NEUROWEAVE: FREQUENCY-FACTORIZED MULTI-SLOT DISCRETE TARGETS FOR MASKED EEG PRE-TRAINING

Anonymous authors
Paper under double-blind review

ABSTRACT

Discrete masked EEG pre-training depends on both target structure and target capacity. A single broadband code must compress heterogeneous rhythms into one discrete decision; increasing its vocabulary does not create independently representable signal components. We introduce NEUROWEAVE, which factorizes each one-second channel patch into delta, theta, and alpha bands and represents each band with multiple discrete slots. Independent tokenizers generate frozen targets, and a shared spatiotemporal Transformer predicts them from masked broadband input. We evaluate target design in stages, separating tokenizer capacity, masked reconstruction, target-band selection, and downstream transfer. Increasing slot count improves tokenizer fidelity more than enlarging a single-slot codebook, while masked reconstruction reveals finite, frequency-dependent useful capacity. Under matched controls, frequency-factorized targets improve low-frequency reconstruction over an equally sized broadband VQ target. Downstream ablations select delta, theta, and alpha, whereas beta and gamma provide no consistent benefit. Across six datasets, the resulting three-band model surpasses single-slot, broadband VQ, and waveform-regression controls on every reported metric and achieves state-of-the-art performance on the reported benchmarks. Code will be released.

1 INTRODUCTION

EEG representation learning must transfer across changes in montage, recording duration, acquisition protocol, population, and label space. The signal itself is heterogeneous: slow deflections, rhythmic activity, and cross-channel coordination coexist at different temporal and spectral scales. Self-supervised pre-training can exploit diverse unlabeled recordings, but transfer depends on whether its target retains the structures needed by later tasks.

Masked EEG models learn temporal context (Banville et al., 2021; Kostas et al., 2021), spectral structure (Wu et al., 2022), and cross-dataset representations (Yang et al., 2023). Discrete targets are attractive because the contextual model predicts compact codes rather than every waveform sample. For example, LABRAM trains a neural tokenizer and predicts its codes for masked channel patches (Jiang et al., 2024). Yet a single broadband code still forces an entire patch into one discrete decision. Enlarging the codebook increases the number of choices for that decision but not the number of signal components that can be represented independently.

This limitation is particularly consequential for EEG. Reconstruction-based foundation models can preferentially encode high-power aperiodic and low-frequency structure while under-representing oscillatory content (Kommineni et al., 2026; Yu et al., 2026). A one-second patch may contain a slow trend together with theta or alpha activity, and their relevance varies across clinical, motor-imagery, and SSVEP tasks. Recent work addresses spectral imbalance through frequency-balanced continuous objectives, transformed reconstruction losses, hybrid raw–band targets, or band-wise VQ tokenization (Yu et al., 2026; Foumani et al., 2025; Kulkarni et al., 2026; Sukhbaatar et al., 2026). However, these approaches leave a basic question unresolved: how much discrete capacity should be assigned within each frequency band, and does improving a band's reconstruction make it a more useful pre-training target?

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

NEUROWEAVE treats target design as two linked choices: the target-band set and the number of discrete slots allocated to each band. Independent tokenizers produce multi-slot codes for delta, theta, and alpha; a shared Transformer predicts the resulting code sequences from masked broadband patches using temporal and cross-channel context. The codes are generated offline and frozen; visible patches provide context for masked targets. Only the backbone is transferred downstream.

We test this design as a sequence of falsifiable questions: Does slot count relieve the tokenizer bottleneck? Do frequency-factorized targets improve masked reconstruction beyond an equally sized broadband target? Which target bands transfer? Do the selected targets generalize across datasets? This staged analysis separates tokenizer fidelity, contextual predictability, target-band selection, and downstream utility. Our contributions are:

- We introduce multi-slot frequency-factorized discrete targets, which allocate several independent codes to each selected EEG band rather than representing a broadband patch with one discrete bottleneck.
- We provide controlled evidence that slot count is the dominant tokenizer-capacity axis, frequency factorization improves masked reconstruction beyond matched broadband VQ targets, and useful capacity is finite and frequency dependent.
- We select delta–theta–alpha targets through downstream ablations and show that the resulting configuration outperforms single-slot, broadband VQ, and direct waveform-regression controls across six datasets.

2 RELATED WORK

EEG foundation models and discrete masked prediction. Task-specific architectures such as EEGNet combine temporal and spatial filtering but are commonly optimized for one task and montage (Lawhern et al., 2018). Self-supervised EEG methods instead learn from unlabeled recordings through relative positioning, contrastive or predictive objectives, and contextual masking (Banville et al., 2021; Kostas et al., 2021). BIOT extends this direction through frequency-domain representations across biosignal datasets (Yang et al., 2023). Masked language and image modeling provide the general contextual-prediction paradigm (Devlin et al., 2019; Bao et al., 2022; He et al., 2022), and vector quantization provides learnable discrete vocabularies (Van Den Oord et al., 2017). LABRAM brings these ideas to EEG by predicting tokenizer codes for masked channel patches (Jiang et al., 2024).

Spectral bias and frequency-aware targets. Recent analyses find that reconstruction-based EEG foundation models can overemphasize aperiodic and low-frequency components (Kommineni et al., 2026; Yu et al., 2026). FAME counters this tendency by reconstructing standardized time–frequency activity with equal band weights; EEG-X applies a dictionary-inspired convolutional transformation before the reconstruction loss; and MANAS-2 combines waveform and spectral-band targets with an energy-dynamics constraint (Yu et al., 2026; Foumani et al., 2025; Kulkarni et al., 2026). BandVQ uses separate VQ tokenizers for five canonical bands, together with band-power and metadata conditioning (Sukhbaatar et al., 2026). These methods motivate frequency-aware EEG targets, but do not isolate the per-band discrete capacity needed for contextual recovery or test whether each target band improves transfer. NEUROWEAVE addresses both questions through multi-slot tokenization and a staged evaluation.

3 NEUROWEAVE

3.1 OVERVIEW AND PROBLEM FORMULATION

Let an EEG segment be $\boldsymbol{X} \in \mathbb{R}^{C\times T}$ with C channels, T samples, and a sampling rate of 200 Hz. Each channel is divided into non-overlapping one-second windows, producing a grid of $N = CW$ patches,

$$\boldsymbol{X} \longrightarrow \{\boldsymbol{x}_{c,w} \in \mathbb{R}^{200} : c = 1, \ldots, C;\ w = 1, \ldots, W\}. \tag{1}$$

As shown in Figure 1, training has two stages. Independent tokenizers first convert every patch into band-specific discrete targets, which are then frozen and cached. A shared Transformer receives a

