)) \in \mathbb{R}^{B \times D \times C \times T}$. Here, $\phi(\cdot)$ denotes an activation function (GELU in our implementation), and $\mathcal{F}_t \in \mathbb{R}^{D \times 1 \times 1 \times k_t}$ is a channel-wise temporal convolution kernel.

Spatial fusion. We fuse cross-channel information using a channel-dependent spatial convolution $\mathcal{F}_s \in \mathbb{R}^{D \times D \times C \times 1}$ along the channel dimension, aggregating multi-channel signals into a compact representation $S = \phi(\mathcal{F}_s(U)) \in \mathbb{R}^{B \times D \times T}$.

Patch embedding with positional encoding. We partition S along the temporal axis with stride P , and map each patch to a token via a learnable linear projection $\mathcal{P} \in \mathbb{R}^{D \times P}$, yielding a token sequence $H = \mathcal{P}(S) \in \mathbb{R}^{B \times L \times D}$, where $L = \lceil T/P \rceil$ denotes the token sequence length. We then add a learnable positional embedding $PE \in \mathbb{R}^{L_{max} \times D}$ to preserve temporal order information:

$$\tilde{H} = H + PE_{[:L]} \in \mathbb{R}^{B \times L \times D} \quad (1)$$

2.2 Closed-Loop Self-Alignment Test-Time Training

Given a token sequence $\tilde{H} \in \mathbb{R}^{B \times L \times D}$, we perform test-time training (TTT) inside each backbone layer by updating a small set of fast parameters with a self-supervised objective. We adopt a fast-slow parameterization: the slow parameters learn stable representation transformations, while the fast parameters W_{fast} are updated in a per-sample inner loop to produce an adaptively self-aligned correction consistent with the current sequence state.

Self-supervised objective. We first project $\tilde{H}$ into multi-head representations $Q, K, V \in \mathbb{R}^{B \times n \times L \times d}$ ($d = D/n$). TTT defines an unlabeled objective on the key-value pair by transforming K with the fast parameters:

$$Z = KW_{fast} \quad (2)$$

We then use Layer Normalization (LN) to fit, in the normalized space, the residual signal induced by (K, V) :

$$\mathcal{L}_{TTT} = \frac{1}{2} \|LN(Z) - (V - K)\|_2^2 \quad (3)$$

This objective defines a token-level self-supervised target, guiding the fast parameters to align with the current sample’s token structure.

One-step fast parameter update. To perform the inner-loop adaptation, we compute $G \triangleq \frac{\partial \mathcal{L}_{TTT}}{\partial Z} \in \mathbb{R}^{B \times n \times L \times d}$ and update the fast parameters with a single SGD step for closed-loop self-alignment:

$$W_{fast} \leftarrow W_{fast} - \eta K^\top G \quad (4)$$

where η denotes a head-wise adaptive step size. After this one-step update, we apply the updated fast parameters to Q to obtain an adaptive transformation:

$$\Delta = QW_{fast} \quad (5)$$

Δ serves as the output of the adaptive representation branch at this layer and is subsequently fused with the main backbone branch. This yields one-shot, full-sequence alignment with minimal test-time overhead, making it suitable for low-latency medical time series classification.

Training and inference usage. During training, TTT is integrated into the forward pass as an in-layer adaptation mechanism. Given an input sample, the model first performs a one-step fast-weight update using the self-supervised objective $\mathcal{L}_{TTT}$, and then uses the updated fast weights to perform the layer forward computation and produce predictions. The outer-loop optimization is driven by the supervised cross-entropy loss $\mathcal{L}_{CE}$, which updates only the slow parameters. At test time, we retain the same in-layer TTT adaptation without label supervision. For each test sample, we apply a single fast-weight update using $\mathcal{L}_{TTT}$ and then run the forward pass for prediction, enabling adaptive inference without annotations or additional fine-tuning.

