CMDA-GCN: Adaptive Graph Convolution with Confidence-Margin Semi-Supervised for Cross-Subject Working-Memory Load Decoding

Congming Tan[1], Zihan Geng[1], Dejing Li[2], Yuhao Jiang[2] & Yin Tian (tianyin@cqupt.edu.cn) [1,2,*]
[1]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications
[2]School of Life Health Information Science and Engineering, Chongqing University of Posts and Telecommunications

Abstract

Accurate assessment of working memory load (WML) is important for cognitive neuroscience and brain-computer interfaces, as WML reflects the recruitment of attentional control and memory resources during encoding, maintenance, and updating. However, WML-related EEG is highly nonstationary: ERP signatures vary in latency and amplitude, and these variations are amplified across subjects, leading to distribution shifts. We propose CMDA-GCN, an adaptive spatiotemporal graph convolutional network based on Confidence-Margin Semi-Supervised Domain Adaptation (CMDA). The model combines multi-scale temporal convolutions with time attention to capture transient ERP dynamics and accommodate latency variability, and uses adaptive graph convolutions to learn task-dependent inter-electrode interactions. To mitigate domain drift, CMDA-GCN integrates multi-source adversarial alignment with confidence-based pseudo-label learning and a hard-sample margin constraint, jointly suppressing pseudo-label noise and sharpening decision boundaries. Experiments on two datasets show consistent performance gains, and visualizations highlight workload-sensitive time windows and distributed interactions, supporting robust and interpretable WML estimation.

Keywords: working memory load; domain adaptation; ERP; graph convolutions

I. Introduction

Cognitive working memory load (WML) reflects the degree to which information-processing resources are engaged during task execution and is a key determinant of performance and reliability (Arvidsson et al., 2020). As high-risk settings such as education, human-computer interaction, and transportation increasingly demand sustained attention and efficiency, both overload and underload can impair performance and elevate error risk, underscoring the need for objective, real-time assessment. Applications have expanded from online learning (S. Kim et al., 2024; Tang et al., 2024) and driving/aviation (Yiu et al., 2025) to auxiliary diagnostic contexts including depression, schizophrenia, and autism spectrum disorder (Cheng et al., 2025; Han et al., 2026; Sarisik et al., 2025). Compared with subjective scales, physiological measures are better suited for continuous monitoring (Li et al., 2025; Tao et al., 2019), and EEG is particularly attractive due to its noninvasiveness, portability, low cost, and high temporal resolution (Das Chakladar & Roy, 2024; Zhou et al., 2022), with demonstrated utility in related decoding tasks such as emotion and sleep monitoring (Tan, Pan, et al., 2025; Tan, Xu, et al., 2025).

Despite its promise, EEG-based WML decoding remains challenging because EEG is weak, noisy, and nonstationary, with pronounced inter-subject variability that limits robust feature learning and cross-subject generalization. Traditional pipelines rely on handcrafted features—e.g., increased theta and decreased alpha power with higher workload (Pergher et al., 2019; Tan et al., 2024), ERP markers such as P300 attenuation (Brouwer et al., 2012), and complexity or fused features (Tian et al., 2019) combined with classifiers (e.g., SVM/LDA), but these approaches are sensitive to temporal variability and depend heavily on feature engineering. End-to-end deep learning offers a more flexible alternative: CNNs can learn hierarchical representations from raw or minimally processed EEG (Chen et al., 2025; Chen et al., 2026), with compact models such as EEGNet showing strong efficiency across paradigms (Lawhern et al., 2018), and later work exploring multi-band 2D mappings (Yuan et al., 2025), 3D convolutions with attention (Zhang et al., 2021), and deeper architectures for applied scenarios (Chen et al., 2026). However, existing approaches still face limitations in joint spatiotemporal modeling: while 1D-CNNs and multi-scale designs capture temporal dynamics effectively (Ding et al., 2020; Zhang et al., 2025), spatial modeling via electrode sequences or 2D image-like mappings often fails to explicitly represent inter-regional functional connectivity and coordinated interactions (Siddharth et al., 2022). Given that higher-order cognition involves both region-specific activations and large-scale cooperation, a framework that jointly models temporal dynamics and spatial topological interactions remains needed.

Motivated by these considerations, we propose CMDA-GCN, an end-to-end adaptive spatiotemporal graph learning framework for cross-subject EEG-based WML decoding (Fig. 1). The framework jointly learns discriminative and domain-invariant representations from temporal and spatial perspectives. Temporally, multi-scale modeling with time attention captures time-varying ERP dynamics and accommodates latency variability; spatially, an adaptive graph explicitly models inter-regional topological interactions. For optimization, multi-source adversarial domain learning is combined with target-domain semi-supervised learning to reduce performance degradation caused by cross-subject distribution shift. Experiments on a public N-back dataset and a self-collected delayed match-to-sample dataset show consistent improvements, and visualization further highlights task-relevant temporal windows and inter-regional interactions, supporting robust

