Supplementary Materials for “DBConformer: Dual-Branch Convolutional Transformer for EEG Decoding”

Ziwei Wang, Hongbin Wang, Tianwang Jia, Xingyi He, Siyang Li, and Dongrui Wu*, *Fellow, IEEE*

I. Results of Cross-Dataset Transfer

To further examine the robustness of DBConformer under distributional shifts, we conducted cross-dataset (CD) transfer experiments, where training and testing were performed on different EEG datasets. This setting is more challenging than within-dataset protocols, as it directly evaluates the model's ability to generalize across heterogeneous recording setups. Specifically, we trained on BNCI2014001 or Blankerz2007 and tested on BNCI2014004. During preprocessing, all EEG channels were aligned to a common subset {C3, Cz, C4}, and the temporal dimension was resampled to 1000 points for consistency.

In addition, we enhanced DBConformer with domain adaptation objectives to further mitigate dataset-level distribution shifts. We incorporated the multi-kernel maximum mean discrepancy loss from deep adaptation network (DAN) [1] and the minimum class confusion (MCC) loss [2], jointly optimized with the standard cross-entropy loss. This combination encourages DBConformer to learn domain-invariant yet discriminative representations.

Table S1 reports the results of cross-dataset transfer experiments. Observe that:

- Across all transfer tasks, DBConformer consistently achieves the best average accuracy without adaptation, outperforming other baselines.
- Incorporating additional cross-dataset data improves generalization for DBConformer, while some baselines such as EEG-Net and EEG Conformer exhibit slight performance degradation, suggesting their weaker robustness under distributional shifts.
- Integrating domain adaptation objectives further enhances performance. Both DAN and MCC lead to gains, with MCC yielding the largest improvements, boosting DBConformer to 70.95% on average.

These results confirm that DBConformer not only generalizes better under dataset-level variability, but also benefits from domain adaptation strategies, highlighting its robustness and suitability for real-world BCI scenarios.

TABLE S1

Average classification accuracies (%) of EEGNet, IFNet, EEG Conformer, and DBConformer on cross-dataset transfer with and without domain adaptation algorithms. “w/o DA” denotes performance without using any domain adaptation algorithms, while “w. DAN” denotes using the DAN loss and “w. MCC” denotes using the MCC loss.

Task	EEGNet			IFNet			EEG Conformer			DBConformer		
	w/o DA	w. DAN	w. MCC	w/o DA	w. DAN	w. MCC	w/o DA	w. DAN	w. MCC	w/o DA	w. DAN	w. MCC
14004 Only	$67.78_{\pm 0.92}$	$67.57_{\pm 1.11}$	$69.88_{\pm 1.15}$	$67.69_{\pm 0.41}$	$67.11_{\pm 0.92}$	$68.82_{\pm 0.47}$	$64.22_{\pm 1.05}$	$64.51_{\pm 1.85}$	$67.99_{\pm 0.59}$	$69.85_{\pm 0.47}$	$70.13_{\pm 0.99}$	$\mathbf{70.69}_{\pm 1.63}$
14001→14004	$65.90_{\pm 0.82}$	$65.93_{\pm 0.71}$	$68.68_{\pm 0.67}$	$68.49_{\pm 0.93}$	$67.62_{\pm 0.62}$	$70.63_{\pm 0.56}$	$64.17_{\pm 0.79}$	$64.79_{\pm 0.52}$	$68.42_{\pm 0.56}$	$70.33_{\pm 0.45}$	$70.52_{\pm 0.97}$	$\mathbf{70.96}_{\pm 0.77}$
Blankerz→14004	$65.56_{\pm 1.24}$	$66.20_{\pm 1.24}$	$65.78_{\pm 1.31}$	$68.56_{\pm 0.58}$	$67.82_{\pm 0.63}$	$67.36_{\pm 0.76}$	$62.66_{\pm 1.26}$	$65.63_{\pm 1.16}$	$65.12_{\pm 1.68}$	$70.37_{\pm 0.17}$	$70.44_{\pm 1.36}$	$\mathbf{71.20}_{\pm 1.42}$
Average	66.41	66.57	68.11	68.25	67.52	68.94	63.68	64.98	67.18	70.18	70.36	**70.95**

II. Additional Evaluation Metrics on MI Classification

To provide a more comprehensive evaluation, we further report three additional metrics alongside accuracy: weighted F1-score, precision, and recall. The weighted F1-score is defined as the class-frequency weighted harmonic mean of precision and recall, ensuring fair evaluation even if class imbalance exists. Precision and recall are computed in the standard way.

