

S1 Datasets and Preprocessing

S1.1 Filtering and epoch construction

Continuous signals were filtered at their source sampling rate and, where required, resampled to 100 Hz before being partitioned into non-overlapping 30-s epochs. Filtering used zero-phase fourth-order Butterworth IIR filters. Dataset-specific filtering and sampling-rate operations are summarized in Table S1. When the requested upper passband edge exceeded the available Nyquist frequency, it was reduced to 0.5 Hz below that frequency.

Table S1: Signal filtering and source sampling-rates.

Dataset	Source sampling rate	Passbands
ANPHY	1000 Hz	EEG 0.3–100 Hz; EOG 0.1–100 Hz; EMG 10–100 Hz
Sleep-EDF20 / EDFx	100 Hz	EEG 0.3–49.5 Hz; EOG 0.1–49.5 Hz
ISRUC-SG1 / SG3	200 Hz	EEG 0.3–99.5 Hz; EOG 0.1–99.5 Hz; EMG 10–99.5 Hz

S1.2 Subject allocation and rotating validation folds

All recordings were grouped by subject before allocation. Subjects were first stably ordered by decreasing retained epoch count and decreasing maximum class count. Let $\mathbf{u}_s \in \mathbb{R}^5$ denote the stage histogram of subject s , let $\mathbf{v}_f \in \mathbb{R}^5$ denote the accumulated histogram of fold f , and let $\mathbf{g} = \sum_s \mathbf{u}_s$ denote the dataset-wide histogram. Among folds whose subject capacity had not been filled, subject s was assigned to the fold minimizing

$$J(s, f) = \frac{1}{5} \sum_{c=1}^5 \left(\frac{v_{f,c} + u_{s,c} - g_c/5}{g_c/5 + 1} \right)^2. \quad (\text{S1})$$

Ties were resolved by the current number of subjects and then by fold index. Fold capacities differed by at most one subject. For split i , fold i served as the test set, fold $(i + 1) \bmod 5$ as the validation set, and the remaining three folds as the training set. Table S2 gives the resulting subject counts.

Table S2: Train/validation/test subject counts for each evaluation split.

Dataset	Split 0	Split 1	Split 2	Split 3	Split 4
ANPHY	16/6/6	16/6/6	17/5/6	18/5/5	17/6/5
Sleep-EDF20	12/4/4	12/4/4	12/4/4	12/4/4	12/4/4
Sleep-EDFx	46/16/16	46/16/16	47/15/16	48/15/15	47/16/15
ISRUC-SG1	39/14/14	40/13/14	41/13/13	41/13/13	40/14/13
ISRUC-SG3	6/2/2	6/2/2	6/2/2	6/2/2	6/2/2

S2 Shared Waveform Encoder

The four convolutions use padding 12, 7, 4, and 3, respectively, and GroupNorm with 4, 4, 8, and 8 groups. Each convolution is followed by GELU and Dropout1d with $p = 0.05$. Adaptive average pooling produces a 64-dimensional vector, which is mapped to $F_m \in \{9, 256\}$ features by a linear projection followed by LayerNorm and dropout $p = 0.1$.

The convolutional trunk contains 55,712 trainable parameters. Including the projection and LayerNorm, the waveform encoder contains 56,315 parameters for $F_m = 9$ and 72,864 parameters for $F_m = 256$. For CNN-2S and CNN-PTFT, the auxiliary epoch-level head concatenates the channel vectors and applies dropout $p = 0.5$, a 64-unit linear layer, and a five-class output layer. This head is discarded after pretraining.

S3 Model-Specific Native Features

GraphSleepNet and ProductGraphSleepNet. The *Native* frontend represents each channel and epoch using differential entropy (DE) computed over nine overlapping frequency bands: 0.5–4, 2–6, 4–8, 6–11, 8–14, 11–22, 14–31, 22–40, and 31–50 Hz. The resulting channel–band features are standardized using statistics estimated from the training recordings and subsequently applied to the validation and test recordings.

STDP-GCN. Following recording-level scalar min–max scaling, Ben’s Spiking Algorithm (BSA) is applied independently to each waveform using a length-20 FIR filter, a normalized cutoff of 0.8, and threshold $\vartheta = 1$. A separate zero-initialized directed STDP weight matrix is constructed for every epoch and weights do not accumulate across epochs. Nonzero final weights define the binary adjacency, and identity self-loops are added.

The *Native* node features are DE values computed over the same nine frequency bands from the BSA residual, using the released fixed FFT length of 7680. These residual features are supplied directly to the graph backbone without subsequent standardization.

S4 Graph-Construction Notes

The evaluated graph-construction rules are shared across the four frontend conditions. Only the node representation supplied to each backbone differs. Thus, feature-dependent adjacencies may change with the supplied representation, but the operations used to construct them remain fixed within each backbone. Table S4 summarizes the normalization and symmetrization choices most relevant to the comparison.

For STDP-GCN, the all-ones adjacency in the released execution path was replaced by the intended BSA/STDP adjacency, which was held fixed across frontend conditions.

Graph-filter recurrence. GraphSleepNet, ProductGraphSleepNet, MSTGCN, and JK-STGCN use the element-wise recurrence in their evaluated formulations. For the model-specific scaled Laplacian $\tilde{\mathbf{L}}$,

$$\mathbf{T}_0 = \mathbf{I}, \quad \mathbf{T}_1 = \tilde{\mathbf{L}}, \quad \mathbf{T}_k = 2\tilde{\mathbf{L}} \odot \mathbf{T}_{k-1} - \mathbf{T}_{k-2}, \quad (\text{S2})$$

where $\odot$ denotes element-wise multiplication. ProductGraphSleepNet, JK-STGCN, and the dynamic MSTGCN branch set $\lambda_{\max} = 2$, giving $\tilde{\mathbf{L}} = \mathbf{L} - \mathbf{I}$. GraphSleepNet estimates $\lambda_{\max}$ separately for each

Table S3: Parameter counts of the node-representation frontends. Counts exclude the auxiliary epoch-classification heads, which are discarded after pretraining. Deterministic feature-extraction procedures contain no learned parameters.

Frontend	F_m	Learned parameters
GraphSleepNet / ProductGraphSleepNet DE	9	0
MSTGCN / JK-STGCN FeatureNet	256	184,288
STDP-GCN residual spectral features	9	0
Shared CNN, 9-dimensional projection	9	56,315
Shared CNN, 256-dimensional projection	256	72,864

Table S4: Graph-construction operations used by the evaluated backbones.

Backbone	Feature-to-adjacency operation	Normalization or symmetrization
Graph SleepNet	Absolute feature differences from the center epoch	Row softmax without ReLU or explicit symmetrization
ProductGraph SleepNet	Mean absolute spatial differences over context positions and temporal differences over channels, followed by ReLU	Column softmax without explicit symmetrization
MSTGCN	Independent exponential feature affinity at each context position, scaled by the feature dimension	Column normalization followed by element-wise $\min(\mathbf{A}, \mathbf{A}^\top)$ before dynamic graph convolution
JK-STGCN	Center-epoch absolute feature differences for the functional graph and a learned context projection for the temporal-information graph	Row softmax for the functional graph; the temporal-information graph is unnormalized
STDP-GCN	Node features do not construct the adjacency	Epoch-specific directed BSA/STDP adjacency with self-loops, fixed across frontend conditions

learned graph using the self-adjoint eigenvalue operation applied to the lower triangle of its learned Laplacian. The static MSTGCN branch uses the largest real eigenvalue of its coordinate-derived Laplacian.

MSTGCN coordinate graph. The static MSTGCN graph is constructed from the available channel names. Recognized single electrodes use the canonical two-dimensional coordinates in Table S5. Bipolar derivations are placed at the mean of their recognized endpoints. EOG channels without recognized endpoints are placed at $(-0.55, 1.12)$ or $(0.55, 1.12)$ according to channel-index parity. For channel index i , EMG channels are placed at

$$(-0.25 + 0.25(i \bmod 3), -1.20).$$

Any otherwise unrecognized channel is placed at angle $2\pi i/N$ on the unit circle.

For channel coordinates $\mathbf{c}_i$, the unnormalized spatial affinity and row-normalized adjacency are

$$\hat{A}_{ij} = \exp\left[-\frac{\|\mathbf{c}_i - \mathbf{c}_j\|_2^2}{2(0.35)^2}\right], \quad A_{ij} = \frac{\hat{A}_{ij}}{\sum_{q=1}^N \hat{A}_{iq}}. \quad (\text{S3})$$

The static branch uses the row-degree Laplacian

$$\mathbf{L} = \text{Diag}(\mathbf{A}\mathbf{1}) - \mathbf{A},$$

scaled using its largest real eigenvalue, followed by the order-3 recurrence in Equation S2.

Table S5: Canonical coordinates used to construct the MSTGCN static graph.

Node	(x, y)		Node	(x, y)		Node	(x, y)	
Fp1	-0.35	0.95	Fp2	0.35	0.95	Fpz	0.00	1.00
F7	-0.75	0.65	F3	-0.38	0.58	Fz	0.00	0.62
F4	0.38	0.58	F8	0.75	0.65	T3/T7	-1.00	0.00
C3	-0.48	0.00	Cz	0.00	0.00	C4	0.48	0.00
T4/T8	1.00	0.00	T5/P7	-0.75	-0.65	P3	-0.38	-0.58
Pz	0.00	-0.62	P4	0.38	-0.58	T6/P8	0.75	-0.65
O1	-0.35	-0.95	Oz	0.00	-1.00	O2	0.35	-0.95
A1/M1	-1.10	-0.12	A2/M2	1.10	-0.12			

S5 Training Objectives and Optimization Configurations

Model-specific objectives. All stage classifiers use unweighted cross-entropy. ProductGraphSleepNet additionally uses spatial and temporal graph-smoothness terms with unit coefficients and unnormalized full-sum reductions. The Frobenius-penalty coefficient is set to zero.

MSTGCN adds graph-smoothness and adjacency-Frobenius penalties, each weighted by 10^{-5} . The functional adjacency is averaged across batch, while the pairwise absolute feature differences are averaged across batch and context positions before the penalties are evaluated. Its subject-domain head predicts the identity of the training subject using a unit-weight cross-entropy loss, while gradient reversal scales the domain gradient reaching the encoder by 0.01. GraphSleepNet, JK-STGCN, and STDP-GCN use only the stage-classification objective.

Auxiliary CNN and FeatureNet pretraining also use unweighted cross-entropy. JK-STGCN FeatureNet pretraining additionally applies an L_2 penalty of 10^{-3} to the initial convolutional kernel of each branch.

Optimizer configurations. Table S6 reports the selected learning rates and the stages at which gradient-value clipping is applied. The pretraining rate applies to a standalone epoch classifier. The joint-encoder rate applies when the waveform encoder is optimized together with the context model. The graph rate applies either to the separate frozen-feature graph stage or to the graph backbone parameter group during joint optimization.

Adam uses $\epsilon = 10^{-7}$ for JK-STGCN and 10^{-8} for the other backbones. The validation-plateau scheduler uses a threshold of 10^{-4} .

S6 Supplementary Results

The absolute performance tables below summarize each metric as the unweighted mean and population standard deviation across the five held-out subject folds. The paired-effect tables use a different, subject-level estimand: all aligned out-of-fold subject contributions are pooled before the metric is recomputed. Effect estimates and confidence intervals in those tables are therefore reported together from the paired analysis.

Table S6: Fixed learning rates and gradient-value clipping stages. All settings were held constant across datasets and cross-validation splits.

Backbone	Condition	Pretraining LR	Joint-encoder LR	Graph LR	Clipping stage
Graph SleepNet	<i>Native</i>	–	–	10^{-3}	–
	CNN-2S	10^{-3}	–	10^{-3}	Pretraining and graph training
	CNN-E2E	–	10^{-3}	10^{-3}	–
	CNN-PTFT	10^{-3}	10^{-3}	10^{-3}	Pretraining only
ProductGraph SleepNet	<i>Native</i>	–	–	10^{-3}	Graph training
	CNN-2S	10^{-3}	–	10^{-3}	Pretraining and graph training
	CNN-E2E	–	10^{-3}	10^{-3}	Joint training
	CNN-PTFT	10^{-3}	10^{-3}	10^{-3}	Pretraining and joint training
MSTGCN	<i>Native</i>	10^{-4}	–	10^{-5}	Pretraining and graph training
	CNN-2S	10^{-3}	–	10^{-3}	Pretraining and graph training
	CNN-E2E	–	10^{-3}	10^{-3}	Joint training
	CNN-PTFT	10^{-3}	10^{-3}	10^{-3}	Pretraining and joint training
JK-STGCN	<i>Native</i>	2×10^{-4}	–	10^{-4}	–
	CNN-2S	2×10^{-4}	–	10^{-4}	Pretraining and graph training
	CNN-E2E	–	2×10^{-4}	2×10^{-4}	Joint training
	CNN-PTFT	2×10^{-4}	2×10^{-4}	2×10^{-4}	Pretraining and joint training
STDP-GCN	<i>Native</i>	–	–	10^{-4}	–
	CNN-2S	10^{-4}	–	10^{-4}	Pretraining and graph training
	CNN-E2E	–	10^{-3}	10^{-3}	Joint training
	CNN-PTFT	10^{-4}	10^{-3}	10^{-3}	Pretraining and joint training

S6.1 Detailed performance results

Class-wise performance. The aggregate class-wise pattern is summarized in Table S7 and complete fold summaries for each dataset are reported in Tables S8–S12.

Table S7: Class-wise F1 across all 25 backbone–dataset combinations. Values are averages of the five-fold means. Parentheses give the number of combinations with higher F1 than the matching *Native* condition. Bold indicates the highest mean for each stage.

Stage	<i>Native</i> mean	CNN-2S mean	CNN-E2E mean	CNN-PTFT mean
W	87.49	89.42 (23/25)	88.72 (20/25)	89.25 (23/25)
N1	46.93	51.62 (21/25)	49.94 (19/25)	50.59 (20/25)
N2	81.09	83.84 (18/25)	83.48 (19/25)	83.07 (16/25)
N3	83.25	84.77 (15/25)	84.32 (15/25)	84.24 (15/25)
REM	80.43	84.49 (22/25)	84.45 (23/25)	84.69 (24/25)

Secondary metrics. Tables S13 and S14 report balanced accuracy and Cohen’s κ for every condition.

Table S8: ANPHY class-wise F1 (%), reported as the unweighted mean $\pm$ population standard deviation across five held-out folds. Bold indicates the best condition for each backbone and stage.