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

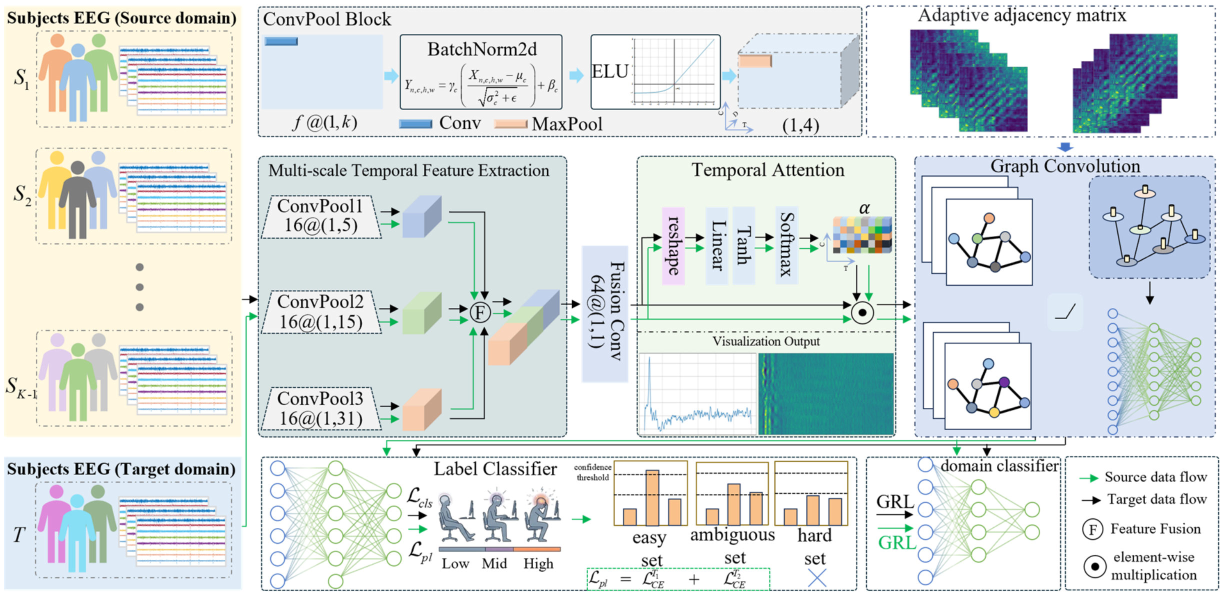

Figure 1: Overall architecture of the proposed CMDA-GCN framework.

and interpretable WML assessment. The main contributions are summarized as follows:

1. Multi-scale temporal modeling with time attention: We develop a temporal module that combines parallel convolutions with different receptive fields and a time-attention mechanism. This design captures workload-related dynamics at multiple time scales and adaptively highlights informative post-stimulus windows, improving both the modeling and interpretability of transient ERP patterns and latency variability.

2. Adaptive graph-based spatial interaction modeling: We propose a data-driven spatial learning scheme that end-to-end optimizes inter-electrode relationships as an adaptive graph. Residual connections and normalization are incorporated after graph propagation to stabilize training and enhance representation capacity, enabling more effective characterization of functional coupling topologies across channels.

3. Multi-source adversarial learning with confidence-margin semi-supervised optimization: We split source subjects into multiple sub-domains and perform multi-source adversarial alignment via gradient reversal to learn domain-invariant features. For the unlabeled target domain, we integrate high-confidence pseudo-label self-training with a hard-sample margin constraint and a warm-up schedule, reducing pseudo-label noise and negative transfer and thereby improving cross-subject generalization.

II. Methodology

A. Overview of CMDA-GCN Model

Let a single-trial EEG segment be denoted as $X \in \mathbb{R}^{B \times C \times L}$, where B is the batch size, C is the number of electrodes, and L is the number of time points. As shown in Fig. 1, CMDA-GCN learns representations that are discriminative for WML while robust to inter-subject variability. The backbone comprises temporal representation learning and spatial topological interaction modeling. Temporally, multi-scale convolutions capture transient dynamics across time scales, and a time-attention module highlights informative windows. Spatially, electrodes are treated as graph nodes; inter-electrode relations are encoded by a learnable adjacency matrix and propagated via graph convolution. During training, supervised learning on labeled source data is complemented with domain-invariant representation learning and target-domain self-training, improving stability under cross-subject evaluation.

B. Multi-scale Temporal Feature Extraction and Temporal Attention

EEG/ERP signals under WML exhibit strong transients, including latency shifts and amplitude fluctuations, and contain dynamics at multiple temporal scales. Therefore, single-scale filtering is often insufficient to capture both brief ERP transients and slower task-related modulations. As shown in Fig. 1, our temporal module employs parallel multi-scale temporal convolution branches with different receptive fields to perform time-domain filtering and downsampling. Let the kernel length of the k-th branch be $s_k \in \{5, 15, 31\}$. A 2D convolution $(1 \times s_k)$ is applied to the input with the kernel sliding only along the time axis, producing the branch-specific feature map $H^{(k)}$:

$$H^{(k)} = Drop\left(Pool(\sigma(BN(X * W^{(k)})))\right) \tag{1}$$

where $*$ denotes convolution, BN is batch normalization, is the ELU activation, $Pool$ is temporal pooling, and $Drop$ is dropout. The outputs of all branches are concatenated along the feature dimension and fused by a 1×1 convolution:

$$H = Merge\left(\left[H^{(1)};H^{(2)};H^{(3)}\right]\right). \tag{2}$$

To explicitly capture the unequal importance of different time windows, we further introduce time attention to adaptively reweight each electrode's temporal sequence. Let the fused features be $H \in \mathbb{R}^{B\times C\times T\times D}$. For node *n*, an attention score and its normalized weight are computed for the feature vector $h_{n,t} \in \mathbb{R}^D$ at time step *t*:

$$e_{n,t} = \tanh(w^T h_{n,t} + b) \tag{3}$$

$$\alpha_{n,t} = \frac{exp(e_{n,t})}{\sum_{\tau=1}^{T} exp(e_{n,\tau})}. \tag{4}$$