As shown in Table S2, DBConformer almost always outperforms baseline models across all four metrics: Accuracy (80.16%), weighted F1 (79.94%), Precision (76.98%), and Recall (86.51%). Since BNCI2014001 is class-balanced, the consistency between weighted F1 and accuracy demonstrates the reliability of our model's predictions. The recall is slightly higher than

Z. Wang, H. Wang, T. Jia, X. He, S. Li, and D. Wu are with Hubei Key Laboratory of Brain-inspired Intelligent Systems, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China.

*Corresponding Author: Dongrui Wu (drwu09@gmail.com).

precision, indicating that DBConformer prioritizes capturing more true positives, while maintaining competitive precision. Overall, these results highlight that DBConformer achieves not only the highest accuracy but also a more favorable balance across complementary performance metrics.

TABLE S2
COMPARISON OF ACCURACY (%), WEIGHTED F1 SCORE (%), PRECISION (%), AND RECALL (%) ON BNCI2014001 DATASET UNDER THE CO SCENARIO.

Model	Accuracy	Weighted F1	Precision	Recall
EEGNet	69.05$_{\pm1.00}$	68.67$_{\pm1.00}$	67.03$_{\pm0.77}$	74.92$_{\pm4.36}$
SCNN	73.57$_{\pm2.36}$	73.35$_{\pm2.35}$	72.34$_{\pm2.30}$	76.35$_{\pm2.77}$
DCNN	59.29$_{\pm1.64}$	50.04$_{\pm1.62}$	58.92$_{\pm2.74}$	33.17$_{\pm3.46}$
FBCNet	68.97$_{\pm1.26}$	67.48$_{\pm1.45}$	67.71$_{\pm2.32}$	75.08$_{\pm2.82}$
ADFCNN	73.73$_{\pm2.26}$	73.44$_{\pm2.25}$	71.61$_{\pm2.27}$	79.52$_{\pm2.21}$
CTNet	73.49$_{\pm2.05}$	72.54$_{\pm1.94}$	74.24$_{\pm3.00}$	69.68$_{\pm4.12}$
EEG Conformer	<u>78.57</u>$_{\pm0.66}$	<u>77.68</u>$_{\pm1.00}$	<u>76.69</u>$_{\pm3.20}$	<u>86.19</u>$_{\pm6.42}$
MSCFormer	75.00$_{\pm1.18}$	74.17$_{\pm1.12}$	73.57$_{\pm1.69}$	82.06$_{\pm3.97}$
TMSA-Net	76.67$_{\pm2.14}$	76.29$_{\pm2.15}$	73.87$_{\pm1.85}$	83.81$_{\pm2.91}$
MSVTNet	74.60$_{\pm3.02}$	74.02$_{\pm3.40}$	71.39$_{\pm4.38}$	84.60$_{\pm2.39}$
SlimSeiz	68.89$_{\pm2.05}$	67.29$_{\pm2.45}$	67.25$_{\pm4.13}$	78.10$_{\pm3.67}$
IFNetV2	77.94$_{\pm0.93}$	77.67$_{\pm1.14}$	76.34$_{\pm0.70}$	81.75$_{\pm2.88}$
DBConformer	**80.16**$_{\pm1.53}$	**79.94**$_{\pm1.73}$	**76.98**$_{\pm1.99}$	**86.51**$_{\pm2.09}$

III. INTERPRETABILITY OF CHANNEL ATTENTION

To investigate the interpretability of the proposed channel attention module, we visualized the attention scores assigned to each EEG channel across 32 trials (a batch) from four MI datasets, illustrated in Fig. S1. We excluded BNCI2014004 from this analysis, as it only contains C3, Cz, and C4 channels and therefore lacks spatial coverage for attention comparison.