Backbone	Condition	W	N1	N2	N3	REM
Graph SleepNet	<i>Native</i>	88.58 $\pm$ 2.57	47.08 $\pm$ 4.30	88.00 $\pm$ 2.41	88.98 $\pm$ 3.80	84.10 $\pm$ 1.83
	CNN-2S	90.56 $\pm$ 1.29	50.98 $\pm$ 2.38	90.08 $\pm$ 1.55	92.23 $\pm$ 2.29	83.34 $\pm$ 3.70
	CNN-E2E	87.54 $\pm$ 3.10	48.82 $\pm$ 6.62	88.76 $\pm$ 1.89	90.73 $\pm$ 1.21	81.71 $\pm$ 5.49
	CNN-PTFT	90.65 $\pm$ 1.66	53.36 $\pm$ 2.26	89.66 $\pm$ 1.81	91.39 $\pm$ 2.55	85.76 $\pm$ 1.98
ProductGraph SleepNet	<i>Native</i>	89.33 $\pm$ 2.91	49.50 $\pm$ 3.51	88.67 $\pm$ 1.76	90.34 $\pm$ 2.15	85.36 $\pm$ 1.98
	CNN-2S	89.83 $\pm$ 1.87	54.42 $\pm$ 1.91	90.97 $\pm$ 0.95	92.71 $\pm$ 1.82	88.08 $\pm$ 1.21
	CNN-E2E	90.10 $\pm$ 1.92	52.32 $\pm$ 2.70	90.24 $\pm$ 0.81	91.25 $\pm$ 1.00	86.62 $\pm$ 0.51
	CNN-PTFT	89.60 $\pm$ 2.07	48.75 $\pm$ 3.92	89.63 $\pm$ 0.65	91.72 $\pm$ 2.81	86.13 $\pm$ 1.26
MSTGCN	<i>Native</i>	87.34 $\pm$ 2.08	49.98 $\pm$ 2.76	88.42 $\pm$ 2.41	91.57 $\pm$ 1.83	82.21 $\pm$ 2.90
	CNN-2S	89.46 $\pm$ 3.19	50.70 $\pm$ 3.79	89.25 $\pm$ 0.87	90.45 $\pm$ 2.13	84.20 $\pm$ 1.40
	CNN-E2E	87.40 $\pm$ 3.03	47.28 $\pm$ 4.33	88.63 $\pm$ 1.21	88.22 $\pm$ 2.41	81.42 $\pm$ 6.10
	CNN-PTFT	89.04 $\pm$ 3.23	50.50 $\pm$ 3.42	89.09 $\pm$ 1.22	90.32 $\pm$ 2.81	83.50 $\pm$ 3.40
JK-STGCN	<i>Native</i>	88.56 $\pm$ 2.64	48.42 $\pm$ 4.93	90.09 $\pm$ 1.56	92.24 $\pm$ 1.70	82.59 $\pm$ 1.68
	CNN-2S	89.18 $\pm$ 2.37	51.80 $\pm$ 2.27	89.91 $\pm$ 0.91	91.88 $\pm$ 2.00	84.88 $\pm$ 2.86
	CNN-E2E	89.17 $\pm$ 2.10	45.81 $\pm$ 4.46	89.12 $\pm$ 1.25	89.53 $\pm$ 1.37	84.25 $\pm$ 2.47
	CNN-PTFT	89.63 $\pm$ 1.92	51.79 $\pm$ 2.70	90.34 $\pm$ 0.60	91.61 $\pm$ 2.43	86.38 $\pm$ 1.37
STDP-GCN	<i>Native</i>	88.11 $\pm$ 3.51	45.46 $\pm$ 7.68	85.23 $\pm$ 1.11	86.40 $\pm$ 5.23	83.74 $\pm$ 1.31
	CNN-2S	89.36 $\pm$ 3.05	51.58 $\pm$ 3.61	89.25 $\pm$ 0.88	91.09 $\pm$ 3.34	85.90 $\pm$ 2.27
	CNN-E2E	89.26 $\pm$ 3.47	50.53 $\pm$ 3.98	88.65 $\pm$ 1.43	89.53 $\pm$ 4.35	85.40 $\pm$ 0.97
	CNN-PTFT	89.42 $\pm$ 2.64	51.08 $\pm$ 3.98	89.70 $\pm$ 0.95	90.26 $\pm$ 3.43	86.80 $\pm$ 1.98

Table S9: Sleep-EDF20 class-wise F1 (%), reported as the unweighted mean $\pm$ population standard deviation across five held-out folds. Bold indicates the best condition for each backbone and stage.

Backbone	Condition	W	N1	N2	N3	REM
Graph SleepNet	<i>Native</i>	85.67 $\pm$ 2.66	39.29 $\pm$ 8.41	84.65 $\pm$ 3.81	85.54 $\pm$ 2.82	78.37 $\pm$ 3.12
	CNN-2S	90.79 $\pm$ 2.17	47.63 $\pm$ 5.44	87.55 $\pm$ 1.73	84.13 $\pm$ 2.09	86.94 $\pm$ 3.06
	CNN-E2E	90.39 $\pm$ 1.53	48.44 $\pm$ 8.83	87.45 $\pm$ 2.19	83.56 $\pm$ 1.71	87.04 $\pm$ 2.44
	CNN-PTFT	90.34 $\pm$ 1.71	47.08 $\pm$ 5.44	86.71 $\pm$ 2.02	82.56 $\pm$ 1.35	87.23 $\pm$ 2.49
ProductGraph SleepNet	<i>Native</i>	87.72 $\pm$ 3.17	43.45 $\pm$ 8.76	87.37 $\pm$ 2.74	83.84 $\pm$ 2.86	84.92 $\pm$ 3.09
	CNN-2S	90.39 $\pm$ 2.36	49.25 $\pm$ 5.42	88.07 $\pm$ 1.79	84.78 $\pm$ 2.45	88.32 $\pm$ 2.87
	CNN-E2E	89.27 $\pm$ 2.51	47.77 $\pm$ 4.14	87.26 $\pm$ 2.13	84.38 $\pm$ 2.90	88.08 $\pm$ 1.66
	CNN-PTFT	90.91 $\pm$ 1.99	46.12 $\pm$ 6.05	87.25 $\pm$ 2.13	83.41 $\pm$ 1.50	87.84 $\pm$ 2.86
MSTGCN	<i>Native</i>	89.46 $\pm$ 2.35	50.66 $\pm$ 5.85	87.66 $\pm$ 2.65	84.30 $\pm$ 2.04	86.61 $\pm$ 2.47
	CNN-2S	90.44 $\pm$ 1.74	46.78 $\pm$ 5.77	87.15 $\pm$ 1.63	83.75 $\pm$ 1.13	86.60 $\pm$ 1.54
	CNN-E2E	90.42 $\pm$ 2.19	47.41 $\pm$ 9.34	87.86 $\pm$ 2.46	82.62 $\pm$ 0.70	87.32 $\pm$ 2.25
	CNN-PTFT	89.78 $\pm$ 2.75	45.95 $\pm$ 5.28	86.77 $\pm$ 2.49	83.06 $\pm$ 1.80	85.37 $\pm$ 3.82
JK-STGCN	<i>Native</i>	90.07 $\pm$ 2.99	48.20 $\pm$ 8.65	88.91 $\pm$ 3.00	88.80 $\pm$ 1.51	86.64 $\pm$ 2.67
	CNN-2S	90.77 $\pm$ 1.32	46.17 $\pm$ 7.10	87.39 $\pm$ 1.78	84.07 $\pm$ 1.86	86.82 $\pm$ 2.64
	CNN-E2E	90.48 $\pm$ 1.71	49.07 $\pm$ 8.82	87.83 $\pm$ 2.96	84.56 $\pm$ 1.47	87.44 $\pm$ 2.87
	CNN-PTFT	90.58 $\pm$ 3.11	49.55 $\pm$ 5.82	87.74 $\pm$ 1.81	85.01 $\pm$ 3.18	86.81 $\pm$ 3.08
STDP-GCN	<i>Native</i>	86.43 $\pm$ 3.44	39.43 $\pm$ 14.64	83.20 $\pm$ 4.44	78.60 $\pm$ 5.16	73.39 $\pm$ 5.14
	CNN-2S	89.66 $\pm$ 2.73	49.13 $\pm$ 8.00	87.78 $\pm$ 1.81	83.57 $\pm$ 2.07	87.76 $\pm$ 2.15
	CNN-E2E	89.39 $\pm$ 2.94	44.63 $\pm$ 12.46	87.48 $\pm$ 2.35	84.45 $\pm$ 2.65	86.07 $\pm$ 1.72
	CNN-PTFT	91.10 $\pm$ 1.68	47.44 $\pm$ 3.72	87.18 $\pm$ 2.08	83.53 $\pm$ 3.59	85.90 $\pm$ 2.73

Table S10: Sleep-EDFx class-wise F1 (%), reported as the unweighted mean $\pm$ population standard deviation across five held-out folds. Bold indicates the best condition for each backbone and stage.

Backbone	Condition	W	N1	N2	N3	REM
Graph SleepNet	<i>Native</i>	88.12 $\pm$ 4.40	39.38 $\pm$ 1.71	82.90 $\pm$ 1.94	75.91 $\pm$ 3.79	74.81 $\pm$ 6.08
	CNN-2S	91.90 $\pm$ 2.50	48.53 $\pm$ 1.26	84.12 $\pm$ 1.62	72.63 $\pm$ 4.67	83.81 $\pm$ 3.94
	CNN-E2E	91.00 $\pm$ 2.69	45.83 $\pm$ 2.30	83.23 $\pm$ 1.55	72.13 $\pm$ 3.36	81.95 $\pm$ 4.09
	CNN-PTFT	91.24 $\pm$ 2.31	45.89 $\pm$ 1.21	82.65 $\pm$ 1.54	71.72 $\pm$ 4.82	82.93 $\pm$ 3.77
ProductGraph SleepNet	<i>Native</i>	89.31 $\pm$ 4.06	47.64 $\pm$ 1.70	84.81 $\pm$ 1.77	76.58 $\pm$ 3.95	81.29 $\pm$ 5.20
	CNN-2S	91.82 $\pm$ 2.68	49.70 $\pm$ 2.17	84.72 $\pm$ 1.18	72.71 $\pm$ 4.53	84.58 $\pm$ 4.08
	CNN-E2E	91.20 $\pm$ 2.53	47.09 $\pm$ 2.06	83.45 $\pm$ 1.98	70.91 $\pm$ 7.99	84.26 $\pm$ 4.06
	CNN-PTFT	91.21 $\pm$ 2.31	48.04 $\pm$ 0.92	83.30 $\pm$ 2.16	71.09 $\pm$ 6.41	83.74 $\pm$ 4.59
MSTGCN	<i>Native</i>	90.35 $\pm$ 3.38	45.65 $\pm$ 2.48	83.47 $\pm$ 1.59	74.62 $\pm$ 3.67	81.89 $\pm$ 5.91
	CNN-2S	90.85 $\pm$ 3.12	45.95 $\pm$ 1.73	82.33 $\pm$ 2.65	73.27 $\pm$ 4.99	81.38 $\pm$ 4.54
	CNN-E2E	91.51 $\pm$ 2.53	47.74 $\pm$ 2.21	83.71 $\pm$ 2.21	73.69 $\pm$ 6.32	83.79 $\pm$ 3.97
	CNN-PTFT	91.12 $\pm$ 2.65	46.59 $\pm$ 0.51	83.03 $\pm$ 1.66	72.02 $\pm$ 6.26	82.67 $\pm$ 4.62
JK-STGCN	<i>Native</i>	91.00 $\pm$ 2.86	47.96 $\pm$ 1.58	85.12 $\pm$ 1.79	78.50 $\pm$ 3.74	82.79 $\pm$ 6.31
	CNN-2S	91.90 $\pm$ 2.64	48.67 $\pm$ 1.62	84.14 $\pm$ 2.03	71.88 $\pm$ 8.00	83.82 $\pm$ 5.15
	CNN-E2E	92.26 $\pm$ 1.98	49.54 $\pm$ 1.05	84.84 $\pm$ 1.56	73.54 $\pm$ 3.39	84.78 $\pm$ 3.46
	CNN-PTFT	91.74 $\pm$ 2.22	47.32 $\pm$ 0.95	83.33 $\pm$ 1.63	72.99 $\pm$ 5.45	83.40 $\pm$ 4.80
STDP-GCN	<i>Native</i>	88.49 $\pm$ 3.20	39.06 $\pm$ 2.68	81.21 $\pm$ 2.18	70.27 $\pm$ 3.54	72.70 $\pm$ 5.81
	CNN-2S	92.02 $\pm$ 2.58	49.82 $\pm$ 1.71	84.46 $\pm$ 1.57	73.60 $\pm$ 3.79	84.57 $\pm$ 3.72
	CNN-E2E	91.37 $\pm$ 2.60	48.86 $\pm$ 1.31	84.02 $\pm$ 1.85	73.16 $\pm$ 5.23	83.11 $\pm$ 4.48
	CNN-PTFT	91.19 $\pm$ 2.41	45.14 $\pm$ 2.43	83.74 $\pm$ 1.64	73.47 $\pm$ 4.13	81.69 $\pm$ 6.06

Table S11: ISRUC-SG1 class-wise F1 (%), reported as the unweighted mean $\pm$ population standard deviation across five held-out folds. Bold indicates the best condition for each backbone and stage.