The temporal features are then reweighted to emphasize informative segments:

$$\tilde{h}_{n,t} = \alpha_{n,t} h_{n,t}, \tilde{H} = \left\{\tilde{h}_{n,t}\right\}_{n,t} \tag{5}$$

The attention weight $\alpha_{n,t}$ contributes to representation learning and also supports interpretability by localizing task-relevant time windows. Finally, each node's attended time-series features are flattened and linearly projected to obtain the node embedding $X_g^{(0)} \in \mathbb{R}^{B\times C\times F}$, which is fed into the graph convolution module.

C. Adaptive Graph Convolution

Along the spatial dimension, each electrode is treated as a graph node, and the topology is learned end-to-end to avoid the limited transferability of fixed adjacencies across subjects. Inter-electrode relations are parameterized by a learnable matrix $P \in \mathbb{R}^{C\times C}$ and symmetrized to model undirected interactions, $\tilde{P} = (P + P^T)/2$. We then apply an exponential mapping and add self-loops to obtain a nonnegative, scale-comparable weight matrix $S = exp(\tilde{P}) + I$, and construct the propagation operator via symmetric degree normalization $D_{ii} = \sum S_{ij}$ to mitigate degree-related scale bias and improve numerical stability during message passing $A = D^{1/2} S D^{-1/2}$. In practice, P is initialized using Pearson correlations to speed up convergence and is further refined in a data-driven manner during training.

Given the high noise level and pronounced inter-subject variability of EEG, we further introduce residual learning and normalization after each propagation layer to stabilize optimization and mitigate representation drift. The *l*-th graph propagation is computed as:

$$\hat{X}_g^{(l+1)} = \sigma(A X_g^{(l)} W^{(l)}) \tag{6}$$

$$X_g^{(l+1)} = LN\left(\hat{X}_g^{(l+1)} + X_g^{(l)}\right) \tag{7}$$

Here, $X_g^{(l)} \in \mathbb{R}^{B\times C\times F}$ denotes the input node representations, $W^{(l)}$ is a learnable parameter matrix, and $\sigma(\cdot)$ is a nonlinear activation function.

Finally, the resulting node representations are aggregated across electrodes and fed into a classification head for workload prediction.

D. Confidence-Margin Semi-Supervised Domain Adaptation

In cross-subject settings, substantial distribution shifts exist between source- and target-domain EEG data. To improve generalization, we jointly enforce class discriminability and domain invariance by combining adversarial alignment with target-domain self-training. A domain discriminator imposes an adversarial constraint on the extracted features to suppress subject-specific information, while unlabeled target samples are optimized via self-training with confidence-based filtering and a margin regularizer to reduce the influence of noisy pseudo-labels on the decision boundary. Formally, let $f \in \mathbb{R}^{B\times d}$ denote the global features extracted by the backbone, and let $q = D(\mathrm{GRL}(f))$ be the output of the domain discriminator $D(\cdot)$. The GRL is an identity mapping in the forward pass, but multiplies the gradients by $-\alpha$ during backpropagation, encouraging domain-indistinguishable representations. The coefficient α is scheduled to increase over training to strengthen alignment at later stages.

For unlabeled target-domain samples, the model constructs pseudo-labels based on the predicted probabilities $p = \mathrm{Softmax}(z)$, with $\hat{y} = \arg\max_c p_c$. A confidence-thresholding strategy is then applied: the cross-entropy loss on pseudo-labels is enforced only when $\hat{p} = \max_c p_c \geq \tau_{\mathrm{high}}$ thereby suppressing noisy supervision during early training and on uncertain samples. Meanwhile, for samples falling into an uncertainty interval $\tau_{low} \leq \hat{p} < \tau_{high}$, we introduce a logit-margin regularization term to sharpen the decision boundary. Let the largest and second-largest logits be $z_{(1)}$ and $z_{(2)}$, respectively, and define the margin as $\Delta = z_{(1)} - z_{(2)}$. The margin constraint can then be formulated as:

$$\mathcal{L}_m = \max(0, m - \Delta), \tag{8}$$

Here, *m* is a predefined margin. This term encourages a larger separation between the top two competing outputs, thereby reducing unstable predictions for hard samples near the decision boundary.