A consistent observation is that higher attention weights are repeatedly assigned to central motor-related channels, specifically C3, Cz, and C4, across all datasets. This aligns well with known neurophysiological principles of MI, where these central electrodes correspond to the sensorimotor cortex contralateral to movement imagination. Notably, even in different experimental setups, the model adaptively focuses on these physiologically relevant channels.

These results indicated that the proposed channel attention module can learn informative spatial priors in a data-driven manner, enhancing both performance and interpretability.

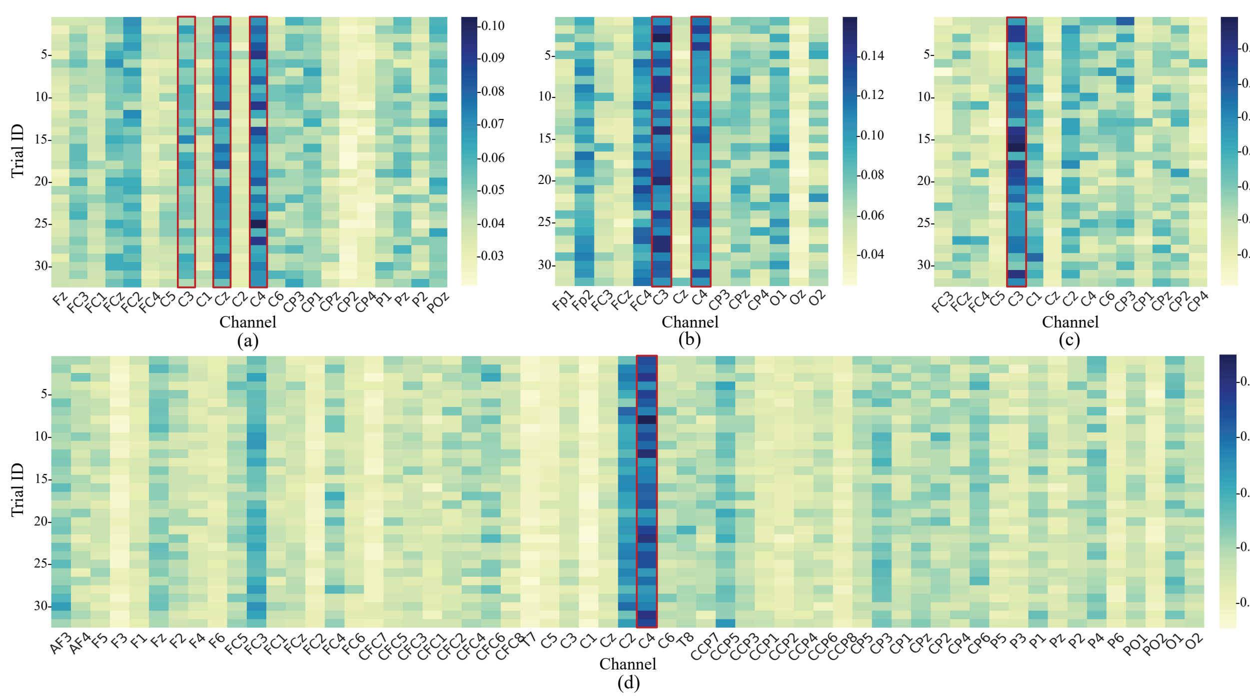

Fig. S1. Attention scores of each EEG channel calculated by the proposed channel attention module on (a) BNCI2014001, (b) Zhou2016, (c) BNCI2014002, and (d) Blankertz2007 datasets. The sum of scores from all channels of each sample is 1.

IV. Causal Interpretability Analysis

As shown in Fig. 6c, we quantified the importance of each EEG channel based on incoming off-diagonal attention weights. The results showed that motor-related electrodes dominate the importance ranking, aligning well with the neuro prior of the MI paradigm.

To further strengthen the interpretability, causal ablation analyses were performed by removing the three most important electrodes (C3, Cz, and C4) from the input EEG signals and re-evaluating performance. We compared DBConformer with representative baselines such as EEGNet, IFNet and EEG Conformer under the CO setting. Results in Table S3 showed that all models experienced a significant performance drop after channel removal, with DBConformer still maintaining a relative advantage over the baselines. This confirms that the channels highlighted in our interpretability analysis are indeed causally linked to decoding accuracy, thereby providing stronger evidence for the plausibility of DBConformer.