Backbone	Condition	W	N1	N2	N3	REM
Graph SleepNet	<i>Native</i>	87.70 $\pm$ 1.61	45.06 $\pm$ 5.60	72.57 $\pm$ 1.60	81.81 $\pm$ 2.52	76.43 $\pm$ 3.05
	CNN-2S	89.30 $\pm$ 2.24	54.52 $\pm$ 2.69	80.10 $\pm$ 0.86	88.89 $\pm$ 0.98	84.11 $\pm$ 2.77
	CNN-E2E	87.73 $\pm$ 3.94	51.17 $\pm$ 3.82	78.45 $\pm$ 1.83	87.31 $\pm$ 0.81	82.31 $\pm$ 3.65
	CNN-PTFT	89.27 $\pm$ 2.21	52.83 $\pm$ 2.30	77.89 $\pm$ 1.03	87.12 $\pm$ 0.81	83.98 $\pm$ 2.73
ProductGraph SleepNet	<i>Native</i>	88.51 $\pm$ 0.94	48.02 $\pm$ 5.52	76.15 $\pm$ 2.45	84.11 $\pm$ 1.59	82.01 $\pm$ 2.10
	CNN-2S	89.99 $\pm$ 1.91	56.03 $\pm$ 2.17	79.76 $\pm$ 1.05	89.09 $\pm$ 0.80	84.80 $\pm$ 3.38
	CNN-E2E	88.25 $\pm$ 2.74	52.10 $\pm$ 3.17	78.96 $\pm$ 1.64	87.85 $\pm$ 0.46	84.77 $\pm$ 1.80
	CNN-PTFT	89.44 $\pm$ 1.91	52.43 $\pm$ 2.81	77.67 $\pm$ 1.14	87.65 $\pm$ 0.86	84.20 $\pm$ 1.60
MSTGCN	<i>Native</i>	87.97 $\pm$ 2.96	52.78 $\pm$ 3.58	78.14 $\pm$ 1.39	86.99 $\pm$ 0.89	82.72 $\pm$ 2.19
	CNN-2S	88.60 $\pm$ 1.59	53.21 $\pm$ 4.10	78.28 $\pm$ 0.65	87.64 $\pm$ 0.96	82.94 $\pm$ 2.25
	CNN-E2E	88.51 $\pm$ 2.68	54.32 $\pm$ 4.13	77.60 $\pm$ 2.91	86.42 $\pm$ 1.19	84.52 $\pm$ 3.15
	CNN-PTFT	89.20 $\pm$ 1.99	53.16 $\pm$ 2.49	77.41 $\pm$ 1.05	87.17 $\pm$ 1.29	84.51 $\pm$ 1.97
JK-STGCN	<i>Native</i>	87.75 $\pm$ 1.10	52.45 $\pm$ 2.72	76.27 $\pm$ 3.89	85.54 $\pm$ 3.45	80.37 $\pm$ 2.69
	CNN-2S	89.37 $\pm$ 2.47	54.33 $\pm$ 4.54	80.34 $\pm$ 1.34	88.61 $\pm$ 0.64	84.04 $\pm$ 3.45
	CNN-E2E	88.99 $\pm$ 2.60	52.77 $\pm$ 5.08	79.27 $\pm$ 1.14	87.57 $\pm$ 0.62	84.21 $\pm$ 3.17
	CNN-PTFT	89.65 $\pm$ 1.89	53.82 $\pm$ 3.75	79.28 $\pm$ 1.06	88.08 $\pm$ 0.83	85.32 $\pm$ 2.15
STDP-GCN	<i>Native</i>	84.72 $\pm$ 2.84	43.72 $\pm$ 2.63	73.45 $\pm$ 2.11	81.26 $\pm$ 1.65	78.82 $\pm$ 2.12
	CNN-2S	89.38 $\pm$ 2.63	54.72 $\pm$ 4.10	79.94 $\pm$ 1.13	88.18 $\pm$ 0.52	85.16 $\pm$ 1.79
	CNN-E2E	88.69 $\pm$ 3.56	53.38 $\pm$ 2.25	77.85 $\pm$ 1.53	87.52 $\pm$ 0.49	83.37 $\pm$ 3.45
	CNN-PTFT	89.18 $\pm$ 2.65	54.68 $\pm$ 3.44	78.84 $\pm$ 0.60	87.82 $\pm$ 0.49	83.99 $\pm$ 2.23

Table S12: ISRUC-SG3 class-wise F1 (%), reported as the unweighted mean $\pm$ population standard deviation across five held-out folds. Bold indicates the best condition for each backbone and stage.

Backbone	Condition	W	N1	N2	N3	REM
Graph SleepNet	<i>Native</i>	81.98 $\pm$ 6.14	47.58 $\pm$ 3.56	69.72 $\pm$ 10.09	83.39 $\pm$ 3.83	78.58 $\pm$ 7.58
	CNN-2S	87.11 $\pm$ 3.99	57.57 $\pm$ 2.87	79.08 $\pm$ 3.30	87.22 $\pm$ 1.85	83.08 $\pm$ 4.93
	CNN-E2E	86.79 $\pm$ 6.10	54.43 $\pm$ 3.50	78.61 $\pm$ 3.73	86.91 $\pm$ 1.88	83.40 $\pm$ 3.17
	CNN-PTFT	86.63 $\pm$ 4.38	57.34 $\pm$ 3.04	78.58 $\pm$ 3.81	87.02 $\pm$ 1.74	83.89 $\pm$ 3.37
ProductGraph SleepNet	<i>Native</i>	82.50 $\pm$ 4.33	45.31 $\pm$ 6.64	66.84 $\pm$ 12.08	76.25 $\pm$ 13.10	77.74 $\pm$ 8.87
	CNN-2S	84.37 $\pm$ 6.81	53.16 $\pm$ 6.29	78.61 $\pm$ 4.30	86.00 $\pm$ 3.77	81.83 $\pm$ 9.27
	CNN-E2E	81.70 $\pm$ 7.69	50.05 $\pm$ 5.54	77.94 $\pm$ 2.22	87.20 $\pm$ 2.08	83.73 $\pm$ 3.98
	CNN-PTFT	84.72 $\pm$ 5.32	51.25 $\pm$ 4.24	76.21 $\pm$ 5.43	87.23 $\pm$ 2.35	84.12 $\pm$ 3.83
MSTGCN	<i>Native</i>	86.25 $\pm$ 5.31	54.60 $\pm$ 5.63	77.65 $\pm$ 4.15	87.89 $\pm$ 1.63	79.09 $\pm$ 5.72
	CNN-2S	85.34 $\pm$ 5.15	54.36 $\pm$ 4.86	77.18 $\pm$ 4.32	87.31 $\pm$ 1.30	81.39 $\pm$ 5.08
	CNN-E2E	83.88 $\pm$ 4.52	49.78 $\pm$ 10.28	79.22 $\pm$ 3.32	88.34 $\pm$ 2.98	85.18 $\pm$ 2.53
	CNN-PTFT	84.55 $\pm$ 5.27	54.39 $\pm$ 5.08	76.93 $\pm$ 4.54	88.42 $\pm$ 2.14	84.11 $\pm$ 4.07
JK-STGCN	<i>Native</i>	87.48 $\pm$ 3.70	56.45 $\pm$ 5.16	76.32 $\pm$ 6.65	84.49 $\pm$ 7.46	79.35 $\pm$ 5.56
	CNN-2S	86.25 $\pm$ 5.74	55.47 $\pm$ 2.50	77.01 $\pm$ 5.10	86.19 $\pm$ 2.96	81.19 $\pm$ 5.11
	CNN-E2E	86.75 $\pm$ 3.46	52.91 $\pm$ 3.88	77.18 $\pm$ 5.32	87.11 $\pm$ 3.78	82.84 $\pm$ 3.08
	CNN-PTFT	86.19 $\pm$ 4.33	55.65 $\pm$ 1.68	77.82 $\pm$ 4.85	86.31 $\pm$ 2.70	85.40 $\pm$ 4.48
STDP-GCN	<i>Native</i>	83.91 $\pm$ 2.94	46.22 $\pm$ 4.01	70.38 $\pm$ 10.24	83.13 $\pm$ 4.65	74.32 $\pm$ 8.73
	CNN-2S	86.92 $\pm$ 3.92	55.94 $\pm$ 2.45	78.56 $\pm$ 3.73	87.50 $\pm$ 1.60	82.59 $\pm$ 7.76
	CNN-E2E	85.88 $\pm$ 3.53	56.45 $\pm$ 4.69	79.37 $\pm$ 3.12	89.46 $\pm$ 1.76	83.75 $\pm$ 4.37
	CNN-PTFT	84.93 $\pm$ 4.52	54.75 $\pm$ 4.74	75.97 $\pm$ 3.08	85.06 $\pm$ 3.45	81.68 $\pm$ 6.38

Table S13: Balanced accuracy (%), reported as the unweighted mean $\pm$ population standard deviation across the five held-out subject folds. Bold indicates the best condition for each backbone and dataset.

Backbone	Condition	ANPHY	EDF20	EDF _x	ISRUC-SG1	ISRUC-SG3
Graph SleepNet	<i>Native</i>	79.26 $\pm$ 1.53	74.64 $\pm$ 2.75	72.07 $\pm$ 0.97	72.91 $\pm$ 1.83	73.19 $\pm$ 4.22
	CNN-2S	81.09 $\pm$ 1.69	79.86 $\pm$ 2.76	75.84 $\pm$ 2.03	79.23 $\pm$ 0.94	79.29 $\pm$ 1.86
	CNN-E2E	79.41 $\pm$ 1.49	79.54 $\pm$ 3.64	75.43 $\pm$ 1.43	77.37 $\pm$ 1.98	79.01 $\pm$ 2.41
	CNN-PTFT	82.27 $\pm$ 1.32	79.59 $\pm$ 2.64	74.97 $\pm$ 2.01	78.19 $\pm$ 0.92	79.16 $\pm$ 1.63
ProductGraph SleepNet	<i>Native</i>	80.61 $\pm$ 1.07	77.11 $\pm$ 3.43	75.02 $\pm$ 1.85	75.34 $\pm$ 2.28	70.23 $\pm$ 4.74
	CNN-2S	83.27 $\pm$ 1.61	80.84 $\pm$ 2.51	76.49 $\pm$ 1.56	80.10 $\pm$ 0.73	77.60 $\pm$ 4.93
	CNN-E2E	82.31 $\pm$ 1.37	80.42 $\pm$ 1.41	75.92 $\pm$ 3.43	78.31 $\pm$ 1.16	77.43 $\pm$ 3.23
	CNN-PTFT	81.46 $\pm$ 1.42	80.03 $\pm$ 2.82	75.29 $\pm$ 2.42	78.25 $\pm$ 0.77	77.65 $\pm$ 2.45
MSTGCN	<i>Native</i>	80.28 $\pm$ 1.05	80.02 $\pm$ 3.25	75.06 $\pm$ 1.55	77.55 $\pm$ 1.39	77.87 $\pm$ 3.14
	CNN-2S	80.85 $\pm$ 1.64	79.49 $\pm$ 2.83	75.00 $\pm$ 1.73	78.23 $\pm$ 1.32	77.77 $\pm$ 2.27
	CNN-E2E	77.93 $\pm$ 3.21	78.87 $\pm$ 4.01	75.90 $\pm$ 2.03	78.57 $\pm$ 1.25	77.97 $\pm$ 1.12
	CNN-PTFT	80.83 $\pm$ 1.49	78.81 $\pm$ 3.22	75.73 $\pm$ 1.99	78.48 $\pm$ 0.98	78.76 $\pm$ 2.38
JK-STGCN	<i>Native</i>	80.40 $\pm$ 2.05	80.50 $\pm$ 2.93	76.40 $\pm$ 1.28	76.03 $\pm$ 1.77	76.53 $\pm$ 3.87
	CNN-2S	81.71 $\pm$ 1.60	79.24 $\pm$ 3.26	75.35 $\pm$ 3.37	79.24 $\pm$ 1.51	77.80 $\pm$ 2.86
	CNN-E2E	79.93 $\pm$ 0.96	79.32 $\pm$ 3.53	76.98 $\pm$ 0.78	78.49 $\pm$ 1.29	78.15 $\pm$ 2.92
	CNN-PTFT	81.90 $\pm$ 0.65	80.40 $\pm$ 2.28	76.32 $\pm$ 1.86	79.22 $\pm$ 0.53	79.20 $\pm$ 2.57
STDP-GCN	<i>Native</i>	78.30 $\pm$ 2.04	72.60 $\pm$ 5.22	70.31 $\pm$ 1.33	72.10 $\pm$ 1.71	71.60 $\pm$ 2.89
	CNN-2S	82.01 $\pm$ 1.36	79.89 $\pm$ 3.10	76.83 $\pm$ 1.17	79.36 $\pm$ 1.26	78.83 $\pm$ 2.51
	CNN-E2E	80.97 $\pm$ 2.13	78.76 $\pm$ 4.10	76.24 $\pm$ 1.84	78.32 $\pm$ 1.24	79.45 $\pm$ 2.44
	CNN-PTFT	81.40 $\pm$ 1.77	79.64 $\pm$ 2.96	75.64 $\pm$ 1.69	78.94 $\pm$ 0.96	77.54 $\pm$ 3.36

Table S14: Cohen’s κ , reported as the unweighted mean $\pm$ population standard deviation across the five held-out subject folds. Bold indicates the best condition for each backbone and dataset.

Backbone	Condition	ANPHY	EDF20	EDF _x	ISRUC-SG1	ISRUC-SG3
Graph SleepNet	<i>Native</i>	0.78 ± 0.03	0.74 ± 0.03	0.71 ± 0.02	0.68 ± 0.01	0.67 ± 0.07
	CNN-2S	0.81 ± 0.02	0.79 ± 0.01	0.75 ± 0.02	0.76 ± 0.01	0.74 ± 0.03
	CNN-E2E	0.79 ± 0.01	0.79 ± 0.02	0.74 ± 0.02	0.74 ± 0.03	0.74 ± 0.04
	CNN-PTFT	0.81 ± 0.02	0.78 ± 0.01	0.73 ± 0.02	0.74 ± 0.02	0.74 ± 0.03
ProductGraph SleepNet	<i>Native</i>	0.80 ± 0.01	0.78 ± 0.03	0.74 ± 0.03	0.72 ± 0.02	0.64 ± 0.07
	CNN-2S	0.83 ± 0.02	0.80 ± 0.01	0.76 ± 0.02	0.77 ± 0.02	0.73 ± 0.05
	CNN-E2E	0.82 ± 0.01	0.79 ± 0.01	0.74 ± 0.02	0.75 ± 0.01	0.72 ± 0.05
	CNN-PTFT	0.81 ± 0.02	0.79 ± 0.01	0.74 ± 0.02	0.75 ± 0.01	0.72 ± 0.04
MSTGCN	<i>Native</i>	0.79 ± 0.02	0.79 ± 0.02	0.74 ± 0.02	0.74 ± 0.02	0.73 ± 0.05
	CNN-2S	0.80 ± 0.01	0.79 ± 0.01	0.73 ± 0.02	0.74 ± 0.01	0.72 ± 0.03
	CNN-E2E	0.78 ± 0.03	0.79 ± 0.02	0.75 ± 0.02	0.74 ± 0.02	0.74 ± 0.02
	CNN-PTFT	0.79 ± 0.02	0.78 ± 0.02	0.74 ± 0.02	0.74 ± 0.02	0.72 ± 0.04
JK-STGCN	<i>Native</i>	0.80 ± 0.02	0.81 ± 0.02	0.76 ± 0.01	0.72 ± 0.02	0.72 ± 0.04
	CNN-2S	0.81 ± 0.01	0.79 ± 0.01	0.75 ± 0.02	0.76 ± 0.02	0.72 ± 0.05
	CNN-E2E	0.79 ± 0.01	0.80 ± 0.03	0.76 ± 0.01	0.75 ± 0.02	0.73 ± 0.05
	CNN-PTFT	0.82 ± 0.01	0.80 ± 0.01	0.74 ± 0.02	0.76 ± 0.01	0.74 ± 0.05
STDP-GCN	<i>Native</i>	0.76 ± 0.03	0.71 ± 0.05	0.69 ± 0.02	0.68 ± 0.02	0.66 ± 0.07
	CNN-2S	0.81 ± 0.02	0.80 ± 0.01	0.76 ± 0.02	0.76 ± 0.01	0.74 ± 0.03
	CNN-E2E	0.80 ± 0.03	0.79 ± 0.02	0.75 ± 0.02	0.74 ± 0.03	0.74 ± 0.03
	CNN-PTFT	0.81 ± 0.02	0.79 ± 0.02	0.74 ± 0.02	0.75 ± 0.02	0.71 ± 0.04

S6.2 Paired subject-level uncertainty

Tables S15–S24 report all 75 predefined paired comparisons and the corresponding class-wise effects. Holm correction was applied across the 75 macro-F1 comparisons as a single family.