E. Loss Function

Combining the above objectives, the overall training loss consists of a supervised classification term on the source domain, an adversarial domain discrimination term, and a target-domain self-training regularization term. Let y_s denote the source-domain labels and z_s the classification logits. The source-domain classification loss is defined as:

$$\mathcal{L}_{cls} = \mathcal{L}_{CE}(z_s, y_s), \tag{9}$$

The domain discrimination loss is formulated as a cross-entropy term, computed on both the source and target domains and then averaged:

$$\mathcal{L}_{dom} = \frac{1}{2}\left(\mathcal{L}_{CE}(D(f_s), d_s) + \mathcal{L}_{CE}(D(f_t), d_t)\right), \tag{10}$$

where d_s and d_t denote the corresponding domain labels. The target-domain pseudo-label loss under confidence thresholding is defined as:

$$\mathcal{L}_{pl} = \mathcal{L}_{CE}(z_t, \hat{y}) \cdot \mathbb{I}(\hat{p} \geq \tau_{\text{high}}) \tag{11}$$

Together with the margin term $\mathcal{L}_{\text{m}}$, this forms the target-domain regularization. The final optimization objective is given by:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_{dom}\mathcal{L}_{dom} + w_{ssl}(\lambda_{pl}\mathcal{L}_{pl} + \lambda_m \mathcal{L}_m), \tag{12}$$

Here, $\lambda_{dom} = 0.1$, $\lambda_{pl} = 1.0$, and $\lambda_m = 0.5$ are weighting coefficients, and $w_{\text{ssl}} = min(1, e/10)$ is a scheduling factor that varies with training progress. This objective maintains source-domain discriminability while mitigating cross-subject distribution shift via adversarial alignment and target-domain self-training regularization. In our implementation, the confidence threshold $\tau_{high} = 0.8$, the uncertainty lower bound $\tau_{low} = 0.6$, and the margin m=0.3.

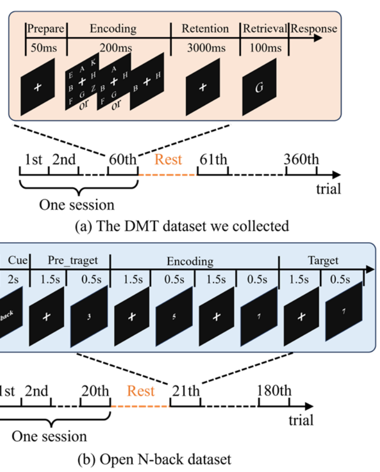

Figure 2: Experimental paradigms of the datasets.

III. Results

A. Datasets and Preprocessing

Twenty healthy participants (normal or corrected-to-normal vision; no psychiatric/neurological history) completed a delayed match-to-sample task (DMT) while EEG was recorded with a 64-channel NeuroScan system. Three workload levels were included: 2 (low), 4 (mid), and 8 (high) items, administered across six sessions (60 trials/session; conditions equiprobable). As shown in Fig. 2(a), each trial consisted of a 50 ms flashing fixation, a 200 ms memory sequence, a 3000 ms retention interval, and a 100 ms probe; participants reported whether the probe appeared in the memory set (F = absent; J = present). We also evaluated on a public N-back dataset with EEG from 26 right-handed participants recorded using 30 electrodes. Participants performed 0-back, 2-back, and 3-back tasks corresponding to low, mid-level, and high workload levels, respectively (Fig. 2(b)); details are provided in (Shin et al., 2018).

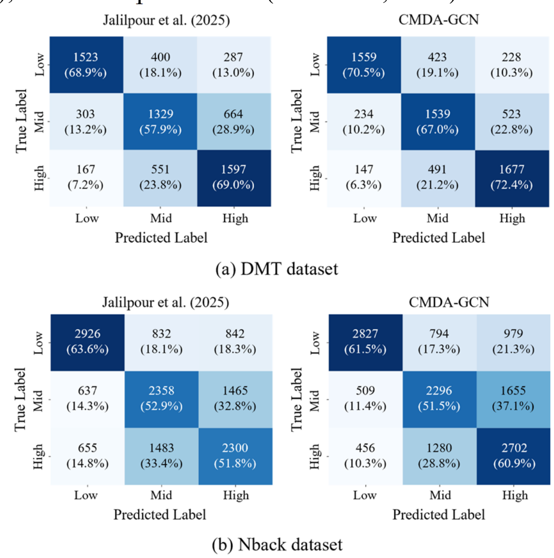

Figure 3: Overall confusion matrices for (a) the DMT dataset and (b) the N-back dataset. The top number denotes trial count and the bottom number denotes the percentage.

B. Baselines and Experimental Setup

We compare the proposed method with widely used deep learning baselines for EEG classification, including models originally developed for ERP decoding, BCI applications, and motor imagery. The baselines are ShallowNet/DeepNet (Schirrmeister et al., 2017), EEGNet (Lawhern et al., 2018), EEG-Inception (Santamaría-Vázquez et al., 2020), JMNet (J. M. Kim et al., 2024), EEGNex (Chen et al., 2024), and Jalilpour et al. (2025). These architectures have been applied to P300, movement-related cortical potentials, sensorimotor rhythms, and other BCI paradigms. Performance was evaluated using subject-independent 10-fold cross-validation and reported in terms of Accuracy (ACC), Cohen's kappa (Kappa), Macro-Sensitivity (MS), Macro-F1 (MF1), and Macro-Precision (MP).