2.3 Gated Convolutional Backbone with CLSA-TTT

To jointly model local dynamics and enable stable adaptation at the token level, we combine a short-range convolution module $\mathcal{F}_{dw}$ with a gated feed-forward structure, and integrate a CLSA-TTT module $\mathcal{M}$ in the main branch to construct an adaptive backbone layer. Given the input token sequence of the k th layer $\mathcal{T}^{(k)}, H^{(k)} \in \mathbb{R}^{B \times L \times D}$, the layer output is defined as

$$H^{(k+1)} = \mathcal{T}^{(k)}(H^{(k)}) = H^{(k)} + \Delta H^{(k)} \quad (6)$$

where $\mathcal{T}^{(k)}$ consists of the short-range convolution module, the CLSA-TTT module, and a gated fusion mechanism.

Short-range convolution module. We introduce a depthwise short-range convolution $\mathcal{F}_{dw}$ in the latent space to capture the short-term waveform variations and local rhythmic cues commonly observed in medical time series. For an input $A^{(k)} \in \mathbb{R}^{B \times L \times 2D}$, the transformation is

$$A^{(k+1)} = \phi(\mathcal{F}_{dw}^{(k)}(A^{(k)})) \in \mathbb{R}^{B \times L \times 2D} \quad (7)$$

Gate-based fusion In the latent space, we first apply a gating projection $W_{in} \in \mathbb{R}^{D \times 4D}$ and split the result into a main branch and a gating branch:

$$\begin{bmatrix} A^{(k)}, R^{(k)} \end{bmatrix} = H^{(k)} W_{in}, \quad A^{(k)}, R^{(k)} \in \mathbb{R}^{B \times L \times 2D} \quad (8)$$

The main branch is processed by the short-range convolution $\mathcal{F}_{dw}^{(k)}$ to obtain $A^{(k+1)}$, which is then fed into the CLSA-TTT module $\mathcal{M}^{(k)}$ to produce a sample-dependent representation transform $\tilde{A}^{(k+1)} = \mathcal{M}^{(k)}(A^{(k+1)}) \in \mathbb{R}^{B \times L \times 2D}$. The gating branch generates token-wise and channel-wise gating coefficients $G^{(k)} = \phi(R^{(k)}) \in \mathbb{R}^{B \times L \times 2D}$. We fuse the branches element-wise and project the result to the original dimensionality with $W_{out} \in \mathbb{R}^{2D \times D}$:

$$\Delta H^{(k)} = \left(G^{(k)} \odot \tilde{A}^{(k+1)} \right) W_{out} \in \mathbb{R}^{B \times L \times D} \quad (9)$$

where $\odot$ denotes element-wise multiplication. Overall, the layer leverages short-range convolutions to model local dependencies, uses CLSA-TTT to enable one-step test-time self-alignment, and employs gating to stabilize information fusion.

3 Experiments

3.1 Experimental settings

Datasets. We evaluate on 4 public clinical datasets, including two EEG datasets (APAVA [3] and ADFTD [13]) and two ECG datasets (PTB [1] and PTB-XL [18]). APAVA is used for Alzheimer’s disease detection with binary labels, while ADFTD targets differential diagnosis among healthy controls, frontotemporal dementia, and Alzheimer’s disease. For ECG, PTB provides binary labels for myocardial infarction detection, and PTB-XL supports multi-class classification of several cardiac conditions. For fair comparison and reproducibility, all datasets are processed and configured following Medformer.

Baselines. We compare our proposed MedTS-TTT against 7 representative state-of-the-art time-series models, including Autoformer [20], Crossformer [23], FEDformer [25], Informer [24], iTransformer [12], PatchTST [14], and the vanilla Transformer [17], as well as 2 MedTS-specific models, Medformer and MedGNN.