Table S15: ANPHY paired subject-level differences relative to *Native*. Macro-F1, accuracy, and balanced-accuracy differences are percentage points; Cohen’s κ differences are reported on their natural scale. Brackets are percentile 95% confidence intervals. p is the two-sided permutation p -value for macro-F1 and p_H its Holm-adjusted value.

Backbone	Condition	Δ Macro-F1	Δ Accuracy	Δ Balanced Accuracy	$\Delta\kappa$	p	p_H
Graph SleepNet	CNN-2S	+2.23 [0.97, 3.47]	+2.07 [0.74, 3.51]	+1.80 [0.38, 3.24]	+0.028 [0.010, 0.047]	0.0026	0.1222
	CNN-E2E	+0.31 [-2.76, 2.75]	+0.48 [-2.40, 2.73]	+0.14 [-2.76, 2.43]	+0.006 [-0.032, 0.036]	0.8266	1.0000
	CNN-PTFT	+2.93 [1.91, 3.98]	+2.35 [1.08, 3.82]	+2.87 [1.78, 4.03]	+0.032 [0.015, 0.051]	0.0002	0.0118
ProductGraph SleepNet	CNN-2S	+2.50 [1.44, 3.51]	+1.98 [1.03, 3.04]	+2.62 [1.57, 3.64]	+0.027 [0.015, 0.041]	0.0004	0.0216
	CNN-E2E	+1.43 [0.29, 2.61]	+1.22 [0.25, 2.24]	+1.51 [0.25, 2.84]	+0.017 [0.003, 0.031]	0.0265	0.9009
	CNN-PTFT	+0.72 [-0.69, 1.91]	+0.77 [-0.32, 1.74]	+1.05 [-0.25, 2.21]	+0.011 [-0.003, 0.023]	0.3001	1.0000
MSTGCN	CNN-2S	+0.63 [-0.63, 2.08]	+0.73 [-0.33, 1.99]	+0.46 [-1.00, 2.09]	+0.009 [-0.005, 0.026]	0.3879	1.0000
	CNN-E2E	-1.49 [-3.27, 0.14]	-0.81 [-2.30, 0.58]	-2.42 [-4.46, -0.48]	-0.014 [-0.034, 0.005]	0.1071	1.0000
	CNN-PTFT	+0.46 [-0.52, 1.47]	+0.45 [-0.49, 1.44]	+0.59 [-0.76, 1.86]	+0.006 [-0.006, 0.019]	0.3868	1.0000
JK-STGCN	CNN-2S	+1.59 [-0.22, 3.21]	+0.74 [-0.53, 1.84]	+1.88 [-0.07, 3.62]	+0.011 [-0.007, 0.026]	0.0919	1.0000
	CNN-E2E	-0.79 [-3.56, 1.45]	-0.49 [-2.95, 1.36]	-0.57 [-3.37, 1.81]	-0.007 [-0.041, 0.018]	0.5547	1.0000
	CNN-PTFT	+1.77 [-0.04, 3.35]	+1.15 [-0.14, 2.25]	+1.76 [-0.16, 3.52]	+0.015 [-0.002, 0.031]	0.0592	1.0000
STDP-GCN	CNN-2S	+3.85 [2.08, 5.31]	+3.53 [1.78, 5.09]	+3.86 [2.32, 5.24]	+0.048 [0.025, 0.068]	0.0005	0.0265
	CNN-E2E	+2.79 [1.33, 4.11]	+2.84 [1.28, 4.43]	+2.59 [1.08, 3.97]	+0.037 [0.016, 0.059]	0.0007	0.0364
	CNN-PTFT	+3.74 [2.39, 5.10]	+3.85 [2.15, 5.60]	+3.26 [1.94, 4.65]	+0.051 [0.029, 0.073]	0.0001	0.0075

S6.3 ISRUC scorer-reference sensitivity

The alternative-reference analysis re-evaluates the fixed out-of-fold predictions from the scorer-1-trained models against scorer 2 at the same retained epoch positions. Table S25 summarizes agreement between the two annotation references, and Tables S26 and S27 report the corresponding CNN-minus-*Native* effects.

S6.4 Fixed-*Native*-adjacency isolation

Macro-F1 results for all 30 fixed-*Native*-adjacency comparisons across the five dataset configurations are reported in Table S28.

Table S16: ANPHY paired subject-level class-wise F1 differences (percentage points) relative to *Native*. Brackets are percentile 95% confidence intervals.

Backbone	Condition	ΔW F1	$\Delta N1$ F1	$\Delta N2$ F1	$\Delta N3$ F1	ΔREM F1
Graph SleepNet	CNN-2S	+2.30	+4.71	+2.01	+2.96	-0.85
		[1.18, 3.45]	[1.44, 7.78]	[0.49, 3.76]	[0.63, 5.60]	[-4.76, 2.61]
	CNN-E2E	-0.22	+1.88	+0.84	+1.46	-2.44
		[-3.77, 2.85]	[-2.62, 5.60]	[-1.39, 2.78]	[-2.22, 4.37]	[-8.35, 1.64]
	CNN-PTFT	+2.30	+6.98	+1.68	+2.20	+1.50
		[1.08, 3.60]	[4.69, 9.26]	[0.25, 3.47]	[0.18, 4.57]	[-0.32, 3.50]
ProductGraph SleepNet	CNN-2S	+0.80	+5.23	+1.99	+2.19	+2.28
		[-0.42, 1.91]	[2.09, 7.76]	[0.87, 3.36]	[0.52, 4.08]	[0.49, 4.44]
	CNN-E2E	+0.90	+3.26	+1.33	+0.70	+0.98
		[-0.42, 2.20]	[0.13, 6.26]	[0.33, 2.45]	[-1.29, 2.47]	[-0.76, 3.32]
	CNN-PTFT	+0.77	+0.27	+0.70	+1.15	+0.71
		[-0.92, 2.28]	[-4.00, 3.57]	[-0.49, 1.89]	[-0.66, 3.09]	[-1.16, 2.79]
MSTGCN	CNN-2S	+1.87	-0.01	+0.44	-1.05	+1.89
		[-0.70, 5.12]	[-2.90, 3.13]	[-0.58, 1.82]	[-2.98, 0.42]	[-1.01, 5.17]
	CNN-E2E	-0.24	-3.34	-0.26	-3.21	-0.38
		[-3.35, 2.93]	[-6.66, 0.06]	[-1.91, 1.36]	[-7.29, -0.32]	[-5.60, 4.02]
	CNN-PTFT	+1.29	+0.19	+0.36	-1.18	+1.64
		[-1.30, 4.50]	[-1.50, 1.92]	[-0.83, 1.63]	[-3.74, 0.53]	[-1.91, 5.13]
JK-STGCN	CNN-2S	+1.15	+4.51	-0.22	-0.45	+2.99
		[-1.33, 3.57]	[0.53, 7.88]	[-1.06, 0.55]	[-1.45, 0.57]	[-0.08, 6.69]
	CNN-E2E	+0.43	-2.58	-1.06	-2.68	+1.95
		[-1.35, 2.25]	[-6.75, 1.21]	[-2.90, 0.38]	[-6.72, 0.10]	[-2.51, 6.46]
	CNN-PTFT	+1.37	+4.13	+0.02	-0.80	+4.14
		[-0.90, 3.71]	[-0.25, 7.71]	[-1.00, 0.99]	[-2.12, 0.47]	[1.31, 7.43]
STDP-GCN	CNN-2S	+1.34	+6.45	+4.09	+4.64	+2.72
		[-0.06, 2.89]	[2.09, 10.23]	[2.13, 6.06]	[2.24, 7.22]	[-0.60, 6.34]
	CNN-E2E	+1.43	+4.62	+3.28	+2.87	+1.74
		[-0.17, 3.22]	[0.77, 8.07]	[1.50, 5.24]	[-0.70, 5.68]	[-1.24, 4.97]
	CNN-PTFT	+1.54	+5.16	+4.64	+3.95	+3.42
		[0.23, 3.08]	[1.84, 8.58]	[2.56, 6.92]	[1.54, 6.39]	[0.56, 6.77]

Table S17: Sleep-EDF20 paired subject-level differences relative to *Native*. Macro-F1, accuracy, and balanced-accuracy differences are percentage points; Cohen’s κ differences are reported on their natural scale. Brackets are percentile 95% confidence intervals. p is the two-sided permutation p -value for macro-F1 and p_H its Holm-adjusted value.

Backbone	Condition	Δ Macro-F1	Δ Accuracy	Δ Balanced Accuracy	$\Delta\kappa$	p	p_H
Graph SleepNet	CNN-2S	+4.37 [2.57, 6.41] +4.16	+3.73 [1.99, 5.58] +3.57	+4.92 [2.93, 7.07] +4.41	+0.053 [0.029, 0.080] +0.051	0.0002	0.0118
	CNN-E2E	[2.21, 6.34] +3.71	[2.01, 5.28] +2.98	[2.35, 6.66] +4.53	[0.029, 0.074] +0.044	0.0008	0.0408
	CNN-PTFT	[1.68, 5.87]	[1.16, 4.83]	[2.44, 6.76]	[0.018, 0.070]	0.0025	0.1200
ProductGraph SleepNet	CNN-2S	+2.70 [0.80, 4.61] +1.69	+1.68 [0.04, 3.35] +0.88	+3.60 [1.86, 5.47] +3.01	+0.025 [0.003, 0.048] +0.015	0.0150	0.5399
	CNN-E2E	[-0.52, 3.70] +1.69	[-1.06, 2.75] +0.95	[1.26, 4.83] +2.71	[-0.011, 0.040] +0.015	0.1425	1.0000
	CNN-PTFT	[-0.44, 3.58]	[-0.97, 2.66]	[0.82, 4.73]	[-0.010, 0.039]	0.1401	1.0000
MSTGCN	CNN-2S	-0.86 [-2.14, 0.41] -0.91	-0.45 [-1.55, 0.71] +0.03	-0.88 [-2.21, 0.50] -1.57	-0.006 [-0.021, 0.010] -0.001	0.2132	1.0000
	CNN-E2E	[-2.00, 0.17] -1.95	[-0.65, 0.70] -1.32	[-2.81, -0.43] -1.73	[-0.011, 0.008] -0.018	0.1425	1.0000
	CNN-PTFT	[-3.72, -0.39]	[-3.09, 0.05]	[-3.66, 0.13]	[-0.042, 0.001]	0.0242	0.8469
JK-STGCN	CNN-2S	-1.77 [-3.66, 0.07] -0.96	-1.15 [-2.63, 0.22] -0.55	-1.73 [-3.82, 0.38] -1.42	-0.015 [-0.036, 0.003] -0.008	0.0916	1.0000
	CNN-E2E	[-2.04, 0.11] -0.74	[-1.49, 0.35] -0.69	[-2.47, -0.17] -0.27	[-0.021, 0.004] -0.008	0.1025	1.0000
	CNN-PTFT	[-2.07, 0.63]	[-1.89, 0.39]	[-1.70, 1.32]	[-0.024, 0.007]	0.2792	1.0000
STDP-GCN	CNN-2S	+7.10 [4.16, 10.29] +6.01	+6.30 [3.69, 9.27] +5.77	+7.25 [4.28, 10.50] +6.02	+0.088 [0.052, 0.129] +0.080	0.0001	0.0075
	CNN-E2E	[3.06, 9.15] +6.59	[3.34, 8.42] +5.75	[3.17, 9.13] +7.07	[0.047, 0.117] +0.081	0.0009	0.0441
	CNN-PTFT	[3.60, 9.87]	[3.23, 8.58]	[4.05, 10.23]	[0.046, 0.120]	0.0008	0.0408

Table S18: Sleep-EDF20 paired subject-level class-wise F1 differences (percentage points) relative to *Native*. Brackets are percentile 95% confidence intervals.