C. Working Memory Detection

On the DMT dataset, as summarized in Table 1, CMDA-GCN achieved 70.10%, 55.10%, 70.12%, 70.27%, and 71.02% in Accuracy, Kappa, MS, MF1, and MP, respectively, representing the best performance among all compared methods. Compared with the strongest baseline (Jalilpour et al.), CMDA-GCN improved these five metrics by 4.85, 7.30, 4.87, 4.92, and 4.88 percentage points, respectively. Further inspection of the confusion matrix in Fig. 3(a) indicates more stable recognition of the mid-level load (4 items) and high-load (8 items) conditions, with class-wise accuracies of 67.0% and 72.4%, respectively. The dominant confusions occur between adjacent load levels, with 22.8% of mid-level load

Table 1: Performance comparison between the proposed method and baseline models on the DMT dataset.

Model	ACC	Kappa	MS	MF1	MP
ShallowNet	59.17±3.41	38.68±5.16	59.21±3.42	58.72±3.55	59.38±3.44
DeepNet	58.19±2.75	37.21±4.14	58.23±2.81	57.64±3.39	59.18±2.59
EEGNet	59.37±2.88	39.02±4.33	59.42±2.93	59.12±3.16	60.29±3.12
EEG-inception	60.64±2.75	40.89±4..15	60.69±2.77	60.57±3.14	61.64±3.03
EEGNeX	62.34±2.99	43.49±4.33	62.47±2.89	61.83±3.45	62.81±2.61
JMNet	60.85±4.38	41.25±6.57	60.99±4.40	60.01±5.17	61.41±3.67
Jalilpour et al.	65.25±2.81	47.80±4.20	65.25±2.80	65.35±2.85	66.14±2.92
ours	70.10±3.33	55.10±4.95	70.12±3.35	70.27±3.35	71.02±3.24

Table 2: Performance comparison between the proposed method and baseline models on the N-back dataset.

Model	ACC	Kappa	MS	MF1	MP
ShallowNet	50.59±3.83	25.62±5.71	50.22±3.77	47.90±4.34	49.98±4.10
DeepNet	49.17±3.21	23.68±4.80	48.98±3.17	47.38±3.58	49.40±3.53
EEGNet	52.82±3.86	29.11±5.76	52.52±3.76	51.26±4.20	52.85±4.25
EEG-inception	50.50±3.80	25.66±5.73	50.27±3.73	49.23±4.36	50.61±4.03
EEGNeX	52.93±4.39	29.21±6.63	52.59±4.42	50.44±5.73	52.77±4.89
JMNet	49.87±3.66	24.63±5.47	49.57±3.62	48.10±4.25	49.45±4.05
Jalilpour et al.	56.79±3.73	35.19±5.59	56.68±3.74	56.72±3.92	57.81±3.84
ours	58.67±4.46	38.08±6.44	58.67±4.42	58.73±4.49	60.80±4.16

trials misclassified as high load and 21.2% of high-load trials misclassified as mid-level load, suggesting improved robustness in discriminating neighboring workload states. Overall, CMDA-GCN shows a clear advantage in identifying the mid- and high-load conditions and effectively suppresses severe cross-level errors, particularly high-load trials misclassified as low load (misclassification rate 6.3%), demonstrating its ability to separate adjacent workload levels more reliably and reduce cross-level confusion.

On the N-back dataset, as reported in Table 2, CMDA-GCN achieved 58.67%, 38.08%, 58.67%, 58.73%, and 60.80% in Accuracy, Kappa, MS, MF1, and MP, respectively, outperforming all competing methods overall. Compared with Jalilpour et al. (2025), CMDA-GCN improved Accuracy, Kappa, and MF1 by 1.88, 2.89, and 2.01 percentage points, respectively. The confusion matrix in Fig. 3(b) further shows that CMDA-GCN attains a 60.9% recognition accuracy for the high-load (3-back) condition, with most errors arising from misclassification as medium load (28.8%) and low load (10.3%). For the mid-level load (2-back) condition, the accuracy reaches 51.5%, and misclassifications are primarily shifted toward the high-load class (37.1%), indicating that some ambiguity remains at the mid–high workload boundary. Nevertheless, the overall discriminative performance and agreement metrics are the best or near-best among all methods. In summary, the confusion-matrix patterns are consistent with the improvements in Table 2, suggesting that CMDA-GCN enhances the separability of medium and high workload states and reduces severe misclassifications in the more challenging cross-subject three-class workload recognition setting, thereby yielding superior overall performance.

D. Visualization

To enhance interpretability, we visualize key representations on the DMT (Fig. 4) and N-back (Fig. 5) datasets. For DMT, Fig. 4(a) presents event-locked whole-brain ERP waveforms and representative channels (Fz, Cz, POz), showing workload-dependent differences with notable latency and amplitude variability in early transients. Fig. 4(b) shows that temporal attention peaks in an early post-stimulus window and then stabilizes, indicating that the model emphasizes informative transient intervals rather than weighting the entire epoch uniformly. Fig. 4(c) further reveals that only a subset of electrodes receives high contributions during this key window, suggesting adaptive selection of workload-relevant spatial cues. Fig. 4(d) visualizes the learned adaptive adjacency, which is sparse yet structured, capturing both local coupling and long-range interactions with partial left-right consistency. Consistent patterns are observed on N-back: electrode-wise weights in Fig. 5(a) emphasize a subset of channels with stronger concentration over anterior regions, temporal attention in Fig. 5(b) again highlights an early window across workload levels, and the adaptive adjacency in Fig. 5(c) exhibits a similarly sparse but structured topology that preserves local coordination while supporting long-range interactions, collectively providing evidence for stable spatiotemporal cues used in cross-subject workload recognition.