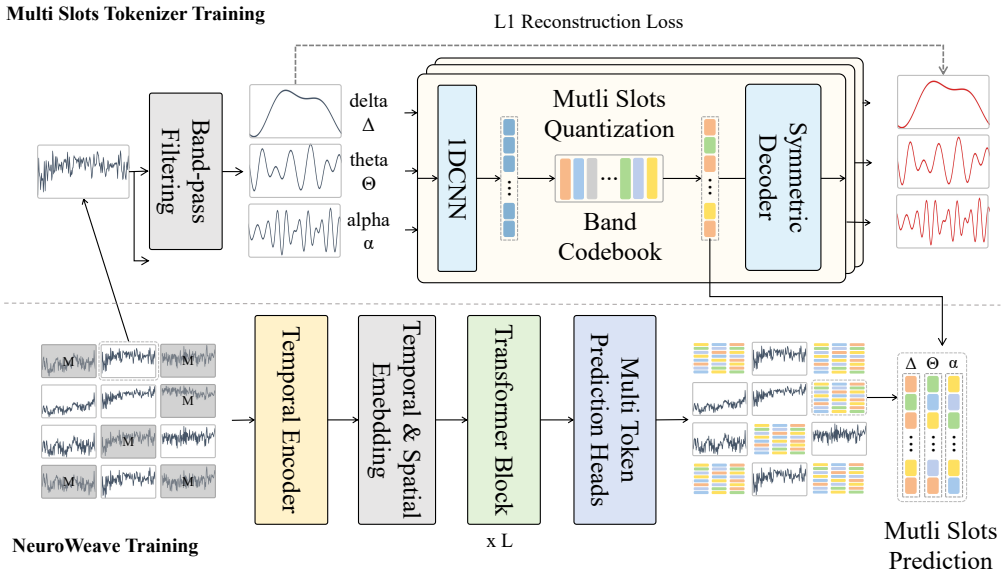


Figure 1: Overview of NEUROWEAVE. Independent tokenizers generate frozen multi-slot codes for delta, theta, and alpha. The online backbone masks raw channel–time patches and predicts their codes from spatiotemporal context. Tokenizers and prediction heads are discarded for downstream transfer.

masked grid of broadband patches and predicts the structured codes at each masked location. The design separates target factorization (the set of bands) from target capacity (the number of slots per band), allowing each decision to be evaluated explicitly.

3.2 CAPACITY-CONTROLLED BAND TOKENIZATION

The selected target bands are

$$\mathcal{B} = \{\delta, \theta, \alpha\}. \tag{2}$$

For each raw patch, a filter bank extracts delta ($\leq$ 3 Hz), theta (4–7 Hz), and alpha (8–13 Hz). A fixed training-set scale s_b bounds each band while retaining amplitude differences between patches,

$$\bar{\boldsymbol{x}}^{(b)} = \tanh\left(\boldsymbol{x}^{(b)}/s_b\right). \tag{3}$$

The scale remains fixed during tokenizer training, pre-training, and inference.

Each band has an independent convolutional encoder, vector quantizer, and decoder. The encoder produces S_b latent vectors, so the waveform is not compressed into one discrete position. We refer to S_b as the *slot count*; it controls how many independently quantized components represent a band waveform, whereas codebook size K controls the number of choices per component. Each latent is replaced by its nearest entry in a band-specific codebook $\boldsymbol{V}^{(b)} \in \mathbb{R}^{K\times D}$,

$$q_s^{(b)} = \arg\min_{k\in\{1,\dots,K\}} \left\| \boldsymbol{z}_s^{(b)} - \boldsymbol{V}_k^{(b)} \right\|_2^2. \tag{4}$$

Section 4.2 varies both S_b and K to isolate their effects on tokenizer capacity. Let $\boldsymbol{z}_q^{(b)}$ denote the selected embedding sequence. Each tokenizer is optimized independently using

$$\mathcal{L}_{\text{tok}}^{(b)} = \left\| \bar{\boldsymbol{x}}^{(b)} - g_b(\boldsymbol{z}_q^{(b)}) \right\|_1 + \mathcal{L}_{\text{code}}^{(b)} + \lambda_{\text{com}}\mathcal{L}_{\text{commit}}^{(b)}, \tag{5}$$

where the last two terms update the codebook and constrain encoder outputs to their selected entries, following VQ-VAE (Van Den Oord et al., 2017).

The target for patch (c, w) concatenates the three code sequences,

$$\boldsymbol{y}_{c,w} = [\boldsymbol{q}^{(\delta)}, \boldsymbol{q}^{(\theta)}, \boldsymbol{q}^{(\alpha)}] \in \{1, \dots, K\}^{S_{\text{tot}}}, \qquad S_{\text{tot}} = \sum_{b \in \mathcal{B}} S_b. \tag{6}$$

Identical integer indices in different bands are unrelated because their codebooks are independent. After tokenizer training, all targets are generated offline and remain fixed during masked pre-training.

This construction separates two bottlenecks that are otherwise easily conflated. Frequency factorization determines which signal components receive explicit supervision, whereas slot count determines how many independently quantized decisions describe each component. Enlarging K refines the choices available at one position but does not create additional positions; increasing S_b changes the structure of the representation itself. Separate codebooks also prevent a numerical index from implying shared semantics across bands. Importantly, the filters and tokenizers are confined to the target branch: the online encoder always receives broadband waveforms. The selected bands therefore guide pre-training without imposing a fixed band-pass constraint on the transferred backbone.

3.3 Masked prediction of structured discrete targets

A temporal convolution maps each visible broadband patch to an embedding. Masked locations use a learned vector, and channel and temporal embeddings identify every grid position,

$$\boldsymbol{e}_{c,w} = \begin{cases} f_{\text{conv}}(\boldsymbol{x}_{c,w}), & (c,w) \notin \mathcal{M}, \\ \boldsymbol{e}^{\text{mask}}, & (c,w) \in \mathcal{M}, \end{cases} + \boldsymbol{e}_c^{\text{ch}} + \boldsymbol{e}_w^{\text{time}}. \tag{7}$$

The CW embeddings are processed by a shared Transformer with global attention across channel and temporal axes. Architectural and optimization settings are reported in Appendix B.

At each masked output $\boldsymbol{h}_{c,w}$, band-specific heads produce S_b categorical distributions over the corresponding codebook. The backbone therefore predicts the structured discrete target sequence at each masked patch, with each component taking one of K entries, rather than regressing waveform samples. This structured objective allows the backbone to use one shared context representation while preserving separate supervision for each band and slot.

3.4 Masked prediction objective and band-balanced loss

For each example, a binary mask $\mathcal{M}$ is sampled with ratio $r = 0.5$. The model predicts the targets at masked locations from the visible patches.

For band b with S_b code slots, the masked loss is

$$\mathcal{L}_{\mathcal{M}}^{(b)} = \frac{1}{|\mathcal{M}| S_b} \sum_{i \in \mathcal{M}} \sum_{s=1}^{S_b} \text{CE}\Big(\hat{\boldsymbol{y}}_{i,s}^{(b)}, y_{i,s}^{(b)}\Big). \tag{8}$$

The total objective averages the slot-normalized losses across bands,

$$\mathcal{L} = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} \mathcal{L}_{\mathcal{M}}^{(b)}. \tag{9}$$

Normalizing by S_b prevents bands with more slots from receiving greater weight. Cached targets avoid running the tokenizers during each pre-training iteration. During downstream adaptation, the tokenizers and prediction heads are removed; the convolutional encoder, positional embeddings, and Transformer form the transferable backbone.

4 Experiments

4.1 Experimental setup

Downstream tasks. We evaluate transfer on six datasets spanning clinical classification, motor imagery, and steady-state visual evoked potential (SSVEP) recognition. TUAB and TUEV provide binary abnormality detection and six-class clinical event classification (Obeid & Picone, 2016); SHU-MI and PhysioNet-MI provide binary and four-class motor imagery (Schalk, 2009); and BETA-SSVEP and Benchmark-SSVEP provide 40-class visual-target recognition (Liu et al., 2020). All

recordings are represented at 200 Hz and divided into non-overlapping one-second patches. Appendix E gives construction and split details.