Backbone	Condition	ΔW F1	$\Delta N1$ F1	$\Delta N2$ F1	$\Delta N3$ F1	ΔREM F1
Graph SleepNet	CNN-2S	+5.47	+7.34	+2.87	-1.59	+7.78
		[2.59, 8.59]	[3.17, 11.71]	[0.94, 4.88]	[-4.74, 1.12]	[4.34, 11.83]
	CNN-E2E	+4.41	+7.91	+2.74	-2.19	+7.94
		[2.19, 6.97]	[2.37, 13.67]	[0.93, 4.64]	[-4.98, 0.50]	[4.18, 12.19]
	CNN-PTFT	+4.77	+6.87	+2.14	-3.31	+8.11
		[2.25, 7.55]	[1.43, 12.28]	[0.09, 4.23]	[-6.94, -0.48]	[4.63, 12.31]
ProductGraph SleepNet	CNN-2S	+2.53	+5.88	+0.69	+0.76	+3.64
		[-0.49, 6.41]	[0.81, 10.67]	[-0.91, 2.15]	[-2.03, 3.64]	[1.87, 5.42]
	CNN-E2E	+0.67	+3.92	-0.03	+0.35	+3.56
		[-2.98, 5.08]	[-2.27, 9.51]	[-2.06, 1.70]	[-3.36, 3.17]	[1.20, 6.01]
	CNN-PTFT	+2.89	+3.16	-0.02	-0.71	+3.13
		[-0.01, 6.60]	[-2.38, 8.02]	[-1.99, 1.64]	[-3.70, 2.39]	[1.44, 4.91]
MSTGCN	CNN-2S	+0.50	-3.72	-0.47	-0.68	+0.07
		[-0.95, 2.19]	[-6.90, -0.88]	[-1.68, 0.67]	[-2.21, 0.79]	[-1.57, 1.81]
	CNN-E2E	+0.47	-4.03	+0.22	-1.85	+0.66
		[-0.77, 2.02]	[-8.39, 0.14]	[-0.57, 0.95]	[-4.07, 0.22]	[-0.70, 1.97]
	CNN-PTFT	-0.89	-5.08	-0.83	-1.15	-1.79
		[-5.34, 2.38]	[-8.27, -1.88]	[-1.80, 0.02]	[-2.80, 0.38]	[-5.12, 0.87]
JK-STGCN	CNN-2S	+1.20	-3.75	-1.51	-4.88	+0.08
		[-0.93, 3.49]	[-9.46, 1.84]	[-2.99, -0.12]	[-8.96, -1.79]	[-1.24, 1.35]
	CNN-E2E	+0.82	-1.10	-1.03	-4.41	+0.91
		[-0.96, 2.57]	[-4.79, 3.12]	[-2.00, -0.08]	[-6.84, -2.37]	[-0.39, 2.36]
	CNN-PTFT	+1.28	-0.28	-1.20	-3.64	+0.12
		[-1.97, 4.11]	[-3.78, 3.61]	[-2.48, -0.01]	[-7.14, -1.33]	[-1.15, 1.43]
STDP-GCN	CNN-2S	+4.14	+7.84	+4.43	+5.39	+13.73
		[0.64, 7.48]	[2.66, 14.28]	[1.89, 7.30]	[-0.08, 11.03]	[9.46, 18.63]
	CNN-E2E	+3.57	+3.98	+4.09	+6.29	+12.12
		[0.54, 7.21]	[-2.31, 10.84]	[1.88, 6.40]	[1.63, 11.06]	[8.28, 16.49]
	CNN-PTFT	+5.34	+6.26	+3.85	+5.68	+11.83
		[2.32, 9.01]	[0.15, 14.44]	[1.59, 6.29]	[1.05, 10.43]	[7.76, 16.36]

Table S19: Sleep-EDFx paired subject-level differences relative to *Native*. Macro-F1, accuracy, and balanced-accuracy differences are percentage points; Cohen’s κ differences are reported on their natural scale. Brackets are percentile 95% confidence intervals. p is the two-sided permutation p -value for macro-F1 and p_H its Holm-adjusted value.

Backbone	Condition	Δ Macro-F1	Δ Accuracy	Δ Balanced Accuracy	$\Delta\kappa$	p	p_H
Graph SleepNet	CNN-2S	+3.85 [2.77, 4.97]	+3.10 [2.12, 4.18]	+3.95 [2.59, 5.40]	+0.044 [0.031, 0.059]	0.0001	0.0075
	CNN-E2E	+2.34 [1.20, 3.59]	+1.90 [0.88, 3.07]	+3.07 [1.75, 4.54]	+0.028 [0.014, 0.045]	0.0003	0.0168
	CNN-PTFT	+2.58 [1.50, 3.75]	+1.67 [0.63, 2.80]	+3.09 [1.77, 4.50]	+0.026 [0.011, 0.041]	0.0001	0.0075
ProductGraph SleepNet	CNN-2S	+0.97 [-0.17, 2.20]	+1.31 [0.30, 2.45]	+1.77 [0.47, 3.18]	+0.019 [0.005, 0.035]	0.1191	1.0000
	CNN-E2E	-0.40 [-1.75, 0.95]	-0.08 [-1.24, 1.16]	+1.07 [-0.30, 2.55]	+0.001 [-0.015, 0.018]	0.5849	1.0000
	CNN-PTFT	-0.23 [-1.39, 0.96]	-0.24 [-1.27, 0.84]	+0.80 [-0.49, 2.21]	-0.001 [-0.015, 0.014]	0.7081	1.0000
MSTGCN	CNN-2S	-0.49 [-1.07, 0.10]	-0.76 [-1.34, -0.22]	-0.06 [-0.70, 0.55]	-0.009 [-0.017, -0.002]	0.1139	1.0000
	CNN-E2E	+1.12 [0.41, 1.85]	+0.62 [-0.03, 1.27]	+1.39 [0.59, 2.14]	+0.010 [0.001, 0.018]	0.0032	0.1472
	CNN-PTFT	-0.05 [-0.65, 0.55]	-0.14 [-0.58, 0.33]	+0.78 [0.14, 1.43]	0.000 [-0.006, 0.006]	0.8821	1.0000
JK-STGCN	CNN-2S	-0.87 [-1.91, 0.18]	-0.43 [-1.27, 0.43]	-0.86 [-1.98, 0.25]	-0.006 [-0.017, 0.005]	0.1158	1.0000
	CNN-E2E	-0.28 [-1.30, 0.73]	+0.17 [-0.66, 1.00]	+0.36 [-0.69, 1.42]	+0.003 [-0.008, 0.014]	0.6130	1.0000
	CNN-PTFT	-1.45 [-2.46, -0.52]	-1.26 [-2.02, -0.53]	-0.02 [-0.93, 0.87]	-0.015 [-0.025, -0.006]	0.0047	0.1974
STDP-GCN	CNN-2S	+6.47 [5.30, 7.71]	+4.96 [3.92, 6.01]	+7.04 [5.72, 8.44]	+0.070 [0.056, 0.085]	0.0001	0.0075
	CNN-E2E	+5.69 [4.61, 6.87]	+4.20 [3.17, 5.34]	+6.33 [5.07, 7.69]	+0.060 [0.046, 0.075]	0.0001	0.0075
	CNN-PTFT	+4.40 [3.26, 5.65]	+3.60 [2.44, 4.85]	+5.30 [4.05, 6.69]	+0.051 [0.036, 0.068]	0.0001	0.0075

Table S20: Sleep-EDFx paired subject-level class-wise F1 differences (percentage points) relative to *Native*. Brackets are percentile 95% confidence intervals.

Backbone	Condition	ΔW F1	$\Delta N1$ F1	$\Delta N2$ F1	$\Delta N3$ F1	ΔREM F1
Graph SleepNet	CNN-2S	+3.36	+8.41	+1.23	-2.84	+9.11
		[1.75, 5.33]	[6.43, 10.49]	[0.35, 2.21]	[-5.35, -0.56]	[6.83, 11.64]
	CNN-E2E	+2.48	+5.62	+0.39	-3.91	+7.11
ProductGraph SleepNet	CNN-2S	[0.98, 4.40]	[3.61, 7.76]	[-0.50, 1.37]	[-6.88, -1.16]	[4.72, 9.55]
		+2.71	+5.94	-0.23	-3.82	+8.32
	CNN-PTFT	[1.08, 4.74]	[4.00, 8.00]	[-1.18, 0.71]	[-6.07, -1.69]	[6.00, 10.90]
MSTGCN	CNN-2S	+2.73	+2.42	-0.03	-3.62	+3.36
		[1.20, 4.66]	[-0.13, 5.23]	[-0.85, 0.86]	[-6.47, -0.96]	[1.50, 5.36]
	CNN-E2E	+1.89	-0.65	-1.38	-4.93	+3.06
JK-STGCN	CNN-2S	[0.38, 3.69]	[-2.85, 1.75]	[-2.54, -0.21]	[-8.79, -1.58]	[1.23, 5.10]
		+1.94	+0.70	-1.53	-4.84	+2.57
	CNN-PTFT	[0.53, 3.69]	[-1.60, 3.17]	[-2.60, -0.53]	[-7.58, -2.29]	[0.75, 4.52]
STDP-GCN	CNN-2S	+0.31	0.00	-1.20	-1.15	-0.38
		[-0.11, 0.79]	[-1.65, 1.71]	[-2.08, -0.43]	[-2.79, 0.36]	[-1.78, 1.16]
	CNN-E2E	+0.63	+2.75	+0.22	+0.01	+2.00
STDP-GCN	CNN-2S	[0.06, 1.32]	[0.87, 4.63]	[-0.56, 0.95]	[-1.78, 1.70]	[0.46, 3.60]
		+0.31	+0.97	-0.44	-1.95	+0.88
	CNN-PTFT	[-0.25, 1.08]	[-0.57, 2.46]	[-0.99, 0.12]	[-3.93, -0.13]	[-0.46, 2.35]
STDP-GCN	CNN-2S	+0.53	+0.29	-1.03	-5.20	+1.04
		[-0.12, 1.26]	[-1.45, 2.06]	[-1.93, -0.10]	[-8.35, -2.17]	[-0.83, 3.10]
	CNN-E2E	+0.67	+1.24	-0.23	-5.23	+2.15
STDP-GCN	CNN-2S	[-0.13, 1.60]	[-0.62, 3.17]	[-1.09, 0.69]	[-8.47, -2.09]	[0.29, 4.27]
		+0.21	-1.06	-1.80	-5.24	+0.63
	CNN-PTFT	[-0.46, 0.95]	[-2.54, 0.44]	[-2.60, -0.98]	[-8.76, -2.11]	[-0.96, 2.33]
STDP-GCN	CNN-2S	+3.39	+9.76	+3.30	+3.83	+12.05
		[2.40, 4.72]	[7.12, 12.26]	[2.22, 4.48]	[0.91, 7.02]	[9.59, 14.65]
	CNN-E2E	+2.61	+8.78	+2.84	+3.74	+10.45
STDP-GCN	CNN-2S	[1.58, 3.66]	[6.87, 10.81]	[1.61, 4.23]	[0.69, 7.10]	[8.07, 12.89]
		+2.31	+4.61	+2.55	+3.64	+8.90
	CNN-PTFT	[1.25, 3.45]	[2.38, 7.02]	[1.26, 4.04]	[0.83, 6.70]	[6.40, 11.42]

Table S21: ISRUC-SG1 paired subject-level differences relative to *Native*. Macro-F1, accuracy, and balanced-accuracy differences are percentage points; Cohen’s κ differences are reported on their natural scale. Brackets are percentile 95% confidence intervals. p is the two-sided permutation p -value for macro-F1 and p_H its Holm-adjusted value.

Backbone	Condition	Δ Macro-F1	Δ Accuracy	Δ Balanced Accuracy	$\Delta\kappa$	p	p_H
Graph SleepNet	CNN-2S	+6.51 [5.09, 7.93] +4.67	+5.88 [4.45, 7.30] +4.18	+6.29 [4.88, 7.73] +4.57	+0.076 [0.058, 0.094] +0.054	0.0001	0.0075
	CNN-E2E	[3.27, 6.09] +5.42	[2.76, 5.61] +4.64	[3.12, 6.05] +5.24	[0.036, 0.073] +0.060	0.0001	0.0075
	CNN-PTFT	[4.07, 6.89]	[3.31, 6.07]	[3.84, 6.77]	[0.043, 0.078]	0.0001	0.0075
ProductGraph SleepNet	CNN-2S	+3.96 [2.64, 5.32] +2.50	+3.26 [2.03, 4.50] +2.02	+4.52 [3.33, 5.87] +2.79	+0.044 [0.028, 0.060] +0.027	0.0001	0.0075
	CNN-E2E	[1.34, 3.82] +2.33	[0.91, 3.27] +1.72	[1.59, 4.22] +2.72	[0.013, 0.044] +0.024	0.0001	0.0075
	CNN-PTFT	[1.21, 3.53]	[0.65, 2.84]	[1.56, 3.98]	[0.010, 0.038]	0.0003	0.0168
MSTGCN	CNN-2S	+0.39 [-0.63, 1.40] +0.71	+0.21 [-0.77, 1.17] +0.06	+0.58 [-0.42, 1.60] +1.11	+0.003 [-0.009, 0.016] +0.002	0.4629	1.0000
	CNN-E2E	[-0.38, 1.77] +0.63	[-1.02, 1.14] +0.18	[0.09, 2.10] +0.94	[-0.012, 0.016] +0.003	0.2116	1.0000
	CNN-PTFT	[-0.38, 1.66]	[-0.75, 1.15]	[-0.11, 2.03]	[-0.009, 0.016]	0.2409	1.0000
JK-STGCN	CNN-2S	+2.76 [1.42, 4.13] +2.15	+3.21 [1.94, 4.52] +2.58	+3.19 [1.76, 4.62] +2.74	+0.041 [0.025, 0.058] +0.033	0.0002	0.0118
	CNN-E2E	[0.75, 3.58] +2.73	[1.19, 3.97] +2.86	[1.32, 4.23] +3.24	[0.016, 0.051] +0.037	0.0045	0.1935
	CNN-PTFT	[1.42, 4.10]	[1.61, 4.17]	[1.86, 4.65]	[0.021, 0.054]	0.0001	0.0075
STDP-GCN	CNN-2S	+7.03 [5.31, 8.84] +5.61	+6.28 [4.62, 8.04] +4.87	+7.19 [5.48, 8.99] +6.07	+0.082 [0.061, 0.105] +0.065	0.0001	0.0075
	CNN-E2E	[4.02, 7.31] +6.37	[3.25, 6.55] +5.53	[4.56, 7.71] +6.73	[0.044, 0.086] +0.073	0.0001	0.0075
	CNN-PTFT	[4.92, 8.02]	[4.08, 7.12]	[5.32, 8.33]	[0.055, 0.093]	0.0001	0.0075

Table S22: ISRUC-SG1 paired subject-level class-wise F1 differences (percentage points) relative to *Native*. Brackets are percentile 95% confidence intervals.