Implementation details. We employ 3 evaluation metrics: accuracy, F1 score (macro-averaged), and AUROC (macro-averaged). Model selection is based on the best validation F1 score, and the selected checkpoint is evaluated on the test set. All experiments are implemented in PyTorch and conducted on a Linux server equipped with an NVIDIA A100 GPU (80GB). For each dataset, we train the model for up to 100 epochs with an early-stopping strategy: training is terminated if the validation macro-F1 does not improve for 10 consecutive epochs, and the checkpoint achieving the best validation macro-F1 is retained for test evaluation. We optimize the model using Adam with a learning rate of 1×10^{-4} . To ensure a consistent and fair comparison across datasets, we use an identical model configuration throughout: the patch size is set to 8, and the backbone consists of 6 stacked layers with a hidden dimension of 128.

3.2 Results

Subject-independent Results. We split the training, validation, and test sets at the subject level, following the same setting as Medformer [19]. Specifically, samples from the same subject appear in only one split to prevent information leakage. This subject-independent setting reflects real-world medical time series diagnosis scenarios, where models are trained on labeled subjects and evaluated on unseen subjects to assess their disease status. In this setting, strong inter-subject variability can lead to noticeable distribution shifts, making test-time training a natural choice for improving robustness when facing new subjects.

Table 1 presents the results of the subject-independent evaluation setup. Overall, our method achieves 11 top-1 results out of 12 evaluations across 4

datasets, 9 baselines, and 3 metrics. Averaged over the four datasets, MedTS-TTT achieves the best overall performance, reaching 75.18% in Accuracy, 70.20% in F1 score, and 86.86% in AUROC, which corresponds to improvements of 0.90%, 0.65%, and 0.83% over the strongest baseline MedGNN, respectively. These results demonstrate that MedTS-TTT provides consistently improved robustness under subject-level distribution shift in subject-independent evaluation.

Table 1. Overall Results of Subject-independent Evaluation. The best results are shown in **bold**, and the second-best results are underlined.

EEG Datasets		APAVA		ADFTD		
Models	Accuracy	F1 score	AUROC	Accuracy	F1 score	AUROC
Autoformer	68.64±1.82	68.06±1.94	75.94±3.61	45.25±1.48	42.59±1.85	61.02±1.82
Crossformer	73.77±1.95	68.93±1.85	72.39±3.33	50.45±2.31	45.50±1.70	66.45±2.03
FEDformer	74.94±2.15	73.51±3.39	83.72±1.97	46.30±0.59	43.91±1.37	62.62±1.75
Informer	73.11±4.40	69.47±5.06	70.46±4.91	48.45±1.96	45.74±1.38	65.87±1.27
iTransformer	74.55±1.66	72.30±1.79	85.59±1.55	52.60±1.59	46.79±1.13	67.26±1.16
PatchTST	67.03±1.65	55.97±3.10	65.65±0.28	44.37±0.95	41.97±1.37	60.08±1.50
Transformer	76.30±4.72	73.75±5.38	72.50±6.60	50.47±2.14	48.09±1.59	67.93±1.59
Medformer	78.74±0.64	76.31±0.71	83.20±0.91	53.27±1.54	50.65±1.51	70.93±1.19
MedGNN	82.60±0.35	80.25±0.16	85.93±0.26	56.12±0.11	55.00±0.24	74.68±0.33
MedTS-TTT	83.03±1.13	81.52±1.45	89.96±1.43	58.10±0.82	55.18±2.27	75.41±1.25
ECG Datasets		PTB		PTB-XL		
Models	Accuracy	F1 score	AUROC	Accuracy	F1 score	AUROC
Autoformer	73.35±2.10	63.69±3.84	78.54±3.48	61.68±2.72	48.85±2.27	82.04±1.44
Crossformer	80.17±3.79	72.75±7.19	88.55±3.45	73.30±0.14	62.59±0.14	90.02±0.06
FEDformer	76.05±2.54	67.14±4.37	85.93±4.31	57.20±9.47	47.89±8.44	82.13±4.17
Informer	78.69±1.68	70.84±3.47	92.09±0.53	71.43±0.32	60.44±0.43	89.65±0.09
iTransformer	83.89±0.71	79.06±1.06	91.18±1.16	69.28±0.22	56.20±0.19	86.71±0.10
PatchTST	74.74±1.62	64.36±3.38	88.79±0.91	73.23±0.25	<u>62.61±0.34</u>	89.74±0.19
Transformer	77.37±1.02	68.47±2.19	90.08±1.76	70.59±0.44	59.05±0.25	88.21±0.16
Medformer	83.50±2.01	79.18±3.31	<u>92.81±1.48</u>	72.87±0.23	62.02±0.37	89.66±0.13
MedGNN	84.53±0.28	<u>80.40±0.62</u>	93.31±0.46	<u>73.87±0.18</u>	62.54±0.20	<u>90.21±0.15</u>
MedTS-TTT	85.04±3.12	80.71±4.81	91.31±2.65	74.56±0.37	63.39±0.70	90.77±0.18