Comparison groups. We compare NEUROWEAVE with randomly initialized CNN–Transformer, ContraWR, SPaRCNet, and ST–Transformer, and with pretrained BIOT (Yang et al., 2023), LABRAM (Jiang et al., 2024), CBRAMOD (Wang et al., 2025), and BRAINOMNI (Xiao et al., 2025), using their public implementations and weights. Three internal controls isolate the target design: S1 uses one slot per band, RawVQ-S32 uses a 32-slot broadband VQ target, and RawRegression uses direct waveform regression. Sn denotes a NEUROWEAVE configuration with $S_b = n$ slots per band. For downstream adaptation, the tokenizers and code-prediction heads are replaced by a task-specific classification head.

Training and evaluation. Final models are pre-trained on the full public EEG collection in Appendix D; diagnostic studies use a common subset to reduce cost. Each downstream model is trained, selected, and evaluated on the fixed train–validation–test partitions. We use binary cross-entropy for TUAB and SHU-MI, cross-entropy otherwise, and report five-seed mean $\pm$ sample standard deviation.

Metrics. For binary classification, we report balanced accuracy, PR-AUC, and AUROC. For multiclass classification, we report balanced accuracy, Cohen's kappa, and weighted F1. Checkpoints are selected using validation AUROC for TUAB, Cohen's kappa for TUEV, and balanced accuracy for the remaining downstream datasets.

The experiments form an evidence chain. We first identify the tokenizer bottleneck, then test masked reconstruction under matched controls, select target bands on TUAB and TUEV, and finally evaluate transfer across all six datasets.

4.2 SLOT COUNT ALLEVIATES THE SINGLE-SLOT TOKENIZER BOTTLENECK

We compare two ways to increase discrete tokenizer capacity. Enlarging the codebook size K gives one slot more candidate embeddings but still compresses an entire band waveform into one index. Increasing the slot count S instead represents the waveform with multiple indices and directly relaxes this structural bottleneck.

Sweep design. The codebook sweep varies $K \in \{32, 64, 128, 256, 512, 1024, 2048, 4096\}$ at $S = 1$, whereas the slot-count sweep varies $S \in \{1, 2, 4, 8, 16, 32, 64, 128\}$ at $K = 1024$. Both sweeps cover delta, theta, alpha, beta, and gamma. To limit computational cost, all configurations are trained on the common diagnostic subset defined above.

Evaluation protocol and metric. We evaluate frozen tokenizers on the complete subject-disjoint TUAB and TUEV validation sets. Inputs and targets follow the registered preprocessing, including fixed input scaling, band-specific filtering, and fixed-scale tanh normalization. Errors are computed after restoring the waveform scale.

For target elements x_i and reconstructions $\hat{x}_i$, we report

$$\mathrm{NMAE} = \frac{\sum_{i=1}^{n} |\hat{x}_i - x_i|}{\sum_{i=1}^{n} |x_i|}. \tag{10}$$

The sums cover all elements in each dataset–band evaluation; lower is better. Figure 2 shows the complete sweeps.

Codebook expansion gives limited gains. With $S = 1$, increasing K from 32 to 4096 reduces NMAE for every band on both datasets, but the curves flatten as K grows. A larger vocabulary therefore improves reconstruction with diminishing returns while retaining the one-index bottleneck.

Additional slots overcome the single-slot bottleneck. With $K = 1024$, $S = 128$ achieves the lowest NMAE for all bands on both datasets, although intermediate changes are not always monotonic. It also outperforms $S = 1$ with the larger $K = 4096$ codebook for every dataset–band

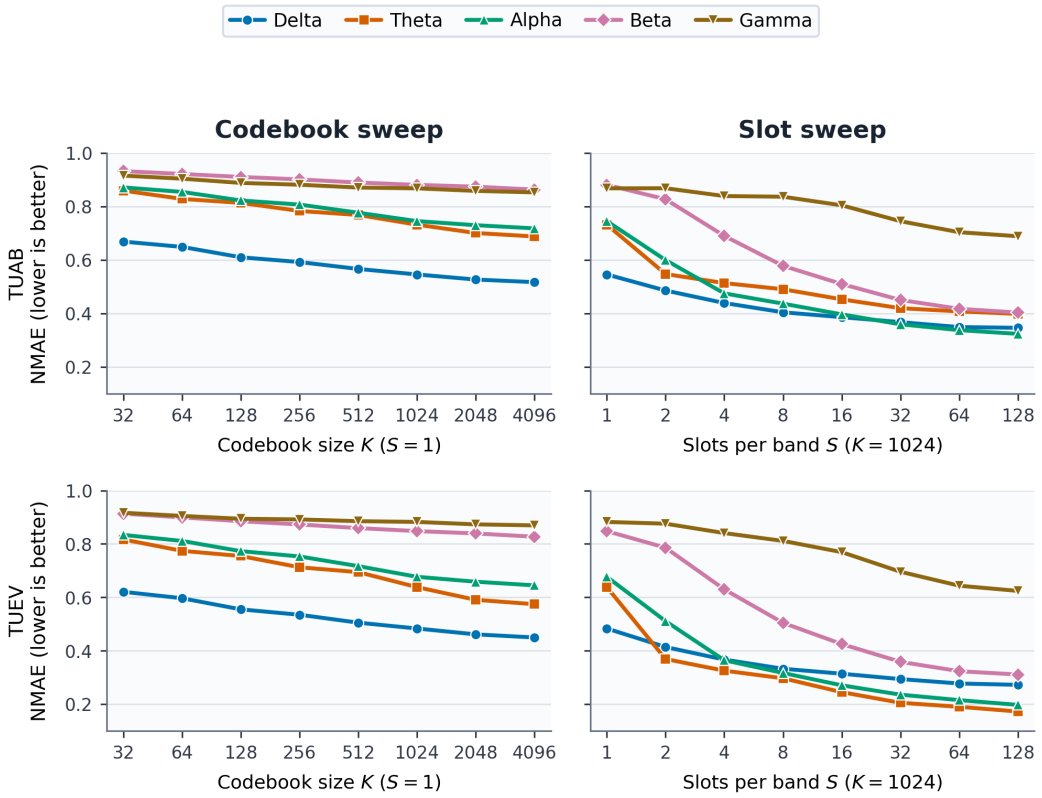


Figure 2: Frozen-tokenizer NMAE on the complete TUAB and TUEV validation sets. Left panels increase codebook size K while retaining a single slot; right panels increase the number of slots S at $K = 1024$. Enlarging the codebook yields progressively smaller gains, whereas additional slots produce substantially larger reductions across the five candidate bands.

pair, showing that multiple indices relieve the bottleneck more effectively than a larger single-slot vocabulary.

Implication. The sweep identifies slot count as the main tokenizer bottleneck but does not set the masked-pre-training operating point: it is not parameter- or compute-matched, uses one seed, and reconstructs visible patches. We therefore next test slot count when every additional target also increases prediction cost.

4.3 Frequency-factorized targets improve masked reconstruction

We next test whether the multi-slot advantage remains when the Transformer infers masked patches from spatiotemporal context, and whether frequency-factorized targets outperform an equally sized broadband target. All configurations use the common diagnostic subset and identical masks on the complete TUAB and TUEV validation sets. We report median patch-level NMAE.