Backbone	Condition	ΔW F1	$\Delta N1$ F1	$\Delta N2$ F1	$\Delta N3$ F1	ΔREM F1
Graph SleepNet	CNN-2S	+1.87	+9.25	+7.39	+6.97	+7.04
		[0.39, 3.47]	[6.80, 11.74]	[5.18, 9.74]	[4.93, 9.14]	[2.80, 11.50]
	CNN-E2E	+0.43	+5.97	+5.82	+5.45	+5.68
ProductGraph SleepNet	CNN-E2E	[-1.88, 2.33]	[3.33, 8.69]	[3.70, 8.06]	[3.33, 7.68]	[2.51, 9.22]
		+1.89	+7.56	+5.22	+5.27	+7.16
	CNN-PTFT	[0.36, 3.44]	[4.93, 10.30]	[3.08, 7.54]	[3.23, 7.56]	[3.10, 11.73]
MSTGCN	CNN-2S	+1.70	+7.62	+3.39	+4.71	+2.37
		[0.19, 3.44]	[5.21, 10.14]	[1.56, 5.21]	[3.17, 6.37]	[-1.71, 6.20]
	CNN-E2E	-0.04	+3.72	+2.70	+3.56	+2.53
JK-STGCN	CNN-E2E	[-2.15, 2.25]	[1.32, 6.21]	[1.38, 4.08]	[1.91, 5.40]	[-0.14, 5.61]
		+1.17	+4.04	+1.34	+3.35	+1.77
	CNN-PTFT	[-0.27, 3.00]	[1.53, 6.66]	[-0.32, 2.95]	[1.70, 5.08]	[-0.95, 4.64]
STDP-GCN	CNN-2S	+0.37	+0.67	+0.08	+0.40	+0.42
		[-1.12, 1.93]	[-1.23, 2.59]	[-1.10, 1.27]	[-0.95, 1.94]	[-1.69, 2.45]
	CNN-E2E	+0.48	+2.02	-0.40	-0.77	+2.19
STDP-GCN	CNN-E2E	[-0.96, 1.89]	[-0.50, 4.48]	[-1.98, 1.21]	[-2.20, 0.55]	[-0.02, 4.57]
		+1.22	+0.61	-0.73	+0.09	+1.95
	CNN-PTFT	[0.14, 2.44]	[-1.42, 2.68]	[-1.94, 0.51]	[-1.26, 1.64]	[-0.02, 4.06]
STDP-GCN	CNN-2S	+1.70	+1.83	+3.71	+3.18	+3.40
		[0.24, 3.21]	[-0.58, 4.23]	[2.01, 5.49]	[1.32, 5.33]	[0.46, 6.39]
	CNN-E2E	+1.55	+0.49	+2.67	+2.23	+3.78
STDP-GCN	CNN-E2E	[0.23, 2.85]	[-1.65, 2.58]	[0.85, 4.58]	[-0.01, 4.65]	[0.98, 6.84]
		+2.02	+1.50	+2.66	+2.69	+4.78
	CNN-PTFT	[0.72, 3.48]	[-0.87, 3.84]	[1.06, 4.41]	[0.97, 4.61]	[2.13, 7.72]
STDP-GCN	CNN-2S	+4.42	+10.93	+6.64	+7.02	+6.13
		[2.66, 6.54]	[7.75, 14.36]	[4.32, 9.22]	[4.47, 9.73]	[3.42, 8.99]
	CNN-E2E	+3.95	+9.42	+4.31	+6.28	+4.10
STDP-GCN	CNN-E2E	[2.13, 6.10]	[6.64, 12.46]	[1.97, 6.78]	[3.77, 8.95]	[1.22, 7.03]
		+4.20	+10.57	+5.35	+6.61	+5.10
	CNN-PTFT	[2.49, 6.35]	[7.80, 13.77]	[3.23, 7.58]	[4.16, 9.30]	[2.67, 7.69]

Table S23: ISRUC-SG3 paired subject-level differences relative to *Native*. Macro-F1, accuracy, and balanced-accuracy differences are percentage points; Cohen’s κ differences are reported on their natural scale. Brackets are percentile 95% confidence intervals. p is the two-sided permutation p -value for macro-F1 and p_H its Holm-adjusted value.

Backbone	Condition	Δ Macro-F1	Δ Accuracy	Δ Balanced Accuracy	$\Delta\kappa$	p	p_H
Graph SleepNet	CNN-2S	+6.43 [2.51, 11.14]	+5.61 [1.99, 9.78]	+5.01 [1.50, 9.55]	+0.071 [0.026, 0.123]	0.0061	0.2501
	CNN-E2E	+5.72 [2.24, 10.01]	+5.26 [1.97, 8.99]	+4.33 [1.24, 8.75]	+0.066 [0.025, 0.114]	0.0076	0.2964
	CNN-PTFT	+6.37 [3.03, 10.51]	+5.42 [1.99, 9.30]	+4.94 [2.24, 8.94]	+0.069 [0.027, 0.117]	0.0040	0.1800
ProductGraph SleepNet	CNN-2S	+5.84 [2.49, 10.17]	+6.33 [2.56, 11.23]	+5.63 [2.31, 10.08]	+0.081 [0.033, 0.143]	0.0067	0.2680
	CNN-E2E	+4.48 [1.33, 9.03]	+5.07 [1.15, 10.36]	+4.24 [1.28, 8.69]	+0.065 [0.017, 0.131]	0.0315	1.0000
	CNN-PTFT	+5.25 [2.02, 9.57]	+5.49 [1.85, 10.09]	+5.32 [2.15, 9.67]	+0.071 [0.025, 0.129]	0.0121	0.4477
MSTGCN	CNN-2S	+0.06 [-2.42, 2.70]	-0.33 [-2.35, 2.04]	-0.22 [-2.85, 2.37]	-0.005 [-0.031, 0.026]	0.9731	1.0000
	CNN-E2E	+0.27 [-3.19, 3.49]	+0.62 [-1.77, 3.10]	+0.14 [-3.36, 3.48]	+0.007 [-0.025, 0.039]	0.8701	1.0000
	CNN-PTFT	+0.32 [-1.69, 2.49]	-0.46 [-2.75, 1.87]	+0.23 [-1.70, 2.41]	-0.005 [-0.034, 0.024]	0.7902	1.0000
JK-STGCN	CNN-2S	-0.21 [-3.57, 3.50]	-0.27 [-3.71, 3.06]	-0.24 [-3.85, 4.09]	-0.003 [-0.047, 0.040]	0.9143	1.0000
	CNN-E2E	-0.46 [-3.31, 3.40]	-0.07 [-3.20, 3.40]	+0.01 [-2.78, 4.29]	0.000 [-0.040, 0.044]	0.8109	1.0000
	CNN-PTFT	+0.60 [-2.14, 4.01]	+0.60 [-2.26, 3.53]	+0.90 [-2.20, 5.19]	+0.008 [-0.029, 0.047]	0.7252	1.0000
STDP-GCN	CNN-2S	+6.55 [3.81, 9.55]	+5.60 [2.50, 8.99]	+5.77 [3.74, 8.25]	+0.072 [0.035, 0.113]	0.0043	0.1892
	CNN-E2E	+6.87 [3.58, 10.29]	+5.92 [1.82, 9.95]	+6.12 [3.62, 9.02]	+0.076 [0.026, 0.125]	0.0102	0.3876
	CNN-PTFT	+4.63 [1.27, 8.13]	+3.37 [-0.40, 7.28]	+3.88 [1.13, 7.05]	+0.044 [-0.002, 0.091]	0.0430	1.0000

Table S24: ISRUC-SG3 paired subject-level class-wise F1 differences (percentage points) relative to *Native*. Brackets are percentile 95% confidence intervals.

Backbone	Condition	ΔW F1	$\Delta N1$ F1	$\Delta N2$ F1	$\Delta N3$ F1	ΔREM F1
Graph SleepNet	CNN-2S	+5.00	+9.38	+7.51	+4.50	+5.77
		[2.10, 7.81]	[5.53, 14.54]	[2.71, 15.09]	[-0.07, 9.49]	[-3.47, 16.84]
	CNN-E2E	+5.47	+5.95	+6.94	+4.20	+6.03
		[1.67, 8.79]	[1.07, 12.39]	[1.93, 15.02]	[-0.16, 8.72]	[-2.49, 16.14]
	CNN-PTFT	+4.88	+9.31	+6.96	+4.05	+6.62
		[1.57, 8.00]	[5.73, 13.75]	[1.61, 15.24]	[-0.66, 8.93]	[-1.72, 16.73]
ProductGraph SleepNet	CNN-2S	+1.90	+6.13	+9.91	+6.23	+5.04
		[-1.08, 5.65]	[1.65, 12.28]	[4.19, 17.53]	[1.79, 13.64]	[-0.04, 10.85]
	CNN-E2E	-1.11	+1.58	+8.78	+7.33	+5.82
		[-8.28, 4.39]	[-4.23, 8.43]	[3.07, 17.50]	[3.42, 14.63]	[0.68, 13.11]
	CNN-PTFT	+2.20	+3.83	+7.80	+6.63	+5.78
		[-2.78, 6.60]	[-1.54, 10.66]	[2.66, 15.15]	[1.21, 15.57]	[0.28, 13.11]
MSTGCN	CNN-2S	-1.39	-0.03	-0.27	-0.44	+2.45
		[-4.04, 0.66]	[-5.40, 5.96]	[-2.86, 2.61]	[-2.61, 1.47]	[-2.81, 6.96]
	CNN-E2E	-3.45	-3.76	+1.69	+0.92	+5.92
		[-6.64, -0.12]	[-15.96, 6.42]	[-0.72, 4.00]	[-1.91, 3.28]	[1.59, 10.80]
	CNN-PTFT	-2.75	-0.66	-0.59	+0.85	+4.76
		[-8.34, 1.36]	[-5.48, 4.10]	[-3.40, 2.20]	[-1.44, 2.87]	[0.99, 8.30]
JK-STGCN	CNN-2S	-0.67	-1.94	+0.09	+0.25	+1.20
		[-2.75, 0.81]	[-7.10, 3.47]	[-4.20, 4.95]	[-2.67, 4.32]	[-5.20, 8.42]
	CNN-E2E	-0.95	-5.00	-0.19	+0.98	+2.86
		[-2.87, 0.73]	[-9.79, 1.15]	[-5.20, 5.61]	[-3.28, 6.55]	[-2.15, 10.27]
	CNN-PTFT	-1.12	-1.96	+0.69	+0.51	+4.90
		[-2.72, 0.39]	[-6.95, 3.25]	[-3.43, 5.60]	[-2.52, 4.72]	[-0.17, 12.07]
STDP-GCN	CNN-2S	+3.46	+8.77	+6.31	+3.86	+10.34
		[1.33, 5.94]	[3.51, 14.76]	[1.53, 13.11]	[0.38, 7.53]	[4.15, 16.67]
	CNN-E2E	+1.63	+9.36	+7.00	+5.61	+10.77
		[-1.22, 4.92]	[4.71, 14.85]	[1.18, 14.32]	[2.01, 9.46]	[5.35, 17.69]
	CNN-PTFT	+1.49	+7.41	+3.43	+1.49	+9.33
		[-0.99, 3.70]	[1.87, 13.48]	[-1.75, 10.55]	[-1.62, 4.95]	[1.86, 17.17]

Table S25: Agreement between the two ISRUC scorers on the epochs retained in the primary analysis. Agreement and class-wise F1 values are percentages. Cohen’s κ is reported on its natural scale. Brackets are subject-bootstrap 95% confidence intervals.

Dataset	Agreement	Cohen’s κ	W	N1	N2	N3	REM
ISRUC-SG1	82.38 [80.96, 83.77]	0.772 [0.754, 0.790]	91.88 [90.16, 93.39]	53.86 [50.87, 56.67]	79.45 [77.41, 81.36]	86.18 [83.96, 88.14]	90.14 [88.52, 91.58]
ISRUC-SG3	88.25 [83.29, 93.53]	0.849 [0.783, 0.917]	94.56 [90.72, 97.39]	74.47 [57.08, 87.95]	86.92 [81.28, 92.65]	91.67 [88.70, 95.09]	90.53 [85.70, 95.23]

Table S26: ISRUC-SG1 scorer-reference sensitivity. Values are CNN-minus-*Native* differences in percentage points when the same scorer-1-trained out-of-fold predictions are evaluated against scorer 1 or scorer 2. Brackets are paired subject-bootstrap 95% confidence intervals.

Backbone	Condition	Δ Macro-F1 scorer 1	Δ Macro-F1 scorer 2	Δ N1 F1 scorer 1	Δ N1 F1 scorer 2
Graph SleepNet	CNN-2S	+6.51 [5.12, 7.98]	+6.08 [4.54, 7.75]	+9.25 [6.77, 11.86]	+8.19 [5.34, 11.17]
	CNN-E2E	+4.67 [3.29, 6.07]	+4.13 [2.59, 5.69]	+5.97 [3.34, 8.66]	+5.11 [2.15, 8.18]
	CNN-PTFT	+5.42 [4.06, 6.89]	+5.22 [3.68, 6.85]	+7.56 [4.98, 10.38]	+7.53 [4.59, 10.45]
ProductGraph SleepNet	CNN-2S	+3.96 [2.64, 5.32]	+3.91 [2.46, 5.38]	+7.62 [5.23, 10.14]	+7.75 [4.65, 11.02]
	CNN-E2E	+2.50 [1.36, 3.77]	+2.16 [1.03, 3.41]	+3.72 [1.36, 6.19]	+2.17 [-0.32, 4.94]
	CNN-PTFT	+2.33 [1.20, 3.51]	+2.21 [0.89, 3.57]	+4.04 [1.59, 6.64]	+4.35 [1.05, 7.98]
MSTGCN	CNN-2S	+0.39 [-0.62, 1.41]	-0.70 [-1.56, 0.15]	+0.67 [-1.27, 2.58]	-1.56 [-3.48, 0.32]
	CNN-E2E	+0.71 [-0.39, 1.76]	-0.33 [-1.34, 0.70]	+2.02 [-0.48, 4.54]	-1.28 [-3.74, 1.23]
	CNN-PTFT	+0.63 [-0.36, 1.64]	-0.37 [-1.15, 0.48]	+0.61 [-1.35, 2.67]	-1.60 [-3.41, 0.36]
JK-STGCN	CNN-2S	+2.76 [1.40, 4.14]	+2.07 [0.81, 3.35]	+1.83 [-0.59, 4.24]	+0.80 [-1.34, 3.02]
	CNN-E2E	+2.15 [0.72, 3.60]	+1.42 [0.11, 2.76]	+0.49 [-1.69, 2.61]	-1.13 [-3.41, 1.25]
	CNN-PTFT	+2.73 [1.42, 4.08]	+2.32 [1.11, 3.61]	+1.50 [-0.83, 3.87]	+1.74 [-0.40, 4.02]
STDP-GCN	CNN-2S	+7.03 [5.31, 8.88]	+6.63 [5.08, 8.24]	+10.93 [7.75, 14.41]	+8.92 [5.76, 12.04]
	CNN-E2E	+5.61 [4.02, 7.31]	+5.80 [4.25, 7.42]	+9.42 [6.69, 12.45]	+9.92 [7.30, 12.66]
	CNN-PTFT	+6.37 [4.89, 7.97]	+5.96 [4.46, 7.49]	+10.57 [7.83, 13.61]	+9.22 [6.46, 12.04]

S6.5 Normalization sensitivity

Normalization-matched *Native* controls. Results for the normalization-matched *Native* controls across all five dataset configurations are reported in Table S29.