Ablation Study. To assess the roles of CLSA-TTT and GCB, we perform ablations on 4 variants: full MedTS-TTT, w/o CLSA-TTT, w/o GCB, and w/ Vanilla TTT. The w/o CLSA-TTT variant removes the CLSA-TTT module from GCB, whereas w/o GCB keeps test-time training but replaces GCB with a standard Transformer FFN block. The w/ Vanilla TTT variant replaces CLSA-TTT with a standard TTT update module [15]. As shown in Fig. 2, the full model performs best across all datasets. Quantitatively, replacing CLSA-TTT with Vanilla TTT reduces average accuracy by 3.63%, supporting the effectiveness of the proposed closed-loop single-step alignment. Removing CLSA-TTT causes the largest overall degradation, with an average accuracy drop of 5.04%, indicating that test-time training is the primary source of robustness gain under subject-level distribution shift. Removing GCB also reduces average accuracy by 4.93%, showing that GCB provides complementary gains.

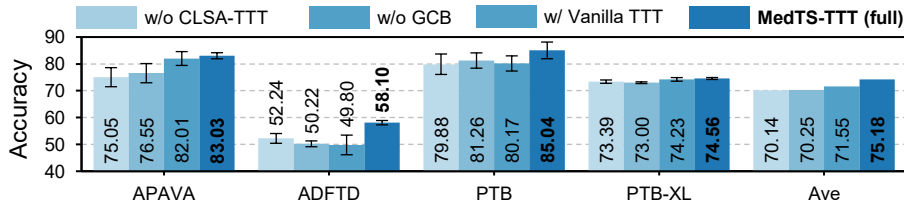


Fig. 2. Performance comparison of full MedTS-TTT and ablated variants.

Efficiency Comparison with the Vanilla TTT Mechanism. To further evaluate CLSA-TTT against the vanilla TTT mechanism, we compare the two methods using runtime measurements. As shown in Table 2, we report mean results over 3 metrics: p50 latency, p95 latency, and throughput. Specifically, CLSA-TTT reduces p50 and p95 latency by 50.09% and 49.76%, respectively, and nearly doubles throughput by 99.90%. These results show that CLSA-TTT substantially improves runtime efficiency over the vanilla TTT mechanism, making it more suitable for latency-sensitive test-time adaptation.

Table 2. Runtime efficiency comparison between CLSA-TTT and Vanilla TTT.

	p50 Latency (ms) ↓	p95 Latency (ms) ↓	Throughput (samples per second) ↑
MedTS-TTT w/ Vanilla TTT	29.35	29.70	4374.23
MedTS-TTT w/ CLSA-TTT	14.65	14.92	8744.17
$\Delta(\%)$	-50.09%	-49.76%	+99.90%

Effect of Test-Time Training. To further understand why CLSA-TTT improves performance, Fig. 3 provides a qualitative analysis of subject-level distribution shift and feature alignment. Fig. 3(left) analyzes APAVA using spectral entropy, a compact measure of spectral complexity. The histogram/CDF reveals a clear subject-level distribution shift even within a single dataset (i.e., under the same acquisition center and device). Consistently, the t-SNE visualizations of features (Fig. 3(right)) show that, without CLSA-TTT, train and test features exhibit a marked distribution shift, whereas enabling CLSA-TTT substantially narrows this gap while preserving class discriminability. This qualitative observation is consistent with the quantitative gains in Fig. 2.