Increasing capacity is not uniformly beneficial. At fixed $K = 1024$, moving beyond one slot generally improves masked reconstruction, but further gains are frequency dependent and may reverse. Thus, enlarging either capacity dimension does not guarantee improvement.

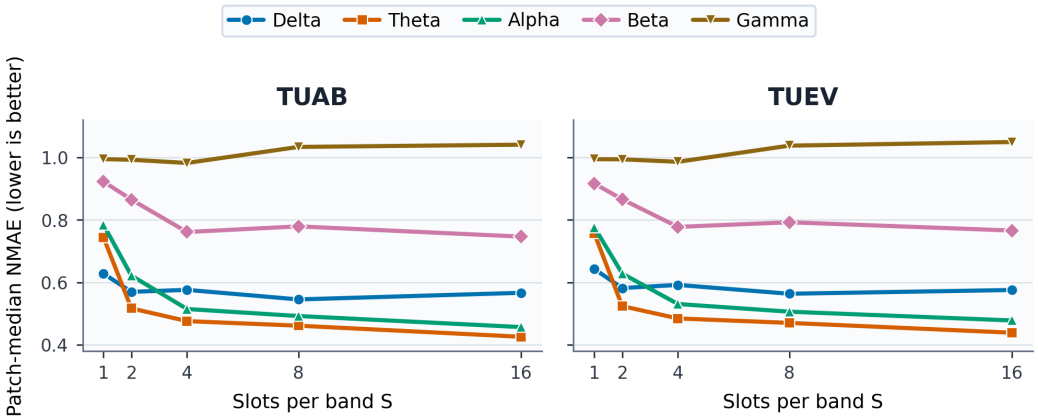


Figure 3: Masked-patch reconstruction across slot counts at $K = 1024$. Curves report median patch-level NMAE using identical masks. Additional slots improve most bands initially, but the optimal slot count is frequency dependent and gamma remains difficult to recover.

Frequency-factorized targets improve low-frequency reconstruction. Table 1 reports the matched target-design comparison. RawVQ-5 and RawVQ-80 reconstruct the broadband waveform with total slot budgets matched to the five-band S1 and S16 configurations. The frequency-factorized models achieve lower NMAE than their matched broadband VQ controls in every dataset–band comparison. Gains are largest for theta and alpha and remain clear for delta, whereas gamma shows little benefit. Separate band targets therefore improve low-frequency reconstruction beyond slot count alone. S16 also outperforms the RawRegression control from delta through beta; neither approach reconstructs gamma reliably.

Together, the results favor a finite multi-slot representation with frequency-factorized targets, particularly for delta, theta, and alpha. Reconstruction alone cannot determine whether every band benefits downstream transfer, which we examine next.

4.4 Target-band selection for downstream transfer

Masked reconstruction shows that useful capacity saturates at a moderate, frequency-dependent slot count. We therefore fix $S = 8$ as a practical trade-off between reconstruction quality and prediction cost, and vary the target bands while keeping per-band capacity fixed. Table 2 compares five nested target sets, from delta alone to all five candidate bands, using five downstream seeds on TUAB and TUEV.

Table 1: Median patch-level NMAE for matched target-design controls. RawVQ-5 and RawVQ-80 use broadband VQ targets with total slot budgets matched to S1 and S16, respectively; RawRegression uses direct waveform regression. Bold marks the lowest value in each dataset–band row.

Dataset	Band	S1	RawVQ-5	S16	RawVQ-80	RawRegression
TUAB	Delta	0.6280	0.6526	**0.5666**	0.6847	0.7668
	Theta	0.7435	0.8905	**0.4259**	0.5724	0.7451
	Alpha	0.7844	0.9770	**0.4569**	0.5995	0.7375
	Beta	0.9226	0.9982	**0.7464**	0.9423	0.9790
	Gamma	**0.9943**	1.0358	1.0406	1.0509	1.0060
TUEV	Delta	0.6424	0.6668	**0.5756**	0.6956	0.7659
	Theta	0.7561	0.8995	**0.4389**	0.5549	0.7342
	Alpha	0.7760	0.9764	**0.4781**	0.6058	0.7373
	Beta	0.9162	0.9986	**0.7658**	0.9487	0.9777
	Gamma	**0.9942**	1.0409	1.0491	1.0625	1.0092

Table 2: Downstream performance of NeuroWeave target-band configurations with $S = 8$. Checkmarks indicate the bands included during pre-training. Results are five-seed mean ± standard deviation; bold denotes the best mean within the table.

δ	θ	α	β	γ	TUAB			TUEV		
					Bal. Acc.	ROC-AUC	PR-AUC	Bal. Acc.	Kappa	W. F1
✓	✓	✓	✓	✓	82.48 ± 0.59	90.63 ± 0.38	90.95 ± 0.60	62.33 ± 2.32	0.6311 ± 0.0141	81.92 ± 0.63
✓	✓	✓	✓	–	82.36 ± 0.51	90.23 ± 0.27	90.54 ± 0.41	$\mathbf{69.70 \pm 2.42}$	0.6905 ± 0.0156	84.76 ± 0.80
✓	✓	✓	–	–	$\mathbf{82.85 \pm 0.06}$	$\mathbf{91.27 \pm 0.11}$	$\mathbf{91.63 \pm 0.11}$	69.68 ± 1.70	$\mathbf{0.7039 \pm 0.0114}$	$\mathbf{85.28 \pm 0.54}$
✓	✓	–	–	–	82.32 ± 0.18	90.77 ± 0.25	91.07 ± 0.16	68.05 ± 2.24	0.6916 ± 0.0218	85.02 ± 0.91
✓	–	–	–	–	81.29 ± 0.13	89.88 ± 0.03	90.47 ± 0.04	67.10 ± 1.62	0.6599 ± 0.0253	83.56 ± 1.05

Adding theta and alpha to delta improves downstream performance across both datasets. In contrast, adding beta provides no consistent benefit: it slightly changes TUEV balanced accuracy but reduces the remaining metrics. Adding gamma causes a clearer degradation, particularly on TUEV. These results suggest that beta and especially gamma can introduce negative transfer rather than useful supervision. We therefore select the delta–theta–alpha target set, which achieves the strongest overall performance across TUAB and TUEV. With both the slot count and target bands fixed, we next assess whether this choice transfers beyond the two selection datasets.

4.5 Transfer across six datasets

We evaluate the selected S8 configuration across the six downstream datasets. TUAB and TUEV serve as target-selection tasks, while SHU-MI, PhysioNet-MI, BETA-SSVEP, and Benchmark-SSVEP test transfer to different paradigms. Tables 3– 8 compare the selected model with randomly initialized task-specific architectures, pretrained EEG models, and the three internal target-design controls. All values are five-seed mean ± standard deviation; bold denotes the highest reported mean in each column.

The task collection stresses diverse aspects of an EEG representation. TUAB and TUEV emphasize clinical morphology and event structure; the two motor-imagery datasets require spatially organized sensorimotor information; and the SSVEP datasets require discrimination of periodic responses across 40 visual targets. Only TUAB and TUEV are used to select the target-band set, so the remaining four datasets test whether the resulting design transfers beyond the paradigms that determined it. The SSVEP tasks are particularly informative: their task-relevant responses need not be confined to the selected pre-training bands, but the downstream backbone remains free to exploit the full broadband input during adaptation.