Available-history normalization sensitivity. Results for the available-history normalization sensitivity analysis across the five dataset configurations are reported in Table S30.

S6.6 Cross-dataset encoder portability

Complete cross-dataset transfer results for both source datasets are reported in Tables S31 and S32. Differences are computed from unrounded values.

S6.7 Training-pipeline timing

Phase-separated and dataset-specific training-pipeline times are reported in Tables S33 and S34, respectively.

Table S27: ISRUC-SG3 scorer-reference sensitivity. Values are CNN-minus-*Native* differences in percentage points when the same scorer-1-trained out-of-fold predictions are evaluated against scorer 1 or scorer 2. Brackets are paired subject-bootstrap 95% confidence intervals.

Backbone	Condition	Δ Macro-F1 scorer 1	Δ Macro-F1 scorer 2	Δ N1 F1 scorer 1	Δ N1 F1 scorer 2
Graph SleepNet	CNN-2S	+6.43 [2.53, 11.18]	+6.16 [2.75, 9.87]	+9.38 [5.48, 14.74]	+7.56 [2.91, 12.93]
	CNN-E2E	+5.72 [2.27, 9.89]	+5.95 [3.05, 8.88]	+5.95 [1.14, 12.23]	+5.64 [-1.98, 13.24]
	CNN-PTFT	+6.37 [2.97, 10.41]	+6.02 [2.65, 9.54]	+9.31 [5.70, 13.75]	+7.49 [1.44, 12.71]
ProductGraph SleepNet	CNN-2S	+5.84 [2.48, 10.21]	+5.25 [1.45, 9.79]	+6.13 [1.68, 12.20]	+3.05 [-2.57, 8.62]
	CNN-E2E	+4.48 [1.32, 8.86]	+4.03 [0.47, 8.32]	+1.58 [-4.35, 8.58]	+0.37 [-5.94, 6.94]
	CNN-PTFT	+5.25 [1.96, 9.56]	+5.39 [1.29, 10.72]	+3.83 [-1.56, 10.61]	+3.45 [-2.54, 9.40]
MSTGCN	CNN-2S	+0.06 [-2.43, 2.71]	+0.01 [-2.76, 2.66]	-0.03 [-5.40, 6.12]	-0.06 [-5.68, 6.38]
	CNN-E2E	+0.27 [-3.23, 3.46]	+0.97 [-3.14, 4.71]	-3.76 [-16.06, 6.45]	-1.93 [-16.02, 9.05]
	CNN-PTFT	+0.32 [-1.72, 2.48]	+0.36 [-1.94, 2.48]	-0.66 [-5.49, 4.06]	-0.49 [-5.22, 4.25]
JK-STGCN	CNN-2S	-0.21 [-3.58, 3.48]	-0.94 [-5.93, 3.28]	-1.94 [-6.92, 3.46]	-5.95 [-12.75, 1.16]
	CNN-E2E	-0.46 [-3.31, 3.48]	-1.03 [-3.87, 2.32]	-5.00 [-9.71, 1.20]	-8.62 [-12.32, -3.65]
	CNN-PTFT	+0.60 [-2.09, 3.97]	+0.89 [-3.47, 5.20]	-1.96 [-6.86, 3.30]	-3.49 [-9.95, 3.67]
STDP-GCN	CNN-2S	+6.55 [3.84, 9.48]	+6.03 [2.80, 9.14]	+8.77 [3.52, 14.64]	+5.81 [-0.54, 11.51]
	CNN-E2E	+6.87 [3.44, 10.32]	+5.45 [2.04, 8.60]	+9.36 [4.75, 15.08]	+4.67 [-1.19, 10.05]
	CNN-PTFT	+4.63 [1.31, 8.19]	+3.25 [-1.95, 7.32]	+7.41 [1.93, 13.36]	+2.47 [-5.73, 9.65]

Table S28: Fixed-*Native*-adjacency isolation results across the five dataset configurations. Endpoint values are five-fold mean macro-F1 (%). Differences are percentage points. Y_{00} uses *Native* nodes and graph, Y_{10} uses CNN nodes with the fold-specific *Native* graph held fixed, and Y_{11} uses CNN nodes with the graph learned from the CNN representation.

Backbone	Dataset	CNN regime	Y_{00}	Y_{10}	Y_{11}	$Y_{10} - Y_{00}$	$Y_{11} - Y_{10}$
GraphSleepNet	ANPHY	CNN-2S	79.35	81.79	81.44	+2.44	-0.35
	EDF20	CNN-2S	74.70	80.18	79.41	+5.48	-0.77
	EDFx	CNN-2S	72.23	76.10	76.20	+3.87	+0.10
	ISRUC-SG1	CNN-2S	72.71	79.16	79.38	+6.45	+0.22
	ISRUC-SG3	CNN-2S	72.25	78.34	78.81	+6.09	+0.47
	ANPHY	CNN-E2E	79.35	80.27	79.51	+0.92	-0.76
	EDF20	CNN-E2E	74.70	78.88	79.38	+4.18	+0.50
	EDFx	CNN-E2E	72.23	74.95	74.83	+2.72	-0.12
	ISRUC-SG1	CNN-E2E	72.71	77.73	77.39	+5.02	-0.34
	ISRUC-SG3	CNN-E2E	72.25	79.08	78.03	+6.83	-1.05
	ANPHY	CNN-PTFT	79.35	81.82	82.16	+2.47	+0.34
	EDF20	CNN-PTFT	74.70	78.35	78.78	+3.65	+0.43
	EDFx	CNN-PTFT	72.23	74.82	74.89	+2.59	+0.07
	ISRUC-SG1	CNN-PTFT	72.71	78.25	78.22	+5.54	-0.03
	ISRUC-SG3	CNN-PTFT	72.25	78.15	78.69	+5.90	+0.54
ProductGraph SleepNet	ANPHY	CNN-2S	80.64	82.85	83.20	+2.21	+0.35
	EDF20	CNN-2S	77.46	79.73	80.16	+2.27	+0.43
	EDFx	CNN-2S	75.93	76.83	76.71	+0.90	-0.12
	ISRUC-SG1	CNN-2S	75.76	80.29	79.93	+4.53	-0.36
	ISRUC-SG3	CNN-2S	69.73	76.39	76.79	+6.66	+0.40
	ANPHY	CNN-E2E	80.64	81.04	82.11	+0.41	+1.06
	EDF20	CNN-E2E	77.46	80.23	79.35	+2.76	-0.88
	EDFx	CNN-E2E	75.93	75.54	75.38	-0.38	-0.16
	ISRUC-SG1	CNN-E2E	75.76	79.48	78.39	+3.72	-1.09
	ISRUC-SG3	CNN-E2E	69.73	76.13	76.12	+6.40	-0.01
	ANPHY	CNN-PTFT	80.64	81.19	81.16	+0.55	-0.02
	EDF20	CNN-PTFT	77.46	78.94	79.11	+1.48	+0.17
	EDFx	CNN-PTFT	75.93	75.42	75.47	-0.51	+0.05
	ISRUC-SG1	CNN-PTFT	75.76	77.67	78.28	+1.91	+0.60
	ISRUC-SG3	CNN-PTFT	69.73	77.93	76.71	+8.20	-1.22

Table S29: Normalization controls across the five dataset configurations. Values are five-fold mean macro-F1 $\pm$ population standard deviation (%). Δ_{M-N} is the difference between the normalization-matched and main-grid *Native* conditions, and Δ_{C-M} is the difference between CNN-2S and the normalization-matched *Native* condition.

Dataset	Backbone	<i>Native</i>	Norm.-matched <i>Native</i>	CNN-2S	Δ_{M-N}	Δ_{C-M}
ANPHY	GraphSleepNet	79.35 ± 1.62	78.21 ± 2.81	81.44 ± 1.52	-1.14	+3.23
	ProductGraphSleepNet	80.64 ± 0.96	79.86 ± 1.80	83.20 ± 1.25	-0.78	+3.34
	MSTGCN	79.91 ± 1.23	79.91 ± 1.23	80.81 ± 1.16	0.00	+0.91
	JK-STGCN	80.38 ± 1.90	78.98 ± 1.96	81.53 ± 1.25	-1.40	+2.55
	STDP-GCN	77.79 ± 2.21	72.12 ± 3.13	81.44 ± 1.62	-5.67	+9.32
Sleep-EDF20	GraphSleepNet	74.70 ± 2.57	75.67 ± 2.83	79.41 ± 1.37	+0.97	+3.74
	ProductGraphSleepNet	77.46 ± 2.88	78.92 ± 3.52	80.16 ± 1.35	+1.46	+1.24
	MSTGCN	79.74 ± 2.16	79.74 ± 2.16	78.94 ± 1.69	0.00	-0.79
	JK-STGCN	80.53 ± 2.47	78.65 ± 2.78	79.04 ± 1.73	-1.87	+0.39
	STDP-GCN	72.21 ± 5.02	68.28 ± 4.43	79.58 ± 1.85	-3.93	+11.30
Sleep-EDFx	GraphSleepNet	72.23 ± 1.56	70.41 ± 2.61	76.20 ± 1.62	-1.81	+5.79
	ProductGraphSleepNet	75.93 ± 2.17	74.15 ± 2.33	76.71 ± 1.43	-1.78	+2.56
	MSTGCN	75.20 ± 1.75	75.20 ± 1.75	74.76 ± 2.08	0.00	-0.44
	JK-STGCN	77.07 ± 1.64	74.13 ± 4.42	76.08 ± 2.67	-2.94	+1.95
	STDP-GCN	70.35 ± 1.37	65.79 ± 2.59	76.90 ± 1.35	-4.55	+11.11
ISRUC-SG1	GraphSleepNet	72.71 ± 1.24	72.13 ± 2.23	79.38 ± 1.08	-0.58	+7.25
	ProductGraphSleepNet	75.76 ± 1.72	75.90 ± 1.69	79.93 ± 1.19	+0.14	+4.03
	MSTGCN	77.72 ± 1.75	77.72 ± 1.75	78.14 ± 1.23	0.00	+0.41
	JK-STGCN	76.47 ± 1.73	77.02 ± 1.89	79.34 ± 1.68	+0.54	+2.32
	STDP-GCN	72.40 ± 1.25	70.31 ± 0.40	79.47 ± 1.15	-2.09	+9.16
ISRUC-SG3	GraphSleepNet	72.25 ± 4.85	73.47 ± 3.22	78.81 ± 2.05	+1.22	+5.34
	ProductGraphSleepNet	69.73 ± 4.88	73.69 ± 2.42	76.79 ± 4.45	+3.96	+3.10
	MSTGCN	77.10 ± 3.23	77.10 ± 3.23	77.12 ± 2.53	0.00	+0.02
	JK-STGCN	76.82 ± 3.31	77.21 ± 2.75	77.22 ± 3.10	+0.39	+0.01
	STDP-GCN	71.59 ± 4.08	66.76 ± 5.36	78.30 ± 2.44	-4.83	+11.54

Table S30: Available-history normalization sensitivity. Values are five-fold mean macro-F1 $\pm$ population standard deviation (%). Parentheses give the change relative to the corresponding full-recording normalization condition in percentage points.

Dataset	Backbone	CNN-2S	CNN-E2E	CNN-PTFT
ANPHY	GraphSleepNet	81.26 $\pm$ 1.73 (−0.17)	81.22 $\pm$ 1.29 (+1.71)	80.70 $\pm$ 1.38 (−1.47)
	ProductGraphSleepNet	82.12 $\pm$ 1.75 (−1.09)	81.76 $\pm$ 1.51 (−0.35)	80.71 $\pm$ 1.13 (−0.45)
	MSTGCN	80.68 $\pm$ 0.67 (−0.13)	78.73 $\pm$ 2.04 (+0.14)	80.99 $\pm$ 0.54 (+0.50)
	JK-STGCN	81.32 $\pm$ 1.05 (−0.21)	80.22 $\pm$ 1.82 (+0.64)	82.47 $\pm$ 1.03 (+0.52)
	STDP-GCN	80.71 $\pm$ 2.71 (−0.72)	81.50 $\pm$ 1.79 (+0.83)	81.58 $\pm$ 2.04 (+0.13)
Sleep-EDF20	GraphSleepNet	78.90 $\pm$ 0.96 (−0.51)	79.52 $\pm$ 2.10 (+0.15)	78.66 $\pm$ 1.90 (−0.12)
	ProductGraphSleepNet	79.66 $\pm$ 1.45 (−0.50)	79.40 $\pm$ 1.06 (+0.05)	79.01 $\pm$ 1.35 (−0.10)
	MSTGCN	77.98 $\pm$ 2.04 (−0.97)	80.11 $\pm$ 1.90 (+0.99)	78.65 $\pm$ 1.70 (+0.46)
	JK-STGCN	79.35 $\pm$ 1.95 (+0.31)	78.91 $\pm$ 2.07 (−0.97)	79.29 $\pm$ 1.31 (−0.64)
	STDP-GCN	79.56 $\pm$ 1.90 (−0.02)	78.86 $\pm$ 2.30 (+0.45)	77.96 $\pm$ 2.22 (−1.07)
Sleep-EDFx	GraphSleepNet	76.83 $\pm$ 1.24 (+0.63)	74.73 $\pm$ 1.83 (−0.11)	75.05 $\pm$ 1.77 (+0.16)
	ProductGraphSleepNet	77.42 $\pm$ 1.38 (+0.71)	76.41 $\pm$ 2.10 (+1.03)	75.73 $\pm$ 1.71 (+0.25)
	MSTGCN	74.19 $\pm$ 1.62 (−0.57)	75.72 $\pm$ 2.11 (−0.37)	74.78 $\pm$ 2.05 (−0.31)
	JK-STGCN	75.82 $\pm$ 1.61 (−0.26)	76.61 $\pm$ 2.76 (−0.38)	75.47 $\pm$ 2.52 (−0.29)
	STDP-GCN	77.05 $\pm$ 1.50 (+0.16)	75.55 $\pm$ 1.68 (−0.56)	75.59 $\pm$ 1.44 (+0.55)
ISRUC-SG1	GraphSleepNet	79.12 $\pm$ 1.23 (−0.26)	77.61 $\pm$ 0.89 (+0.22)	77.88 $\pm$ 1.05 (−0.33)
	ProductGraphSleepNet	79.46 $\pm$ 1.25 (−0.47)	78.82 $\pm$ 1.50 (+0.44)	77.63 $\pm$ 0.55 (−0.65)
	MSTGCN	77.99 $\pm$ 0.39 (−0.14)	78.61 $\pm$ 0.49 (+0.33)	78.90 $\pm$ 0.44 (+0.61)
	JK-STGCN	79.81 $\pm$ 0.78 (+0.47)	79.16 $\pm$ 0.79 (+0.60)	79.90 $\pm$ 0.49 (+0.67)
	STDP-GCN	79.38 $\pm$ 1.28 (−0.09)	79.01 $\pm$ 1.07 (+0.84)	78.59 $\pm$ 1.29 (−0.31)
ISRUC-SG3	GraphSleepNet	78.87 $\pm$ 1.66 (+0.05)	78.43 $\pm$ 2.19 (+0.40)	79.83 $\pm$ 1.64 (+1.14)
	ProductGraphSleepNet	76.69 $\pm$ 1.83 (−0.10)	76.86 $\pm$ 3.44 (+0.74)	76.01 $\pm$ 3.50 (−0.70)
	MSTGCN	76.19 $\pm$ 2.69 (−0.93)	78.40 $\pm$ 2.86 (+1.12)	76.30 $\pm$ 2.51 (−1.38)
	JK-STGCN	75.79 $\pm$ 1.99 (−1.43)	77.27 $\pm$ 3.13 (−0.09)	76.97 $\pm$ 2.58 (−1.31)
	STDP-GCN	78.29 $\pm$ 1.81 (−0.01)	74.88 $\pm$ 4.94 (−4.10)	76.83 $\pm$ 2.97 (+0.35)