TABLE S3
Average classification accuracies (%) of EEGNet, IFNet, EEG Conformer, and DBConformer on three MI datasets under the CO setting. "All" denotes performance using all EEG channels, while "Remove" denotes performance after removing the three most important electrodes (C3, Cz, and C4). "Diff. " represents the accuracy drop caused by this ablation.

Dataset	EEGNet			IFNet			EEG Conformer			DBConformer		
	All	Remove	Diff.	All	Remove	Diff.	All	Remove	Diff.	All	Remove	Diff.
BNCI2014001	69.05$\pm$1.00	68.68$\pm$2.45	0.37	77.94$\pm$0.93	76.11$\pm$1.69	1.83	78.57$\pm$0.66	76.35$\pm$0.59	2.22	**80.16**$\pm$1.82	77.14$\pm$2.26	1.91
Zhou2016	80.13$\pm$3.35	73.04$\pm$6.92	7.09	81.70$\pm$2.08	75.89$\pm$0.98	5.81	73.87$\pm$4.51	70.24$\pm$2.41	3.63	**82.49**$\pm$2.40	76.86$\pm$2.44	5.63
BNCI2014002	66.07$\pm$2.76	64.64$\pm$1.94	1.43	78.29$\pm$1.68	76.79$\pm$0.87	1.50	76.21$\pm$1.46	76.00$\pm$1.92	0.21	**79.14**$\pm$0.17	77.43$\pm$1.46	1.71
Average	71.75	68.79	2.96	79.31	76.26	3.05	76.22	74.20	2.02	**80.57**	77.14	3.08

V. Effect of Fusion Strategies

To further evaluate the influence of fusion strategies between temporal and spatial representations, we compared four alternatives:

- Summation: element-wise addition of temporal and spatial embeddings.
- Hadamard product: element-wise multiplication of embeddings to enhance multiplicative interactions, similar to the operation in IFNet [3].
- Linear projection: concatenation of embeddings followed by a linear transformation.
- Concatenation: direct concatenation of temporal and spatial embeddings without additional transformation.

As shown in Table S4, all strategies achieve competitive performance, with concatenation consistently yielding the best results across both datasets. Overall, the superiority of concatenation highlights that preserving both temporal and spatial information in their raw form before classification is most effective, validating our design choice in DBConformer.

TABLE S4
Results of DBConformer with different fusion strategies for features from T-Conformer and S-Conformer on the BNCI2014001 dataset.

Fusion strategy	Accuracy (%)	Weighted F1 (%)	Precision (%)	Recall (%)
Summation	<u>79.52</u>$\pm$0.82	<u>79.06</u>$\pm$0.72	<u>76.72</u>$\pm$1.14	85.87$\pm$3.34
Hadamard product	78.25$\pm$1.34	77.44$\pm$1.57	74.46$\pm$2.26	**87.94**$\pm$3.27
Linear projection	78.97$\pm$1.42	78.33$\pm$1.64	76.03$\pm$1.77	86.19$\pm$3.08
Concatenation	**80.16**$\pm$1.53	**79.94**$\pm$1.73	**76.98**$\pm$1.99	<u>86.51</u>$\pm$2.09

VI. Effect of Single-Branch Variants of DBConformer

To further assess the contribution of each branch in DBConformer, we conducted an additional experiment on BNCI2014001 under the CO setting, comparing three variants: T-Conformer only, S-Conformer only, and the full dual-branch DBConformer. Table S5 summarizes the results in terms of parameter counts, training/inference efficiency, and classification performance.

Results showed that the S-Conformer alone yields substantially lower accuracy (55.87%), while T-Conformer alone achieves moderate performance, and the full DBConformer provides the best accuracy. This gap highlights the importance of temporal information for MI decoding: EEG signals are inherently rhythmic and dominated by temporal dynamics such as event-related desynchronization/synchronization. Spatial-only modeling discards fine-grained oscillatory patterns, leading to significant information loss. However, spatial relations remain complementary, as incorporating the S-Conformer alongside the temporal branch yields clear gains over T-Conformer alone. These results validate the necessity of the dual-branch design, where temporal rhythms are modeled as the primary source of discriminative features, and spatial inter-channel interactions serve as crucial auxiliary cues.