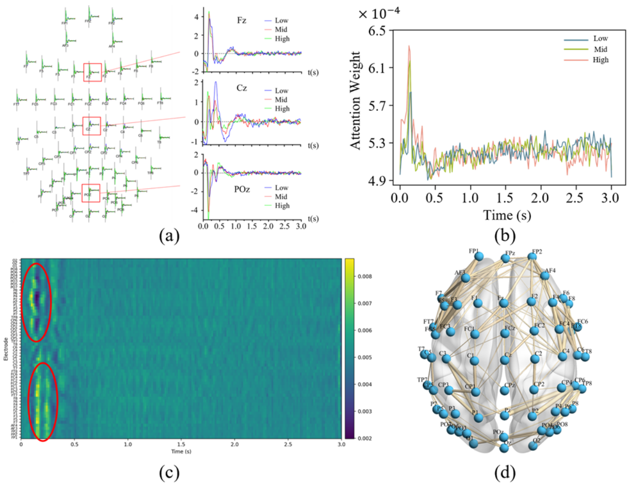

Figure 4: Visualization results on the DMT dataset.

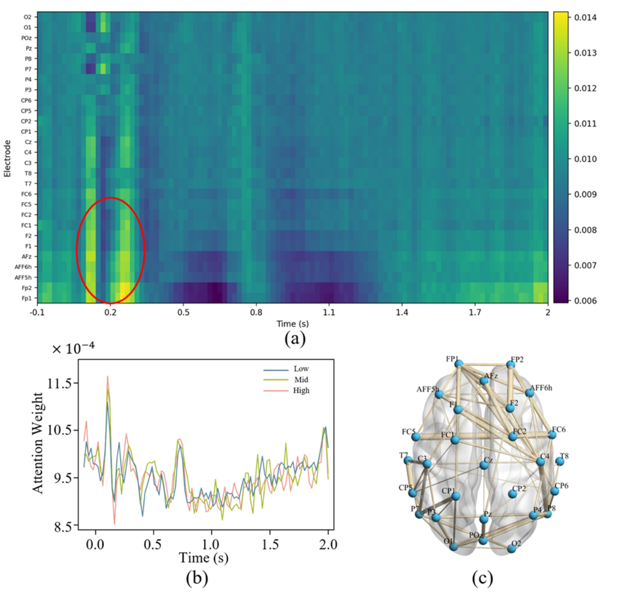

Figure 5: Visualization results on the N-back dataset.

IV. Discussion

ERP evidence and the attention visualizations provide a convergent temporal account of the model's decision cues. In DMT (Fig. 4(a)), workload effects are prominent in early components (around the P200/N200 range) and later positive activity (P300), consistent with reports that cognitive load modulates both early perceptual-attentional processing and later updating-related processes (Ren et al., 2023; Yu et al., 2024). Notably, temporal attention in Fig. 4(b) concentrates around ~200 ms, suggesting that the model preferentially weights an early post-stimulus window rather than integrating the entire epoch uniformly. This is a plausible cross-subject strategy because early sensory–attentional operations tend to exhibit more stable timing across individuals than later components, which can be more sensitive to strategy and inter-subject variability. A similar early-window emphasis is observed in N-back (Fig. 5(b)), supporting the view that early transient responses provide more transferable discriminative cues across subjects.

Spatial visualizations further indicate that workload-relevant information is spatially sparse. In DMT, the electrode-weight map (Fig. 4(c)) assigns higher contributions to a subset of channels within the key interval, implying that discriminative information is concentrated in localized sensor clusters rather than being globally distributed. The N-back weights show a comparable pattern (Fig. 5(a)), with stronger emphasis over anterior electrodes, consistent with greater engagement of control-related processes in updating demands (Huang et al., 2025; Ren et al., 2023).. Finally, the adaptive adjacency patterns reveal task-consistent interaction structures. DMT exhibits denser local coupling within frontal and posterior regions (Fig. 4(d)), compatible with a functional division in delayed match-to-sample processing in which frontal systems support maintenance/control and posterior systems support representational fidelity and matching-related comparisons (Li et al., 2022). In contrast, N-back shows more frontal-dominant strengthening (Fig. 5(c)), aligning with sustained monitoring, interference control, and online updating requirements (Huang et al., 2025). In both tasks, the emergence of sparse yet structured graphs is consistent with network accounts emphasizing efficient cognition through a balance of local specialization and selective long-range integration (Palma-Espinosa et al., 2025). Overall, Figs. 3-5 suggest that CMDA-GCN's gains likely derive from emphasizing transferable early temporal signatures, focusing on informative channel subsets, and learning task-consistent structured interactions that support robust spatial propagation under cross-subject shifts.