Table S31: Cross-dataset transfer using Sleep-EDFx as the source dataset. Macro-F1 values are reported as the mean $\pm$ population standard deviation across five target folds. Parentheses give the percentage-point difference between CNN-2S-XD and the corresponding target-dataset baseline. Positive values indicate higher macro-F1 for CNN-2S-XD.

Target	Backbone	CNN-2S-XD	<i>Native</i>	Target CNN-2S
ANPHY	GraphSleepNet	79.43 $\pm$ 2.10	79.35 $\pm$ 1.62 (+0.08)	81.44 $\pm$ 1.52 (−2.01)
	ProductGraphSleepNet	80.97 $\pm$ 1.29	80.64 $\pm$ 0.96 (+0.33)	83.20 $\pm$ 1.25 (−2.23)
	MSTGCN	80.53 $\pm$ 1.19	79.91 $\pm$ 1.23 (+0.63)	80.81 $\pm$ 1.16 (−0.28)
	JK-STGCN	82.67 $\pm$ 1.45	80.38 $\pm$ 1.90 (+2.29)	81.53 $\pm$ 1.25 (+1.14)
	STDP-GCN	79.47 $\pm$ 3.02	77.79 $\pm$ 2.21 (+1.69)	81.44 $\pm$ 1.62 (−1.96)
ISRUC-SG1	GraphSleepNet	76.76 $\pm$ 0.69	72.71 $\pm$ 1.24 (+4.04)	79.38 $\pm$ 1.08 (−2.63)
	ProductGraphSleepNet	77.16 $\pm$ 0.50	75.76 $\pm$ 1.72 (+1.40)	79.93 $\pm$ 1.19 (−2.77)
	MSTGCN	77.41 $\pm$ 1.09	77.72 $\pm$ 1.75 (−0.31)	78.14 $\pm$ 1.23 (−0.73)
	JK-STGCN	78.77 $\pm$ 0.78	76.47 $\pm$ 1.73 (+2.29)	79.34 $\pm$ 1.68 (−0.57)
	STDP-GCN	77.87 $\pm$ 1.05	72.40 $\pm$ 1.25 (+5.48)	79.47 $\pm$ 1.15 (−1.60)
ISRUC-SG3	GraphSleepNet	78.51 $\pm$ 1.90	72.25 $\pm$ 4.85 (+6.26)	78.81 $\pm$ 2.05 (−0.30)
	ProductGraphSleepNet	74.29 $\pm$ 1.42	69.73 $\pm$ 4.88 (+4.57)	76.79 $\pm$ 4.45 (−2.50)
	MSTGCN	77.81 $\pm$ 1.14	77.10 $\pm$ 3.23 (+0.71)	77.12 $\pm$ 2.53 (+0.69)
	JK-STGCN	77.80 $\pm$ 3.12	76.82 $\pm$ 3.31 (+0.98)	77.22 $\pm$ 3.10 (+0.58)
	STDP-GCN	77.73 $\pm$ 1.71	71.59 $\pm$ 4.08 (+6.14)	78.30 $\pm$ 2.44 (−0.57)

Table S32: Cross-dataset transfer using ISRUC-SG1 as the source dataset. Macro-F1 values are reported as the mean $\pm$ population standard deviation across five target folds. Parentheses give the percentage-point difference between CNN-2S-XD and the corresponding target-dataset baseline. Positive values indicate higher macro-F1 for CNN-2S-XD.

Target	Backbone	CNN-2S-XD	<i>Native</i>	Target CNN-2S
Sleep-EDF20	GraphSleepNet	79.98 $\pm$ 1.45	74.70 $\pm$ 2.57 (+5.27)	79.41 $\pm$ 1.37 (+0.57)
	ProductGraphSleepNet	80.47 $\pm$ 1.69	77.46 $\pm$ 2.88 (+3.01)	80.16 $\pm$ 1.35 (+0.31)
	MSTGCN	78.85 $\pm$ 2.03	79.74 $\pm$ 2.16 (−0.89)	78.94 $\pm$ 1.69 (−0.09)
	JK-STGCN	78.76 $\pm$ 2.95	80.53 $\pm$ 2.47 (−1.76)	79.04 $\pm$ 1.73 (−0.28)
	STDP-GCN	79.91 $\pm$ 2.52	72.21 $\pm$ 5.02 (+7.70)	79.58 $\pm$ 1.85 (+0.33)
Sleep-EDFx	GraphSleepNet	74.51 $\pm$ 2.50	72.23 $\pm$ 1.56 (+2.28)	76.20 $\pm$ 1.62 (−1.69)
	ProductGraphSleepNet	76.60 $\pm$ 1.68	75.93 $\pm$ 2.17 (+0.67)	76.71 $\pm$ 1.43 (−0.11)
	MSTGCN	73.51 $\pm$ 2.61	75.20 $\pm$ 1.75 (−1.69)	74.76 $\pm$ 2.08 (−1.25)
	JK-STGCN	74.96 $\pm$ 3.23	77.07 $\pm$ 1.64 (−2.11)	76.08 $\pm$ 2.67 (−1.12)
	STDP-GCN	75.24 $\pm$ 2.74	70.35 $\pm$ 1.37 (+4.90)	76.90 $\pm$ 1.35 (−1.65)
ISRUC-SG3	GraphSleepNet	80.79 $\pm$ 1.48	72.25 $\pm$ 4.85 (+8.54)	78.81 $\pm$ 2.05 (+1.98)
	ProductGraphSleepNet	77.12 $\pm$ 2.24	69.73 $\pm$ 4.88 (+7.39)	76.79 $\pm$ 4.45 (+0.33)
	MSTGCN	79.38 $\pm$ 2.00	77.10 $\pm$ 3.23 (+2.28)	77.12 $\pm$ 2.53 (+2.26)
	JK-STGCN	78.76 $\pm$ 3.59	76.82 $\pm$ 3.31 (+1.94)	77.22 $\pm$ 3.10 (+1.54)
	STDP-GCN	80.14 $\pm$ 1.80	71.59 $\pm$ 4.08 (+8.54)	78.30 $\pm$ 2.44 (+1.83)
ANPHY	GraphSleepNet	79.60 $\pm$ 1.89	79.35 $\pm$ 1.62 (+0.25)	81.44 $\pm$ 1.52 (−1.84)
	ProductGraphSleepNet	81.83 $\pm$ 1.38	80.64 $\pm$ 0.96 (+1.19)	83.20 $\pm$ 1.25 (−1.37)
	MSTGCN	79.72 $\pm$ 1.75	79.91 $\pm$ 1.23 (−0.18)	80.81 $\pm$ 1.16 (−1.09)
	JK-STGCN	82.08 $\pm$ 0.75	80.38 $\pm$ 1.90 (+1.70)	81.53 $\pm$ 1.25 (+0.55)
	STDP-GCN	80.87 $\pm$ 2.04	77.79 $\pm$ 2.21 (+3.08)	81.44 $\pm$ 1.62 (−0.57)

Table S33: Phase-separated training-pipeline time in minutes per fold. Values are the mean $\pm$ population standard deviation across dataset-fold runs. Feature preparation includes frontend training and, where applicable, frozen-representation generation and storage.

Backbone	Condition	Feature preparation	Model training	Total
Graph SleepNet	<i>Native</i>	–	4.33 ± 4.57	4.33 ± 4.57
	CNN-2S	4.15 ± 3.47	3.83 ± 4.12	7.98 ± 7.46
	CNN-E2E	–	11.22 ± 7.52	11.22 ± 7.52
	CNN-PTFT	3.72 ± 3.11	11.70 ± 9.65	15.42 ± 12.68
ProductGraph SleepNet	<i>Native</i>	–	35.13 ± 34.29	35.13 ± 34.29
	CNN-2S	4.07 ± 3.41	25.35 ± 22.34	29.42 ± 25.63
	CNN-E2E	–	36.32 ± 26.39	36.32 ± 26.39
	CNN-PTFT	3.72 ± 3.16	33.10 ± 27.14	36.82 ± 30.18
MSTGCN	<i>Native</i>	4.22 ± 3.33	7.91 ± 7.85	12.13 ± 11.05
	CNN-2S	3.02 ± 2.24	8.44 ± 7.55	11.46 ± 9.71
	CNN-E2E	–	17.07 ± 13.44	17.07 ± 13.44
	CNN-PTFT	3.14 ± 2.55	18.78 ± 15.57	21.92 ± 18.07
JK-STGCN	<i>Native</i>	7.28 ± 6.47	7.24 ± 7.20	14.52 ± 13.52
	CNN-2S	4.14 ± 3.32	7.38 ± 7.59	11.52 ± 10.82
	CNN-E2E	–	17.10 ± 14.34	17.10 ± 14.34
	CNN-PTFT	3.50 ± 2.95	13.97 ± 10.76	17.46 ± 13.64
STDP-GCN	<i>Native</i>	–	5.03 ± 6.66	5.03 ± 6.66
	CNN-2S	5.76 ± 5.26	2.99 ± 3.36	8.76 ± 8.47
	CNN-E2E	–	10.37 ± 6.86	10.37 ± 6.86
	CNN-PTFT	5.22 ± 4.90	10.56 ± 7.33	15.79 ± 11.84

Table S34: Training-pipeline time by dataset, in minutes per fold. Values are the five-fold mean $\pm$ population standard deviation.

Backbone	Condition	ANPHY	Sleep-EDF20	Sleep-EDFx	ISRUC-SG1	ISRUC-SG3
Graph SleepNet	<i>Native</i>	1.3 ± 0.1	3.0 ± 1.0	12.7 ± 2.6	4.1 ± 1.1	0.6 ± 0.1
	CNN-2S	4.3 ± 0.8	4.8 ± 0.6	22.1 ± 3.0	7.5 ± 0.7	1.2 ± 0.2
	CNN-E2E	11.1 ± 1.4	6.2 ± 0.6	24.1 ± 2.5	12.5 ± 1.4	2.3 ± 0.4
	CNN-PTFT	11.1 ± 0.9	7.4 ± 0.7	38.9 ± 3.1	16.5 ± 2.0	3.1 ± 0.9
ProductGraph SleepNet	<i>Native</i>	15.4 ± 3.4	21.1 ± 3.6	96.9 ± 21.3	38.2 ± 5.8	4.1 ± 1.5
	CNN-2S	13.7 ± 1.6	20.4 ± 6.0	76.9 ± 9.5	30.1 ± 4.9	6.0 ± 1.0
	CNN-E2E	29.3 ± 6.5	22.2 ± 3.4	81.7 ± 15.0	40.8 ± 6.6	7.5 ± 2.0
	CNN-PTFT	28.4 ± 7.0	20.2 ± 1.5	93.0 ± 4.9	36.6 ± 2.9	5.8 ± 0.5
MSTGCN	<i>Native</i>	5.7 ± 0.6	7.7 ± 0.5	32.1 ± 4.1	13.3 ± 4.8	1.8 ± 0.1
	CNN-2S	6.8 ± 1.0	7.6 ± 1.1	29.6 ± 2.5	11.6 ± 1.9	1.8 ± 0.1
	CNN-E2E	14.2 ± 2.5	9.8 ± 0.9	41.4 ± 6.5	17.0 ± 1.9	2.9 ± 0.4
	CNN-PTFT	18.3 ± 3.6	12.4 ± 1.7	55.3 ± 6.7	20.2 ± 2.3	3.3 ± 0.3
JK-STGCN	<i>Native</i>	6.5 ± 0.6	9.5 ± 1.3	40.0 ± 4.7	14.4 ± 1.3	2.3 ± 0.4
	CNN-2S	5.8 ± 0.4	7.5 ± 1.0	32.1 ± 3.6	10.5 ± 0.9	1.7 ± 0.1
	CNN-E2E	17.0 ± 5.1	8.7 ± 2.3	41.2 ± 11.6	15.6 ± 2.9	3.0 ± 0.6
	CNN-PTFT	13.6 ± 1.3	11.0 ± 1.8	42.3 ± 5.8	17.6 ± 2.1	2.9 ± 0.5
STDP-GCN	<i>Native</i>	1.2 ± 0.2	2.4 ± 0.7	17.9 ± 3.2	3.1 ± 0.9	0.6 ± 0.1
	CNN-2S	4.5 ± 0.6	5.5 ± 1.1	24.7 ± 3.6	7.8 ± 1.3	1.2 ± 0.1
	CNN-E2E	10.5 ± 1.0	4.9 ± 1.0	21.7 ± 2.2	12.3 ± 1.8	2.4 ± 0.3
	CNN-PTFT	14.0 ± 1.8	8.0 ± 1.3	36.4 ± 5.4	17.6 ± 2.8	2.9 ± 0.7