4 Conclusion

In this paper, we presented MedTS-TTT, a test-time training framework for medical time series modeling that mitigates subject-level distribution shift under cross-subject real-world settings. It combines CLSA-TTT for efficient single-step adaptation with a Gated Convolutional Backbone for improved token-level sequence modeling. Experiments on four public datasets with subject-independent splits show consistent gains over competitive baselines.

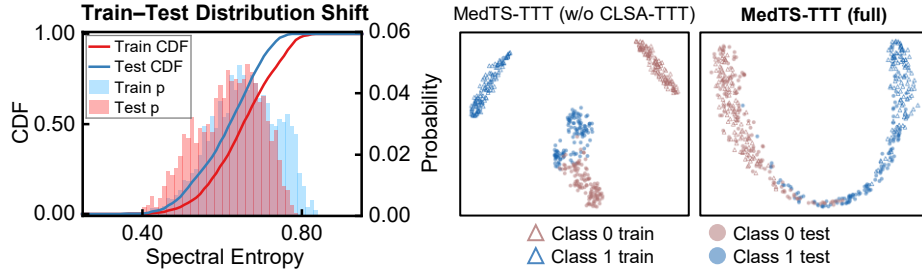


Fig. 3. Distribution-Shift and Feature Alignment under Test-Time Training.

Acknowledgments This work was supported by the National Key Research and Development Program of China (2024YFE0203100), the Guangdong Provincial Key Laboratory of Ultra High Definition Immersive Media Technology (Grant No. 2024B1212010006), and the Shenzhen Science and Technology Program (JCYJ20230807120800001).

Disclosure of Interests The authors declare no competing interests.

References

1. Bousseljot, R.D., Kreiseler, D., Schnabel, A.: Nutzung der EKG-signal-datenbank CARDIODAT der PTB über das internet. *Biomedizinische Technik/Biomedical Engineering* **40**(Ergänzungsband 1), 317–318 (1995)
2. Chang, H., Liu, B., Zong, Y., Lu, C., Wang, X.: EEG-based Parkinson’s disease recognition via attention-based sparse graph convolutional neural network. *IEEE Journal of Biomedical and Health Informatics* **27**(11), 5216–5224 (2023)
3. Escudero, J., Abásolo, D., Hornero, R., Espino, P., López, M.: Analysis of electroencephalograms in Alzheimer’s disease patients with multiscale entropy. *Physiological Measurement* **27**(11), 1091–1106 (2006)
4. Fan, W., Fei, J., Guo, D., Yi, K., Song, X., Xiang, H., Ye, H., Li, M.: Towards multi-resolution spatiotemporal graph learning for medical time series classification. In: *Proceedings of the ACM on Web Conference 2025*. pp. 5054–5064. ACM (2025)
5. Gandelsman, Y., Sun, Y., Chen, X., Efros, A.A.: Test-time training with masked autoencoders. In: *Advances in Neural Information Processing Systems*. vol. 35, pp. 29374–29385 (2022)
6. Gui, Y., Chen, M., Su, Y., Luo, G., Yang, Y.: Eegmamba: Bidirectional state space model with mixture of experts for eeg multi-task classification. *arXiv preprint arXiv:2407.20254* (2024)
7. Han, D., Li, Y., Li, T., Cao, Z., Wang, Z., Song, J., Cheng, Y., Zheng, B., Huang, G.: ViT³: Unlocking test-time training in vision. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 51–61 (2026), oral presentation
8. Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y.: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine* **25**(1), 65–69 (2019)