The internal controls provide the primary test of the proposed target design: they use the same broadband inputs, backbone family, masking procedure, and downstream protocol, while varying discrete capacity or target structure. The external pretrained models instead assess competitiveness in the wider EEG foundation-model landscape. Because their pre-training corpora, architectures, and adaptation pipelines are not identical, we interpret these comparisons descriptively and base claims about the contribution of multi-slot, frequency-factorized supervision on the controlled ablations.

Table 3: Clinical downstream performance on TUAB and TUEV.

Model	TUAB			TUEV		
	BAcc	AUROC	PR-AUC	BAcc	Kappa	W. F1
CNN–Transformer	78.52 ± 0.45	87.26 ± 1.14	87.22 ± 1.30	39.95 ± 3.21	0.4107 ± 0.0232	71.46 ± 0.90
ContraWR	77.80 ± 0.92	85.78 ± 0.77	85.72 ± 1.04	46.34 ± 3.47	0.4830 ± 0.0208	75.12 ± 0.65
SPaRCNet	80.22 ± 0.47	87.78 ± 0.56	86.26 ± 1.30	47.15 ± 5.42	0.4995 ± 0.0703	75.04 ± 3.27
ST–Transformer	81.04 ± 0.61	88.75 ± 0.69	88.12 ± 0.40	43.07 ± 3.44	0.4332 ± 0.0197	71.20 ± 0.84
BIOT	78.79 ± 0.57	87.43 ± 0.35	87.21 ± 0.66	46.01 ± 1.43	0.4600 ± 0.0268	73.07 ± 1.27
LaBraM	81.43 ± 0.30	90.53 ± 0.13	90.27 ± 0.11	66.10 ± 0.98	0.6626 ± 0.0199	82.92 ± 0.85
CBraMod	78.07 ± 0.83	86.31 ± 1.03	86.85 ± 0.77	57.87 ± 2.18	0.5688 ± 0.0201	78.64 ± 0.89
BrainOmni	81.19 ± 0.43	89.08 ± 0.32	89.33 ± 0.53	59.70 ± 0.74	0.5966 ± 0.0176	79.90 ± 1.13
NeuroWeave (S8)	$\mathbf{82.85 \pm 0.06}$	$\mathbf{91.27 \pm 0.11}$	$\mathbf{91.63 \pm 0.11}$	$\mathbf{69.68 \pm 1.70}$	$\mathbf{0.7039 \pm 0.0114}$	$\mathbf{85.28 \pm 0.54}$
NeuroWeave (S1)	82.54 ± 0.48	90.92 ± 0.31	91.16 ± 0.53	56.16 ± 10.42	0.6018 ± 0.0162	80.09 ± 1.35
RawVQ-S32	81.66 ± 0.20	90.33 ± 0.24	90.38 ± 0.24	68.81 ± 1.49	0.6808 ± 0.0200	83.81 ± 0.98
RawRegression	80.87 ± 0.10	89.46 ± 0.28	89.24 ± 0.26	55.82 ± 1.43	0.4883 ± 0.0106	73.92 ± 0.56

Table 4: Motor-imagery performance on SHU-MI and PhysioNet-MI.

Model	SHU-MI			PhysioNet-MI		
	BAcc	AUROC	PR-AUC	BAcc	Kappa	W. F1
CNN–Transformer	61.65 ± 3.17	68.89 ± 4.12	69.43 ± 4.24	31.39 ± 3.85	0.0853 ± 0.0513	25.06 ± 7.90
ContraWR	64.18 ± 0.39	71.97 ± 0.71	73.38 ± 0.92	35.76 ± 7.51	0.1437 ± 0.1001	31.92 ± 10.76
SPaRCNet	65.19 ± 0.92	72.89 ± 1.68	73.25 ± 1.41	56.04 ± 1.07	0.4136 ± 0.0142	55.82 ± 1.21
ST–Transformer	63.22 ± 1.94	70.44 ± 0.97	72.29 ± 0.60	57.60 ± 1.04	0.4344 ± 0.0139	57.44 ± 1.27
BIOT	61.99 ± 1.21	66.13 ± 0.50	65.61 ± 0.80	31.46 ± 0.65	0.0862 ± 0.0086	31.12 ± 0.87
LaBraM	61.63 ± 0.63	65.68 ± 0.58	66.37 ± 0.61	51.46 ± 3.50	0.3527 ± 0.0467	49.31 ± 4.47
CBraMod	64.24 ± 2.88	68.75 ± 4.37	66.86 ± 5.00	59.32 ± 1.38	0.4576 ± 0.0184	59.17 ± 1.49
BrainOmni	65.98 ± 1.17	72.32 ± 0.51	73.20 ± 0.79	56.48 ± 0.82	0.4196 ± 0.0110	56.50 ± 0.78
NeuroWeave (S8)	$\mathbf{66.71 \pm 2.08}$	$\mathbf{73.59 \pm 2.43}$	$\mathbf{74.11 \pm 1.83}$	$\mathbf{59.80 \pm 0.70}$	$\mathbf{0.4640 \pm 0.0093}$	$\mathbf{59.95 \pm 0.81}$
NeuroWeave (S1)	66.32 ± 2.00	72.92 ± 2.90	73.72 ± 2.59	58.30 ± 2.38	0.4440 ± 0.0318	58.53 ± 2.27
RawVQ-S32	64.89 ± 1.19	70.97 ± 1.41	71.42 ± 1.94	56.75 ± 3.55	0.4233 ± 0.0475	56.98 ± 3.92
RawRegression	63.86 ± 1.95	67.60 ± 2.61	65.46 ± 2.66	58.42 ± 0.51	0.4456 ± 0.0068	58.53 ± 0.48

Across all six datasets, S8 outperforms S1, RawVQ-S32, and RawRegression on every reported metric, supporting the joint value of multi-slot quantization and frequency-factorized targets. Gains over S1 are modest but consistent on TUAB and substantially larger on TUEV; RawVQ-S32 approaches S8 on TUEV balanced accuracy but remains lower on kappa and weighted F1. Thus, matching a broadband target budget does not reproduce the benefit of factorization.

The same ordering extends to motor imagery and SSVEP. Although margins on SHU-MI and PhysioNet-MI are smaller, S8 improves every reported metric; on both SSVEP datasets it also exceeds all three internal controls. These results show that supervision in delta, theta, and alpha can shape a transferable broadband encoder without precluding adaptation to other task-relevant spectral structure.

The controls also distinguish the sources of this gain. S1 tests whether additional discrete positions matter, RawVQ-S32 tests whether simply matching the total target budget is sufficient, and RawRegression tests whether direct continuous reconstruction can replace discrete supervision. Their consistent ordering shows that neither target count nor waveform reconstruction alone explains the improvement; frequency-factorized multi-slot targets are jointly more expressive for the tokenizers and more useful for contextual pre-training.

5 Conclusion

NeuroWeave frames discrete target design as two separable choices: which frequency components receive supervision and how much discrete capacity each component receives. Multiple slots relieve the single-code bottleneck more effectively than a larger single-slot vocabulary, but useful capacity remains finite and frequency dependent. Frequency factorization improves reconstruction over matched broadband VQ, and downstream ablations select delta, theta, and alpha. The resulting three-band configuration outperforms all internal controls and achieves state-of-the-art results on the evaluated benchmarks.