TABLE S5
COMPARISON OF SINGLE-BRANCH T-CONFORMER, S-CONFORMER, AND THE DUAL-BRANCH DBCONFORMER ON BNCI2014001 UNDER THE CO SETTING.

Structure	# Params	Training Time (s)	Inference (s)	Accuracy (%)	Weighted F1 (%)	Precision (%)	Recall (%)
T-Conformer only	52,331	57.35	$5.3\,e^{-3}$	78.57	77.58	73.92	85.71
S-Conformer only	48,746	60.33	$5.5\,e^{-3}$	55.87	49.04	47.43	58.73
DBConformer	92,066	67.74	$8.5\,e^{-3}$	**80.16**	**79.94**	**76.98**	**86.51**

VII. MODEL CONFIGURATION

Due to the specific characteristics of datasets, scenarios, and tasks, we adopted distinct model configurations accordingly, as summarized in Table S6.

TABLE S6
MODEL CONFIGURATIONS OF DBCONFORMER FOR EACH DATASET.

Task	Dataset	Patch Size	Embed. Size	# T-Trans. Layers	# T-Trans. Heads	# S-Trans. Layers	# S-Trans. Heads
CO/CV	BNCI2014001	125	40	2	2	2	2
	BNCI2014004	125	40	2	2	2	2
	Zhou2016	125	40	2	2	2	2
	Blankertz2007	125	40	2	2	2	2
	BNCI2014002	128	40	4	2	4	2
	OpenBMI	32	40	4	2	4	2
LOSO	BNCI2014001	125	40	2	2	2	2
	BNCI2014004	125	40	2	2	2	2
	Zhou2016	125	40	6	2	6	2
	Blankertz2007	125	40	6	2	6	2
	BNCI2014002	128	40	6	2	6	2
	OpenBMI	32	40	6	2	6	2
	CHSZ	100	60	2	2	2	2
	NICU	128	60	2	2	2	2

VIII. EVALUATION METRICS

1) Evaluation Metrics: Accuracy is employed to evaluate the performance on MI and SSVEP. Accuracy is the overall proportion of correctly classified trials, regardless of class:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \tag{1}$$

where *TP* is the number of correctly classified positive trials, *TN* the number of correctly classified negative trials, *FP* the number of negative trials misclassified as positive, and *FN* the number of positive trials misclassified as negative.

To evaluate the classification performance on class-imbalanced seizure datasets, three additional metrics are included in seizure detection.

- Area under the receiver operating characteristic curve (AUC) is a threshold-independent metric that reflects the model's ability to distinguish between seizure and non-seizure trials across varying decision thresholds.
- F1 Score is the harmonic mean of precision and recall, indicating the balance between false positives and false negatives:

$$F1 = \frac{2TP}{2TP + FP + FN} \times 100\%. \tag{2}$$

- Balanced classification accuracy (BCA) is the average of sensitivity and specificity, which accounts for class imbalance:

$$BCA = \frac{sensitivity + specificity}{2} \times 100\%, \tag{3}$$

where $sensitivity = \frac{TP}{TP+FN}$ is the proportion of correctly detected seizure trials, and $specificity = \frac{TN}{TN+FP}$ is the proportion of correctly detected normal trials.

REFERENCES

[1] M. Long, Y. Cao, J. Wang, and M. I. Jordan, “Learning transferable features with deep adaptation networks,” in *Proc. Int’l Conf. on Machine Learning*, Lille, France, Jul. 2015, pp. 97–105.
[2] Y. Jin, X. Wang, M. Long, and J. Wang, “Minimum class confusion for versatile domain adaptation,” in *Proc. European Conf. on Computer Vision*, Glasgow, UK, Aug. 2020, pp. 464–480.
[3] J. Wang, L. Yao, and Y. Wang, “IFNet: An interactive frequency convolutional neural network for enhancing motor imagery decoding from EEG,” *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1900–1911, 2023.