9. Kachare, P., Puri, D., Sangle, S.B., Al-Shourbaji, I., Jabbari, A., Kirner, R., Alameen, A., Migdady, H., Abualigah, L.: Lcadnet: a novel light CNN architecture for EEG-based Alzheimer disease detection. *Physical and Engineering Sciences in Medicine* **47**(3), 1037–1050 (2024)
10. Kim, D., Lee, K.R., Lim, D.S., Lee, K.H., Lee, J.S., Kim, D.Y., Sohn, C.B.: A novel hybrid CNN-transformer model for arrhythmia detection without R-peak identification using Stockwell transform. *Scientific Reports* **15**(1), 7817 (2025)
11. Li, J., Li, X., Mao, Y., Yao, J., Gao, J., Liu, X.: Classification of Parkinson’s disease EEG signals using 2D-MDAGTS model and multi-scale fuzzy entropy. *Biomedical Signal Processing and Control* **91**, 105872 (2024)
12. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: iTransformer: Inverted transformers are effective for time series forecasting. In: *International Conference on Learning Representations* (2024)
13. Miltiadous, A., Tzamourta, K.D., Afrantou, T., Ioannidis, P., Grigoriadis, N., Tsilikakis, D.G., Angelidis, P., Tsipouras, M.G., Glavas, E., Giannakeas, N., Tzallas, A.T.: A dataset of EEG recordings from: Alzheimer’s disease, Frontotemporal dementia and healthy subjects (2024)
14. Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers. In: *International Conference on Learning Representations* (2023)
15. Sun, Y., Li, X., Dalal, K., Xu, J., Vikram, A., Zhang, G., Dubois, Y., Chen, X., Wang, X., Koyejo, S., Hashimoto, T., Guestrin, C.: Learning to (learn at test time): RNNs with expressive hidden states. In: *Proceedings of the 42nd International Conference on Machine Learning*. *Proceedings of Machine Learning Research*, vol. 267, pp. 57503–57522. PMLR (2025)
16. Tadesse, G.A., Javed, H., Weldemariam, K., Zhu, T.: A spectral-longitudinal model for detection of heart attack from 12-lead electrocardiogram waveforms. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. pp. 6009–6012 (2020)
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: *Advances in Neural Information Processing Systems*. vol. 30, pp. 5998–6008 (2017)
18. Wagner, P., Strodthoff, N., Bousseljot, R.D., Kreiseler, D., Lunze, F.I., Samek, W., Schaeffter, T.: PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data* **7**(1), 154 (2020)
19. Wang, Y., Huang, N., Li, T., Yan, Y., Zhang, X.: Medformer: A multi-granularity patching transformer for medical time-series classification. In: *Advances in Neural Information Processing Systems*. vol. 37, pp. 36314–36341 (2024)
20. Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In: *Advances in Neural Information Processing Systems*. vol. 34, pp. 22419–22430 (2021)
21. Ye, J., Zhang, W., Li, Z., Li, J., Tsung, F.: Medspaformer: A transferable transformer with multi-granularity token sparsification for medical time series classification. *Proceedings of the AAAI Conference on Artificial Intelligence* **40**(33), 27791–27799 (2026)
22. Zhang, Y., Li, J.: Application of heartbeat-attention mechanism for detection of myocardial infarction using 12-lead ECG records. *Applied Sciences* **9**(16), 3328 (2019)
23. Zhang, Y., Yan, J.: Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In: *International Conference on Learning Representations* (2023)

24. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* **35**(12), 11106–11115 (2021)
25. Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., Jin, R.: FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In: *Proceedings of the 39th International Conference on Machine Learning. Proceedings of Machine Learning Research*, vol. 162, pp. 27268–27286. PMLR (2022)
26. Zhou, Y., Zhao, S., Wang, J., Jiang, H., Li, S., Luo, B., Li, T., Pan, G.: Personalized sleep staging leveraging source-free unsupervised domain adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence* **39**(13), 14529–14537 (2025)
27. Zini, S., Barbera, T., Bianco, S., Napoletano, P.: Alzheimer’s disease classification from EEG using a multiscale temporal deep network. *Biomedical Signal Processing and Control* **114**, 109321 (2026)