AI USE STATEMENT

During the preparation of this manuscript, GPT-5 was used to assist with literature discovery and summarization, as well as the drafting and language refinement of selected passages. All AI-assisted content was critically reviewed and independently verified by the authors, who take full responsibility for the accuracy, originality, and integrity of the final manuscript.

REFERENCES

Hubert Banville, Omar Chehab, Aapo Hyv"arinen, Denis-Alexander Engemann, and Alexandre Gramfort. Uncovering the structure of clinical eeg signals with self-supervised learning. *Journal of Neural Engineering*, 18(4):046020, 2021.

Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. BEiT: BERT pre-training of image transformers. In *International Conference on Learning Representations*, 2022.

BCI Competition IV. BCI Competition IV: Data Set 1. Official dataset description, 2008. URL `https://www.bbci.de/competition/iv/desc_1.html`.

BCMI Laboratory, Shanghai Jiao Tong University. SEED-FRA and SEED-GER. SEED dataset portal. URL `https://bcmi.sjtu.edu.cn/home/seed/`.

Hsiang-Yun Sherry Chien, Hanlin Goh, Christopher M. Sandino, and Joseph Y. Cheng. MAEEG: Masked auto-encoder for EEG representation learning. *arXiv preprint arXiv:2211.02625*, 2022.

Paolo Detti, Giampaolo Vatti, and Guillermo Zabalo Manrique de Lara. Siena scalp EEG database. PhysioNet, version 1.0.0, 2020. URL `https://physionet.org/content/siena-scalp-eeg/1.0.0/`.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019.

Navid Mohammadi Foumani, Soheila Ghane, Nam Nguyen, Mahsa Salehi, Geoffrey I. Webb, and Geoffrey Mackellar. EEG-X: Device-agnostic and noise-robust foundation model for EEG. *arXiv preprint arXiv:2511.08861*, 2025.

Haaglanden Medisch Centrum. Haaglanden medisch centrum sleep staging database. PhysioNet, version 1.1. URL `https://physionet.org/content/hmc-sleep-staging/1.1/`.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16000–16009, 2022.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=QzTpTRVtrP`.

Aditya Kommineni, Emily Zhou, Kleanthis Avramidis, Simon Bock Segaard, Jeppe Roden M"unster, Andreas Peter Juhl Hansen, Takfarinas Medani, Tiantian Feng, Richard Leahy, and Shrikanth Narayanan. Aperiodic and low-frequency spectral bias in reconstruction based EEG foundation models. *arXiv preprint arXiv:2605.26434*, 2026.

Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data. *Frontiers in Human Neuroscience*, 15:653659, 2021.

Arvasu Kulkarni, Aditya Ray Mishra, Jeet Bandhu Lahiri, Mahir Jain, Parshva Runwal, Lakshya Saini, Siddharth Panwar, and Sandeep Singh. MANAS-2: Constrained reconstruction for EEG foundation models. *arXiv preprint arXiv:2609.13717*, 2026.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: A compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.

Bingchuan Liu, Xiaoshan Huang, Yijun Wang, Xiaogang Chen, and Xiaorong Gao. BETA: A large benchmark database toward SSVEP-BCI application. *Frontiers in Neuroscience*, 14:627, 2020. doi: 10.3389/fnins.2020.00627.

Matthew D. Luciw, Ewa Jarocka, and Benoni B. Edin. Multi-channel EEG recordings during 3,936 grasp and lift trials with varying weight and friction. *Scientific Data*, 1:140047, 2014. doi: 10.1038/sdata.2014.47.

Iyad Obeid and Joseph Picone. The temple university hospital eeg data corpus. *Frontiers in Neuroscience*, 10:196, 2016.

PhysioNet. Sleep-EDF database expanded. Version 1.0.0. URL `https://physionet.org/content/sleep-edfx/1.0.0/`.

Gerwin Schalk. EEG Motor Movement/Imagery Dataset. PhysioNet, 2009. Version 1.0.0.

Jamiyan Sukhbaatar, Satoshi Imamura, and Toshihisa Tanaka. BandVQ: Band-wise vector-quantized EEG foundation model. *arXiv preprint arXiv:2605.24921*, 2026.

Temple University. Temple university hospital EEG corpus: TUAR, TUEP, TUSL, and TUSZ. Official corpus descriptions. URL `https://isip.piconepress.com/projects/nedc/html/tuh_eeg/`.

Logan Trujillo. Raw EEG data, 2020.

Logan T. Trujillo, Candice T. Stanfield, and Ruben D. Vela. The effect of electroencephalogram (EEG) reference choice on information-theoretic measures of the complexity and integration of EEG signals. *Frontiers in Neuroscience*, 11:425, 2017. doi: 10.3389/fnins.2017.00425.

Aaron Van Den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A criss-cross brain foundation model for eeg decoding. In *International Conference on Learning Representations*, 2025.

Di Wu, Siyuan Li, Jie Yang, and Mohamad Sawan. neuro2vec: Masked fourier spectrum prediction for neurophysiological representation learning. *arXiv preprint arXiv:2204.12440*, 2022.

Qinfan Xiao, Ziyun Cui, Chi Zhang, SiQi Chen, Wen Wu, Andrew Thwaites, Alexandra Woolgar, Bowen Zhou, and Chao Zhang. BrainOmni: A brain foundation model for unified EEG and MEG signals. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. BIOT: Biosignal transformer for cross-data learning in the wild. In *Advances in Neural Information Processing Systems*, 2023.

Junjie Yu, Zihan Deng, Jianyu Zhang, Junrong Mu, Jiahui An, Wenxiao Ma, Ziling Lu, Yue Wang, Yan Zhu, Kexin Lou, and Quanying Liu. Understanding and correcting low-frequency bias in EEG foundation model. *arXiv preprint arXiv:2608.01898*, 2026.

Tongtian Yue, Shuning Xue, Xuange Gao, Yepeng Tang, Longteng Guo, Jie Jiang, and Jing Liu. EEGPT: Unleashing the potential of EEG generalist foundation model by autoregressive pre-training. *arXiv preprint arXiv:2410.19779*, 2024.

Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant: Foundation model for intracranial neural signal. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

A More Related Works

Self-supervised EEG representation learning predates recent foundation-model efforts. Banville et al. (Banville et al., 2021) learn latent structure from clinical EEG without labels, while BENDR combines a convolutional front end, Transformer context modeling, and a contrastive objective (Kostas et al., 2021). Neuro2Vec shifts the prediction target to masked Fourier spectra (Wu et al., 2022), and MAEEG studies masked reconstruction with a Transformer encoder–decoder (Chien et al., 2022). These works establish that temporal context and spectral structure can both support transfer, but they generally use continuous reconstruction targets.

Recent foundation models focus on scaling across heterogeneous recordings. BIOT converts channels into a common token sequence to accommodate mismatched montages, sequence lengths, and missing values (Yang et al., 2023). LaBraM introduces discrete neural-spectrum codes for masked channel-patch prediction (Jiang et al., 2024); CBraMod instead separates spatial and temporal interactions through criss-cross attention (Wang et al., 2025). EEGPT uses electrode-wise autoregressive pre-training to support diverse electrode configurations (Yue et al., 2024), while BrainOmni extends tokenized pre-training across EEG and MEG by encoding sensor properties alongside neural activity (Xiao et al., 2025). Beyond scalp EEG, Brant combines temporal, spatial, and frequency-domain modeling for large-scale intracranial recordings (Zhang et al., 2023). These studies primarily address data heterogeneity, backbone structure, or predictive direction; they do not isolate the number of independent discrete decisions needed within an EEG frequency component.