V. Conclusion

This paper proposes CMDA-GCN for cross-subject working memory load recognition and validates it on two paradigms: delayed match-to-sample and N-back tasks. CMDA-GCN integrates multi-scale temporal modeling with time attention, adaptive spatial graph propagation, and a unified optimization strategy that combines multi-source adversarial alignment with target-domain semi-supervised learning, enabling stable and transferable representations under noisy EEG and pronounced inter-subject variability. Experiments show that CMDA-GCN outperforms all baselines on both datasets and reduces confusions between adjacent workload levels, indicating improved cross-subject generalization. Visualization further supports the neurophysiological plausibility of the learned cues: temporal attention highlights workload-sensitive time windows, electrode-wise weights emphasize informative spatial clusters, and adaptive adjacency patterns are task-consistent—showing stronger frontal–posterior local coupling in DMT and more frontal connectivity in N-back, consistent with maintenance/matching versus cognitive control and online updating demands. Overall, CMDA-GCN provides an effective end-to-end solution that improves both performance and interpretability, and future work will explore source-space analyses and broader multi-device/multi-site validation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant W2411084; the Chongqing Natural Science Fund Innovation and Development Joint Fund (Municipal Education Commission) project under Grant CSTB2024NSCQ-LZX0121 and the funding of Institute for Advanced Sciences of Chongqing University of Posts and Telecommunications under Grant E011A2022327.

References

Arvidsson, H., Larsson, G., Larsolle, A., Neely, G., & Hansson, P.-A. (2020). Easily Applicable Methods for Measuring the Mental Load on Tractor Operators. *Journal of agricultural safety and health*, *26*(1), 5-14. https://doi.org/10.13031/jash.13223

Brouwer, A. M., et al. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, *9*(4), 14, Article 045008. https://doi.org/10.1088/1741-2560/9/4/045008

Chen, J. C., Cui, Y. G., Wei, C. F., Polat, K., & Alenezi, F. (2026). Driver fatigue detection using EEG-based graph attention convolutional neural networks: An end-to-end learning approach with mutual information-driven connectivity. *Applied Soft Computing*, *186*, 15, Article 114097. https://doi.org/10.1016/j.asoc.2025.114097

Chen, X., Teng, X. B., Chen, H., Pan, Y. F., & Geyer, P. (2024). Toward reliable signals decoding for electroencephalogram: A benchmark study to EEGNeX. *Biomedical Signal Processing and Control*, *87*, 12, Article 105475. https://doi.org/10.1016/j.bspc.2023.105475

Cheng, C., Liu, W. Z., Jia, Z. Y., & Luo, W. B. (2025). A Multi-stage Hemisphere Asymmetry Fusion Network Inspired by the brain for EEG depression detection. *Information Fusion*, *124*, 14, Article 103342. https://doi.org/10.1016/j.inffus.2025.103342

Das Chakladar, D., & Roy, P. P. (2024). Cognitive workload estimation using physiological measures: a review. *Cognitive Neurodynamics*, *18*(4), 1445-1465. https://doi.org/10.1007/s11571-023-10051-3

Ding, Y., et al. (2020). TSception: A deep learning framework for emotion detection using EEG. *2020 international joint conference on neural networks (IJCNN)*, 1-7,

Han, L., Huang, Z. M., Li, C. S., & Chen, J. C. (2026). Diagnosis of autism spectrum disorder using EEG microstates and chaos game representation. *Biomedical Signal Processing and Control*, *112*, 11, Article 108722. https://doi.org/10.1016/j.bspc.2025.108722

Huang, S. J., et al. (2025). Exploring the n-back task: insights, applications, and future directions. *Frontiers in Human Neuroscience*, *19*, 27, Article 1721330. https://doi.org/10.3389/fnhum.2025.1721330

Jalilpour, S., & Müller-Putz, G. (2025). A framework for Interpretable deep learning in cross-subject detection of event-related potentials. *Engineering Applications of Artificial Intelligence*, *139*, 14, Article 109642. https://doi.org/10.1016/j.engappai.2024.109642

Kim, J. M., et al. (2024). A learnable continuous wavelet-based multi-branch attentive convolutional neural network for spatio-spectral-temporal EEG signal decoding. *Expert Systems with Applications*, *251*, 10, Article 123975. https://doi.org/10.1016/j.eswa.2024.123975

Kim, S., et al. (2024). Characteristic Behaviors of Elementary Students in a Low Attention State During Online Learning Identified Using Electroencephalography. *Ieee Transactions on Learning Technologies*, *17*, 619-628. https://doi.org/10.1109/tlt.2023.3289498

Lawhern, V. J., et al. (2018). EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, *15*(5), 17, Article 056013. https://doi.org/10.1088/1741-2552/aace8c

Li, C. B., et al. (2025). EEG-Based Emotion Monitoring and Regulation System by Learning the Discriminative Brain Network Manifold. *Ieee Transactions on Neural Networks and Learning Systems*, *36*(10), 17751-17765. https://doi.org/10.1109/tnnls.2025.3576182

Li, X. Q., O'Sullivan, M. J., & Mattingley, J. B. (2022). Delay activity during visual working memory: A meta-analysis of 30 fMRI experiments. *Neuroimage*, *255*, 19, Article 119204. https://doi.org/10.1016/j.neuroimage.2022.119204

Palma-Espinosa, J., Orellana-Villota, S., Coronel-Oliveros, C., Maidana, J. P., & Orio, P. (2025). The balance between integration and segregation drives network dynamics maximizing multistability and metastability. *Scientific Reports*, *15*(1), 13, Article 18811. https://doi.org/10.1038/s41598-025-01612-z