Target construction provides a complementary design axis. VQ-VAE introduces learned discrete codebooks (Van Den Oord et al., 2017), and BEiT demonstrates masked prediction of tokenizer-produced visual codes (Bao et al., 2022). For EEG, frequency-aware objectives can rebalance spectral supervision through standardized time–frequency targets, transformed reconstruction spaces, or joint waveform–band objectives (Yu et al., 2026; Foumani et al., 2025; Kulkarni et al., 2026). BandVQ most closely shares the use of band-specific VQ tokenizers (Sukhbaatar et al., 2026). NeuroWeave differs by treating target factorization and per-band slot count as separate, experimentally testable choices: the former determines which components are supervised, whereas the latter controls how many discrete positions represent each selected component.

B Hyperparameter Settings

Tables 5–7 list the training settings for the band-specific tokenizers, NeuroWeave masked pre-training, and downstream fine-tuning. Unless stated otherwise, all values correspond to the selected delta–theta–alpha S8 configuration. The diagnostic studies vary codebook size K, slot count S, or the target-band set, as described in Section 4.2 and Sections 4.3–4.4. Entries marked "–" are implementation settings to be specified in the final release configuration.

Table 5: Band-specific tokenizer training hyperparameters.

Group	Hyperparameter	Value
Input	Sampling rate	200 Hz
	Patch length	200 samples (1 second)
	Delta filter	Low-pass at 3 Hz
	Theta filter	4–7 Hz band-pass
	Alpha filter	8–13 Hz band-pass
	Band normalization	Fixed-scale $\tanh(\boldsymbol{x}^{(b)}/s_b)$
Tokenizer	Encoder / decoder	Band-specific 1D CNNs
	Latent slots per band, S_b	8 (selected configuration)
	Codebook size, K	1024
	Embedding dimension, D	16
	Reconstruction loss	ℓ_1
Optimization	Batch size	1024
	Optimizer	AdamW, $(\beta_1, \beta_2) = (0.9, 0.99)$
	Peak / minimum learning rate	5×10^{-5} / 1×10^{-5}
	Learning-rate scheduler	Cosine
	Weight decay	10^{-4}
	Training epochs / warmup epochs	100 / 10

Table 6: NEUROWEAVE masked pre-training hyperparameters.

Group	Hyperparameter	Value
Input	Sampling rate / patch length	200 Hz / 200 samples
	Maximum channels / temporal windows	128 / 16
Targets	Target bands	Delta, theta, alpha
	Slots per band / total targets	8 / 24
	Codebook size	1024
Temporal encoder	Architecture	Three-layer 1D CNN
	Input length	200 samples (1 second)
	Input channels	$\{1, 8, 8\}$
	Output channels	$\{8, 8, 8\}$
	Kernel sizes	$\{15, 3, 3\}$
	Strides	$\{8, 1, 1\}$
	Paddings	$\{7, 1, 1\}$
	Output dimension	200
Transformer	Blocks	12
	Attention heads	10
	MLP ratio	4
	Positional embeddings	Channel and temporal
Objective	Mask ratio, r	0.5
	Masking scheme	Random patch masking
	Prediction loss	Cross-entropy over VQ codes
	Loss aggregation	Slot-normalized and band-balanced
Optimization	Batch size	512
	Optimizer	AdamW
	Peak / minimum learning rate	5×10^{-4} / 1×10^{-5}
	Learning-rate scheduler	Cosine
	Weight decay	10^{-4}
	Training epochs / warmup epochs	50 / 5

Table 7: Downstream fine-tuning hyperparameters.

Group	Hyperparameter	Value
Optimization	Batch size	512
	Peak learning rate	5×10^{-4}
	Minimum learning rate	1×10^{-6}
	Learning-rate scheduler	Cosine
	Optimizer	AdamW
	Adam (β_1, β_2)	$(0.9, 0.999)$
	Weight decay	0.05
	Total epochs	50
	Warmup epochs	5
	Drop path	0.1
	Layer-wise learning-rate decay	0.65
Objective	Label smoothing (multi-class classification)	0.1

C Additional Downstream Results

Table 8 reports the detailed SSVEP results omitted from the main paper for space. The same models, preprocessing, and five-seed evaluation protocol are used as for the clinical and motor-imagery results.

Table 8: SSVEP performance on BETA-SSVEP and Benchmark-SSVEP.

Model	BETA-SSVEP			Benchmark-SSVEP		
	BAcc	Kappa	W. F1	BAcc	Kappa	W. F1
CNN–Transformer	7.00 ± 9.82	0.0462 ± 0.1007	4.76 ± 10.16	23.50 ± 19.90	0.2154 ± 0.2041	21.18 ± 20.30
ContraWR	22.77 ± 18.66	0.2079 ± 0.1914	21.62 ± 19.51	57.02 ± 2.23	0.5592 ± 0.0228	56.24 ± 2.83
SPaRCNet	4.93 ± 0.73	0.0249 ± 0.0075	3.01 ± 0.66	25.46 ± 13.95	0.2355 ± 0.1431	23.52 ± 14.37
ST–Transformer	48.32 ± 1.20	0.4700 ± 0.0123	48.45 ± 1.18	75.90 ± 1.88	0.7528 ± 0.0193	76.05 ± 1.94
BIOT	36.23 ± 1.12	0.3460 ± 0.0115	36.05 ± 0.76	61.88 ± 1.65	0.6090 ± 0.0169	61.44 ± 1.78
LaBraM	37.48 ± 1.01	0.3588 ± 0.0104	36.82 ± 0.96	75.88 ± 1.51	0.7526 ± 0.0155	75.93 ± 1.51
CBraMod	39.14 ± 5.65	0.3758 ± 0.0579	39.30 ± 6.18	73.52 ± 6.42	0.7284 ± 0.0659	74.66 ± 6.02
BrainOmni	15.00 ± 7.66	0.1282 ± 0.0786	12.20 ± 7.63	73.81 ± 1.29	0.7314 ± 0.0132	74.09 ± 1.27
NeuroWeave (S8)	$\mathbf{49.09 \pm 1.37}$	$\mathbf{0.4778 \pm 0.0141}$	$\mathbf{49.33 \pm 1.48}$	$\mathbf{78.96 \pm 1.85}$	$\mathbf{0.7842 \pm 0.0189}$	$\mathbf{79.14 \pm 1.87}$
NeuroWeave (S1)	43.50 ± 4.15	0.4205 ± 0.0425	43.74 ± 4.12	77.94 ± 2.69	0.7737 ± 0.0276	78.09 ± 2.66
RawVQ-S32	43.62 ± 3.79	0.4218 ± 0.0389	43.00 ± 4.15	76.58 ± 1.89	0.7598 ± 0.0194	76.65 ± 1.93
RawRegression	40.70 ± 2.72	0.3918 ± 0.0279	40.99 ± 2.67	65.90 ± 18.77	0.6502 ± 0.1926	64.79 ± 21.01

D Pre-Training Datasets

The full pre-training collection combines EEG from motor, cognitive, resting-state, affective, event-related, clinical, and sleep recordings. Acquisition details below describe the source datasets; all retained EEG is resampled to 200 Hz before entering the common pipeline.

The final downstream models use the complete collection. To make the tokenizer capacity and masked-reconstruction studies tractable, Sections 4.2 and 4.3 use the same reduced subset: Raw EEG Data, Resting State EEG Data, TUSL, TUSZ, Grasp and Lift, Inria BCI, and Brain Invaders. This shared diagnostic corpus preserves motor, cognitive, resting-state, event-related, and clinical recordings while holding the data source fixed across target designs.

BCI Competition IV-1. This dataset contains continuous motor-imagery EEG acquired from 59 channels at 1,000 Hz using BrainAmp MR plus amplifiers. Participants perform two imagery tasks selected from left-hand, right-hand, and foot movement, with additional idle periods. We include its EEG recordings in the motor-imagery portion of the corpus (BCI Competition IV, 2008).

Grasp and Lift EEG Challenge. This dataset contains recordings from 12 participants performing grasp-and-lift trials. EEG was acquired from 32 channels at 500 Hz with a BrainAmp ampli-

fier, providing activity associated with object manipulation in addition to the imagined-movement recordings (Luciw et al., 2014).

Inria BCI Challenge. The dataset contains P300 spelling recordings from 26 participants. EEG was collected from 56 channels at 600 Hz using Ag/AgCl sensors in a VSM–CTF-compatible recording system. We use the EEG signals without the spelling-task labels (Jiang et al., 2024).

Raw EEG Data. This collection contains 64-channel EEG sampled at 256 Hz during information-integration and multidimensional rule-based categorization tasks, and contributes cognitive-task recordings to the corpus (Trujillo, 2020).

Resting State EEG Data. This dataset contains 22 participants, each with four minutes of eyes-closed and four minutes of eyes-open resting activity. EEG was recorded from 64 channels at 256 Hz using active Ag/AgCl electrodes (Trujillo et al., 2017).

SEED-FRA and SEED-GER. These emotion datasets contain recordings from eight French and eight German participants, respectively. Positive, neutral, and negative states are elicited using video stimuli. The source EEG comprises 62 channels acquired at 1,000 Hz using an ESI NeuroScan system. We use EEG without emotion labels or eye-movement measurements (BCMI Laboratory, Shanghai Jiao Tong University).

Siena Scalp EEG Database. This database contains scalp EEG from 14 patients with epilepsy. The recordings used in the corpus contain 31 channels sampled at 512 Hz and were acquired with EB Neuro and Natus Quantum LTM amplifiers, providing clinical EEG collected outside the TUH environment (Detti et al., 2020).

Target Versus Non-Target. This dataset contains 50 participants performing the Brain Invaders visual P300 task under an oddball paradigm. EEG was acquired from 32 channels at 512 Hz using a g.USBamp amplifier and g.GAMMAcap. Target and non-target event labels are not used as pre-training targets (Jiang et al., 2024).

TUH EEG Artifact Corpus (TUAR). TUAR contains annotations for five artifact categories, including ocular, muscle, and electrode-related artifacts. The version represented in the corpus uses 23-channel EEG sampled at 256 Hz and contributes variability typical of clinical acquisition (Obeid & Picone, 2016; Temple University).

TUH EEG Epilepsy Corpus (TUEP). TUEP contains recordings from 100 individuals with epilepsy and 100 without epilepsy, with diagnoses determined by a neurologist. The represented data contain 19–23 channels sampled at 256 Hz; diagnostic labels are not used during pre-training (Obeid & Picone, 2016; Temple University).

TUH EEG Seizure Corpus (TUSZ). TUSZ provides clinical EEG with annotations of seizure onset, offset, involved channels, and seizure type. The represented recordings contain 19–23 channels sampled at 256 Hz and are used for self-supervised learning rather than supervised seizure detection (Obeid & Picone, 2016; Temple University).

TUH EEG Slowing Corpus (TUSL). TUSL provides slowing-event annotations and contains 23-channel recordings at 256 Hz. It adds clinical slowing patterns to the range of EEG activity in the corpus; its annotations are not used for pre-training supervision (Obeid & Picone, 2016; Temple University).

Haaglanden Medisch Centrum Sleep Staging Database. HMC contains overnight polysomnographic recordings from patients referred for sleep examinations. Although the public database includes EEG, other physiological measurements, and sleep-stage annotations, we use only EEG. The prepared pool contains 129 participants (Haaglanden Medisch Centrum).

Sleep-EDF Expanded. Sleep-EDF comprises the Sleep Cassette and Sleep Telemetry collections, with 153 and 44 polysomnographic recordings, respectively. The source EEG was sampled at 100 Hz and includes manually scored sleep stages. We construct a 1,000-hour EEG pool by taking equal-duration, temporally centered segments from the 197 recordings and resampling them to 200 Hz. Equal-duration selection balances contributions across recordings, rather than across sleep-stage classes (PhysioNet).

Source-specific temporal windows are used before conversion to the common one-second patch representation. The acquisition rates above characterize the original recordings; resampling does not increase their effective bandwidth.

E Downstream Dataset Splits

We use one fixed train–validation–test partition for each downstream dataset and retain the same partition across models and training seeds. Except for SHU-MI, the partitions are subject-disjoint. The sample counts below refer to the task-level examples presented to the downstream models, before their internal division into one-second patches.

TUAB (abnormality detection). TUAB is a binary normal-versus-abnormal EEG classification task derived from the Temple University Hospital EEG resources (Obeid & Picone, 2016). Each example contains 10 seconds of EEG. We use a subject-level manifest that assigns 248,180 examples to training, 76,056 to validation, and 36,945 to testing; subjects do not overlap between the three partitions.

TUEV (clinical event classification). TUEV contains six classes of annotated clinical EEG events from the Temple University Hospital EEG resources (Obeid & Picone, 2016). Each example spans 5 seconds. A subject-level manifest produces 68,656 training, 14,414 validation, and 28,305 test examples, with disjoint subjects across the three partitions.

SHU-MI (motor imagery). SHU-MI is a binary motor-imagery task using 32-channel, 4-second trials. Because the registered protocol is trial-based rather than cross-subject, trials are stratified by class into an 80/10/10 train–validation–test split. This yields 9,590 training, 1,199 validation, and 1,199 test examples. The same stratified partition is used for every model and seed.

BETA-SSVEP (visual target recognition). The BETA database contains 70 participants performing a 40-class SSVEP spelling task with 64-channel EEG (Liu et al., 2020). We construct 2-second examples and partition participants in acquisition order: S01–S56 for training, S57–S63 for validation, and S64–S70 for testing. The resulting partitions contain 8,960, 1,120, and 1,120 examples, respectively.

PhysioNet-MI (motor imagery). PhysioNet-MI is constructed from the EEG Motor Movement/Imagery Dataset and uses four motor-imagery classes, 64 channels, and 4-second examples (Schalk, 2009). Its 109 participants are assigned in acquisition order to training (S001–S087), validation (S088–S097), and testing (S098–S109), corresponding to 87, 10, and 12 participants. After example construction, the three partitions contain 7,830, 948, and 1,059 examples.

Benchmark-SSVEP (visual target recognition). Benchmark-SSVEP contains 35 participants, 40 stimulus classes, and 64-channel EEG. For each trial, we retain the 5-second interval from 0.5 to 5.5 seconds after stimulus onset, yielding 240 examples per participant. Participants S01–S28 form the training set, S29–S31 the validation set, and S32–S35 the test set. These partitions contain 6,720, 720, and 960 examples, respectively.