Pergher, V., Wittevrongel, B., Tournoy, J., Schoenmakers, B., & Van Hulle, M. M. (2019). Mental workload of young and older adults gauged with ERPs and spectral power during N-Back task performance. *Biological Psychology*, *146*, 8, Article 107726. https://doi.org/10.1016/j.biopsycho.2019.107726

Ren, M., et al. (2023). Neural signatures for the n-back task with different loads: An event-related potential study. *Biological Psychology*, *177*, 12, Article 108485. https://doi.org/10.1016/j.biopsycho.2023.108485

Santamaría-Vázquez, E., Martínez-Cagigal, V., Vaquerizo-Villar, F., & Hornero, R. (2020). EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-Based Brain-Computer Interfaces. *Ieee Transactions on Neural Systems and Rehabilitation Engineering*, *28*(12), 2773-2782. https://doi.org/10.1109/tnsre.2020.3048106

Sarisik, E., et al. (2025). EEG-based Signatures of Schizophrenia, Depression, and Aberrant Aging: A Supervised Machine Learning Investigation. *Schizophrenia Bulletin*, *51*(3), 804-817. https://doi.org/10.1093/schbul/sbae150

Schirrmeister, R. T., et al. (2017). Deep Learning With Convolutional Neural Networks for EEG Decoding and

Visualization. *Human Brain Mapping*, *38*(11), 5391-5420. https://doi.org/10.1002/hbm.23730

Shin, J., et al. (2018). Data Descriptor: Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. *Scientific Data*, *5*, 16, Article 180003. https://doi.org/10.1038/sdata.2018.3

Siddharth, Jung, T. P., & Sejnowski, T. J. (2022). Utilizing Deep Learning Towards Multi-Modal Bio-Sensing and Vision-Based Affective Computing. *Ieee Transactions on Affective Computing*, *13*(1), 96-107. https://doi.org/10.1109/taffc.2019.2916015

Tan, C., Pan, T., & Tian, Y. (2025). GPNet: Granularity-Aware Pyramid Network with Graph Aggregation for Sleep Staging and Face-Emotional Recognition Speed Prediction. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *47*. 923-930, https://escholarship.org/uc/item/4366b688

Tan, C., Xu, J., Hu, L., & Tian, Y. (2025). FusSADGCNN: Decoding the Impact of Transcranial Electrical Stimulation on Neuromodulation in Emotion Recognition and Emotion Elicitation. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2025.3586906

Tan, E. D., et al. (2024). Theta activity and cognitive functioning: Integrating evidence from resting-state and task-related developmental electroencephalography (EEG) research. *Developmental Cognitive Neuroscience*, *67*, 18, Article 101404. https://doi.org/10.1016/j.dcn.2024.101404

Tang, H., Dai, M., Du, X., Hung, J.-L., & Li, H. (2024). An EEG study on college students' attention levels in a blended computer science class. *Innovations in Education and Teaching International*, *61*(4), 789-801.

Tao, D., et al. (2019). A Systematic Review of Physiological Measures of Mental Workload. *International Journal of Environmental Research and Public Health*, *16*(15), 23, Article 2716. https://doi.org/10.3390/ijerph16152716

Tian, Y., Zhang, H. L., Jiang, Y. H., Li, P. Y., & Li, Y. (2019). A Fusion Feature for Enhancing the Performance of Classification in Working Memory Load With Single-Trial Detection. *Ieee Transactions on Neural Systems and Rehabilitation Engineering*, *27*(10), 1985-1993. https://doi.org/10.1109/tnsre.2019.2936997

Yiu, C. Y., Ng, K. K. H., Li, Q. B., & Yuan, X. (2025). Keeping Pilots in the Loop: An Explainable Spatiotemporal EEG-Driven Deep Learning Framework for Adaptive Automation in Cruising Flight Phase. *Ieee Transactions on Intelligent Transportation Systems*, *26*(7), 9838-9851. https://doi.org/10.1109/tits.2025.3567987

Yu, S., Tian, L., Wang, G., & Nie, S. (2024). Which ERP components are effective in measuring cognitive load in multimedia learning? A meta-analysis based on relevant studies. *Frontiers in psychology*, *15*, 1401005.

Yuan, Q. L., et al. (2025). TF-MEET: A Transferable Fusion Multi-Band Transformer for Cross-Session EEG Decoding. *Caai Transactions on Intelligence Technology*, *10*(6), 1799-1812. https://doi.org/10.1049/cit2.70056

Zhang, J., et al. (2025). Working-memory load decoding model inspired by brain cognition based on cross-frequency coupling. *Brain Research Bulletin*, *221*, 12, Article 111206. https://doi.org/10.1016/j.brainresbull.2025.111206

Zhang, Y. S., et al. (2021). An end-to-end 3D convolutional neural network for decoding attentive mental state. *Neural Networks*, *144*, 129-137. https://doi.org/10.1016/j.neunet.2021.08.019

Zhou, Y. Y., et al. (2022). Cognitive Workload Recognition Using EEG Signals and Machine Learning: A Review. *Ieee Transactions on Cognitive and Developmental Systems*, *14*(3), 799-818. https://doi.org/10.1109/tcds.2021.3090217